Jameel M. Al-Khayri · Shri Mohan Jain
Dennis V. Johnson *Editors*

Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools

Volume 1

Springer

Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools

mrivas@fagro.edu.uy

Jameel M. Al-Khayri • Shri Mohan Jain
Dennis V. Johnson
Editors

Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools

Volume 1

Springer

Editors
Jameel M. Al-Khayri
Department of Agricultural Biotechnology
King Faisal University
Al-Hassa, Saudi Arabia

Shri Mohan Jain
Department of Agricultural Sciences
University of Helsinki
Helsinki, Finland

Dennis V. Johnson
Cincinnati, OH, USA

ISBN 978-3-319-22520-3 ISBN 978-3-319-22521-0 (eBook)
DOI 10.1007/978-3-319-22521-0

Library of Congress Control Number: 2015959794

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

Thus far conventional plant breeding methods have been successfully used for sustainable food production worldwide. Human population is increasing at an alarming rate in developing countries, and food availability to feed the additional mouths could gradually become a serious problem. Moreover, agriculture production is being adversely affected as a result of environmental pollution, rapid industrialization, water scarcity, erosion of fertile topsoil, limited possibility of expanding arable land, lack of improvement of local plant types, erosion of genetic diversity, and dependence on a relatively few crop species for the world's food supply. According to FAO, 70 % more food must be produced over the next four decades to feed the projected nine billion people by the year 2050. Only 30 plant species are used to meet 95 % of the world's food requirements, which are considered as the *major crops*. The breeding programs of these crops have been very much dependent on the ready availability of genetic variation, either spontaneous or induced. Plant breeders and geneticists are under constant pressure to sustain and expand food production by using innovative breeding strategies and introducing minor crops which are well adapted to marginal lands and provide a source of nutrition as well as crops having abiotic and biotic stress tolerances. In traditional breeding, introgression of one or a few genes in a cultivar is carried out via backcrossing for several generations. Now, new innovative additional plant breeding tools, including molecular breeding and plant biotechnology, are available to plant breeders, which have a great potential to be used along with the conventional breeding methods for sustainable agriculture. With the development of new molecular tools such as genomics, molecular marker-assisted backcrossing has made possible rapid introgression of transgenes, reduction of linkage drag, and manipulation of genetic variation for the development of improved cultivars. For example, molecular breeding has great potential to become a routine standard practice in the improvement of several crops. However, a multidisciplinary approach of traditional plant breeding, plant biotechnology, and molecular biology would be strategically ideal for developing new improved crop varieties worldwide to feed the world. This book highlights the recent progress in the development of plant biotechnology, molecular tools, and their usage in plant breeding.

The basic concept of this book is to examine the use of innovative methods augmenting traditional plant breeding toward the development of new crop varieties, grown under different environmental conditions, to achieve sustainable food production.

This book consists of two volumes: Volume 1 subtitled *Breeding, Biotechnology and Molecular Tools* and Volume 2 subtitled *Agronomic, Abiotic and Biotic Stress Traits*. This first volume comprises 21 chapters separated into four parts: Part I, Genetic Resource Utilization and Conservation; Part II, Breeding Strategies; Part III, In Vitro Culture and Transgenic Approaches; and Part IV, Molecular Breeding. Topics covered include plant domestication and germplasm utilization, conventional breeding techniques, and the role of biotechnology. In addition, various biotechnological applications in plant breeding including functional genomics, mutations and methods of detection, and molecular markers are covered. Moreover, in vitro techniques and their applications in plant breeding are discussed with emphasis on embryo rescue, somatic cell hybridization, and somaclonal variation. Other chapters cover haploid breeding, transgenics, cryogenics, and bioinformatics.

Each chapter begins with an introduction covering related background and provides in-depth discussion of the subject supported with high-quality color photos, illustrations, and relevant data. The chapter concludes with prospects for future research directions and a comprehensive list of pertinent references to facilitate further reading.

The book is an excellent reference source for plant breeders and geneticists engaged in breeding programs involving biotechnology and molecular tools together with traditional breeding. It is suitable for both undergraduate and postgraduate students specializing in agriculture, biotechnology, and molecular breeding, as well as for agricultural companies.

Chapters were written by internationally reputable scientists and subjected to a review process to assure quality presentation and scientific accuracy. We greatly appreciate all chapter authors for their contributions toward the success and quality of this book. We are proud of this diverse collaborative undertaking, especially since the two volumes represent the efforts of 105 scientists from 29 countries. We are also grateful to Springer for giving us an opportunity to compile this book.

Al-Hassa, Saudi Arabia — Jameel M. Al-Khayri
Helsinki, Finland — Shri Mohan Jain
Cincinnati, OH, USA — Dennis V. Johnson

Contents

Contributors

Mohammad Abhary Department of Biology, Taibah University, Al-Madinah, Saudi Arabia

George Acquaah Department of Natural Sciences, College of Arts and Sciences, Bowie State University, Bowie, MD, USA

Toyoaki Anai Laboratory of Plant Genetics and Breeding, Faculty of Agriculture, Saga University, Saga, Japan

Tariq Ahmad Bhat Department of Education, Government of Jammu and Kashmir, Jammu and Kashmir, India

N. Manikanda Boopathi Department of Plant Molecular Biology and Bioinformatics, Tamil Nadu Agricultural University, Coimbatore, India

Department of Fruit Crops, Horticultural College and Research Institute, Tamil Nadu Agricultural University, Periyakulam, India

Alberto Camacho-Henriquez Rural Development and Agriculture, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, Bonn, Germany

Anja Christinck Seed4change–Research and Communication, Gersfeld, Germany

Federico Condón Germplasm Bank, National Agricultural Research Institute (INIA), Colonia, Uruguay

Ponnusami Dachinamoorthy Department of Plant Molecular Biology and Bioinformatics, Tamil Nadu Agricultural University, Coimbatore, India

Aayudh Das Department of Biology & Microbiology, College of Agriculture and Biological Sciences, South Dakota State University, Brookings, SD, USA

E. Lamalakshmi Devi Department of Genetics and Plant Breeding, College of Agriculture, Tripura, Lembucherra, West Tripura, India

Gea Galluzzi Bioversity International, Maccarese (Fiumicino), Italy

Mohd Gulfishan Assistant Professor, Faculty of Life and Allied Health Sciences, Glocal University, Saharanpur, India

Stef de Haan Global Program Genetic Resources, International Potato Center (CIP), Lima, Peru

D. Gavin Humphreys Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, Canada

Matthias Jäger Decision and Policy Analysis (DAPA), Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia

Estelle Jaligot CIRAD – UMR DIADE, Montpellier, France

Joanna Jankowicz-Cieslak Plant Breeding and Genetics Laboratory, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, IAEA Laboratories Seibersdorf, International Atomic Energy Agency, Vienna International Centre, Vienna, Austria

Guo-Liang Jiang Agricultural Research Station, Virginia State University, Petersburg, VA, USA

Deming Jin MOA Key Laboratory of Crop Ecophysiology and Farming System in the Middle Reaches of the Yangtze River, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China

Meenakshi Joshi Department of Genetics and Plant Breeding, G.B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, India

Ponnaikoundar Kavitha Department of Plant Molecular Biology and Bioinformatics, Tamil Nadu Agricultural University, Coimbatore, India

Haeng Hoon Kim Department of Well-Being Resources, Sunchon National University, Suncheon, Jeonnam, Republic of Korea

Ron E. Knox Swift-Current Research and Development Centre, Agriculture and Agri-Food Canada, Swift Current, Canada

Friederike Kraemer Rural Development and Agriculture, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, Bonn, Germany

Tondi Yacouba Nassirou MOA Key Laboratory of Crop Ecophysiology and Farming System in the Middle Reaches of the Yangtze River, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China

Mohammad Oves Center of Excellence in Environmental Studies, King Abdul Aziz University, Jeddah, Saudi Arabia

Bimal Paudel Department of Biology & Microbiology, College of Agriculture and Biological Sciences, South Dakota State University, Brookings, SD, USA

Elena Popova Gosling Research Institute for Plant Preservation, Department of Plant Agriculture, University of Guelph, Guelph, Canada

Rajasekar Ravikesavan Department of Millets, Tamil Nadu Agricultural University (TNAU), Coimbatore, India

Adel Rezk Department of Agricultural Biotechnology, College Agriculture and Food Science, King Faisal University, Al-Hassa, Saudi Arabia

Alain Rival CIRAD – UMR DIADE, Montpellier, France

Mercedes Rivas Department of Plant Biology, Faculty of Agronomy, Eastern Regional University Center, University of the Republic, Rocha, Uruguay

Jai S. Rohila Department of Biology & Microbiology, College of Agriculture and Biological Sciences, South Dakota State University, Brookings, SD, USA

Selvam Sathish Department of Plant Molecular Biology and Bioinformatics, Tamil Nadu Agricultural University, Coimbatore, India

Praveen K. Saxena Gosling Research Institute for Plant Preservation, Department of Plant Agriculture, University of Guelph, Guelph, Canada

Mukund Shukla Gosling Research Institute for Plant Preservation, Department of Plant Agriculture, University of Guelph, Guelph, Canada

Ankita Singh Department of Genetics and Plant Breeding, G.B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, India

Sakshi Singh Department of Molecular and Human Genetics, Faculty of Science, Banaras Hindu University, Varanasi, India

Brahma D. Singh School of Biotechnology, Banaras Hindu University, Varanasi, India

Vinay K. Singh Centre for Bioinformatics, School of Biotechnology, Faculty of Science, Banaras Hindu University, Varanasi, India

Anil K. Singh Department of Genetics and Plant Breeding, Institute of Agriculture Science, Banaras Hindu University, Varanasi, India

College of Agriculture and Research Station, Korea, Chhattisgarh, India

Ram Kushal Singh Centre for Sugarcane Biotechnology, Sugarcane Research Institute (U.P. Council of Sugarcane Research), Shahjahanpur, India

Sujeet Pratap Singh Centre for Sugarcane Biotechnology, Sugarcane Research Institute (U.P. Council of Sugarcane Research), Shahjahanpur, India

Acram Taji School of Agricultural, Computational and Environmental Sciences, The University of Southern Queensland, Toowoomba, Australia

Tanya Tapingkae Faculty of Agricultural Technology, Chiang Mai Rajabhat University, Chiang Mai, Thailand

Bradley J. Till Plant Breeding and Genetics Laboratory, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, IAEA Laboratories Seibersdorf, International Atomic Energy Agency, Vienna International Centre, Vienna, Austria

Bingliang Wan Hubei Key Laboratory of Food Crop Germplasm and Genetic Improvement, Food Crop Research Institute, Hubei Academy of Agricultural Science, Hubei Collaborative Innovation Center for Grain Industry, Wuhan, China

H. Dayton Wilde Department of Horticulture and the Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA, USA

Aysen Yumurtaci Faculty of Science and Letters, Department of Biology, University of Marmara, Istanbul, Turkey

Zul Zulkarnain Faculty of Agriculture, University of Jambi, Jambi, Indonesia

Part I
Genetic Resource Utilization and Conservation

Chapter 1
Plant Domestication and Utilization: The Case of the Pampa Biome

Mercedes Rivas and Federico Condón

Abstract The domestication of plants, as a bio-cultural process, is a continuous phenomenon intrinsically associated with the use of plants. Traditional and scientific knowledge constitute the basis of the various uses of plants from in situ harvesting to complete domestication of crops. One of the most important challenges of our time is to achieve the conservation and sustainable use of plant genetic resources of landraces, species in the process of domestication and species used in situ. The in situ conservation of agricultural biodiversity is a basic element for the development of more sustainable agroecosystems, the adaptation to climate change, the conservation of ecosystem services and to ensure local food security—a conception that is strongly linked to the local development and the protection of cultural and biological diversity. Through case studies from the Pampa Biome we will discuss the valorization of plant genetic resources through new domestication, the promotion of the use of scientifically developed best management practices for in situ conservation, the widening of the germplasm base for breeding programs, plant breeding for stress tolerance, the development of participatory plant breeding programs and the development of high quality products.

Keywords Conservation • In situ utilization • Incipient domestication • Landraces • Pampa Biome • Plant genetic resources • Semidomestication

M. Rivas (✉)
Department of Plant Biology, Faculty of Agronomy, Eastern Regional University Center, University of the Republic, Rocha, Uruguay
e-mail: mrivas@fagro.edu.uy

F. Condón
Germplasm Bank, National Agricultural Research Institute (INIA), Colonia, Uruguay
e-mail: fcondon@inia.org.uy

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_1

1.1 Introduction

The beginnings of agriculture date back to the Neolithic Period, only about 10,000–15,000 years ago (Gepts et al. 2012), while during the previous 2 million years humans lived exclusively as hunters, gatherers and fishermen. In their harvesting activities, humans accumulated extensive knowledge relative to plants, which allowed them to identify and use a variety of important plant resources (Harlan 1975), an activity that has continued throughout history in vast regions of the planet. The knowledge of plant diversity and its use has been and remains the basis for the various uses of plants, ranging from in situ utilization to their introduction into cultivation and ultimately their use in the modern breeding process.

The effect of human action as a promoter of variation and change on plants and animals was first reported by Darwin (1859), and starting from the twentieth century, the construction of the concept of domestication has been based on archaeological, ethnobotanical, ecological and genetic studies.

The *primitive professionals of domestication* (Sereno et al. 2008) developed a number of techniques; they harvested a large number of species and they accumulated an extensive knowledge in relation to the life cycles of plants. Of the approximately 250,000 species of higher plants, only a 0.04 % can be considered domesticated (Leakey 2012). The numbers are much larger when it comes to species used by humans, being in the order of 7000 species (Myers 1983). There are eight plant families in which a high proportion domesticated species are concentrated: Poaceae, Leguminosae, Rosaceae, Solanaceae, Asteraceae, Myrtaceae, Malvaceae and Cucurbitaceae (Evans 1996).

A large portion of the research into domestication of plants has been focused on species that exhibit the so-called *domestication syndrome*, which is defined by the presence of traits like non-shattering seeds, a trend towards annual life cycles, increased seed size, reduced tillering and increase in selfing. The species that acquired these traits during the process of domestication are considered truly domesticated. Major research issues today include the dating of archaeological remains to determine when the changes that define domestication occurred, if these traits were quickly fixed or if it was a slow process, if these traits are determined by a few or by many genes and in which geographical location they occurred (Meyer and Purugganan 2013). Also, with the development of sequencing technology, new genes related to domestication are being found and selection imprints studied. The characters of the domestication syndrome, analyzed in various grain crops in the Middle East, show that there are several morpho-physiological traits that differentiate cultigens from wild ancestors, but nevertheless there is the need to establish for each crop the existence of a clear dichotomy for each trait between wild and domesticated forms (Abbo et al. 2014). These authors point out that this approach is applicable to other traits and crops, such as the absence of glycoalkaloids in potato, absence of bitterness in almond and in cucurbits and lack of dormancy in domesticated lupins, soybeans and *Vigna* species. But not all differences between the ancestral species and cultivated are the result of domestication, many of these differences

arose post-domestication. The traits that evolved post-domestication represent a continuum of phenotypes between the wild and the domesticated gene pool, marking the dynamism of these genetic entities (Abbo et al. 2014), and distinguishing domestication genes from improvement genes (Olsen and Wendel 2013).

The processes of assemblage of domesticated phenotypes occur mainly at the population level, which implies that unless wild populations disappear or reproductive barriers isolate them, there are periods of coexistence between domesticated and wild populations, a factor that also contributes to generate a continuum of diversity. Those cases in which the final results of domestication are new species, such as those derived from inter-specific hybridization and polyploidy, are exceptions to the rule, although in such cases it could also be possible to exchange genes with their parental populations (Clement et al. 2010).

This chapter will consider a broad concept of domestication, in the sense of a continuous process that can recognize different degrees of dependence of plants of human activities. Domestication is a co-evolutionary process between plants and humans, so there are domestication gradients ranging from human use of the populations in the wild to typically domesticated populations completely dependent on human action to be able to exist and complete their life cycle. Also we will bear in mind that domestication does not necessarily apply to an entire species, but that intraspecific variation can be found in the degree of domestication (Pickersgill 2007).

This conceptual framework allows the understanding of domestication as a gradual process ranging from the use of natural populations in their original environments to the monoculture of a single genotype, through several intermediate situations or different genetic changes (Clement 1999). Gathering was long regarded as a Harvest of Nature, but now many ethnobotanical studies have documented that this activity would have involved the existence of social arrangements, the use of special tools and the development of strategies with different complexity (Casas et al. 1997). Harris (1989) proposed that the categories of wild and domesticated are merely human constructions and do not necessarily reflect the real world; while Bettinger (2012) states that there is no real break between collection (gathering) and agriculture. Domestication in that sense should be understood as a continuous process in time and space, which is still practiced by farmers and breeders.

Research in recent decades has established that prehistoric humans modified their environments before or during the process of domestication (Ellis et al. 2013). The domestication of landscapes is the consequence of the cultural processes associated with environmental manipulation, and its results are modified environments appropriate for the management of populations of plants and animals. Just as there is a gradient of domestication in plants, there are gradients of landscape domestication. Understanding the processes of co-domestication of crops and landscapes has been identified as relevant for understanding the management and domestication of plants (Clement et al. 2009; Harris 2012; McKey et al. 2012; Lins Neto et al. 2014). So it is possible to separate the use of the term domestication from a biological/genetic point of view as associated with the domestication syndrome, from the use

of the term from a cultural perspective, associated with human decisions to choose certain species for their promotion and use.

In today's world we face many challenges related to conservation, sustainable use and equitable sharing of benefits derived from the use of plant genetic resources and their associated traditional knowledge as recognized in the Convention on Biological Diversity (FAO 2010; UN 1992). The strategies related to food security and adaptation and mitigation of the effects caused by climate change, also rely heavily on the use of a greater diversity of plant genetic resources (Chakraborty and Newton 2011; Jarvis et al. 2008). Crop diversification and broadening of the genetic base used in breeding programs are part of this strategy as well as the conservation and use of landraces (FAO 2010; Salhuana and Pollak 2006).

Furthermore, the prioritization of research and development of so-called orphan or neglected crops has been reported as deficient in the current strategy by Padulosi et al. (2013). Orphan crops related to neglected and underutilized species, including domesticated, semidomesticated and wild, with a usage that is typically local and traditional. In many areas these crops are being lost together with the traditional knowledge associated with their use.

Although there are available to society new biotechnology and scientific tools and a more systematic knowledge, *new domestications* are considered very scarce by some authors. Among these we find listed strawberries, cranberries, macadamia nuts and pecans (Diamond 2012); mango and lychee (Leakey 2012); kiwi (Ferguson and Huang 2007); oil palm (Zeven 1972); jojoba (Benzioni 2006) and murtilla (Chilean guava) (Pastenes et al. 2003), among others. The success of these domestications depends not only on the development of plants adapted to agriculture, but also in the development of the culture of use and adoption of these *new* crops. The insertion of new plant products into potential markets, either occupying niches similar those of other crops (i.e. a new nut) or generating a completely new niche (identification and education of consumers, development of recipes) is a key factor for the development of new domesticated plants.

The challenges related to the conservation and sustainable use of biodiversity occur simultaneously in an era of enormous technological development in genomics, bioinformatics and biological systems, in which the theory leads to the proposal of the replacement of traditional breeding with the engineering plants or superdomestication (Vaughan et al. 2007).

In this chapter we intend to illustrate and discuss the situation of different species found at different stages of the domestication process, both based on knowledge generated by genetic and archaeological studies and based on the experience generated by the current use and handling plants and plant genetic resources. South American examples will be used, with emphasis on species from the Pampa Biome (southern Brazil, Uruguay and eastern Argentina), also known by the name of Río de la Plata Grasslands (Fig. 1.1).

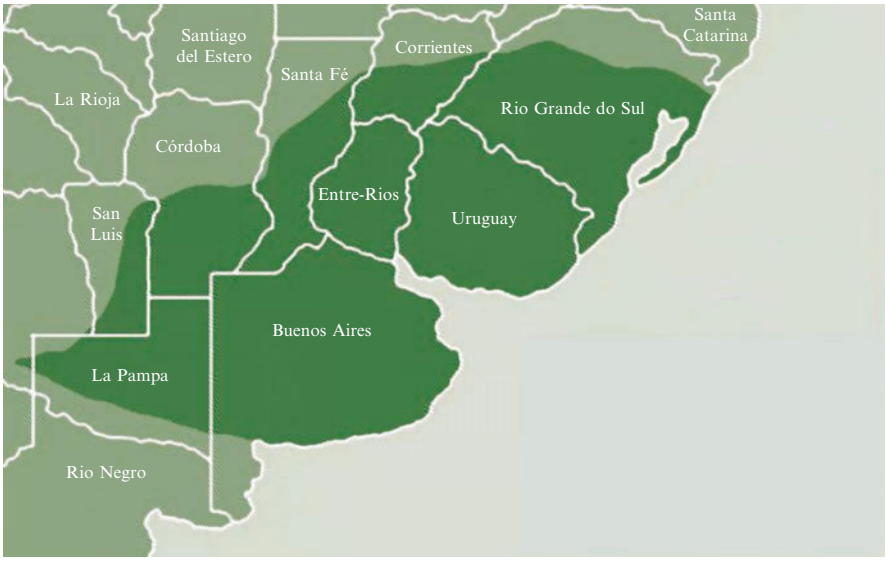

Fig. 1.1 Geographic location of the Pampa Biome (*dark green*)

1.2 Domestication and Use of Plant Genetic Resources in the Pampa Biome

The Pampa biome is one of the main areas of natural temperate-subtropical grasslands of the world. It comprises 760,000 km^2 of fields and plains of central Argentina, southern Brazil and Uruguay (Soriano 1991). The natural vegetation is a mosaic of grasslands (*campos*), wetlands, shrublands, forests and woodlands. Ecologically it occupies a transition zone between tropical and temperate climates, with hot summers, cold winters and no dry season. The geological, geomorphological and soil diversity present in the area have led to a diversity of types of rangelands, both in their physiognomy and in their flora (Lezama et al. 2011; Overbeck et al. 2007; Rosengurtt 1943). In this land we can find an outstanding group of plant genetic resources with value for forage, fruit, medicinal and ornamental use, among others (Berretta et al. 2007; Clausen et al. 2010; Rivas et al. 2014; SNAP 2013; Stumpf et al. 2009).

These are lowlands that were the scenario of diverse human experiences and management practices that transformed the environment at different scales and intensities at different periods of prehistory. The traditional view about the Pampa Biome was that it was only inhabited by hunter-gatherers in the pre-Hispanic era, A situation that is being changed based on results from paleoecological and botanical data on the presence of crops in archaeological contexts found in mound complexes in the lowlands of southern Uruguay (Iriarte et al. 2004). Among the main plant resources identified in archaeological contexts of Uruguay were silicophytoliths and starches among which stand out *Phaseolus* sp. 3050 yr BP (Iriarte et al. 2001),

Cucurbita spp 4600 yr BP (del Puerto and Inda 2008; López Mazz 2000), maize 3660 yr BP (Iriarte et al. 2001, 2004), *Canna* sp. 3660 yr BP (Campos et al. 2001; Iriarte et al. 2001), *Calathea* sp. 3051 yr BP (Iriarte et al. 2001), *Butia odorata* 4600 yr BP (del Puerto and Inda 2008; López Mazz 2000), *Syagrus romanzoffiana* 3660 yr BP (Campos et al. 2001) and *Passiflora* sp. (del Puerto and Inda 2005). These data are consistent with those of the Amazon, the Matto Grosso Pantanal (Schmitz and Beber 2000), the Paraná Delta (Bonomo et al. 2011; Sánchez et al. 2013) and the Atlantic Coast of Brazil for the middle and late Holocene period. The data presented above correspond to a broad region of subtropical-temperate lowlands. It has traditionally been considered that the tropical lowlands have been the exclusive location of domestication, although as Piperno (2012) points out, future research is needed in southern South America.

1.3 Landraces

The post-domestication stage is characterized by the appearance of new diversity and the assemblage of *better* phenotypes clearly selected for the environment and the preferences of the farmers. This process gave and continues giving way to cultivated populations selected by farmers or *landraces*. These landraces all belong to the same species but with clearly different phenotypes, they are the closer gene pool to the elite gene pool used by the modern breeding process and therefore they are the closer source of genes to be used by breeding. Landraces represent the form that domesticated plants took for production until modern plant breeding began to substitute them in the twentieth century as a source of seed. Landraces therefore can be considered the most modern expression of domestication, natural selection and farmer selection, and hold an important wealth of genetic diversity.

The Southern Cone of South America, as a subtropical lowland environment, has been the center of encounter of very different human cultures-indigenous peoples (Barbieri et al. 2014) that were in different cultural development stages by the time the Spanish and Portuguese arrived in the 1500s. Much later in history, a significant number of immigrants and slaves arrived to this region of the world: Spaniards, Italians, Germans, Russians, Basques, Africans, among others. The different cultures brought their own seeds and cultural knowledge to the region, generating a process of adaptation of those seeds kept in farms and home gardens, starting a new adaptation process to the climate and ecology of the region and eventually generating a secondary diversity center for a group of species.

Maize appears as an example of post-domestication diversity increase in this area, as a crop that arrived to the region not from Europe, but from America. Freitas et al. (2003), based on the study of DNA from archaeological remains, found that lowland maize entered southern South America from the northeast through the basin of the Paraguay River, also finding evidence of the presence of a high rate of diversification. Yang et al. (2013) reports the finding of new genetic variation that caused the attenuation of the photoperiod sensitivity in maize. This new functional

genetic change occurred post-domestication due to the insertion of a transposable element within the ZmCCT promoter, and most likely generated the possibility of the expansion of maize to longer-day environments like the Southern Cone of South America, and into new environments, different from its tropical origin, into what today is known as the Pampa Biome.

For the case of maize, landraces are maintained by farmers, and these landraces have very different grain characteristics associated with different uses; a single maize race has been the source for more than one landrace. A great wealth of maize races had been studied and classified (Brieger et al. 1958; Paterniani and Goodman 1977) and reclassified (Gutiérrez et al. 2003). Numbers indicate that more than 300 races have been identified in South America (Paterniani and Goodman 1977). Vilaró (2011) studied the distribution of maize races in the Southern Cone (Brazil, Argentine, Chile, Bolivia, Paraguay and Uruguay) identifying a higher richness of maize races in southern Bolivia, southern Paraguay, Uruguay, the area north of the Buenos Aires Province and the center of Chile, corresponding to mean latitude-low elevation sites except in Bolivia. In this sense, long-day adapted maize races constitute a gene pool not adapted to regeneration in tropical conditions, in which male and female flowering times differ too much to ensure a good representation of the original population alleles. These results agree with what Freitas et al. (2003) postulated based on DNA data.

Maize landraces are always under the threat of their replacement by modern improved cultivars or simply disappearance if the farmers go out of business (Berretta et al. 2007; Porta et al. 2013; Vidal et al. 2011). Collection also has been an important initiative to try to conserve this diversity ex situ, by the Latin American Maize Project (LAMP 1997) one of the major efforts in this sense, with the objective of conserving and giving access to this gene pool; finally leading to the formation of the Germplasm Enhancement of Maize project (GEM 2003). As a strategy to sort and use the genetic diversity, core collections have been proposed for different countries (Abadie et al. 1999; Malosetti and Abadie 2001; PROCISUR 2005) based on morphological characteristics and grain characteristics (Burle et al. 2002).

For other cultivated crops for which landraces have been reported in the region, like horticultural species in Uruguay and Brazil, the culturally diverse immigration to the region still has a role in the maintenance and use of these genetic resources for human consumption (Barbieri et al. 2014). Landraces have been identified for onions (Barbieri et al. 2005; Monteverde et al. 2014; Porta et al. 2014); garlic, sweet pepper and carrot (Galván et al. 2005), for different pumpkin species (Barbieri et al. 2014; Fischer et al. 2012; Priori et al. 2013) and species of *Capsicum* (Neitzke et al. 2011; Vasconcelos et al. 2014; Villela et al. 2014). Landraces have also been identified for forage species like oats and birdsfoot trefoil (Rebuffo et al. 2005). Also, wheat can be cited as a crop in which local landraces have made a significant contribution to the improvement of the crop, contributing disease-resistance genes (Condón and Germán 2005) and adaptation genes. These landraces are today in many cases being conserved in local genebanks like EMBRAPA (Barbieri et al. 2014) and INIA Uruguay (Berretta et al. 2007). They are also being utilized to

generate locally-adapted varieties and to introgress genetic resistance genes and adaptation genes to modern improved varieties (Monteverde et al. 2014).

Landraces of all crops are considered under threat of replacement by commercial varieties (Barbieri et al. 2014) or by genetic contamination through pollen exchange with GMO varieties and modern varieties. Landraces constitute a 24 % of the total accessions conserved ex situ (FAO 2010), which indicate that they are considered important plant genetic resources by the formal research system. The Second Report on the State of the World Plant Genetic Resources for Food and Agriculture (FAO 2010) also reports that on-farm conservation of landraces has improved since the first report. Landraces are always under threat of disappearing either due to replacement with modern varieties or because farmers go out of business; this is even more important in the case of neglected crops because landraces are the only form in which they survive (Galluzzi and López Noriega 2014). The deployment of a conservation strategy for landraces can be also affected by legal issues (López Noriega et al. 2013). Finally, social, cultural and environmental aspects should also be taken into account to design a strategy to ensure the process of natural and anthropological selection of plants while conserving the diversity that is comprised in landraces, that are still evolving and could become a key factor in the response to global climate change (Mercer and Perales 2010; Tapia et al. 2015).

1.4 Populations of Incipient Domestication and Semidomestication

Ethnobotanical information has shown that the gathering and management of numerous plant species remains an important mechanism of plant utilization. A diversity of species are propagated and managed by human actions, but do not depend on it for their reproduction, they are the so-called semidomesticated species or species in a stage of incipient domestication.

These two groups of plants are in early stages of domestication. The cultivated or managed populations have not fixed any traits that distinguishes them from wild populations, but they may differ in allele frequencies for traits subject to selection (Pickersgill 2007). The distinction between species of incipient domestication and semidomesticated species is that the former would only be promoted with a phenotypic average that does not differ from wild populations, while the semidomesticated species would be handled with a phenotypic mean that differs from that of wild populations (Clement 1999). Populations in these stages maintain the phenotypic diversity found in wild populations, although Clement (1999) notes that semidomesticated populations may have a greater diversity and new forms are present in them. As diversity may be due exclusively to natural selection, it is necessary to distinguish the occurrence of anthropogenic activities to refer to incipient domestication activities. Studies of incipient domestication also have a high value for

understanding the origin and dissemination of the experience of handling and domesticating plants (Lins Neto et al. 2014).

In Mesoamerica various management strategies have been identified as practiced by traditional communities such as gathering, incipient management and cropping. Management practices may include strategies and community agreements for the use of forest products, deliberate clearing, burning and even irrigation; planting or seeding and/or reduction of competition from unused plants. Ultimately it comes to management strategies that include tolerance (let-standing), protection (encouraging growing) and promotion of stocks or plants (Casas et al. 2007).

Artificial selection can occur at different levels of intensity (Blancas et al. 2009; Casas et al. 2007). It simply can occur in the systematic in situ gathering when people distinguish plants with favorable characteristics and harvest the best products. Artificial selection acquires greater significance in handling of let-standing, growing encouragement and protection (Casas et al. 2007). In all the cases analyzed by these authors they demonstrated that artificial selection occurs with in situ management and it has a significant impact on plant populations. Desirable characteristics of plants are preserved and promoted by management practices (Lins Neto et al. 2014); people in Mexico use 5000–7000 plant species and are domesticating more than 200 plants that co-occur with wild relatives (Casas et al. 2007). Among them are domesticated species (in the form of landraces) of global importance, others of regional importance and many of only local importance.

For the Amazon, there are indicated to be at least 138 species with some degree of domestication before the European conquest. A higher number of fruit and nuts trees stand out from this group at different stages of domestication, especially incipient. Root and tuber crops are another outstanding group, including cassava (Clement et al. 2010).

In the Pampa Biome there are still a few research studies about semidomesticated and species in an incipient stage of domestication. Fruit trees are probably the most prominent group in which both people and scientists perform activities to achieve domestication. A paradigmatic case is *Acca sellowiana* (Berg.) Burret (Myrtaceae) known as *guayabo del país* in Uruguay, *goiaba-serrana* in Brazil and *feijoa* internationally (Fig. 1.2). It is native to southern Brazil and northeastern Uruguay (Ducroquet et al. 2000). The feijoa tree has value as a fruit, ornamental and medicinal species. The fruit is mainly consumed fresh, although there are some examples of juices and marmalade production. It is characterized by a distinctive flavor, in addition to nutritional characteristics (Thorp and Bieleski 2002). It stands out due to its high iodine content and the presence of bio-active compounds with antibacterial, anticancer, anti-inflammatory and antioxidant action (Weston 2010) and also for a high carotene content (Clerici and Carvalho-Silva 2011).

In its region of origin, feijoa is traditionally used in family farming production systems, from the harvesting of wild trees in situ, isolated trees grown in gardens or rural houses, or small-scale production in fruit farms, mostly with individuals originated through sexual reproduction, a process in which there has been a certain selection history (Donazzolo 2012; Santos et al. 2009; Vignale and Bisio 2005). Moreover, *Acca sellowiana* is grown in New Zealand, California, the Caucasus

Fig. 1.2 *Acca sellowiana* in the Quebrada de los Cuervos Gorge (Uruguay) (**a**) fruits harvested from wild plants; (**b**) wild tree managed for fruit production; (**c**) local farmers in an activity of participative selection; (**d**) details of the flowers

republics of Georgia and Azerbaijan and in Colombia and Chile (Amarante and Santos 2011; Thorp and Bieleski 2002). It is possible to find it in home gardens of Florida (United States) and in Mediterranean countries, especially Italy and Israel (Ducroquet et al. 2000). With the crop development there has been the generation of a range of products different to those traditionally known in the region of origin. In New Zealand it is used in desserts, ice cream, drinks like feijoa wine and dried fruit chips in mixtures with cereals.

Some feijoa plants could be considered as incipiently domesticated in the region of origin as they are maintained and promoted by farmers (Calvete 2013; Donazzolo 2012; Lins Neto et al. 2014; Nodari et al. 2008; Santos et al. 2009, 2013; Vignale and Bisio 2005). The superiority of phenotypic averages of these populations relative to wild populations was also noted by Mattos (1986), Vignale and Bisio (2005) and Donazzolo (2012). In Uruguay feijoa plants appear in gardens of old rural settlements accompanied with quince, pear and orange trees. There are also numerous fruit plants in the southern region, some of them centuries old, evidence that this

species was a component of the fruit trees grown for generations in certain areas of the country.

Studies have been conducted in Brazil and Uruguay on the phenotypic and genetic diversity, both in wild populations and in selected materials (Baccino 2011; Nodari et al. 1997; Puppo 2008; Quezada 2008; Rivas et al. 2007; Santos 2005; Santos et al. 2008), a list of descriptors has been developed (Puppo et al. 2014), progress has been made in the knowledge of its reproductive biology (Finatto et al. 2011; Santos et al. 2007) and recently the first genetic map of the species has been generated (Quezada et al. 2014).

Both Brazil and Uruguay have made progress in recent years related to the domestication and use of feijoa, establishing field collections that have allowed characterization and evaluation of materials, conducting of crossbreeding and selection of materials. In Brazil, some cultivars have been released to growers (Nodari et al. 2008).

Feijoa has also been given a place in participatory domestication programs or participatory plant breeding (Calvete 2013; Santos et al. 2013), in which work is carried out with a participatory approach that integrates the knowledge of local actors and researchers for conservation in situ, material selection and development of handcrafted products.

Another example of a species in an incipient domestication or semidomesticated stage, is *Bromus auleticus* Trinius ex-Nees, a perennial grass with a winter cycle and of recognized value as a forage species. It is a natural component of the campos rangelands of the Pampa Biome, but due to cattle overgrazing it is suffering genetic erosion. It has been the subject of research with the intention of domestication, both for its reintroduction into natural areas for inclusion in cultivated pastures. From the work done since the 1980s some cultivars have been selected from wild populations (Bemhaja 2001; Olmos 1993; Rivas 2001) with superior characteristics in dry matter yield and forage quality. However, like most native forages, the main problems to solve linked to its domestication is seed production, planting and establishment of crops (Pinget and De Battista 2007; Ré et al. 2006).

Finally we want to point that besides a valuable group of fruit trees and native forage, other species from the Pampa Biome are in different stages of domestication. Yerba mate (*Ilex paraguarensis*), algarrobo (*Prosopis nigra* and *P. affinis*), congorosa (*Maytenus ilicifolia*), marcela (*Achyrocline satureioides*), and a variety of ornamental species are just a few of the other examples to consider (Barbieri et al. 2014; Clausen et al. 2010; SNAP 2013; Stumpf et al. 2009).

1.5 *In Situ* Utilization

In the Pampa Biome the co-evolution of grasses and herbivores has generated considerable diversity in species and genetic resources. The grasslands of the Río de la Plata are internationally recognized as a primary center of diversity for forage grasses and legumes. This natural wealth sustains the livestock production that takes

place on them, where the harvesting (gathering) is carried out by cattle and sheep, grazing according to the management decided by farmers. Is this domestication? The answer depends on the intention with which farmers manage grazing. As a matter of fact, an adequate grazing in what is related to stock rate and relief periods clearly favors the best species and genotypes from a grazing point of view (Jaurena et al. 2012; Millot et al. 1987). In our conceptual framework, this can be treated as a form of landscape domestication, because certain species and genotypes are promoted. This topic is relevant not only from an academic point of view, but even becomes more important due to the intensification of land use in the Pampa Biome (Laterra and Rivas 2005; Overbeck et al. 2007; Paruelo et al. 2006). The substitution of natural grassland areas with forestation, agricultural crops and mining increases the risk of losing valuable plant genetic resources and underlines the need to prioritize biodiversity conservation in the grasslands (Faber-Langendoen and Josse 2010), the forgotten biome when it refers to the design of conservation policies (Bond and Parr 2010; Heidenreich 2009; Overbeck et al. 2007).

Another different example of utilization by gathering (extraction) refers to *Butia odorata* (Barb. Rodr.) Noblick, one of the southernmost palm trees in the world, with distribution in southern Brazil and southeastern Uruguay. This palm (locally called *butiá*) referred to as a *palmar* when forming extensive plant communities in which the palms are virtually the only tree component on a matrix of natural grassland. Its utilization has been documented in prehistoric time through both archaeological remains and the presence of artifacts developed to break the hard fruit endocarps to allow the consumption of the seeds. Among the many traditional uses is the consumption of fresh fruits and liqueurs, jellies and marmalades. Currently, some micro-enterprises have developed a diversified and high quality product-line based on butiá which is starting to be commercialized in the formal and gourmet products market. Moreover, palm-tree communities present serious conservation problems due to the lack of regeneration, creating a situation in which actual palm trees are hundreds years old (Rivas 2005). This has been the subject of several research efforts in Brazil and Uruguay, aimed to generate an in situ management strategy combining conservation with local development (Rivas 2013; Rivas et al. 2012). Recommendations for grazing management and harvesting of fruits of butiá have been compiled in a good practices manual (Rivas and Barbieri 2014) (Fig. 1.3).

An important group of species, particularly for fruit and medicinal use, are utilized in situ in the Pampa Biome in a variety of management situations. The sustainability of such utilization deserves the attention of researchers, both for generating adequate management practices as well as for the development of products that can improve the income of farmers (Rivas et al. 2010).

Fig. 1.3 (**a**) Landscape of a *Butia odorata* palm grove (Rocha, Uruguay); (**b**) infructescence; (**c**) fruits

1.6 Plant Breeding and Genetic Resources for Food Security in the Context of Climate Change

Thus far we have referred to the conservation and use of plant genetic resources in the context of domestication of new crops, valorization of landraces and sustainable management of plants populations used in situ. In this section we will refer to plant genetic resources as an essential raw material for breeding programs within the context of food insecurity and climate change.

In today's world agriculture faces great challenges related to world population growth, pressures on natural resources—in particular the expansion of the agricultural frontier—and climate change. To meet these challenges, breeding programs have largely reoriented their objectives towards stress tolerance, both biotic and abiotic (Borém et al. 2012).

The climate change scenario presents an uncertainty of both the rate and the absolute value of climatic changes. In this context, climate may influence many aspects relative to agriculture, including among other things, yield reductions, potential changes in diseases and pest range distribution, changes in virulence, and the possibility of the appearance of new pests, as reviewed by Garret et al. (2014). It can be argued that this climate change will test the ability of humankind to adapt and evolve since a future can be foreseen as a scenario in which food production will take place under a harsher environment and with fewer natural resources, and

smaller agricultural area per capita. To achieve genetic gain, breeding programs will need to have the availability of genetic variants from which to select, paired with the ability to identify, select and introgress adaptation traits, such as root development, heat tolerance, submergence tolerance, translocation rate, or eventually other traits that we cannot predict to help in dealing with climate change. Plant breeding strategies should include actions to conserve and maintain genetic diversity that can allow identification and introduction of genes into cultivated species from wild relatives (Forte et al. 2014) and landraces and will require the adjustment of breeding objectives to cope with abiotic stresses determined by a changing environment in a given place.

A change in environmental conditions will affect other organisms not under direct human control and they may evolve freely to adapt to new conditions: plant pathogens, for example, can change and behave differently. This situation has been modelled in different pathogen-species model such as for the wheat-*Fusarium* pathosystem (Zhang et al. 2014), leading in many cases to scenarios of increases in the frequency or severity of diseases. To cope with biotic stresses, the identification of durable resistance genes that confer a broad spectrum resistance has also been suggested to cope with the effect of climate change (Mundt 2014), an approach that is already being applied to improve genetic resistance in wheat to rust in Uruguay (Silva et al. 2015).

In relation to abiotic stress, breeding has developed cultivars with resistance to heat or less cold-hour requirements, such as in apple, carrot and cauliflower in Brazil (Borém et al. 2012), making their cultivation possible in regions were the crops were not adapted previously. Breeding for resistance to saline soils and soils with high levels of aluminium is a challenge essentially linked to the expansion of the agricultural frontier. Understanding the genetics of tolerance to these stresses and the identification of tolerant plants are the first steps towards obtaining cultivars with these phenotypes. In all cases breeding for abiotic stresses has led to advancement of the knowledge on the physiology and genetics of stresses in recent years, although it is still necessary to continue this work for plant breeding to effectively select materials with adequate tolerance or resistance (Cavatte et al. 2012). For example, it has been stablished in maize that there is no relationship between nutritional efficiency and tolerance to nutritional stresses, these two characteristics being determined by two different groups of genes, which in practical terms means that both traits can be selected simultaneously (Fritsche-Neto and DoVale 2012). Gene cloning for aluminum tolerance is being carried out by different research groups and the expectation is that their results will be useful for genetic improvement (Borém et al. 2012).

In this challenging context, plant genetic resources fulfill a fundamental strategic role for the broadening of the genetic base plant breeding programs and their associated biotechnology. Conservation, characterization and use of plant genetic resources are the pillars on which stand the possibility of continuing in the process of domestication and breeding of plants and crops to cope with climate change and ensure food security.

1.7 Conclusions and Prospects

The domestication of crops is not an issue of the past; domestication in a broader sense is an endless process that continues through efforts by farmers and plant breeders. These processes of utilization and selection of plants are made to a large number of species with traits of interest in different regions of the world. They are evident in populations used in situ, for others with some degree of domestication, and even for landraces. In the words of McGuire and Qualset (2012) the total domestication is not necessary.

The *real* new domestications are scarce, probably due to the difficulty of finding markets to justify the feasibility and investment programs for domestication and breeding (McGuire and Qualset 2012; Rivas et al. 2012). This situation is exacerbated for perennial species due to the time required for domestication and the demands of today's markets (Clement et al. 2005). For new domesticate to become a crop it must be adopted by farmers, have high quality planting material with good agronomic performance and gain market access. The implementation of a new crop at worldwide scale is a very elusive task, unless it has the support and financing of large companies.

As mentioned above, new domestications or simply promoting plant populations that are being utilized in situ or have some degree of domestication, appear to have a better opportunity for local developments and for the production of high quality products. Internationally there is a growing interest in new foods and other products that can contribute to human health and nutrition, as well as cosmetic products, medicines and other uses. The development of markets that benefit farmers of such new products is proposed as a priority line of action for neglected and underutilized crops (Padulosi et al. 2013).

The in situ conservation of agricultural biodiversity is a basic element for the development of more sustainable agroecosystems, the adaptation to climate change, the conservation of ecosystem services and to ensure peoples food security—a conception that is strongly linked to local development and the protection of cultural diversity. It requires strong support from research and development systems that must integrate conservation and development plans based on the valorization of plant genetic resources. The future of domestication and plant breeding will continue to rely on changing the architecture, metabolism and physiology of crops, especially within the context of climate change and food security challenges. To achieve increases in production and generate new products and uses derived from plants—pharmaceuticals, cosmetics and industrial products—it will be necessary to use a combination of applied and theoretical approaches (Vaughan et al. 2007). This is about promoting interdisciplinary programs involving geneticists, breeders, biochemists, sociologists, anthropologists, chemical engineers, agronomists as well as marketing specialists, among others. And especially to work with the people who possess the traditional knowledge and know-how to utilize plants, in that sense seems that participatory breeding programs seem adequate for achieving the desired results (Leakey 2012).

References

Abadie T, Cordeiro CM, Andrade RV et al (1999) Desenvolvimento da coleção nuclear de germoplasma de milho do Brasil. Avances de investigación en recursos genéticos en el Cono Sur. PROCISUR/IICA, Montevideo, pp 101–107

Abbo S, van-Oss Pinhasi R, Gopher A et al (2014) Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. Trends Plant Sci 19:351–360

Amarante CVT, Santos KL (2011) Goiabeira serrana (*Acca sellowiana*). Rev Bras Frut 33(1):i–ii

Baccino ME (2011) Estructura genética de cuatro poblaciones silvestres de *Acca sellowiana* (Berg) Burret situadas en el noreste de Uruguay. Tesis Lic. Ciencias Biológicas, Universidad de la República, Uruguay

Barbieri RL, Leite DL, Choer E et al (2005) Divergência genética entre populações de cebola com base em marcadores morfológicos. Ciên Rur 35:303–308

Barbieri RL, Costa Gomes JC, Alercia A et al (2014) Agricultural biodiversity in Southern Brazil: integrating efforts for conservation and use of neglected and underutilized species. Sustainability 6:741–757

Bemhaja M (2001) Gramínea nativa perenne invernal para suelos arenosos: *Bromus auleticus* cv. INIA Tabobá. In: Rivas M, Oliveira JC (eds) Los recursos fitogenéticos del género *Bromus* en el Cono Sur, Diálogo 56. PROCISUR/IICA, Montevideo, pp 103–104

Benzioni A (2006) Jojoba research as basis for domestication of jojoba in Israel. Israel J Plant Sci 54:157–167

Berretta A, Condón F, Rivas M (2007) Segundo informe país sobre el estado de los recursos fitogenéticos – Uruguay. FAO, Rome. http://www.fao.org/docrep/013/i1500e/Uruguay.pdf. Accessed 29 Jan 2015

Bettinger R (2012) Early steps in agricultural domestication. In: Gepts P, Bettinger R, Brush S et al (eds) Biodiversity in agriculture: domestication, evolution and sustainability. Cambridge University Press, Cambridge, pp 19–20

Blancas J, Casas A, Lira R et al (2009) Traditional management and morphological patterns of *Myrtillocactus schenckii* in the de Tehuacán Valley, Central México. Econ Bot 63:375–387

Bond WJ, Parr CL (2010) Beyond the forest edge: ecology, diversity and conservation of the grassy biomes. Biol Cons 143:2395–2404

Bonomo M, Politis G, Gianotti García C (2011) Montículos, jerarquía social y horticultura en las sociedades indígenas del Delta del río Paraná (Argentina). Lat Amer Antiq 22:297–333

Borém A, Ramalho MAP, Fritsche-Neto R (2012) Abiotic stresses: challenges for plant breeding in the coming decades. In: Fritsche-Neto R, Borém A (eds) Plant breeding for abiotic stress tolerance. Springer, Berlin, pp 1–12

Brieger FG, Gurjel JTA, Patterniani E et al (1958) Races of maize in Brazil and other eastern South American countries. Pub 593, National Academies of Science. National Research Council, Washington, DC

Burle ML, Abadie T, Alves RBN et al (2002) Análise geográfica da coleção de germoplasma de milho em SIG: distribuição da diversidade e aplicação de descritores ecológicos. In: Resumos expandidos XXIV Congresso Nacional de Milho e Sorgo, Embrapa Milho e Sorgo, Florianópolis, 1–5 Sept 2002

Calvete A (2013) Contribución al mejoramiento genético participativo de guayabo del pais *Acca sellowiana* Berg. (Burret) en el paisaje protegido Quebrada de los Cuervos. Tesis Ing. Agr. Universidad de la República, Uruguay

Campos S, del Puerto L, Inda H (2001) Opal phytoliths analysis: its application to the archaeobotanical record in the East of Uruguay. In: Meunier JD, Colin F (eds) Phytoliths: applications in earth sciences and human history. Balkema, Rotterdam, pp 129–142

Casas A, Caballero J, Mapes C et al (1997) Manejo de la vegetación, domesticación de plantas y origen de la agricultura em Mesoamerica. Bol Soc Bot Mex 61:31–47

Casas A, Otero-Arnaiz A, Pérez-Negrón E, Valiente-Banuet A (2007) *In situ* management and domestication of plants in Mesoamerica. Ann Bot 100:1101–1115

Cavatte PC, Martins SCV, Morais LE et al (2012) The physiology of abiotic stresses. In: Fritsche-Neto R, Borém A (eds) Plant breeding for abiotic stress tolerance. Springer, Berlin, pp 21–51
Chakraborty S, Newton AC (2011) Climate change, plant diseases and food security: an overview. Plant Pathol 60:2–14
Clausen A, Ferrer M, Rosso B (2010) Los recursos fitogenéticos del Cono Sur – regiones ecológicas y principales especies. In: Berretta A (ed) Estrategia en los recursos fitogenéticos para los países del Cono Sur. PROCISUR/IICA, Montevideo, pp 21–40
Clement CR (1999) 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and human population decline. Econ Bot 53:188–202
Clement CR, Lleras E, van Leeuwen J (2005) O potencial das palmeiras tropicais no Brasil: acertos e fracassos nas últimas décadas. Agrocien 9(1–2):67–71
Clement CR, Borém A, Lopez MTG (2009) Da domesticação ao melhoramento de plantas. In: Borém A, Lopez MTG, Clement CR (eds) Domesticação e melhoramento: espécies Amazônicas. Universidade Federal de Viçosa, Viçosa, pp 11–38
Clement CR, De Cristo-Araújo M, Coppens D'Eeckenbrugge G et al (2010) Origin and domestication of native Amazonian crops. Diversity 2:72–106
Clerici MTPS, Carvalho-Silva LB (2011) Nutritional bioactive compounds and technological aspects of minor fruits grown in Brazil. Food Res Inter 44:1658–1670
Condón F, Germán S (2005) Casos de uso de recursos genéticos en programas de mejoramiento de cereales de invierno en Uruguay. Agrocien 9:215–219
Darwin C (1859) On the origin of species. John Murray, London
del Puerto L, Inda H (2005) Paleoetnobotánica de los constructores de cerritos del noreste de Uruguay: análisis de silicofitolitos de la estructura monticular Yale27 y su entorno. Traball Arqueol Paisaxe 36:109–122
del Puerto L, Inda H (2008) Estrategias de subsistencia y dinámica ambiental: análisis de silicofitolitos en sitios arqueológicos de la cuenca de Laguna de Castillos, Rocha, República Oriental del Uruguay. In: Zucol AF, Osterrieth M, Brea M (eds) Fitolitos: estado actual de su conocimiento en América del Sur. Universidad Nacional de Mar del Plata, Mar del Plata, pp 221–236
Diamond J (2012) The local origins of domestication. In: Gepts P, Bettinger R, Brush S et al (eds) Biodiversity in agriculture. Domestication, evolution and sustainability. Cambridge University Press, Cambridge, pp 9–17
Donazzolo J (2012) Conservação pelo uso e domesticação da feijoa na Serra Gaúcha RS. Tese (Doutorado em Ciencias) – Centro de Ciências Agrárias, Universidade Federal de Santa Catarina, Florianópolis
Ducroquet JPHJ, Hickel ER, Nodari RO (2000) Goiabeira-serrana (*Feijoa sellowiana*). Série Frutas Nativas 5, FUNEP, Jaboticabal
Ellis EC, Kaplan JO, Fuller DQ et al (2013) Used planet: a global history. Proc Nat Acad Sci 110(20):7978–7985
Evans LT (1996) Crop evolution, adaptation and yield. Cambridge University Press, Cambridge
Faber-Langendoen D, Josse C (2010) World grasslands and biodiversity patterns. A report to IUCN ecosystem management programme. Nature Reserve, Arlington
FAO (2010) Food and Agriculture Organization of the United Nations. Commission on Genetic Resources for Food and Agriculture. Second report on the state of the world's plant genetic resources for food and agriculture. FAO, Rome
Ferguson AR, Huang H (2007) Genetic resources of kiwifruit: domestication and breeding. Hort Rev 33:1–121
Finatto T, Santos KL, Steiner N et al (2011) Late-acting self-incompatibility in *Acca sellowiana* (Myrtaceae). Austral J Bot 59(1):53–60
Fischer S, Peil R, Neitzke R et al (2012) Landraces of ornamental pumpkins and squashes cultivated in southern Brazil. Acta Hort 937:403–407
Forte P, Virili ME, Kuzmanović L et al (2014) A novel assembly of *Thinopyrum ponticum* genes into the durum wheat genome: pyramiding *Fusarium* head blight resistance onto recombinant

lines previously engineered for other beneficial traits from the same alien species. Mol Breed 34(4):1701–1716
Freitas FO, Bendel G, Allaby RG et al (2003) DNA from primitive maize landraces and archaeological remains: implications for the domestication of maize and its expansion into South America. J Archaeol Sci 30:901–908
Fritsche-Neto R, DoVale JC (2012) Breeding for stress-tolerance or resource – use efficiency? In: Fritsche-Neto R, Borém A (eds) Plant breeding for abiotic stress tolerance. Springer, Berlin, pp 13–19
Galluzzi G, López Noriega I (2014) Conservation and use of genetic resources of underutilized crops in the Americas-a continental analysis. Sustainability 6:980–1017
Galván G, González H, Vilaró F (2005) Estado actual de la investigación en poblaciones locales de hortalizas en Uruguay y su utilización en el mejoramiento. Agrocien 9:115–122
Garrett KA, Thomas-Sharma S, Forbes GA et al (2014) Climate change and plant pathogen invasions. In: Ziska LH, Dukes JS (eds) Invasive species and global climate change 4. CABI, Oxfordshire, pp 22–44
GEM (2003) Germplasm enhancement of maize. http://www.public.iastate.edu/~usda–gem/index.htm. Accessed 29 Jan 2015
Gepts P, Bettinger R, Brush S et al (2012) Introduction: the domestication of plants and animals: ten unanswered questions. In: Gepts P, Bettinger R, Brush S et al (eds) Biodiversity in agriculture. Domestication, evolution and sustainability. Cambridge University Press, Cambridge, pp 1–8
Gutiérrez L, Franco J, Crossa J, Abadie T (2003) Comparing a preliminary racial classification with a numerical classification of the maize landraces of Uruguay. Crop Sci 43:718–727
Harlan JR (1975) Crops & man. American Society of Agronomy. Crop Science Society of America, Madison
Harris DR (1989) An evolutionary continuum of people-plant interaction. In: Harris DR, Hillman GC (eds) Foraging and farming: the evolution of plant exploitation. Unwin Hyman, London, pp 11–26
Harris DR (2012) Evolution of agroecosystems: biodiversity, origins and differential development. In: Gepts P, Bettinger R, Brush S et al (eds) Biodiversity in agriculture. Domestication, evolution and sustainability. Cambridge University Press, Cambridge, pp 21–56
Heidenreich B (2009) What are global temperate grasslands worth? A case for their protection. A review of current research on their total economic value. Temperate Grassland Conservation Initiative, Vancouver
Iriarte J, Holst I, López Mazz JM et al (2001) Subtropical wetland adaptations in Uruguay during the Mid-Holocene: an archaeobotanical perspective. In: Purdy B (ed) Enduring records. The environmental and cultural heritage of wetlands. Oxbow Books, Oxford, pp 61–70
Iriarte J, Hols I, Marozzi O et al (2004) Evidence for cultivar adoption and emerging complexity during the mid-Holocene in the La Plata basin. Nature 432:614–617
Jarvis A, Upadhyaya H, Gowda C et al (2008) Climate change and its effect on conservation and use of plant genetic resources for food and agriculture and associated biodiversity for food security: thematic background study. FAO, Rome
Jaurena M, Lezama F, Cruz P (2012) Perennial grasses traits as functional markers of grazing intensity in basaltic grasslands of Uruguay. Chilean J Agr Res 72(4):541–549
LAMP (1997) Latin American maize project. In: Salhuana W, Sevilla R, Eberhart SA (eds). http://www.ars.usda.gov/sp2UserFiles/Place/50301000/Reference_Documents/LAMP-Final-Report-1997.pdf. Accessed 29 Jan 2015
Laterra P, Rivas M (2005) Bases y herramientas para la conservación *in situ* y el manejo integrado de los recursos naturales en los campos y pampas del Cono Sur. Agrocien 9(1–2):169–178
Leakey RRB (2012) Participatory domestication of indigenous fruit and nut trees: new crops for sustainable agriculture in developing countries. In: Gepts P, Bettinger R, Brush S et al (eds) Biodiversity in agriculture. Domestication, evolution and sustainability. Cambridge University Press, Cambridge, pp 479–501

Lezama F, Altesor A, Pereira M et al (2011) Descripción de la heterogeneidad florística en los pastizales naturales de las principales regiones geomorfológicas de Uruguay. In: Altesor A, Ayala W, Paruelo JM (eds) Bases ecológicas y tecnológicas para el manejo de pastizales, vol 26, Serie FPTA-INIA. INIA, Montevideo, pp 15–32
Lins Neto EMF, Peroni N, Casas A et al (2014) Brazilian and Mexican experiences in the study of incipient domestication. J Ethnobiol Ethnomed 10:33
López Mazz JM (2000) Trabajos en tierra y complejidad cultural en las tierras bajas del Rincón de los Indios. In: Durán A, Bracco R (eds) Arqueología de las tierras bajas. Ministerio de Educación y Cultura /Universidad de la República, Montevideo, pp 271–286
López Noriega I, Halewood M, Galluzzi G et al (2013) How policies affect the use of plant genetic resources: the experience of the CGIAR. Resources 2(3):231–269
Malosetti M, Abadie T (2001) Sampling strategy to develop a core collection of Uruguayan maize landraces based on morphological traits. Genet Res Crop Evol 48:381–390
Mattos JR (1986) A goiabeira serrana. "AP" 19. Instituto de Pesquisas de Recursos Naturais Renováveis, Porto Alegre
McGuire PE, Qualset CO (2012) Uses of biodiversity and new and future domestications. In: Gepts P, Bettinger R, Brush S et al (eds) Biodiversity in agriculture. Domestication, evolution and sustainability. Cambridge University Press, Cambridge, pp 475–477
McKey DB, Elias M, Pujol B et al (2012) Ecological approaches to crop domestication. In: Gepts P, Bettinger R, Brush S et al (eds) Biodiversity in agriculture. Domestication, evolution and sustainability. Cambridge University Press, Cambridge, pp 479–501
Mercer KL, Perales HR (2010) Evolutionary response of landraces to climate change in centers of crop diversity. Evol Appl 3:480–493
Meyer RS, Purugganan MD (2013) Evolution of crop species: genetics of domestication and diversification. Nat Rev Genet 14:840–852
Millot JC, Methol R, Risso D (1987) Relevamiento de pasturas naturales y mejoramientos extensivos en áreas ganaderas del Uruguay. FUCREA-CHPA, Montevideo
Monteverde E, Galván GA, Speranza P (2014) Genetic diversification of local onion populations under different production systems in Uruguay. Plant Genet Res. Charact Util 1:9. doi:http://dx.doi.org/10.1017/S1479262114000963
Mundt CC (2014) Durable resistance: a key to sustainable management of pathogens and pests. Infect Genet Evol 27:446–455
Myers N (1983) A wealth of wild species: storehouse for human welfare. Westview Press, Boulder
Neitzke RS, Barbieri RL, Vasconcelos CS et al (2011) Diversity in *Capsicum* landraces cultivated in Brazil. Acta Hort 918:531–536
Nodari RO, Guerra MP, Ducroquet JPHJ (1997) Genetic variability of *Feijoa sellowiana* germoplasm. Acta Hort 452:41–46
Nodari RO, Santos KL, Ducroquet JPHJ et al (2008) Goiabeira-serrana: domesticação. In: Barbieri RL, Stumpf ERT (eds) Origem e evoluçao de plantas cultivadas. Embrapa, Brasilia, pp 407–428
Olmos F (1993) *Bromus auleticus.* Serie técnica 35, INIA, Montevideo
Olsen KM, Wendel JF (2013) A bountiful harvest: genomic insights into crop domestication phenotypes. Ann Rev Plant Biol 64:47–70
Overbeck GE, Muller SC, Fidelis A et al (2007) Brazil's neglected biome: the South Brazilian Campos. Persp Plant Ecol Evol Syst 9:101–116
Padulosi S, Thompson J, Rudebjer P (2013) Fighting poverty, hunger and malnutrition with neglected and underutilized species (NUS): needs, challenges and the way forward. Bioversity International, Rome
Paruelo JM, Guerschman JP, Piñeiro G et al (2006) Cambios en el uso de la tierra en Argentina y Uruguay: marcos conceptuales para su análisis. Agrocien 10(2):47–61
Pastenes C, Santa-Maria E, Infante R et al (2003) Domestication of the Chilean guava (*Ugni molinae* Turcz.), a forest understorey shrub, must consider light intensity. Sci Hort 98:71–84
Paterniani E, Goodman M (1977) Races of maize in Brazil and adjacent areas. CIMMYT, Mexico

Pickersgill B (2007) Domestication of plants in the Americas: insights from Mendelian and molecular genetics. Ann Bot 100:925–940

Pinget AD, Ré AE, De Battista JP (2007) Variabilidad genética en cebadilla chaqueña (*Bromus auleticus* Trin. ex Ness). In: XXXVI Cong Arg de Genética. J Basic Appl Genet 18 (Suppl):S-143–144

Piperno DR (2012) New archaeobotanical information on early cultivation and plant domestication involving microplant (phytolith and starch grain) remains. In: Gepts P, Bettinger R, Brush S et al (eds) Biodiversity in agriculture. Domestication, evolution and sustainability. Cambridge University Press, Cambridge, pp 136–159

Porta B, Antúnez MJ, Olaizola J et al (2013) Identificación y análisis de diversidad de variedades criollas de maíz conservadas *in situ-on farm* en Tacuarembó. In: IX SIRGEALC (Simposio de Recursos Genéticos para América Latina y El Caribe), CENTA, Acajutla, Nov 2013. http://www.centa.gob.sv/index.php?option=com_content&view=article&id=%20692&Itemid=84. Accessed 29 Jan 2015

Porta B, Rivas M, Gutiérrez L, Galván GA (2014) Variability, heritability, and correlations of agronomic traits in an onion landrace and derived S1 lines. Crop Breed Appl Biotech 14:29–35

Priori D, Barbieri RL, Castro CM et al (2013) Diversidade genética de *Cucurbita pepo*, *C. argyrosperma* e *C. ficifolia* empregando marcadores microssatélites. Hort Bras 31:361–368

PROCISUR (2005) Desarrollo de colecciones núcleo de maíz en el Cono Sur de América Latina: Argentina, Bolivia, Brasil, Chile, Paraguay y Uruguay. REGENSUR, PROCISUR/IICA, Montevideo

Puppo M (2008) Prospección y caracterización de poblaciones silvestres de *Acca sellowiana* (Guayabo del país). Tesis Ing. Agr. Universidad de la República, Uruguay

Puppo M, Rivas M, Franco J et al (2014) Propuesta de descriptores para *Acca sellowiana* (Berg.) Burret. Rev Brasil Fruit 36(4):957–970

Quezada M (2008) Estudio de la diversidad genética de una colección de *Acca sellowiana* (Berg) Burret con alto potencial agronómico mediante el uso de marcadores moleculares RAPD. Tesis Lic. Universidad de la República, Uruguay

Quezada M, Pastina MM, Ravest G et al (2014) A first genetic map of *Acca sellowiana* based on ISSR, AFLP and SSR markers. Sci Hort 169:138–146

Ré AE, De Battista JP, Costa MC (2006) Variabilidad de caracteres asociados al vigor de plántula en cebadilla chaqueña (*Bromus auleticus* Trin.). 1. Variabilidad fenotípica, genética, heredabilidad y progreso esperado por selección. In: 26 Congreso Argentino de Producción Animal (RAPA 26) Mar del Plata, 18–20 de October http://www.aapa.org.ar/congresos/2006/PpPdf/PP16.pdf. Accessed 29 Jan 2015

Rebuffo M, Condón F, Alzugaray R (2005) Variedades criollas de forrajeras templadas: conservación y uso en mejoramiento genético. Agrocien 9:105–114

Rivas M (2001) El cultivar Potrillo de *Bromus auleticus*. In: Rivas M, Oliveira JC (eds) Los recursos fitogenéticos del género *Bromus* en el Cono Sur, Diálogo 56. PROCISUR/IICA, Montevideo, pp 105–108

Rivas M (2005) Desafiaos y alternatives para la conservacion in situ de los plamares de butia capitita (Mart). Becc Agrocien (Uruguay) 9(1–2):161–168

Rivas M (2013) Conservação e uso sustentável de palmares de *Butia odorata* (Barb. Rodr.) Noblick. PhD dissertation, Universidade Federal de Pelotas, Pelotas

Rivas M, Barbieri RL (2014) Boas práticas de manejo para o extrativismo sustentável do butiá. Embrapa, Brasilia

Rivas M, Vignale B, Camussi G et al (2007) Los recursos genéticos de *Acca sellowiana* (Berg.) Burret en Uruguay. Avances de Investigación en Recursos Genéticos del Cono Sur II. PROCISUR/IICA, Montevideo, pp 103–112

Rivas M, Clausen A, León-Lobos P (2010) Conservación *in situ* de recursos fitogenéticos de importancia para la agricultura y la alimentación. In: Berretta A (ed) Estrategia en los Recursos Fitogenéticos para los Países del Cono Sur. PROCISUR/IICA, Montevideo, pp 59–74

Rivas M, Barbieri RL, Maia LCD (2012) Plant breeding and *in situ* utilization of palm trees. Ciênc Rur 42(2):261–269
Rivas M, Jaurena M, Gutiérrez L et al (2014) Diversidad vegetal del campo natural de *Butia odorata* (Barb.Rodr.) Noblick en Uruguay. Agrocien (Uruguay) 18(2):14–27
Rosengurtt B (1943) Estudios sobre praderas naturales del Uruguay. Tercera contribución. Barreiro y Ramos, Montevideo
Salhuana W, Pollak L (2006) Latin American maize project (LAMP) and germplasm enhancement of maize (GEM) project: generating useful breeding germplasm. Maydica 51:339
Sánchez J, Colobig M, Zucol A et al (2013) Primeros resultados sobre el uso prehispánico de los vegetales en el sitio arqueológico Los Tres Cerros 1 (Victoria, Entre Ríos, Argentina): Análisis del registro biosilíceo. Darwin (nueva serie) 1(2):201–219
Santos KL (2005) Bases genéticas de características de importancia agronômica em goiabeira-serrana (*Acca sellowiana*). Tesis Msc. Universidade Federal de Santa Catarina, Florianópolis
Santos KL, Lenzi M, Capestrano C et al (2007) Evidencia da atuaçao do sistema de autoincompatibilidade tardia em *Acca sellowiana* (Berg.) Burret. (Myrtaceae). Rev Brasil Frut 29(1):120–123
Santos KL, Santos MO, Laborda PR et al (2008) Isolation and characterization of microsatellite markers in *Acca sellowiana* (Berg) Burret. Mol Ecol Res 8(5):998–1000
Santos KL, Peroni N, Guries RP et al (2009) Traditional knowledge and management of *Feijoa* (*Acca sellowiana*) in Southern Brazil. Econ Bot 63(2):204–214
Santos KL, Peroni N, Guries RP et al (2013) Participatory domestication of the fruit tree species feijoa (*Acca sellowiana*) in Brazil. In: De Boef WS, Subedi A, Peroni N et al (eds) Community biodiversity management. Promoting resilience and the conservation of plant genetic resources. Routledge, New York, pp 240–245
Schmitz PI, Beber VM (2000) Aterros no pantanal de Mato Grosso do Sul, Brasil. In: Durán A, Bracco R (eds) Arqueología de las tierras bajas. Ministerio de Educación y Cultura, Montevideo, pp 65–70
Sereno MJCM, Wietholter P, Terra F et al (2008) Domesticaçao das plantas. In: Barbieri RL, Stumpf ERT (eds) Origem e evoluçao de plantas cultivadas. Embrapa, Brasilia, pp 39–58
Silva P, Calvo-Salazar V, Condón F et al (2015) Effects and interactions of genes Lr34, Lr68 and Sr2 on wheat leaf rust adult plant resistance in Uruguay. Euphytica. doi:10.1007/s10681-014-1343-6
SNAP (2013) Sistema nacional de áreas protegidas. Especies prioritarias para la conservación en Uruguay. MVOTMA-SNAP-MEC, Montevideo
Soriano A (1991) Río de la Plata grasslands. In: Coupland RT (ed) Natural grasslands. Introduction and western hemisphere. Elsevier, Amsterdam, pp 367–407
Stumpf ERT, Barbieri RL, Heiden G (eds) (2009) Cores e formas no Bioma Pampa – plantas ornamentais nativas. Embrapa Clima Temperado, Pelotas
Tapia C, Torres E, Parra-Quijano M (2015) Searching for adaptation to abiotic stress: ecogeographical analysis of highland Ecuadorian maize. Crop Sci 55:262–274
Thorp G, Bieleski R (2002) Feijoas: origins, cultivation and uses. HortResearch, Auckland
UN (United Nations) (1992) Convention on biological diversity. Secretariat of the Convention on Biological Diversity, Montreal
Vasconcelos CS, Barbieri RL, Neitzke R et al (2014) Distância genética entre variedades crioulas de *Capsicum chinense*. Magistra 26:178–185
Vaughan D, Balazs E, Heslop-Harrison J (2007) From crop domestication to super-domestication. Ann Bot 100:893–901
Vidal R, Porta B, Alessandri G (2011) Conservación de las variedades locales de maíz Blanco Dentado en Uruguay. IN VIII Simposio Internacional sobre Recursos Genéticos para América Latina y El Caribe, Quito, pp 190–191. http://www.iniap.gob.ec/sitio/images/stories/descargas/sirgealc/SIRGEALC_2011(Resumenes).pdf. Accessed 29 Jan 2015
Vignale B, Bisio L (2005) Selección de frutales nativos en Uruguay. Agrocien 9(1–2):35–39

Vilaró M (2011) Estudio de la diversidad genética de colecciones de maíz (*Zea mays* L.) del Cono Sur de América. Masters degree Dissertation Magister en Ciencias Ambientales. Facultad de Ciencias, Universidad de la República, Montevideo. http://ambiente.fcien.edu.uy/tesis/Tesis%20M%20Vilaro%2019-9-11.pdf. Accessed 29 Jan 2015

Villela JC, Barbieri RL, Castro CM et al (2014) Caracterização molecular de variedades crioulas de pimentas (*Capsicum baccatum*) com marcadores microssatélites. Hort Bras 32:131–137

Weston RJ (2010) Bioactive products from fruit of the feijoa (*Feijoa sellowiana*, Myrtaceae): a review. Food Chem 121:923–926

Yang Q, Li Z, Li W et al (2013) CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. PNAS 110:16969–16974

Zeven A (1972) The partial and complete domestication of the oil palm (*Elaeis guineensis*). Econ Bot 26:274–279

Zhang X, Halder J, White RP et al (2014) Climate change increases risk of *Fusarium* ear blight on wheat in central China. Ann Appl Biol 164(3):384–395

Chapter 2
Decentralized Collaborative Plant Breeding for Utilization and Conservation of Neglected and Underutilized Crop Genetic Resources

Alberto Camacho-Henriquez, Friederike Kraemer, Gea Galluzzi, Stef de Haan, Matthias Jäger, and Anja Christinck

Abstract Neglected and underutilized species (NUS) are referred to as species with under-exploited potentials with regard to food and nutrition security, health, income generation and environmental services. In the past, most breeding activities have concentrated on major crops, leading to overdependence on a few plant species of worldwide economic importance. The reasons for NUS being underutilized are complex and entail economic and agro-ecological constraints as well as lack of knowledge, awareness and supportive policies. *Contextualization, decentralization* and *collaboration* are proposed as important design principles for breeding strategies focusing on NUS. Four case studies illustrate how they are practically implemented in breeding programs. Two case studies focus on improving NUS for adaptation to climate variability and change: participatory breeding programs for sorghum and pearl millet in West Africa and Andean potatoes in Peru are taken as examples. Further case studies concentrate on improving agronomic and nutritional

A. Camacho-Henriquez (✉) • F. Kraemer
Rural Development and Agriculture, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, Godesberger Allee 119, 53175 Bonn, Germany
e-mail: alberto.camacho-henriquez@giz.de; friederike.kraemer@giz.de

G. Galluzzi
Bioversity International, via dei Tre Denari 472/a, 00057 Maccarese (Fiumicino), Italy
e-mail: geagalluzzi@gmail.com

S. de Haan
Global Program Genetic Resources, International Potato Center (CIP), Apartado 1558, Lima 12, Peru
e-mail: s.dehaan@cgiar.org

M. Jäger
Decision and Policy Analysis (DAPA), Centro Internacional de Agricultura Tropical (CIAT), P.O. Box 6713, Cali, Colombia
e-mail: m.jager@cgiar.org

A. Christinck
Seed4change–Research and Communication, Gichenbach 34, D-36129 Gersfeld, Germany
e-mail: mail@seed4change.de

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_2

properties of local quinoa and amaranth varieties in Peru, and sensory quality of a less common vegetable crop in Germany. We discuss breeding strategies for NUS based on their objectives, the genetic resources used, breeding processes and techniques, as well as links to value chain development and seed systems. We conclude that breeding strategies for NUS should increasingly be embedded in multi-stakeholder strategies for collaborative learning and to expand the focus of interest from enhancing varieties towards building innovative institutions and partnerships that help develop sustainable food and farming systems and are linked to human needs.

Keywords Agrobiodiversity • Environmental adaptation • Landraces • Multi-stakeholder strategies • Neglected crops • Participatory plant breeding • Sensory quality • Underutilized crops • Value chains

2.1 Introduction

2.1.1 Neglected and Underutilized Genetic Resources

The term *neglected and underutilized species* (NUS) applies to a wide range of different crops, including less common grain crops, roots, tubers, legumes, leafy vegetables, fruits and spices. In this chapter, NUS refers to species with under-exploited potentials with regard to food security, health, income generation and environmental services, following a general definition given by Jaenicke and Höschle-Zeledon (2006). Terms such as *orphan*, *minor*, *traditional* or *niche* crops have also been used for these crops, which tend to be outside the focus of mainstream agricultural research and development. However, not only entire species can be neglected and underutilized, but also sub-species, cultivar groups or local varieties of major crops, such as maize, rice or wheat. Thus, from a more practical point of view, nearly all landraces of crops, as well as cultivated, semi-wild and wild species that are used for food, fodder, fiber, fuel or health purposes and are of only local or regional importance, can be considered as NUS.

Although research on breeding strategies for NUS may appear as a highly specialized field of knowledge, NUS are in fact the vast majority of all known crops. Overdependence on a few plant species of worldwide economic importance intensifies many acute difficulties faced by countries and communities (Khoury et al. 2014) with regard to food and nutrition security, ecosystem sustainability and adaptation to climate change. Therefore, enhancing the use of NUS to better tap the abovementioned potentials has been identified as an important strategic element for developing more productive, sustainable and resilient agricultural production and food systems in the future (Jaenicke and Höschle-Zeledon 2006; Padulosi et al. 2013; Thies 2000).

In order to raise the profile and contributions of NUS towards these goals, a wide range of activities are necessary, including maintaining genetic diversity on farms; developing better varieties; enhancing cultivation practices and postharvest technologies; improving market access; providing evidence of nutritional benefits and promoting their use; and developing supporting policies at national and international level (Padulosi et al. 2013). Plant breeding could thus be considered as one building block of a broader strategy for promoting the conservation and use of the worlds underutilized plant genetic resources.

NUS are taxonomically a heterogeneous group, and also differ in their reproduction biology (e.g. self-pollinating versus cross-pollinating; sexually versus clonally propagated). Hence, NUS cannot easily be treated as one group from a biological point of view. Rather than centering attention on specific species and the biotechnical aspects of breeding, we will thus focus on common properties of NUS and resulting implications for breeding programs. We will then suggest important design principles for breeding strategies focusing on NUS and illustrate these with examples, so that it becomes more tangible how one can effectively account for them in breeding activities focusing on specific crops. We will conclude with a discussion of the case studies and related issues, and finally present our conclusions.

2.1.2 Common Properties of NUS

NUS have some common properties that help explain why their wider utilization as crops is hampered in spite of their potential, for example regarding food and nutrition security, or farming system resilience. Understanding these common properties is important for appropriately considering them in breeding strategies.

2.1.2.1 Limited Scale of Distribution

In general, the importance of NUS is restricted to certain localities or regions, or to certain population groups that depend upon them. This results in the species being cultivated and marketed only locally, in areas where they are well known, a fact that limits their economic importance. However, NUS may be of great importance, also economically, for those who actually use and maintain them.

Sorghum and millet in West Africa could serve as examples. The importance of these species for local food supply is actually greater in this region compared with major staple crops, such as maize or rice. Similarly, pearl millet is the predominant staple food in some parts of northwest India, and more important than wheat or rice. However, sorghum and millet are of minor importance globally, and preferences and needs with regard to agronomic and postharvest traits vary widely across the regions where these crops are used. The geographical distribution area or number of users may even be much more limited for certain underutilized fruits, herbs, vegetables and fodder crops.

2.1.2.2 Economic Constraints

As a result of this situation, the full market potential for these crops or the products derived from them tends not to be adequately seized, e.g. due to lack of market access or absence of organized and efficient value chains. Investment in postharvest handling, processing, quality norms and standards, packaging, labeling and distribution is deficient. Very often, the products are unknown beyond the original area of distribution, so that marketing at a larger scale would require a considerable effort to increase the level of awareness and to stimulate demand. However, there is evidence showing that, despite the weak development of value chains, considerable additional income from processing and marketing could be generated, once a global demand for these crops beyond their center of origin is established and value chain upgrading and marketing strategies are developed and implemented (Padulosi et al. 2014).

2.1.2.3 Change of Food and Farming Systems

NUS are basically a heritage of past and present generations of farmers and rural communities. Looking deeper into the driving factors for agrobiodiversity loss, we find that variety replacement is of primary importance, followed by land use change and environmental degradation. Variety replacement, or displacement of crops by other species, is closely related to the rapid and ongoing change in food and farming systems in general that has taken place on a global scale since the 1960s. Moreover, population growth, economic pressure, policy impact and civil strife have all contributed to agrobiodiversity loss and rapid change in conditions faced by farmers according to the Food and Agriculture Organization of the United Nations (FAO 1997). These factors can combine and result in former landraces and local crops becoming less adapted and competitive in contemporary food and farming systems. Some may also lose important adaptive traits if fewer farmers continue to grow them, or do so on a smaller area.

2.1.2.4 Importance for Nutrition and Health

NUS are important for food and nutrition security of local populations, not only due to the quantity they contribute to food supply in a given region, but also in view of nutrition and health aspects. It is widely accepted that diversified diets – including fruits, leafy vegetables, legumes, roots, tubers, herbs and spices – are essential for human nutrition, and complement diets based on common staple crops such as rice or maize (Kahane et al. 2013; Keding et al. 2013; see also examples in Burlingame and Dernini 2012). Many leguminous crops, e.g. cowpea and winged bean, are excellent sources of protein and micronutrients. Tropical fruits – including mangoes

and lychees – have high vitamin and mineral contents. These are even exceeded by some semi-wild fruits collected in forests and rangelands, a famous example being the fruits of the baobab tree being very rich in Vitamin C (FAO 1992). Wild fruits can contribute considerably to reducing nutritional risks in vulnerable population groups, as demonstrated, inter alia, by examples, from Kenya, Ghana and India (Jaenicke and Pasiecznik 2009; Kuhnlein 2012).

The various squashes, orange-fleshed sweet potato varieties and some leafy vegetables contain significant amounts of *β*-carotene, a natural vitamin A precursor. Even if grown on small areas only, these crops can be important for improving the diets of children or ill persons; for example, the health status of people affected by chronic diseases or HIV/Aids can be influenced positively through diverse, nutrient-rich diets based on the aforementioned crops (Gari 2004). Germplasm collections of various NUS, including beans, cassava and sweet potato, show considerable variability in key micronutrient contents (Tumwegamire 2011), pointing to their untapped potentials for increasing nutritional value from NUS through targeted breeding and selection.

2.1.2.5 Importance for Resilience of Food and Farming Systems

Frequently, NUS are associated with complex low-external-input farming systems and may be adapted to a highly specific set of agro-ecological conditions. Under such conditions, resistance to typical and variable stress conditions, such as drought, poor soil fertility, extreme heat, flooding, etc., are of great importance. Certain NUS also serve as traditional emergency foods if other crops fail. At the same time, they may have agronomic constraints that limit their wider use, e.g. low yield potential or high labor demand for processing (Padulosi et al. 2013).

Diversification appears to be a key factor for improving resilience and resource-use efficiency in farming systems, with NUS playing an important role in this regard (Kahane et al. 2013; Padulosi et al. 2013). NUS are often multipurpose crops that provide additional services to farmers or farming systems, besides supplying harvestable food. Among the most important, particularly under low input conditions, may be biological nitrogen fixation and mobilization of nutrients from the subsoil. However, fodder yield and quality are also of high importance to farmers in mixed crop-livestock systems, both economically and in view of efficient nutrient cycling.

2.1.2.6 Lack of Knowledge, Awareness and Resources

There is often a lack of well-documented scientific knowledge on NUS, for example relating to the range and properties of existing varieties, cultivation practices, current and potential uses, commercially valuable traits, nutritional value and consumer preferences. At the same time, a rich body of local knowledge may exist that is based on a long tradition of rural communities growing and using these crops.

Therefore, a lack of scientific knowledge can lead to the potential of these crop species being undervalued or ignored by researchers, breeders and policymakers alike. NUS-based strategies are thus more often than not neglected when policies and actions aiming at larger issues, such as food security, health, agricultural development or adaptation to climate change, are discussed. In the best case, they may be treated as a separate issue, but links to the aforementioned policy and action fields remain weak. However, the importance of agrobiodiversity and NUS for food and nutrition security has been repeatedly recognized according to the M S Swaminathan Research Foundation in India (MSSRF 2005) as well as other researchers (Fanzo and Mattei 2012; Kahane et al. 2013).

The same applies also to resources made available for research and for value chain development, including all stages from breeding and seed systems, cultivation and postharvest practices, processing, and up to final consumption. Given the limited economic importance and geographical distribution of each individual NUS, other major crops may regularly appear as having greater priority for investment, for government and the private sector alike.

2.1.3 Important Implications for Plant Breeding Strategies

The general properties of NUS described above have important implications for designing appropriate breeding strategies for NUS. These refer to the broader context and the importance of diversity in this regard, as well as to the availability of scientific information.

2.1.3.1 Understanding the Context

As stated earlier, NUS are often associated with traditional, low-external-input, farming systems, in which they serve multiple purposes. Farmers manage these low-external-input farming systems based on observations of complex factors they basically cannot control. These factors are the result of the high temporal and spatial variability of environmental conditions paired with unfavorable socioeconomic circumstances. In contrast, farmers operating in high-input systems aim at controlling such environmental factors, such as by applying irrigation, fertilizers or pesticides (Kaufmann et al. 2013).

In order to develop breeding strategies for NUS, basic insights into the ecological, economic and social context are thus necessary to understand how plant breeding could effectively contribute to the farmers' own capacity to manage and improve low-external-input farming systems in a sustainable way.

Moreover, the multi-functionality of NUS in low-external-input farming systems needs to be carefully explored. It is not only harvestable yield, but also postharvest traits, non-food uses, market access, consumer preferences and cultural aspects that

all need to be taken into consideration. By striving to achieve a deeper understanding of these multi-functional aspects, the risk of improving certain traits simply at the expense of other, less obvious but still important ones, can be reduced. Moreover, these insights could help breeders identify and address trade-offs between different competing objectives and suggest new, additional options that complement the existing ones.

2.1.3.2 Understanding the Links Between Diversity and Adaptation

For many NUS, a high level of diversity may exist on geographically small scales (e.g. various local landraces). Furthermore, the genetic diversity within these local varieties also tends to be high. Experience shows that farmers working under low-external-input conditions often rely on this diversity as an important means of adaptation to adverse and variable climatic conditions, thus improving the resilience of their farming system. For example, farmers flexibly adapt the variety choice, the proportion of different varieties sown, and seed mixtures to the time of sowing and the soil conditions.

Decision rules farmers use in such situations were described, for example, by Werner et al. (2010) for Fulani pastoralists' choice of millet varieties in the Mopti region of Mali. Their choice as to which variety to sow on which area is based on several criteria, including the stock available in the household, the date of onset of rains, the availability of field preparation tools and the experience of past years (Kaufmann et al. 2013). This means that farmers working under low-input conditions often require a range of varieties with complementary traits; serving this need, and developing appropriate schemes for testing and evaluation, is not normally the focus of formal breeding programs.

In recent years, the diversity *within* varieties has been increasingly understood as an important mechanism for adaptation to adverse and unpredictable climatic conditions. Considering how individual plants and plant populations cope with unpredictable climatic conditions is essential for effectively breeding for yield stability under such conditions (Haussmann et al. 2012). Important pathways to yield stability include (1) *phenotypic plasticity*, meaning that individual plants show a plastic response to environmental conditions, e.g. developing fewer tillers in stress conditions and more under favorable conditions; (2) *population buffering*, which means diversity exists within a variety for key traits, such as flowering dates, resulting in only a portion of plants being susceptible to stress at a given time or place, and all remaining plants able to *escape* the adverse conditions; and (3) *adaptive traits,* directly conferring tolerance or resistance to stresses arising due to climatic variation, such as flooding, drought, or pests and diseases. Thus, by using local landraces as breeding partners, and by breeding variety types that allow a high level of diversity to be maintained, breeding strategies for NUS can actively build on diversity rather than reducing or replacing it.

2.1.3.3 Understanding and Valuing Practitioners' Knowledge

Farmers and farming communities have developed, maintained and improved crop diversity over a long time. Experts, such as plant breeders and genetic resources conservation specialists, are historically rather new actors in this field, which is even more the case for most NUS. Scientific knowledge may not be easily available for many issues relating to NUS, and that is why involving practitioners in a breeding program can be of great value.

However, important differences may exist in the ways that scientists and practitioners create, understand and evaluate knowledge. For example, practitioners tend to observe the performance of items under variable *real life* conditions at one location over many years. They make their observations while farming, and in most cases their observations are restricted to what can be observed by using one's senses. Scientists, in contrast, tend to reduce variability in their trial designs and observe small samples over short periods, but across different locations, and often with the help of highly-sophisticated instruments and evaluation methods that open up new possibilities for observation (Hoffmann et al. 2007).

Hence, in breeding programs, farmers can not only share views and observations in interviews but also contribute knowledge embedded in local varieties, or selection skills that follow their own ways of observation. Plant breeders, in contrast, have access to new breeding material and can make use of contemporary breeding techniques (such as molecular marker-assisted selection), if appropriate. By uniting the expertise of both, new paths can be followed. Targeted Allele Introgression is a practical approach in this direction which could also be relevant for NUS. It was developed at the International Maize and Wheat Improvement Center (CIMMYT), Mexico, in order to provide small-scale farmers with maize that better corresponds to their own local maize varieties, yet also expresses additional traits desired by farmers, but not currently accessible. In order to add these traits to farmers' local maize varieties, maize populations containing the desired traits were identified for crossing with farmers' varieties (Badstue et al. 2012). Understanding and valuing the different contributions of scientists and practitioners could thus be an important entry point for fruitful collaboration between farmers and plant breeders in breeding programs focusing on NUS.

2.2 Design Principles for Breeding Strategies Focusing on NUS

In this section, we propose three key elements or design principles for breeding strategies focusing on NUS: contextualization, decentralization and collaboration. Each of these design principles will be briefly described in the following subsections.

2.2.1 Contextualization

Recalling the abovementioned properties of NUS, and the implications for breeding programs, it is obvious that careful diagnostic work will be necessary to improve understanding of the context. This requires approaches and tools for assessment that are not normally applied in breeding programs.

Besides borrowing expertise from other scientific disciplines, e.g. for socioeconomic assessments, a large number of tools has been suggested and described in manuals that facilitate dialogue between farmers and scientists (see, for example, Christinck et al. 2005; CIP-UPWARD 2003; Sthapit et al. 2006). By applying them systematically, a joint understanding can be achieved of the broader context and the ways that farmers use agrobiodiversity. These include classification, mapping, scoring and ranking exercises adapted to the context of natural resource and agrobiodiversity management and breeding, as well as analytic and strategic planning tools, such as SWOT analysis or scenario work.

Studying and screening local germplasm can also be a way to learn about the context in a very hands-on approach. As a result of local cultivation, use and selection practices, knowledge relating to environmental adaptation, postharvest processing and nutrition requirements may be embedded in the local varieties, and can be made explicit by joint evaluation work. The same applies to evaluating exotic germplasm for new agronomic and potentially commercially valuable traits that are not yet known to farmers.

Furthermore, diagnostic studies on seed systems are necessary, as formal seed systems for NUS are usually absent or underdeveloped. Seed of NUS is usually saved from the farmer's harvest or disseminated via social networks. Informal seed systems are often efficient in maintaining traditional landraces and securing farmers' basic requirements for seed, but may be less so for the dissemination of new varieties. First, because maintaining the variety identity may not be possible under farmers' management without additional measures being taken (depending on the crop and variety type); and second, because social networks may not be extensive enough to allow seed to travel rapidly over large distances, and across social groups. This means that strategic planning is not only required for the breeding process itself, but also for seed production and dissemination of newly developed varieties.

2.2.2 Decentralization

Given the often limited areas of distribution, along with highly diverse environmental conditions as well as preferences and needs of farmers and consumers, decentralizing the breeding process appears to be a promising option for breeding strategies focusing on NUS.

A method known as *evolutionary plant breeding* has been proposed since the 1950s (Suneson 1956). In this approach, landraces, populations or lines of different origins are assembled and bulked to enhance spontaneous or facilitated cross-pollination, resulting in diversified composite populations. Over several generations, the resulting progenies are propagated in bulk and can be subjected to natural and human selection under various agro-ecological conditions. If such *evolutionary populations* are allowed to evolve in different target environments, they can adapt to new agronomic techniques, new uses and eventually new climates (Ceccarelli 2009). Depending on the intensity of human selection applied, either diverse populations can be maintained, or lines can be derived and tested as pure lines where found appropriate (Ceccarelli 2014). In recent years, this approach has awakened new interest from researchers involved in breeding for organic farming, low-input conditions and NUS (Döring et al. 2011; Galluzzi et al. 2014; Murphy et al. 2005).

Decentralized selection, defined as selection in target environments, was first used as a term by Simmonds (1984). It aims at improving performance of varieties in specific environments, rather than the mean performance *across* different environments. In the past two decades, decentralized selection has been more systematically conceptualized with a view to improving selection gains for marginal, low-input farming systems (Ceccarelli et al. 1996; Cooper et al. 1996). These researchers came to the conclusion that selection and testing of varieties for marginal production conditions and resource-poor farmers should be carried out more intensively in the target environments, that is, in farmers' fields. By identifying the best genotypes for the respective target environment, the interaction between plant populations and specific environmental conditions (G × E interaction) can be positively exploited. This concept is now largely accepted, for which methodology descriptions and practical examples are available (Annichiarico 2009; Ceccarelli 2012; Ceccarelli et al. 2000; Chapman et al. 2002, 2003; Dawson et al. 2008).

Innovative trial designs, and advances in methods for statistical data analysis focusing on the objectives of decentralized and participatory research designs, helped support the approach (Barahona and Levy 2002; Bellon and Reeves 2002). In particular, the *mother and baby* (m&b) trial design has become widely accepted and applied, as it provides the possibility to test and statistically evaluate subsets (baby trials) of a main set of entries (mother trial) at various locations (Snapp 2002).

Decentralization can thus be applied not only for optimizing selection, but also as a more general design principle for breeding programs. It allows evaluation of a relatively large number of lines at different locations under farmer management. A well-documented example is the decentralized barley breeding program of ICARDA, implemented in Syria since 1991 (Ceccarelli et al. 2013a). Breeding projects for NUS could build on existing experience, taking into account the specific context and objectives of each individual project.

2.2.3 *Collaboration*

As stated earlier, plant breeding for NUS can be part of a broader strategy to promote NUS, one that should also rely on awareness and capacity building, marketing and policy development. This broader strategy could be denominated a holistic value chain development strategy, as tested and documented by Padulosi et al. (2014). Depending on the goals of each individual initiative and the existing potentials and constraints, multiple stakeholders, including farmers, farmer organizations, researchers, NGO staff, commercial enterprises and government agencies, may all be involved.

Participatory Plant Breeding (PPB) is a proven concept focusing on the collaboration between farmers and plant breeders. Its original aims were to better target the needs of resource-poor farmers, increasing the level of biodiversity in farmers' fields, increasing both research efficiency and farmer empowerment (Weltzien et al. 2003). It is grounded in the broader context of participatory research, learning and action concepts; the basic idea is that by integrating complementary knowledge and skills, and applying them for shared goals, progress can be achieved more efficiently than by each partner alone (Hoffmann et al. 2007). Ashby and Lilja (2004) drew attention to distinguishing functional and empowering participation: functional participation addresses the objective of improving the efficiency of research processes by involving prospective users of the results. They tend, however, to leave the balance of power for decision-making in the research essentially unchanged: i.e. researchers make most of the critically important decisions. Empowering participation, in contrast, changes the balance of power in decision-making in the research process, usually by giving users a more important role in key decisions about the end product, as well as in how the research is carried out. In practice, both aspects are present in participatory plant breeding projects to varying degrees. Successful examples of such forms of collaboration have been reported from a number of crops and geographical regions, and in some cases with impact on a relatively broad scale (see, for example, Conroy et al. 2009; Humphries et al. 2005; Mustafa et al. 2006)

Other stakeholders, beyond farmers and plant breeders, have been involved occasionally, but scientific documentation is scarce. In some cases, not only researchers and farmers, but also their market partners were involved, for example where the process aimed at developing new products and value chains. Recently, collaborative multi-stakeholder platforms were applied as a tool to bridge the gap between supply and demand and integrate stakeholders into an organized and functional value chain in a project focusing on underutilized native chili pepper (*Capsicum* spp.) varieties in Bolivia and Peru (Jäger et al. in prep.).

The project aimed, inter alia, at developing commercial, institutional and technological innovations that help reduce transaction cost, create trust and capture valuable market opportunities based on commercially valuable traits identified in the collections assembled in both countries. Towards this end, gene banks, small-scale producers and farmers' associations, processors, national research institutes, foundations, private companies, international research and development organizations,

government officials, restaurant chefs and their respective associations, national and international universities, and donors were brought together to jointly identify existing bottlenecks along the chain and to develop enhanced strategies.

As an important outcome, 39 promising accessions identified by the project in Peru were officially adopted by the Horticulture Plant Breeding Program at INIA in order to obtain elite varieties in the future. Selection aimed at balancing two main criteria: (1) market opportunities based on promising commercially valuable traits identified; and (2) optimal representation of diversity present in the national *Capsicum* collection.

To summarize, breeding strategies for NUS can benefit from the experience gained with collaboration in PPB programs, but will require expanding the concept towards collaborative learning and collective action involving multiple stakeholders operating at various levels of value chains to address the specific challenges associated with NUS.

2.3 Examples of Successful Case Studies

In this section, we present four case studies focusing on different objectives, species and geographical regions. Furthermore, the examples demonstrate how breeding programs for specific NUS can relate to the socio-cultural, ecological and economic context, and how decentralization and collaboration can be realized in breeding programs at different scales and in different institutional settings.

2.3.1 Breeding Sorghum and Pearl Millet for Improved Productivity under Conditions of Climate Variability and Change in West Africa

2.3.1.1 Background and Objectives

Pearl millet (*Pennisetum glaucum*) and sorghum (*Sorghum bicolor*) are the most important staple crops cultivated under rainfed conditions in the dryland areas of West Africa, and are thus most important for food security in the region. However, breeding progress has remained limited for these crops in the past, and farmers' adoption of newly bred varieties, particularly those representing plant types different from the local guinea landraces, was low. If farmers adopted new varieties, these were mostly purified guinea-race sorghum landraces selected from local materials. They gave little yield advantages and mainly provided slightly earlier maturity (Weltzien et al. 2006).

Semiarid and sub-humid regions of West Africa are characterized by highly variable climatic conditions, with soil fertility and rainfall being the most important factors influencing crop yields. Variability of rainfall between years has always

been high, and this applies to the total amount of rainfall as well as to its distribution. The beginning of the rainy season is extremely variable and unpredictable, whereas the end of the season tends to be slightly more stable. Both drought conditions and excess of rainfall, including flooding of fields, are commonly experienced by West African farmers. Thus, the uncertainty of climatic conditions is an important factor shaping the options available and the choices made by farmers (Haussmann et al. 2012). These factors have important consequences for setting breeding priorities and developing crop improvement strategies that favor adaptation to marginal farming systems under conditions of climate variability and change.

Decentralized, farmer participatory breeding work was initiated in the region by plant breeders from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) Regional Hub in Bamako, Mali. From the outset, diagnostic studies accompanied the practical breeding and variety evaluation activities, in order to better understand the specific constraints under which farmers in the target area work. Over the years, the work has evolved and now includes pre-breeding, breeding and variety evaluation as well as seed production activities in Mali, Niger and Burkina Faso, in cooperation with the national breeding programs of all three countries, and farmer organizations.

2.3.1.2 Collecting and Characterization of Breeding Material

For plant breeding to be successful, it is essential that suitable genetic variation exists for relevant traits and trait combinations, including sources of resistance, so that gains from selection can be achieved. By the scientists of this program, it was thus understood as a specific long-term task to conserve, evaluate and develop further the adapted diversity available in a region in order to secure sustainable breeding progress in the future. Studying the adaptation mechanisms of landraces to the highly variable climatic conditions has become an important element of research.

Some years ago, plant breeders and farmers jointly developed the *dynamic gene pool management* approach. By crossing genetically and/or geographically diverse material, new broad-based populations were created. Small seed lots of these base populations are distributed to the participating farmers in contrasting sites of various target regions. Natural as well as farmer selection lead to the formation of sub-populations, which provide excellent sources of variation for specific adaptive traits, while also ensuring that the farmers' agronomic and quality requirements are met. Through periodic recombination of selected materials, new trait combinations can occur. The sub-populations can be used for the development of contrasting varieties with differing adaptive capacities or specific traits (Haussmann et al. 2012).

This approach takes up the concept of evolutionary breeding described in Sect. 2.2; it is more difficult to apply in sorghum, which is basically a self-pollinating crop. Here, a certain proportion of male sterile lines can be mixed into the seed lots, so that crosses occur on these plants. In pearl millet, however, cross pollination occurs spontaneously.

The dynamic gene pool management approach thus offers opportunities to creatively manage and use genetic diversity for farmers and breeders alike. It is also an option for linking in situ management of plant genetic resources with ex situ conservation; ex situ conserved germplasm can periodically be grown in farmers' fields under the local climatic and management conditions in order to maintain its adaptive capacities.

2.3.1.3 Breeding Process and Outcomes

The breeding and variety evaluation follows a decentralized and participatory approach. Farmers participate in all stages of the breeding program, but the participatory variety evaluation trials are at the core of the program's activities. In participating villages, farmers can observe the growth of a set of 32 test populations, managed on 2 sites of the village by 2 farmers, and grown with 2 replications per site. This work at the village level is supported by a farmer organization, with a technical adviser being paid for by the breeding program. The evaluation workshops use participatory methods, as described by Weltzien et al. (2005) (see Fig. 2.1). Interested farmers can then grow a subset of 3–5 test populations in their own fields for further evaluation.

Fig. 2.1 Farmers in Mali evaluate a set of sorghum test cultivars (Source: E. Weltzien/ICRISAT)

Following this pattern, promising populations are being grown and tested in the target environments and subjected to key local management practices. Postharvest, processing and food quality traits are part of the evaluation. Farmers get exposed to the materials the plant breeders are working with, experience it themselves, provide feedback and select for further testing, evaluation and variety registration, as well as for further use in breeding.

The result of 10 years of developing and institutionalizing the approach is that newly developed test varieties are continuously being offered to farmers through the variety evaluation trials. Improved varieties based on local guinea landraces show modest yield superiority over common landraces, but are preferred and adopted by farmers due to other advantages, e.g. early maturity of food quality. Very promising with regard to yield are hybrids based on landraces. These show on-farm grain yield superiority of 17–37 % over a landrace check, with the three top yielding hybrids showing 30 % yield advantages across productivity levels. With absolute average yield advantages ranging from 380 kg/ha under lower to 660 kg/ha under higher productivity conditions, this yield progress could really make a difference with regard to food access and income (Rattunde et al. 2013).

2.3.1.4 Link to Seed Systems

In large parts of Africa, commercial seed markets are poorly developed, with hybrids of maize and some cash crops (e.g. cotton) being important exceptions. Hence, the majority of farmers use sorghum and pearl millet seed from their own production, or obtain it through their informal social networks. Experience from Mali shows that the local seed systems provide good quality seed of local landraces, but diffusion of improved varieties is slow (Siart 2008). Therefore, a way for seed of newly developed varieties to be made accessible to farmers needs to be considered and addressed by the breeding program.

Seed fairs, farmer-managed seed production, and marketing by emerging private seed enterprises were options explored in this project, and all show promise for improving access to new, improved varieties preferred by farmers. They need, however, financial, technical and organizational support to become effective. With private seed companies covering only part of the farmers' seed requirements, plant breeders in this project linked up with existing farmer cooperatives for the production and dissemination of seed.

The starting point is the participatory variety evaluation, through which the farmers know the range of varieties offered by the breeding program and how they may complement the existing (local) portfolio of varieties. They can decide which of these varieties they consider worthwhile for seed production and dissemination. Seed production and commercialization schemes were developed together with breeders, farmer organizations and national seed services (see vom Brocke et al. 2011, describing the process in Burkina Faso). The number of farmer organizations producing seed has expanded and seed production has flourished. In Burkina Faso,

for example, production increased from 2.55 mt produced by 10 groups in 2006, to 51.63 mt produced by 26 groups in 2009 (vom Brocke et al. 2012).

The decentralized production and marketing of certified seed through farmer organizations has evolved over the last years, with several approaches for communication and marketing being tested at various scales, including cooperation with mass media (radio) and mobile seed shops on motor cycles. A *mini-pack* approach, distributing small quantities of seed for testing purpose, has shown promise for enabling the customers to observe new varieties in their own fields with minimal risk.

2.3.1.5 Link to Value Chains

Sorghum and pearl millet are staple food crops, and farmers generally keep the harvest for their own consumption and sell the surplus as grain. Marketing takes place locally or via the same farmer organizations that are involved in the seed production. Grain prices have remained high since 2008, so that selling grain is attractive for farmers. Even though initial trials selling processed products such as flour or couscous have been made by some farmer groups, building up value chains for processed food has not been a focus of the program so far.

2.3.1.6 Outlook

The main motivation for the breeders involved in this project is developing varieties that suit farmers various needs, and thereby increasing the options farmers have to cope with variable and changing conditions. Thus, the main foci are on breeding varieties that are adapted to climate variability, poor soil fertility, particularly low phosphorus availability and with improved nutritional quality.

Previous diagnostic work revealed that the fertility status of women's fields is often particularly poor. Even though local sorghum landraces are generally adapted to low soil fertility, variability exists that can be made use of by targeted selection (Leiser et al. 2012). Moreover, women and their young children are the most vulnerable groups with regard to nutrition deficiencies in rural West Africa. This is why there is ongoing work on breeding *biofortified* varieties with higher iron content linked to other activities aiming to develop improved methods for postharvest processing and to increase nutritional knowledge.

Progress towards commercial sustainability of the seed production and marketing activities will be the other major challenge in the years to come. Towards this end, developing a strong vision will be required among the stakeholders for a seed market addressing farmer needs and farmer co-operatives as important players (Diakité et al. 2008; Kaboré et al. 2010).

2.3.2 *Decentralized Breeding and Participatory Varietal Selection for Climate Change Adaptation Using Potato*

2.3.2.1 Background and Objectives

Quechua and Aymara farmers practicing agriculture above 3800 m elevation in the high Andes are particularly hard-hit by climate change as a consequence of increasingly frequent and extreme unseasonal weather events (hail, frosts, drought) and the upward expansion of pests and diseases formerly a problem only at lower elevations, including potato blight *(Phytophthora infestans)* and potato moth *(Phthorimaea operculella)* (Giraldo et al. 2010; Kroschel et al. 2013; Perez et al. 2010). Households in the high Andes depend on the potato as a staple food and typically grow both landraces and bred varieties in different spatial compositions (De Haan and Juarez 2010).

Even though potato is the world's major non-cereal food crop (FAO 2008) and might not be considered a NUS sensu stricto, it is well known that globally it has a narrow genetic base involving predominantly one out of multiple domesticated species (Hawkes 1979; Wang 2011). The long-day adapted *Solanum tuberosum* ssp. *tuberosum* is commonly used by global breeding programs, deriving from an original gene pool of about 200 landraces of Chilean origin. However, the Andean gene pool, which includes diploid to pentaploid forms, can be considered underutilized in global crop improvement. The International Potato Center (CIP) has developed several breeding populations based on Andean species, including the *Phytophthora infestans*-resistant B1C5 population based on pure tetraploid *Solanum tuberosum* ssp. *andigena* crosses and primarily selected for yield, culinary quality and diversity in shape and skin color. These populations are used in decentralized and participatory selection with small-scale farmers in the high Andes.

2.3.2.2 Collecting and Characterization of Breeding Material

Solanum tuberosum ssp. *andigena* is extremely rich in intraspecific diversity compared with all the other potato (sub)species. Initially more than 2000 landraces were screened for *Phytophthora infestans* resistance in intentional exposure trials. As a result, 57 potato landraces with high levels of *P. infestans* resistance were selected and subsequently crossed among themselves. The best clones, based on recurrent selection for resistance, uniformity and yield, were prioritized for further crosses during five cycles (B1C0 to B1C4). Cycles B1C0 to B1C3 were selected under late blight exposure in the highlands of Ecuador, and from cycle B1C4 onwards the process was continued in Peru.

In 1999 CIP breeders made the first crosses to develop the so-called B1C5 population based on a pure *Solanum tuberosum* ssp. *andigena* pedigree. The population was initially screened and evaluated under intentional exposure trials in multiple environments so as to determine performance and stability. A diverse set of 110

advanced clones was subsequently proposed and made available for decentralized participatory selection throughout Peru. These clones maintained many of the characteristics of their original landrace parents, including strong skin colors, diverse shapes and culinary quality. Yet, they also showed enhanced traits such as *Phytophthora infestans* resistance, high yields under low-external-input use, yield stability in extreme environments, relative earliness (120 instead of 180 days), among other characteristics.

2.3.2.3 Breeding Process and Outcomes

After the formal breeding process and within 3 years after crosses were made, diverse sets of 20–30 advanced clones were distributed for decentralized participatory selection in remote highland communities, based on demand. Consortia consisting of farmer communities, municipalities, NGOs, universities and Peru's national potato program (INIA) were formed. Coordinating members of each consortium were trained in the m&b trial design adapted for potato (Fonseca et al. 2011). This methodology allows for structural, bottom-up and on-farm selection of diverse clones under recommended management (mother trials) and under prevailing farmer practices (baby trials).

Evaluations are conducted at flowering, harvest and postharvest stages. Farmer selection criteria at each of the three stages are free-listed and subsequently prioritized through weighted ranking exercises using seeds (maize kernels for women and beans for men). At flowering and harvest stage, the three preferred clones are individually ranked by each participant by depositing grains in closed containers positioned in the center of each replicate plot.

Furthermore, at harvest stage the following exercises are conducted: (i) standard evaluation of yield (weight, number of tubers), (ii) sensory evaluation (appearance, texture and taste) and (iii) sharing and discussion of results with all farmers present. Figure 2.2 shows a group of women farmers evaluating test clones for sensory quality. Participatory postharvest selection involves storage evaluation (weight loss, dormancy, sprouting behavior) and semi-structured surveys with users (consumers or processors). The adapted m&b trial design is gender sensitive, easily applied with illiterate participants, and clearly differentiates between recommended and farmer management practices. To date, four so-called novo-andigena varieties have been officially released as a result (De Haan and Bastos 2012): *Pallay Poncho* (INIA 311, 2007, Cusco region), *Puca Lliclla* (INIA 312, 2007, Cusco region), *Altiplano* (INIA 317, 2010, Puno region) and *Kawsay* (INIA 321, 2013, Huancavelica region).

2.3.2.4 Link to Seed Systems

Peru's potato seed system is predominantly informal and farmer-based. INIA's national potato program typically offers certified seed shortly after official variety releases. However, the volume of formal tubers seed is usually limited. Furthermore,

Fig. 2.2 Women farmers evaluate test clones of potatoes for sensory quality (Source: S. de Haan/CIP)

smallholders can rarely afford the high prices of tuber seed resulting from the costly in vitro >>greenhouse >>open field production pipeline. Therefore, decentralized participatory selection has generally relied on the relative strengths of informal seed systems in the Andes, such as acceptable quality, efficient farmer-to-farmer diffusion, guarantees based on trust, and accessibility to the poor and remote (Thiele 1999). Some of the consortia have promoted adapted practices within the framework of farmer seed systems, e.g. roguing and positive selection.

2.3.2.5 Link to Value Chains

Farmers from the highland communities where novo-Andigena varieties were selected have generally taken advantage of early access and local tuber seed stocks to set-up small-scale businesses that serve regional demand for seed and consumption potatoes.

2.3.2.6 Outlook

Contradictions between formal variety release procedures and decentralized participatory selection pose a real challenge to actors involved in rural development in farmer communities in centers of crop origin. Formal release procedures commonly require trials on relatively large areas that surpass small-scale farmer capacity. Furthermore, formal procedures typically only allow for the release of a single new variety while farmers participating in decentralized selection frequently prefer a combination or mixture.

Sometimes this can be resolved. For example, in the case study of a *Solanum tuberosum* ssp. *andigena* participatory selection process in southern Peru, male and female farmers had different preferences; after communication of research results to INIA's variety release committee it was decided to release two varieties (*Pallay Poncho* and *Puca Lliclla*). However, such processes require mediation.

Farmers in the communities and direct vicinity where novo-andigena cultivars were released now enjoy higher levels of food security compared with the initial situation. Initial impact studies by CIP have shown high rates of adoption. The decentralized participatory selection process has added diversity to prevailing variety portfolios and thereby increased options for farmers. Currently, CIP is expanding the application of decentralized participatory selection of Andean-type biofortified clones and the m&b trials design to poverty pockets in Bangladesh, Nepal and Bhutan.

2.3.3 Conservation Breeding: Promoting the Conservation and Use of Underutilized Andean Crops by Enhancing Yield Potential and Seed Flow

2.3.3.1 Background and Objectives

Peru is one of the world's ten mega-diverse countries, hosting a rich diversity of ecosystems, plant and animal species and crops. A major part of the population, especially that living in rural areas, depends on the use of local biodiversity and the ecosystem services it supplies, e.g. in terms of water supply and soil fertility. Many of Peru's native crops, some of which are shared with other Andean countries, have a strong significance in local culture and traditions, being used as food, or for healing or ritual purposes. Among the species originating from the Andean region are tubers, pseudo-cereals, vegetables and fruit species; some of these are today relatively underutilized for a number of reasons, including poor productivity, especially in marginal agricultural areas and the consequent tendency to abandon their cultivation (Pastor et al. 2008).

In 2010, numerous activities organized in the frame of the UN Year of Biodiversity significantly contributed to focusing attention on the country's rich agricultural heritage; opportunities and practical steps forward to improve it conservation and

sustainable use were discussed in the forum Aprovechando la Agrobiodiversidad del Perú [Making the most of Peru's Agrobiodiversity], organized jointly by the Ministries of Environment and Agriculture, INIA and Bioversity International.

Here, a breeding program targeting some of Peru's underutilized crops was discussed and implemented in four marginal mountain areas of the country (Ayacucho, Cusco, Junín and Puno). The methodological approach derived from an earlier, successful experience carried out in maize (Oscanoa et al. 2004; Sevilla 2008). The method, termed *conservation breeding* (*mejoramiento conservativo* in Spanish), combines elements of evolutionary and participatory breeding, thus laying emphasis on maximizing the use of crop biodiversity and the incorporation of local farmers' preferences and needs. Five focus crops were prioritized for an initial pilot phase of the program, based on a combined assessment of their relevance for local farming systems, their potential for strengthening local food security and livelihoods, and the existence of previous or on-going research in INIA's decentralized research stations.

The five focus crops were the pseudo-cereals quinoa (*Chenopodium quinoa*), *cañihua* (*Chenopodium pallidicaule*) and amaranth (*Amarantus caudatus*), Andean lupin or *tarwi* (*Lupinus mutabilis*) and popping varieties of common bean (locally called *frijol ñuña*). Figure 2.3 shows an amaranth field in the Andes.

Notwithstanding their increasing tendency towards being abandoned in cultivation, a number of local landraces continue to exist in the target areas. However,

Fig. 2.3 Amaranth field in the Andes (Source: S. Padulosi/Bioversity International)

intra-specific diversity available to individual farmers has been decreasing as a result of reduced planting areas and limited seed flow, thus leading to reduced yield potential and variability for important adaptive traits.

The aim of the breeding effort was to re-introduce diversity from other farmers in the area, hoping to enhance yields, enable improved adaptation to environmental variability, and hopefully identify types with improved nutritional quality or other use-related important traits. The breeding methodology was developed for all five focus crops, although early results have only been collated for quinoa and amaranth so far (Galluzzi et al. 2014).

2.3.3.2 Collecting and Characterization of Breeding Material

The first phase of the conservation breeding approach focuses on collecting germplasm of each species in the target regions in order to capture a representative sample of the existing diversity, including those alleles that occur at lower frequencies. Based on theoretical considerations, and given the predominantly self-pollinating nature of the target species, the decision was taken to collect approximately 100 seeds for each landrace in each of the target regions. The sources included farmers' fields, grain stocks and local grain markets.

Upon collection, the material received preliminary characterization and was grouped based on morphological traits, taking farmer classifications into consideration. The collections were then sown at INIA's experimental stations (one in each target region) and the standing crop again characterized in order to either confirm or re-define the groups.

Collecting, characterizing and evaluating crop genetic resources are key steps not only for carefully identifying the material for the subsequent breeding cycles, but also for establishing a baseline status for the species' on-farm conservation in a defined region, allowing for future assessments and monitoring of genetic erosion (Padulosi and Dulloo 2012).

2.3.3.3 Breeding Process and Outcomes

Equal numbers of seed (around 100) from each identified group were pooled and sown together in INIA's stations, to allow for spontaneous recombination and to obtain a composite population for each landrace. The composites for each landrace were then subjected to several more cycles of spontaneous within-composite recombination and selection. The theoretical background is that composite populations contain more diversity than any population harvested from single farmer's fields, and thus offers more potential for achieving gains from selection.

After the first recombination cycle, at the time of flowering, any damaged or diseased inflorescence was eliminated, and at harvesting time seed was selected from the most representative plants, i.e. those which best expressed the key morphological traits of their racial group, according to what was reported by farmers and

observed during collection. The harvested seed was then made available to interested farmers for further recombination and selection cycles in their fields, assisted by INIA technicians. For these recombination cycles, the composite was sown flanked by a pollinating population consisting of a mixture of seeds from the best performing 20 % of individual plants identified within the composite in the previous cycle. This procedure was repeated over several cycles.

Initial results show that, over 2 or 3 cycles, consistent yield improvements were observed in both quinoa and amaranth, for which preliminary results are available. After 2 years, the yield gains of 24 quinoa and 14 amaranth composites tested in Ayacucho were 8.75 % and 8.17 %, respectively; for the quinoa landraces Negras and Chullpi in the Puno region, yield gains of 11.3 % and 10.4 %, respectively, were observed over three cycles. These gains are measured against the average of the original composite.

Longer-term benefits with regard to yield stability are expected, but need observation and analyses over more cycles. An early characterization of the nutritional profile of the composites revealed high protein, fat and mineral (Fe, Ca) content compared with other staple crops, in line with the values reported in scientific literature for these Andean grains. To consistently enhance and stabilize nutrition-relevant traits, however, a more targeted breeding approach may be needed in the future.

2.3.3.4 Link to Seed Systems

In the course of the project, farmers were trained in seed selection, multiplication and storage practices, in order to make high quality seed of the gradually improved composites available in their communities.

This form of farmer-based multiplication of the materials derived from each selection cycle is an important contribution to strengthening informal or local seed systems and fostering diffusion of gradually improved materials, thus immediately enhancing the impact of the breeding program. Moreover, improving the seed-flow can also be a measure to maintain the circulation of diversity among farmers, and thus supporting continued in situ conservation and counterbalancing the effects of genetic drift.

Making available the seed of improved varieties beyond the immediate participants and their communities will require a broader effort in the future, and is indeed one of the major technical and organizational bottlenecks in most participatory breeding programs. As in most other countries, formal requirements exist in Peru for variety registration and seed certification. However, under the *common-seed* category, seed can be distributed with minimal registration requirements (as long as certain basic quality standards are fulfilled), providing a potential opening for dissemination of PPB-derived varieties. Were the breeding effort to be expanded further, possibly more appropriate schemes for registration, certification and distribution of local genetic materials would need to be explored, taking issues such as benefit sharing and farmers' rights into account.

2.3.3.5 Link to Value Chains

In Peru, marketing options and value chains exist for quinoa and amaranth, both in local as well as (to some degree) in international markets. The focus of the conservation breeding project for the moment has been to improve the performance and competitiveness of the Andean grains at local levels, so that they remain an integral part of local farming and food systems, and their genetic diversity is preserved through sustainable use. However, the collection of the on-farm diversity and its characterization offer important opportunities to identify interesting types for agrobiodiversity-based value chain development (Kruijssen et al. 2009).

2.3.3.6 Outlook

Conservation breeding based on a combination of evolutionary and participatory breeding practices is an interesting approach to rescuing genetic resources and promoting the value of underutilized crops. It aims at reconciling agrobiodiversity conservation and variety improvement, with farmers playing a key role in the process.

Technical issues still need to be improved. Experts have observed a higher than expected out-crossing rate in the target species described here, which are considered predominantly inbreeding; more systematic determination of the extent of out-crossing in the target species will be important, since the out-crossing rate determines the effectiveness of recombinations. This assessment could be greatly aided by the application of molecular markers. The introduction of controlled crosses, including training farmers to carry them out, could also be explored, as an option for further enhancing recombination among desired types.

In addition to continuing the effort in the target crops described here, it would be interesting to expand the conservation breeding approach to other underutilized crops, by developing new protocols and exploring additional diversity hotspots in the country, possibly linking the breeding effort with market and value chain development, as additional incentives for farmers.

2.3.4 Kultursaat: Establishing Innovative Institutional Structures and Partnerships to Support Farmer Breeding and Add Value to Vegetable Crops

2.3.4.1 Background and Objectives

Kultursaat is a farmer-driven not-for-profit organization based in Germany. It was founded in 1994 out of an existing informal group of biodynamic farmers and gardeners. Their joint concerns were that the varieties available for commercial vegetable production, increasingly hybrids, were not well adapted to the conditions of

organic farming systems, and were also found to be inferior in sensory quality compared with open-pollinating varieties formerly used.

These organic farmers had started experimenting with developing breeding and selection methods for various vegetable crops on their farms, and exchanged experiences at regular meetings. However, in order to be able to work more effectively towards maintaining and breeding open-pollinating varieties for organic vegetable farming, they felt the need to create a legal entity to assume responsibility for administrative work and fundraising, so that individual farmers could concentrate on breeding, besides the normal work on their farms.

It is a constitutive element of Kultursaat that the associated breeders receive funds from the organization to support their breeding work. In return, they commit themselves to transfer their property rights to Kultursaat. By doing so, it is made sure that the varieties developed serve the public good and are held in trust by Kultursaat. The organization does not particularly focus on NUS, but aims to offer varieties of a range of vegetable crops to commercial organic vegetable producers. Hence, some less common species have been included in order to tap existing market potential for niche crops.

One relatively well documented example is the parsnip project of Kultursaat. Accessions from various sources had been assembled and preliminarily screened by students of a vocational school and several farmer breeders over 6 years. Then the organization successfully raised funds from a government scheme to promote research and knowledge of organic farming for a project aiming to establish an assortment of improved parsnip varieties and to develop further their breeding and selection methodologies. This project was implemented between 2004 and 2006 (Bauer et al. 2006). Some activities were continued and further developed between 2006 and 2008 (Horneburg et al. 2009).

Parsnip (*Pastinaca sativa*) is a robust root vegetable suitable for storage that has been traditionally consumed during the winter season in many parts of Europe. Wild forms of the same species also play a role as fodder for livestock and wild animals. Since the eighteenth century, it has increasingly been replaced by potatoes and carrots. However, it has a high content of micronutrients, particularly minerals, e.g. potassium, calcium, magnesium and zinc, and vitamins such as vitamins K, C, E and folate (USDA 2014). It further contains essential oils that are said to improve immune defense, and the nitrate content is generally low. Due to its sweet taste and soft consistency, it is sometimes used as baby food, and in other respects as soup and salad vegetable, in cooked and raw form. However, only few consumers nowadays continue to have knowledge of the crop and its use. In Germany, it is found mainly in organic food shops and at vegetable markets.

Many varieties have become heterogeneous, and seed of only one open-pollinating variety (Halblange Weiße) is widely available in Germany. Similar to other varieties, it shows considerable variability with regard to sensory quality. Single roots develop acerbic taste and a strong smell while cooking, which is perceived as unpleasant by some people. This was found to be an important constraint on wider use (Bauer et al. 2006). Parsnip is a biennial plant that is harvested for

consumption after the first year, but requires storing and replanting of roots to harvest seed in the second year.

2.3.4.2 Collecting and Characterization of Breeding Material

Assembling the assortment of varieties and different proveniences was done over several years and by several people. Some accessions were obtained from gene banks and botanical associations in various European countries, and from other breeders. The vast majority of these are no longer in commercial vegetable production. The project could build on screening and initial breeding activities conducted previously at four different locations. In the framework of the project, 71 provenances of 21 varieties were screened and described systematically in the years 2004–2006.

2.3.4.3 Breeding Process and Outcomes

Ten promising accessions were grown at three locations in the first and fifth locations in the second and third years, with three replications each. These locations are the farms of different farmer breeders associated with Kultursaat. The breeding activities are generally embedded in the commercial organic vegetable production of the farms. This means, for example, that fields are subjected to the usual crop rotation pattern, and the same management measures.

Field observations were taken for assessing plant stand, vigor, disease incidence and leaf morphology. At harvesting, the farmer breeders harvested all plants of a plot and described them with regard to the shape of the roots and the characteristics of the peel. Then all under- and oversized as well as blemished plants were removed and leaves from remaining parsnip roots cut to determine the marketable yield. The test for sensory quality included the traits of sweetness, aroma, consistency, color and smell. Developing and improving the methodology for the sensory quality test and some other methods was a major focus of the project, and was described in detail by Bauer et al. (2006). Chemical assessment was done based on sub-samples at the University of Kassel (see Fig. 2.4 for sampling method); these analyses included dry matter and sugar contents, as well as nitrate and mineral contents. After the storage period, a final test for sensory quality was again applied before transplanting the selected roots to produce seed. The details of the breeding process varied across farms; in general, selected roots were grown in isolation and seed from these individual plants again sown and screened for the above-described traits. In the last year, the best individuals of a variety were planted in groups to produce seed (Bauer et al. 2006).

The results suggest that targeted selection can enhance sensory quality in many cases. Yield and trait expression varied across locations and years. However, selections of five varieties were identified that combined good scores for sensory quality

Fig. 2.4 Sampling parsnip roots for the sensory quality test (Source: B. Horneburg)

and acceptable yield levels across locations. Several other selections showed potential for specific purposes (Bauer et al. 2006).

2.3.4.4 Link to Seed Systems

For seed production and marketing, an independent seed company was established: the Bingenheimer Saatgut AG. This company contracts organic farmers to multiply seed of varieties bred by Kultursaat breeders, and of some other open-pollinating varieties that are of interest for commercial organic vegetable farmers. The company is owned by shareholders, among them many of the Kultursaat breeders and the seed producers (Kultursaat 2014). Currently, seed is available of three selected and improved parsnip varieties, and can be ordered from an online shop: Aromata, Halblange Weiße and White Gem (Bingenheimer Saatgut AG 2014).

2.3.4.5 Link to Value Chains

The breeding activities were mainly linked to the farmers' own marketing activities via farm stores and vegetable markets. Linking up with other market partners was not the focus of the project, but the documentation of quality parameters is regarded as an important basis for establishing such links in the future. The improved parsnip varieties offer new opportunities for use as processed baby food, to serve individual dietary needs and more generally for marketing as health food (Bauer et al. 2006).

2.3.4.6 Outlook

In the past 20 years Kultursaat and its partners have successfully established an institutional structure for farmer-led vegetable breeding, including for variety registration, seed production and commercialization, which corresponds to the legal framework for these activities in Germany and other European countries. Kultursaat is successful in coordinating the breeding activities at several locations, and in raising funds for breeding and related costs. The organization could thus serve as a model for developing new partnerships that support farmer breeding, including NUS.

Kultursaat does not claim royalty fees but strives to develop innovative funding instruments that aim at increasing awareness that not only farmers benefit from breeding, but also all other market partners along the seed and food value chains. The FAIR-BREEDING® partnership, for example, is based on voluntary contributions from some organic food traders. They have committed themselves to donate 0.3 % of their annual turnover from fruits and vegetables to Kultursaat for a period of 10 years without claiming any rights. With the associated seed enterprise and two others in Switzerland and Austria, a contribution to variety development, maintenance and official fees was agreed upon under private law. Kultursaat further receives funds from individuals, private foundations and government schemes.

The spread of the varieties, however, is still limited due to the lower yield level of open-pollinating varieties compared with the hybrids commonly used. As vegetables are generally paid for by weight, lower yields reduce the income of farmers. Even though consumers in Germany are used to paying a premium price for organically produced vegetables, it is not common to pay more for certain varieties. Marketing vegetables under a variety denomination is not usual, so that consumer awareness is generally low concerning the properties of different varieties. With the sweet and aromatic carrot variety Rodelika, the flagship of Kultursaat, new marketing approaches are being explored, and other crops and varieties could follow (Fleck 2013).

2.4 Discussion and Further Research Needs

The case studies show how breeding strategies for NUS are contextualized and develop different foci and priorities accordingly. In two case studies, adaptation to climate variability and change, or harsh climatic conditions in general, were important objectives (Sects. 3.1 and 3.2, above). In the other cases, broadening the genetic base and improving agronomic and nutrition-related traits (Sect. 3.3), or improving sensory quality to address obstacles relating to broader utilization (Sect. 3.4), were the main targets. However, the case studies show that even though main objectives varied between case studies, the range of traits evaluated was broader than in most formal breeding programs. Food quality was, for example, an important aspect also in case studies that did not particularly focus on it.

In all case studies, assembling and screening genetic resources, conserved either in situ or ex situ, provided the basis for the subsequent breeding activities, and local varieties of different provenances, varieties that were neglected previously or landraces, played an important role in providing a starting point for successful breeding and selection.

The breeding progress documented in all four case studies was achieved within a relatively short time, and was based on simple mass selection methods, mostly for traits that are easily observed in the field, or that do not require sophisticated equipment for assessing them (as in the case of the sensory quality test, Sect. 3.4 above). This finding suggests that breeding progress may in many cases depend on effective contextualization, decentralized selection and collaboration, rather than on *high-end* breeding technology – factors that appear to be undervalued in many formal breeding programs.

However, the case study of sorghum and pearl millet breeding in West Africa (Sect. 3.1) demonstrates how technical knowledge provided by a large formal breeding program can lead to considerable progress, if strategically oriented towards the main constraints faced by farmers, and backed up with capacity building and institutional structures supporting collaboration in the longer term. In this case, advanced breeding technologies, such as hybrid breeding and use of male sterile lines, were combined with the participatory approach. Biotechnological breeding techniques such as marker-assisted selection and genomic breeding could also play an increasing role for enhancing NUS in the future. A project that has shown early success in this field of research included marker-assisted selection for resistance to the parasitic weed striga in pearl millet (Haussmann et al. 2004). Marker-assisted selection can to a large extent simplify and speed up the selection work of breeders, yet to date it is only feasible with a small number of relevant traits, and for few crops.

Applying biotechnical breeding techniques that were originally developed for major crops for the genetic enhancement of NUS was proposed for example by the International Atomic Energy Agency (IAEA 2004) and Ochatt and Jain (2007). The role of biotechnology for NUS has been assessed in a study by Dawson and Jaenicke (2006). They conclude that even though some cases exist where application of biotechnology has proven relevant, a lack of integrated thinking limits the applicability of results on a larger scale. Moreover, high investments may be required initially, e.g. for equipment and training. The authors identify five key issues that should be addressed in order to make biotechnology beneficial for NUS: (1) integrating biotechnology within the wider context of species promotion, (2) identifying clearly how biotechnology application relates to the needs of (poor) farmers, (3) providing suitable indicators for impact, (4) giving due consideration to the role of partnerships in promoting NUS, and (5) giving proper consideration to property rights.

The case studies further show that where formal breeding programs do pay attention to biodiversity and NUS, they tend to concentrate on less underutilized species that are regionally important or show considerable marketing potential to be worth the effort. In spite of the innovation potential arising from effective collaboration between farmers and researchers, it should not be ignored that there will never be enough plant breeders or formal breeding programs to take care of the diversity of

food crops and varieties rural communities use and need for achieving food and nutrition security (Chaudhary and Sthapit 2013). Therefore other, complementary, approaches and institutions are also needed to take on responsibility for agrobiodiversity management, breeding activities and seed diffusion, and in which farmers and farming communities play a leading role.

The evolutionary plant breeding approach offers potential for crop improvement in cases where the involvement of formal institutions is weak or resources are limited; once assembled, evolutionary populations can be handled quite easily by farmers alone. It thus makes farmers less dependent on institutions, while still allowing for some breeding progress for relevant traits. It could also be a path to follow after more intensive collaborative breeding projects on NUS come to an end, as suggested by Ceccarelli et al. (2013b).

Strengthening grass-root institutions in a coordinated way could be a further option for leveraging the potentials of NUS in the longer term. This approach is being pursued by the Community Biodiversity Management (CBM) Program in South Asia, coordinated by LI-BIRD, an NGO based in Nepal. The CBM approach is being implemented at 29 sites across four countries: Nepal, India, Sri Lanka and Bangladesh. Here, strengthening community capacities and local institutions relating to agrobiodiversity management is the main concern. Consequently, communities develop their own priorities relating to agrobiodiversity, crop improvement and seed diffusion, generate funds and monitor and evaluate the outcomes of their activities. These activities include seed exchange, home gardens, participatory plant breeding, value addition through postharvest processing (e.g. juice from a semi-wild plant with high vitamin C content) and diversity fairs or blocks (Development Fund 2012). Approaches to and practical examples of CBM were further compiled by de Boef et al. (2013). One example presented refers to value chain development and regional branding of Kalajeera, a local rice variety of the Jeypore region, India: an approach that relies, inter alia, on community-based participatory breeding and seed production (Chaudhary and Swain 2013).

Linking breeding of NUS to markets and value chain development was less focused on in the case studies presented, with Sect. 3.4 being an exception, where farmer-breeders implicitly addressed market requirements and consumer preferences that were already known to them. However, much potential could still arise from tying breeding activities to larger approaches for raising the profile and marketing potential of NUS, as demonstrated in the case study of the *Capsicum* project, funded by the German Federal Ministry for Economic Cooperation and Development (BMZ) via Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), coordinated by Bioversity International (Jäger et al. in prep.).

Methodologies and good practices for building value chains for NUS were summarized by Will (2006). Building clear visions among stakeholders of what should be achieved (for whom), and which leverage points need to be addressed by which actions, are crucial for success, and could by tied to strategic work on breeding objectives for NUS in a collaborative approach. The proposed steps for value chain development include selection of NUS that merit wider promotion, mapping and analyzing value chains, assessing opportunities, possible interventions and entry

points, and developing and implementing a strategy for upgrading (Will 2006). A holistic value chain framework – covering interventions starting from collecting, conserving and characterizing underutilized genetic resources, breeding, developing good agricultural, postharvest and processing practices to final consumption, nutrition awareness campaigns, new product development and creation of an enabling policy environment – was recently developed and tested in Bolivia, Peru and India in the framework of an IFAD-funded project IFAD NUS, coordinated by Bioversity International (Padulosi et al. 2014).

An issue mentioned in all our case studies and of crucial importance for breeding strategies that aim at promoting NUS are variety release and seed regulations. In his report to the United Nations General Assembly, the UN Special Rapporteur on the Right to Food, Olivier de Schutter, highlighted the importance of informal seed systems for food security of vulnerable groups. He emphasized that informal seed systems will continue to be marginalized if states do not take targeted measures to strengthen them. A focus on protection of intellectual property rights of plant breeders and commercial enterprises can impede, rather than enhance innovation in variety development and seed provision (De Schutter 2009). Bentley et al. (2011) point in the same direction, stating that *the formal seed system is actually a marriage between the government and the private sector*. Thus, the needs, contributions and rights of farmers and local communities have so far seldom been considered in developing a political vision and designing the regulatory frameworks for variety release and seed diffusion.

In most countries, the traditional informal institutions for creating, maintaining and developing agrobiodiversity have been weakened, and the commercial sector covers only a part of these activities. At the same time, new players may enter the scene, such as farmer organizations, farmer-consumer associations or other NGOs concerned about promoting and maintaining agrobiodiversity. In the future, innovative institutional arrangements could be based on aligning these emerging forces. The Kultursaat case study (Sect. 3.4) shows, for example, how farmer breeding can contribute to developing new varieties for purposes that are neglected by the formal sector in a highly regulated context, by building innovative partnerships based on mutual trust with seed companies and traders. New forms of partnership between farmers and scientists also formed the basis of a participatory maize breeding project in China that resulted not only in improved maize production, but also in organizational and policy change with regard to agrobiodiversity conservation, plant breeding and rural extension (Yiching and Vernooy 1999).

In view of international conventions and treaties, namely the Convention on Biological Diversity (CBD), the International Treaty for Plant Genetic Resources for Food and Agriculture (ITPGRFA), and the Right to Food, governments have a primary responsibility to work towards the progressive realization of the goals agreed upon. On the one hand, they should provide sources of funding and supporting legal frameworks for breeding and seed production initiatives that serve these goals. On the other hand, existing policies should be revised and obstacles removed. For example, variety registration rules and seed legislations too often focus only on the commercial breeding and seed sector, and ignore that other, complementary

structures are also needed. Innovative initiatives that are based on agrobiodiversity, NUS and decentralized seed production, and that serve societal goals and common goods, such as food and nutrition security or ecosystem resilience, should be encouraged, rather than restricted, so that they can effectively complement the existing set of formal institutions and activities focusing on major crops.

2.5 Conclusions and Prospects

In order to design breeding strategies for NUS, the multiple functions many NUS fulfill in their specific socio-ecological and economic context need to be taken into consideration, and require careful diagnostic studies from the outset. Not any particular breeding technology, but effective targeting towards the needs of farmers and their market partners, and selecting among methodological and technology options accordingly, appears to be a key success factor. This means that breeding programs need to be decentralized and require, depending on the objectives, complementary capacity and multi-disciplinary skills from social scientists, economists, nutrition or food processing practitioners, traders and farmers. Multi-stakeholder approaches, focusing on collaborative learning, innovation and institutional development, should be envisioned. In the longer term, centering breeding strategies occasionally on a single NUS will not be sufficient. What will be required are comprehensive strategies and innovative institutional arrangements that deal with agrobiodiversity, farming system resilience, income generation and food and nutrition security in an inclusive and holistic way. In order to achieve this, the attention of plant breeders will need to increasingly expand from species and varieties towards people and their needs.

Acknowledgements We would like to express our gratitude for Dr Salvatore Ceccarelli's helpful comments on the manuscript. We further thank Dr Eva Weltzien (ICRISAT), as well as Dietrich Bauer and Michael Fleck (Kultursaat e.V.), for their willingness to share information and for clarifying questions regarding the case study examples. Last but not least, Thor Lawrence's careful editing work is gratefully acknowledged.

References

Annichiarico P (2009) Coping with and exploiting genotype-by-environment interactions. In: Ceccarelli S, Guimarães EP, Weltzien E (eds) Plant breeding and farmer participation. FAO, Rome, pp 519–564

Ashby JA, Lilja N (2004) Participatory research: does it work? Evidence from participatory plant breeding. In: New directions for a diverse planet. Proceedings of the 4th international crop science congress, Brisbane, Australia, 26 Sept–1 Oct 2004

Badstue LB, Hellin J, Berthaud J (2012) Re-orienting participatory plant breeding for wider impact. Afr J Agr Res 7(4):523–533

Barahona C, Levy S (2002) How to generate statistics and influence policy using participatory methods in research. Statistical Services Centre working paper. University of Reading, Reading

Bauer D, Kern M, Horneburg B, Fleck M (2006) Erstellung, Erweiterung und qualitative Verbesserung des Pastinakensortiments für den Ökologischen Anbau unter besonderer Berücksichtigung des Geschmacks; Weiterentwicklung der entsprechenden Zuchtmethoden. Final report. http://orgprints.org/15268/1/15268-03OE481-kultursaat-fleck-2006-pastinake.pdf

Bellon MR, Reeves J (eds) (2002) Quantitative analysis of data from participatory methods in plant breeding. CIMMYT, Mexico

Bentley JW, van Mele P, Reece JD (2011) How seed works. In: van Mele P, Bentley JF, Guéi RG (eds) African seed enterprises. Sowing the seeds of food security. CAB International, Wallingford, pp 8–24

Bingenheimer Saatgut AG (2014) Online catalogue. http://www.bingenheimersaatgut.de/de/Gemuese/Wurzel-Knollengemuese/Pastinaken

Burlingame B, Dernini S (eds) (2012) Sustainable diets and biodiversity, directions and solutions for policy, research and action. In: Proceedings of the international scientific symposium biodiversity and sustainable diets united against hunger, FAO Headquarters, Rome. 3–5 Nov 2010. FAO, Rome

Ceccarelli S (2009) Evolution, plant breeding and biodiversity. J Agr Envir Int Dev 103(1/2):131–145

Ceccarelli S (2012) Plant breeding with farmers – a technical manual. ICARDA, Aleppo

Ceccarelli S (2014) Drought. In: Jackson M, Ford-Lloyd B, Parry M (eds) Plant genetic resources and climate change. CAB International, Wallingford, pp 221–235

Ceccarelli S, Grando S, Booth RH (1996) International breeding programmes and resource-poor farmers: crop improvement in difficult environments. In: Eyzaguirre P, Iwanaga M (eds) Participatory plant breeding. Proceedings of a workshop on participatory plant breeding, Wageningen, 26–29 Jul 1995. IPGRI, Rome, pp 99–116

Ceccarelli S, Grando S, Tutwiler J et al (2000) A methodological study on participatory barely breeding. 1. Selection phase. Euphytica 111:91–104

Ceccarelli S, Galie A, Grando S (2013a) Participatory breeding for climate change-related traits. In: Cole (ed) Genomics and breeding for climate-resilient crops, Vol. 1. Springer, Berlin/Heidelberg, pp 331–376

Ceccarelli S, Grando S, Winge T (2013b) Participatory barley breeding in Syria. In: Andersen R, Winge T (eds) Realising farmers' rights to crop genetic resources: success stories and best practices. Routledge, London/New York, pp 97–116

Chapman SC, Hammer GL, Podlich DW, Cooper M (2002) Linking bio-physical and genetic models to integrate physiology, molecular biology and plant breeding. In: Kang MS (ed) Quantitative genetics, genomics and plant breeding. CABI, Wallingford, pp 167–187

Chapman S, Cooper M, Podlich D, Hammer G (2003) Evaluating plant breeding strategies by simulating gene action and dryland environment effects. Agron J 95:99–113

Chaudhary P, Sthapit B (2013) Agricultural biodiversity conservation and food security for small farmers. In: Christinck A, Padmanabhan M (eds) Cultivate diversity, A handbook on transdisciplinary approaches to agrobiodiversity research. Margraf Publishers, Weikersheim, pp 53–69

Chaudhary SS, Swain S (2013) Value chain development and the regional branding of Kalajeera rice in Jeypore, India. In: De Boef WS, Subedi A, Peroni N et al (eds) Community biodiversity management. Promoting resilience and the conservation of plant genetic resources, Issues in agricultural biodiversity series. Routledge, London, pp 201–206

Christinck A, Weltzien E, Hoffmann V (eds) (2005) Setting breeding objectives and developing seed systems with farmers, A handbook for practical use in participatory plant breeding projects. Margraf Publishers, Weikersheim

CIP-UPWARD (2003) Conservation and sustainable use of agricultural biodiversity: a sourcebook. Three volumes. International Potato Center (CIP)/Users Perspectives With Agricultural Research and Development (UPWARD), Los Baños

Conroy C, Witcombe JR, Virk DS et al (2009) New upland rice varieties for India. Rainfed Agriculture Impact Study No. 1. Monitoring impact assessment and learning component (MIL) 2.2 of the Research Into Use Programme (RIU). DFID, London

Cooper M, Stucker RE, De Lacy IH, Harch BD (1996) Wheat breeding nurseries, target environments, and indirect selection for grain yield. Crop Sci 37:1168–1176

Dawson IK, Jaenicke H (2006) Underutilised plant species: the role of biotechnology. Position Paper No. 1. International Centre for Underutilised Crops (ICUC), Colombo

Dawson JC, Murphy KM, Jones SS (2008) Decentralized selection and participatory approaches in plant breeding for low-input systems. Euphytica 160:143–154

De Boef WS, Subedi A, Peroni N et al (eds) (2013) Community biodiversity management. Promoting resilience and the conservation of plant genetic resources, Issues in agricultural biodiversity series. Routledge, London

De Haan S, Bastos C (2012) Catálogo de nuevas variedades de papa: sabores y colores para el gusto peruano. Instituto Nacional de Innovación Agraria (INIA), Red LatinPapa, Lima, Peru

De Haan S, Juarez H (2010) Land use and potato genetic resources in Huancavelica, central Peru. J Land Use Sci 5(3):179–195

De Schutter O (2009) Seed policies and the right to food: enhancing agrobiodiversity and encouraging innovation. Report of the Special Rapporteur on the Right to Food to the General Assembly of the United Nations, 23 Jul 2009. UNO, New York

Development Fund (2012) Community-based biodiversity management – South Asia Programme (CBM-SA), evaluation report. Norwegian Development Fund, Oslo

Diakité L, Sidibé A, Smale M, Grum M (2008) Seed value chains for sorghum and millet in Mali, a state-based system in transition. IFPRI discussion paper 00749. IFPRI, Washington DC

Döring T, Knapp S, Kovacs G et al (2011) Evolutionary plant breeding in cereals – into a new era. Sustainability 3(10):1944–1971

Fanzo J, Mattei F (2012) Ensuring agriculture, biodiversity and nutrition remains central to addressing the MDG hunger target. In: Burlingame B, Dernini S (eds) Sustainable diets and biodiversity. FAO, Rome, pp 44–53

FAO (1992) Forests, trees and food. Forestry Department, FAO, Rome. http://www.fao.org/docrep/006/u5620e/U5620E03.htm

FAO (1997) The state of the world's plant genetic resources for food and agriculture. FAO, Rome. http://apps3.fao.org/wiews/docs/swrfull.pdf

FAO (2008) New light on a hidden treasure: an end-of-year review. International year of the potato 2008. FAO, Rome. http://www.fao.org/potato-2008/pdf/IYPbook-en.pdf

Fleck M (2013) Kultursaat – frei finanzierte gärtnerische Züchtung. Ein Konzept wird 20 Jahre alt. Lebendige Erde 2013(4):12–14

Fonseca C, De Haan S, Miethbauer T et al (2011) Selección participativa de variedades de papa en Peru. In: Thiele G, Quirós CA, Ashby J et al (eds) Métodos participativos para la inclusión de los pequeños productores rurales en la innovación agropecuaria. Alianza Cambio Andino, Lima, Peru, pp 169–184 http://cipotato.org/publications/pdf/005888.pdf

Galluzzi G, Estrada R, Apaza V et al (2014) Participatory breeding in the Peruvian highlands: opportunities and challenges for promoting conservation and sustainable use of underutilized crops. Renew Agr Food Syst [online]. doi:10.1017/S1742170514000179

Gari JA (2004) Plant diversity, sustainable rural livelihoods and the HIV/Aids crisis. FAO/UNDP, Rome/Bangkok

Giraldo D, Juarez H, Pérez W et al (2010) Severity of the potato late blight (*Phytophthora infestans*) in agricultural areas of Peru associated with climate change. Rev Peruana Geo-Atmos (RPGA) 2:56–67

Haussmann BIG, Hess DE, Omanya GO et al (2004) Genomic regions influencing resistance to the parasitic weed *Striga hermonthica* in two recombinant inbred populations of sorghum. Theor Appl Genet 109(5):1005–1016

Haussmann BIG, Rattunde HF, Weltzien-Rattunde E et al (2012) Breeding strategies for adaptation of pearl millet and sorghum to climate variability and change. J Agr Crop Sci 198(5):327–339

Hawkes JG (1979) Genetic poverty of the potato in Europe. In: Zeven AC, van Harten AM (eds) Proceedings of the conference on broadening the genetic base of crops, Wageningen, 3–7 Jul 1978. Centre for Agricultural Publishing and Documentation, Wageningen, pp 19–27

Hoffmann V, Probst K, Christinck A (2007) Farmers and researchers: how can collaborative advantages be created in participatory research and technology development? Agr Human Val 24:355–368

Horneburg B, Bauer D, Bufler G (2009) Züchterische Verbesserung der sensorischen Qualität der Pastinake (*Pastinaca sativa* L.) im Praxisbetrieb. E-proceedings of the 10th scientific conference on organic farming, Zurich, Switzerland, 11–13 Feb 2009. http://orgprints.org/14145/1/Horneburg_14145.pdf

Humphries S, Gallardo O, Jimenez J, Sierra F (2005) Linking small farmers to the formal research sector. Lessons from a participatory bean breeding program in Honduras. AgREN network paper 142. DFID, London

IAEA (2004) Genetic improvement of neglected and under-utilized crops in low income food deficit countries through irradiation and related techniques. In: Proceedings of a final research coordination meeting organized by the joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture and held in Pretoria, South Africa, 19–23 May 2003 (IAEA-TECDOC-1426). IAEA, Vienna, Austria

Jaenicke H, Hoeschle-Zeledon I (eds) (2006) Strategic framework for underutilized plant species research and development, with special reference to Asia and the Pacific, and to Sub-Saharan Africa. ICUC/GFU, Bogor/Rome

Jaenicke H, Pasiecznik N (2009) Making the most of underutilized crops. LEISA Mag 25(1):11–12

Jäger M, Amaya K, Oros R et al (in prep.) Linking genebanks and small farmers to high-value markets through collective action: the case of chili peppers in Bolivia and Peru. Submitted for publication to J Dev Stud

Kaboré R, Dabat MH, Brocke K (2010) Coordination et durabilité des nouvelles formes de production semencière vivrière au Burkina Faso. In: Seiny-Boukar L, Boumar P (eds) Actes du colloque Savanes africaines en développement: innover pour durer. Garoua, Cameroun, 20–23 April 2009. CIRAD, Montpellier

Kahane R, Hodgkin T, Jaenicke H et al (2013) Agrobiodiversity for food security, health and income. Agron Sust Dev 33(4):671–693

Kaufmann B, Arpke H, Christinck A (2013) From assessing knowledge to joint learning. In: Christinck A, Padmanabhan M (eds) Cultivate diversity! A handbook on transdisciplinary approaches to agrobiodiversity research. Margraf Publishers, Weikersheim, pp 115–141

Keding G, Schneider K, Jordan I (2013) Production and processing of foods as core aspects of nutrition-sensitive agriculture and sustainable diets. Food Sec 5(6):825–846

Khoury CK, Bjorkman AD, Dempewolf H et al (2014) Increasing homogeneity in global food supplies and implications for food security. Proc Nat Acad Sci USA 111(11):4001–4006

Kroschel J, Sporleder M, Tonnang HEZ et al (2013) Predicting climate-change-caused changes in global temperature on potato tuber moth *p* (Zeller) distribution and abundance using phenology modeling and GIS mapping. Agr For Meteor 170:228–241

Kruijssen F, Giuliani A, Sudha M (2009) Marketing underutilized crops to sustain agrobiodiversity and improve livelihoods. Acta Hort 806(2):415–422

Kuhnlein HV (2012) Biodiversity and sustainability of indigenous peoples' foods and diets. In: Burlingame B, Dernini S (eds) Sustainable diets and biodiversity. Directions and solutions for policy, research and action. Proceedings of the international scientific symposium biodiversity and sustainable diets united against hunger, FAO Headquarters, Rome, 3–5 Nov 2010. FAO, Rome, pp 118–124

Kultursaat (2014) www.kultursaat.org

Leiser WL, Rattunde HFW, Piepho HP et al (2012) Selection strategy for sorghum targeting phosphorus-limited environments in West Africa: analysis of multi-environment experiments. Crop Sci 52(6):2517–2527

MSSRF (2005) UN millennium development goals – five years later. Agricultural biodiversity and the elimination of hunger and poverty: the Chennai platform for action. MSSRF, Chennai, India. http://www.mssrf.org/bd/ebooks/ebook7.pdf

Murphy K, Lammer D, Lyon S et al (2005) Breeding for organic and low-input farming systems: an evolutionary-participatory breeding method for inbred cereal grains. Renew Agr Food Syst 20:48–55

Mustafa Y, Grando S, Ceccarelli S (2006) Assessing the benefits and costs of participatory and conventional barley breeding programs in Syria. ICARDA, Aleppo

Ochatt S, Jain SM (2007) Breeding of neglected and under-utilized crops, spices and herbs. Science Publishers, Enfield

Oscanoa C, Vilchez J, Narro T, Sevilla R (2004) Participatory breeding and decentralized seed production in maize in the Central Highlands of Peru. CIMMYT, El Batán

Padulosi S, Dulloo EM (2012) Towards a viable system for monitoring agrobiodiversity on-farm: a proposed new approach for Red Listing of cultivated species. In: Padulosi S, Bergamini N, Lawrence T (eds) Proceedings of the international conference on-farm conservation of neglected and underutilized species: status, trends and novel approaches to cope with climate change, Frankfurt, 14–16 June 2011. Bioversity International, Rome, pp 171–199

Padulosi S, Thompson J, Rudebjer P (2013) Fighting poverty, hunger and malnutrition with neglected and underutilized species (NUS): needs, challenges and the way forward. Bioversity International, Rome

Padulosi S, Amaya K, Jäger M et al (2014) A holistic approach to enhance the use of neglected and underutilized species: the case of Andean grains in Bolivia and Peru. Sustainability 6(3):1283–1312

Pastor S, Fuentealba B, Ruiz M (2008) Cultivos subutilizados en el Perú. Análisis de las políticas públicas relativas a su conservación y uso sostenible. Sociedad Peruana de Derecho Ambiental (SPDA), Lima

Perez C, Nicklin C, Dangles O et al (2010) Climate change in the high Andes: implications and adaptation strategies for small scale farmers. Int J Env Cult Econ Soc Sustain 6:71–88

Rattunde HFW, Weltzien E, Diallo B et al (2013) Yield of photoperiod-sensitive sorghum hybrids based on guinea-race germplasm under farmers' field conditions in Mali. Crop Sci 53:2454–2461

Sevilla R (2008) Mejoramiento poblacional del maíz en la Sierra. In: Proceedings of the international symposium investigación hacia un desarrollo sostenible, La Molina University, Lima, 16–17 Sept 2008

Siart S (2008) Strengthening local seed systems: options for enhancing diffusion of varietal diversity of sorghum in southern Mali. Margraf Publishers, Weikersheim

Simmonds NW (1984) Decentralized selection. Sugar Cane 6:8–10

Snapp S (2002) Quantifying farmer evaluation of technologies: the mother and baby trial design. In: Bellon MR, Reeves J (eds) Quantitative analysis of data from participatory methods in plant breeding. CIMMYT, Mexico, pp 8–16

Sthapit BR, Shrestha P, Upadhyay MP (eds) (2006) On-farm management of agricultural biodiversity in Nepal: good practices. NARC, LI-BIRD and Bioversity International, Pokhara

Suneson CA (1956) An evolutionary plant breeding method. Agron J 48:188–191

Thiele G (1999) Informal seed systems in the Andes: why are they important and what should we do with them? World Dev 27(1):83–99

Thies E (2000) Promising and underutilized species, crops and breeds. GTZ, Eschborn

Tumwegamire S (2011) Genetic variation, diversity and genotype by environment interactions of nutritional quality traits in East African sweet potato. PhD thesis, Makerere University, Kampala

USDA (2014) Nutrient content of parsnip, raw. National nutrient database for standard reference, Release 26. United States Department of Agriculture. http://ndb.nal.usda.gov/ndb/foods/show/3093

vom Brocke K, Trouche G, Hocdé H, Bonzi N (2011) Sélection variétal au Burkina Faso: un nouveau partenariat entre chercheurs et agriculteurs. Grain de Sel 52-53:20–21

vom Brocke K, Zougmoré R, Barro C et al (2012) Using indigenous knowledge and participatory plant breeding to improve crop production and regenerate crops and tree species in West Africa. In: Padulosi S, Bergamini N, Lawrence T (eds) On-farm conservation of neglected and underutilized species: status, trends and novel approaches to cope with climate change. Proceedings of an international conference, Frankfurt, 14–16 Jun 2011. Bioversity International, Rome, pp 116–122

Wang F (2011) Genetic diversity of the selected 64 potato germplasm revealed by AFLP markers. Mol Plant Breed 2:22–29

Weltzien E, Smith ME, Meitzner LS, Sperling L (2003) Technical and institutional issues in participatory plant breeding – from the perspective of formal plant breeding: a global analysis of issues, results, and current experience. CIAT, Cali

Weltzien E, vom Brocke K, Rattunde F (2005) Planning plant breeding activities with farmers. In: Christinck A, Weltzien E, Hoffmann V (eds) Setting breeding objectives and developing seed systems with farmers, A handbook for practical use in participatory plant breeding projects. Margraf Publishers/CTA, Weikersheim/Wageningen, pp 123–152

Weltzien E, Christinck A, Touré A et al (2006) Enhancing farmers' access to sorghum varieties through scaling-up participatory plant breeding in Mali, West Africa. In: Almekinders C, Hardon J (eds) Bringing farmers back into breeding. Experiences with participatory plant breeding and challenges for institutionalisation, vol 5, Agromisa special. Agromisa, Wageningen, pp 61–72

Werner M, Diarra L, Hülsebusch C, Kaufmann B (2010) Securing food supply by adapting millet growing to climate variability: decision-making rules of Fulani agro-pastoralists in the Mopti region, Mali. In: Proceedings of Deutscher Tropentag World Food System – a contribution from Europe. Zurich, 14–16 Sept 2010. ETH/DITSL, Zurich/Witzenhausen. http://www.tropentag.de/2010/abstracts/links/Werner_jmrIOtsF.php

Will M (2006) Promoting value chains of neglected and underutilized species for pro-poor growth and biodiversity. Guidelines and good practices. GFU, Rome

Yiching S, Vernooy R (eds) (1999) Seeds and synergies: innovating rural development in China. Practical Action Publishing, Bourton on Dunsmore, UK and the International Development Research Centre (IDRC), Ottawa

Chapter 3
Plant Cryopreservation for Biotechnology and Breeding

Elena Popova, Mukund Shukla, Haeng Hoon Kim, and Praveen K. Saxena

Abstract Plant biodiversity is crucial for sustaining human life on our planet. More than 50,000 species are used globally for food, feed, fiber, medicine and horticulture. A wide range of plant-based biotechnological systems such as isolated root cultures, embryogenic cell and tissue cultures and cell suspensions are used in breeding programs, forestry and the production of pharmaceuticals. Cryopreservation is an essential tool for conservation and long-term maintenance of diverse germplasms with minimal requirements for cost and labor and a low risk of loss of preserved samples. However, large-scale use of cryogenic storage to back-up plant genetic collections is hampered by unavailability of effective methodology and genotype-specific responses of diverse specimens to cryoprotective treatments. Newly developed techniques such as droplet-vitrification are more effective and user-friendlier than classical methods of cryopreservation. Cryopreservation has been successfully employed for preserving several different types of plant materials. In this chapter we review various approaches to develop and improve cryopreservation protocols for diverse plant species. Applications of modern cryopreservation methods in biotechnology-based industry as well as breeding programs are also discussed.

Keywords Cryopreservation • Gene banks • Germplasm collections • Liquid nitrogen • Long-term storage • Secondary metabolites

E. Popova (✉) • M. Shukla • P.K. Saxena
Gosling Research Institute for Plant Preservation, Department of Plant Agriculture, University of Guelph, Guelph, ON N1G2W1, Canada
e-mail: epopova@uoguelph.ca; mshukla@uoguelph.ca; psaxena@uoguelph.ca

H.H. Kim
Department of Well-Being Resources, Sunchon National University, Suncheon, Jeonnam 540-742, Republic of Korea
e-mail: cryohkim@sunchon.ac.kr

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_3

3.1 Introduction

It has long been recognized that preserving plant biodiversity is essential for classical and modern plant breeding programs including genetic engineering (Cruz-Cruz et al. 2013; Cyr 2000; Wang et al. 2012; Withers 1987). Easy access to diverse plant germplasm is a prerequisite for breeding more productive cultivars, which in turn enhances food security (Wang et al. 2014a). The traditional approach to conservation of valuable clones and cultivars is based on plants grown in the fields. These collections are vulnerable to pests and pathogens and can be completely lost due to natural disasters such as hurricanes, floods or droughts (Panis and Lambardi 2006). Ex situ techniques including in vitro cultivation, slow growth storage and cryopreservation are now internationally recognized as major back-up options to support field collections (Engelmann 1997; FAO 2014). Recent advances in plant biotechnology and its broad application in modern agricultural and horticultural industries have raised a new demand for conserving plants and plant tissues such as undifferentiated cell and embryogenic cultures and transgenic or normal root cultures (Engelmann 2014). These cultures can be utilized for the large-scale production of pharmaceuticals and their precursors (Paek et al. 2005, 2009; Verpoorte et al. 2000; Wink et al. 2005) and can also serve as a powerful tool for the production of genetically superior clonal lines of forest and fruit trees (Cyr et al. 2001).

Cryopreservation, the storage of living materials at cryogenic temperatures (below −130 °C), is an alternative to conventional field and in vitro germplasm collection which enables plant genetic resources to be conserved safely and cost-effectively for decades with minimal requirements of space and routine maintenance (Engelmann 2004; Pence 2011; Volk et al. 2014b). More importantly, cryopreservation eliminates the need for regular renewal of the collection thus reducing the risk of genetic erosion caused by pests, diseases, weather conditions, pollution and genetic variations (Panis and Lambardi 2006; Wang et al. 2014a). It also prevents the decline in culture productivity, which is usually associated with culture duration (Reinhoud et al. 2000). Such advantages are due to the effect of extremely low temperatures (usually the materials are stored in liquid nitrogen (LN) or its vapor phase at temperatures ranging from −140 °C to −196 °C) that arrest nearly all cell division and metabolic activities of the cells. Thus, the plant samples can be stored unaltered and they remain viable for a theoretically unlimited duration (Benson 2008; Harding 2004; Wang et al. 2014a).

Cryopreservation serves modern breeding programs by providing long-term storage and easy international access to various genetic materials such as seeds, pollen and meristematic apices and buds. The consistently evolving area of phytochemical production via biotechnological methods is also supported by cryobanking of root cultures, embryogenic and non-embryogenic cell lines to ensure their genetic and biochemical stability (Fig. 3.1).

Cryopreservation of seeds is the most efficient option for many species due to ease of application and the amount of diversity conserved (Engelmann 2004; Pritchard and Nadarajan 2008). The status of seed cryobanking has been reviewed extensively (Pence 1995; Pritchard 2007; Pritchard and Nadarajan 2008; Stanwood

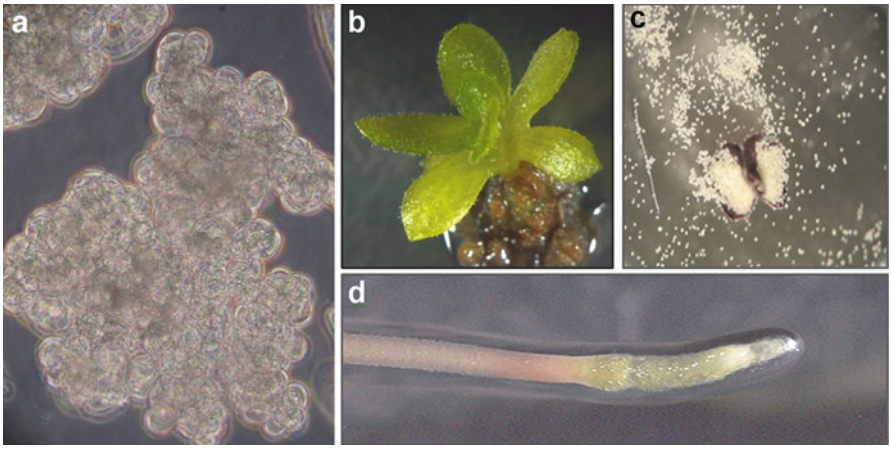

Fig. 3.1 Plant materials at various stages of cryopreservation. (**a**) Undifferentiated cells of American elm in liquid medium before adding a cryoprotectant solution; (**b**) Chrysanthemum shoot tip showing regrowth after cryopreservation; (**c**) mature pollen grains freshly dehisced from anthers of an American elm tree suitable for cryopreservation; (**d**) hairy root of *Rubia akane* showing regrowth after cryopreservation

and Bass 1981; Touchell and Dixon 1994). Cryopreservation is often addressed as the best conservation option for seedless species, species with non-orthodox seeds and for cultivars with unique attributes that can be reproduced only by vegetative propagation (Engelmann 2014; Reed 2001). Some cryopreservation techniques can be specifically tailored to eliminate pathogens from plant materials (cryotherapy) thus improving crop health and quality (Wang et al. 2009).

This chapter covers the progress of cryopreservation to conserve horticultural and some agricultural crops focusing mainly on clonally propagated species. We also give a brief review of the application of cryogenic storage to conserve cell and root cultures for industrial production of bioactive compounds. Using these examples we provide an insight into cryopreservation methodology, its evolution and technical challenges.

3.1.1 Cryopreservation of In Vitro Propagated Plant Materials: Methodology and Challenges

The theoretical bases of plant germplasm cryopreservation were created in the 1970s (Lyons et al. 1979; Mazur 1984; Meryman 1974). In the 1980s, they were translated into a methodology, which allowed cryopreservation of a significant number of species, mostly of temperate origin (Withers 1985). Following the *vitrification revolution* in the 1990s (Benson 2004), the range of cryogenic techniques increased remarkably resulting in successful cryopreservation of many tropical crops and species with

recalcitrant seeds (Engelmann and Takagi 2000; Sakai and Engelmann 2007). The number of reports on large-scale implementation of cryostorage to conserve plant genetic collections in national and industrial repositories is rising steadily (see Cruz-Cruz et al. 2013; Reed 2001, 2008). Recent examples of successful cryobanking of germplasm collections include *Musa* (Panis and Thinh 2001; Panis et al. 2005), *Allium* (Ellis et al. 2006; Kim et al. 2012a), mulberry (Atmakuri et al. 2009; Fukui et al. 2011), *Malus* (Towill et al. 2004) and potato (Kaczmarczyk et al. 2011; Sakai and Engelmann 2007). Thus, within the past three decades, substantial progress has been achieved in both the practical application of cryopreservation and the understanding of biophysical processes underlying cell response to cryogenic temperatures and cryoprotection. Various aspects of such fundamental and applied research have been described earlier (Engelmann and Takagi 2000; Fahy et al. 1990; Fuller et al. 2004; Harding 2004; Kartha 1985; Reed 2008; Towill and Bajaj 2002; Uemura and Steponkus 1992, 1999). In addition, the potential of cryopreservation in sustained production and preservation of economically important medicinal and aromatic plants has also been reviewed (Bajaj 1995; Dixit et al. 2004; Popova et al. 2011).

However, despite the promising results and achievements cited above, the success in the large-scale utilization of cryopreservation for long-term conservation of the existing germplasm collections has been limited. The major difficulty lies in the response of plant material to pre- and post-cryopreservation treatments, which is usually genotype and material specific. To date, there is no uniform method that can be applied with no or only minor modifications to a broad range of plant materials. It is therefore not surprising that some researchers are skeptical about the possibility of developing such method(s) for taxonomically unrelated or even related species.

The most popular and widely applicable cryopreservation methods developed so far include programmed (or slow) freezing, vitrification, encapsulation-dehydration, encapsulation-vitrification and the recently developed droplet-vitrification (see: Reed 2008; Sakai and Engelmann 2007 for method description). These methods exploit different approaches to explant dehydration, which is an essential step in cryopreserving hydrated living materials (Mazur 1984). Sufficient dehydration is required to achieve vitrification of protoplasm, i.e. transition of water from the liquid phase directly into an amorphous or glassy state, thereby avoiding the lethal formation of intracellular ice (Fahy et al. 1984).

3.1.2 Programmed Freezing

Programmed freezing is based on freeze-induced dehydration. Samples pretreated in cryoprotectants (usually dimethylsulfoxide [DMSO], ethylene glycol [EG] and sucrose alone or in low-concentration mixtures) are dehydrated while frozen slowly (0.3–1 °C/min) to −40 °C to −70 °C, then plunged directly into LN. The major disadvantages of this method are the requirement of an expensive program freezer and relatively long exposure of samples to subzero temperatures, which can be deleterious for cold-sensitive species.

3.1.3 Vitrification

The vitrification method originally developed by Sakai in the late 1980s has been successfully applied to more than 200 plant species (Matsumoto and Niino 2014). In vitrification, samples are dehydrated osmotically in a sequence of *loading* (LS) and *vitrification* (VS) solutions on ice or at ambient temperatures. The loading step, also known as *osmoprotection*, is performed via sample exposure to cryoprotectant solutions of moderate concentrations (35–45 %), usually composed of glycerol and sucrose. This step serves to prepare samples to the following extensive dehydration with highly concentrated (up to 100 %, w/v) vitrification solutions (*cryoprotection* step). The most commonly used vitrification solutions are PVS2 (Plant vitrification solution 2) which consists of, w/v, 30 % glycerol, 15 % ethylene glycol, 15 % DMSO and 0.4 M sucrose (Sakai et al. 1990) and PVS3 comprising, w/v, 50 % glycerol and 50 % sucrose (Nishizawa et al. 1993). Modifications of these two basic solutions have also been tested with promising results (Kim et al. 2009; Suzuki et al. 2008; Turner et al. 2001). Dehydrated samples are sealed in cryo-ampoules or polypropylene straws and rapidly immersed in LN. Thus, in contrast to a programmed freezing, in vitrification the most important step is pre-freezing osmotic dehydration/cryoprotection rather than freezing per se.

3.1.4 Droplet Vitrification

For droplet vitrification, samples are osmoprotected as in vitrification but cryopreserved on aluminum plates or foil stripes in microliter drops of vitrification solutions. This ensures better contact of explants with liquid nitrogen thus facilitating cooling and rewarming of samples (Towill and Bonnart 2003) which results in improved regrowth for a number of species, for example, chrysanthemum (Halmagyi et al. 2004), garlic (Kim et al. 2006a) and potato (Kaczmarczyk et al. 2011).

3.1.5 Encapsulation

Based on the technology of artificial seeds, encapsulation (Fabre and Dereuddre 1990) employs encapsulation of plant materials in Ca-alginate beads followed by air desiccation (encapsulation-desiccation) or dehydration VS (encapsulation-vitrification) combined with rapid immersion in LN. Though more laborious than *classic* vitrification, this technique is useful when cryopreserving small, fragile or extremely desiccation-sensitive materials (Escobar-Pérez 2005; Gupta 2014; Hirata et al. 2002; Matsumoto and Sakai 1995; Mikula 2006).

Generally, regardless of the cooling procedure, rewarming of samples rapidly in a water bath at approx. 40 °C was beneficial for their recovery (Sakai and Engelmann 2007).

3.2 Developing a Cryopreservation Protocol: Empirical and Systematic Approaches

Cryopreservation of clonally-propagated plant material is a complicated process comprising multiple steps such as explant selection and excision, preconditioning, osmoprotection and cryoprotection with LS and VS, freezing, storage, rewarming and recovery (Fig. 3.2). For successful regrowth, not only each step of the cryopreservation protocol should be optimized but also the physiological state of explants should be considered. Manipulation with growth regulators, cold acclimation and the addition of ABA to the culture medium are the most common procedures to enhance physiological tolerance of donor cultures and explants to stress conditions associated with cryopreservation (Burchett et al. 2006; Chetverikova 1999; Popova et al. 2009; Uemura and Steponkus 1999). Concentrations of penetrating and non-penetrating cryoprotectants as well as duration and temperature of cryoprotection treatment can be varied in order to dehydrate samples, while avoiding or reducing cytotoxicity (Kim et al. 2009; Sakai and Engelmann 2007). Preculture of plant materials with amino acids (proline), osmotically active chemicals such as sugars and sugar alcohols (mannitol, sorbitol) is a common and a very effective way to improve post-cryopreservation regrowth (Baskakova et al. 2003; Butenko et al. 1984; Carpentier et al. 2007; Popova et al. 2010; Ramon et al. 2002; Reinhoud et al. 2000; Zhu et al. 2006) though it needs to be tested and optimized for every species.

Oxidative stress is believed to be one of the major factors causing cell injury in cryopreserved samples (Benson 1990; Benson and Bremner 2004). A relatively new approach to enhance regrowth of cryopreserved tissues is the addition of antioxidants (reduced glutathione, ascorbic acid, melatonin, etc.) to cryoprotective solutions or to the culture medium used for pretreatment and recovery. For example, the addition of glycine betaine, ascorbic acid, glutathione and ABA into PVS2 improved post-cryopreservation regrowth of *Arabidopsis* seedlings (Ren et al. 2014). Vitamin E and ascorbic acid added during the pretreatment, loading, unloading and regrowth steps enhanced regrowth and reduced malondialdehyde formation in shoot tips of cryopreserved blackberry (Uchendu et al. 2010a). Lipoic acid, glutathione and glycine betaine also showed beneficial effect when used with the same plant under similar conditions (Uchendu et al. 2010b). Melatonin supplied at preculture and recovery steps significantly improved regrowth of American elm shoot tips cryopreserved by vitrification (Uchendu et al. 2013). Survival of cryopreserved *Rhodiola crenulata* callus increased following a 5-day pretreatment with 0.1 μM melatonin and was associated with enhanced peroxidase and catalase activity of melatonin-treated cells compared to the control (Zhao et al. 2011). All these findings suggest that the antioxidants may play an essential role in reducing oxidative stress associated with the cryopreservation process even with thoroughly optimized steps of the protocol.

On account of the abundance of protocols available worldwide for various plant species and materials (Engelmann 2014; Reed 2008), one will have a broad selection of options while attempting to cryopreserve a new species, genotype or a cell line. A

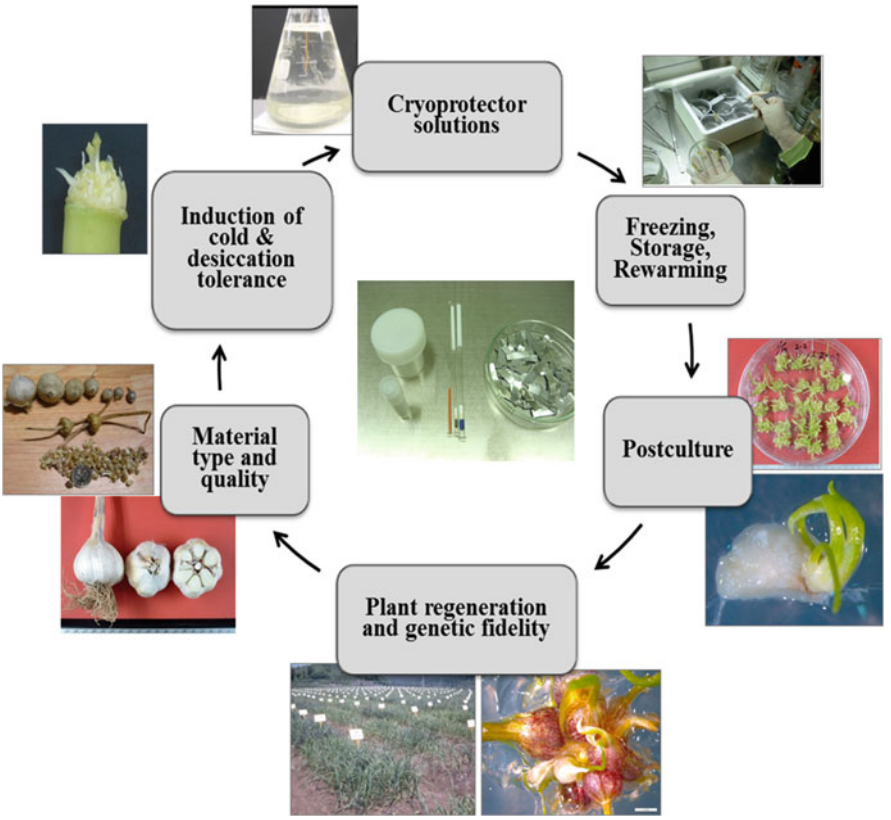

Fig. 3.2 A simplified scheme of cryopreserving clonally propagated germplasm: cryobanking of *Allium* collection at National Agrobiodiversity Center, South Korea as an example. Photograph in the center of the illustration represents various containers/holders, ampoules, straws and aluminum foil strips used for cryopreservation

simple, time-saving approach would involve the use of an existing protocol developed for a closely related species with further modification and optimization of major steps, if needed, to improve post-cryopreservation recovery. Unfortunately, in many studies none of the existing protocols results in desirable regrowth after cryogenic storage. In this case, the researcher has two options. The first obvious and commonly used strategy is to develop the protocol de novo by optimizing preculture conditions, composition and concentration of cryoprotectant solutions, cooling and rewarming rates, and conditions during recovery with a hope that the process will enhance the regrowth. Also referred to as the *experimental* or *trial-and-error* approach, this blind screening of multiple conditions is time and labor consuming. Often this approach is impractical and too expensive to justify, particularly for the curators of small but valuable collections in botanical gardens or private companies. A common consequence of failures in this approach is discontinuation of cryopreservation research on such recalcitrant species, cell lines or cultivars (Popov, Redington, personal communication).

An alternative to *blind* screening is a systematic evaluation of the impact of each step of the cryopreservation protocol on the recovery and physiological state of the samples. This includes cytological and ultrastructural investigation of cells (Bachiri et al. 2000; Miao et al. 2005; Micuła et al. 2005; Salma et al. 2014), measuring the kinetics of cryoprotectant influx/efflux (Kim et al. 2004a, b), determination of the amount and physical state of intracellular water (Fang et al. 2009; Volk and Walters 2006), thermal analysis for understanding water and lipid behaviour during cooling and rewarming (Kim et al. 2009; Towill and Bonnart 2003), biochemical assays (Ramon et al. 2002; Zhu et al. 2006; etc.) and gene expression (Ren et al. 2013; Volk et al. 2014a). In the most desirable scenario these analyses are performed systematically using selected model systems which exhibit different tolerance to cryopreservation treatment as shown for shoot tips of *Arabidopsis*, garlic and mint and root culture of madder (Kim et al. 2012a, b, 2014; Volk et al. 2014a). A representative example of such a multi-stage investigation is the protocol development for large-scale cryopreservation of *Allium* field collection at the National Agrobiodiversity Center in Suwon, South Korea. Major steps of this process are shown in Table 3.1 and Fig. 3.2. As a result of this study, a total of 1158 *Allium* accessions have been cryopreserved during 2005–2010 using the droplet-vitrification technique with a mean post-cryo regeneration percentage of 65.9 % (Kim et al. 2012a).

It is quite obvious from the scope of a systematic development of a cryopreservation protocol (Table 3.1) that this approach is ambitious, extensive and must be reinforced by long-term planning and strong financial support. The cryobanking of the whole national *Allium* collection (Fig. 3.2) was achieved through combined efforts and investments on behalf of enthusiastic members of the cryopreservation team and field conservatory groups, government support and over 10 years of extensive research employing a range of biochemical and molecular tools including analytical chemistry (HPLC, Differential Scanning Calorimetry), electron microscopy, RT-PCR, immuno-assay, etc. These methods require specialized equipment, chemical reagents of high purity and qualified personnel, which may be financially and technically restrictive for small laboratories, particularly in developing countries (Reed 2001).

In view of these limitations we designed a simplified approach to protocol development based on differentiation of samples according to their size and permeability to cryoprotectants as well as their sensitivity to chemical and osmotic toxicity of vitrification solutions modified from those commonly employed in PVS2 and PVS3 (Kim et al. 2009). The plant material is classified according to its response to a number of standardized treatments with progressively increasing total concentration of cryoprotectants. The consequent steps are rapid freezing of samples in drops of vitrification solution on aluminum foil strips (droplet vitrification), rewarming by plunging into pre-heated (40 °C) 1.2 M sucrose solution and further unloading for a proper duration depending on the materials. The treatments with better survival results can be incorporated into standard protocol which can be further adapted for different genotypes.

Table 3.1 Systematic development of a cryopreservation protocol for cryobanking *Allium* collection at the National Agrobiodiversity Center (Suwon, South Korea) using clove apices and bulbil primordia from immature inflorescences

Step	Factors investigated	Year
1. Revealing the most critical factors that affect regeneration after cryostorage; developing the *standard protocol*	Origin and size of starting material; explant size	2001–2002
	Cold acclimation and preconditioning	2001–2003
	Cryoprotectant treatment: duration, composition of cryoprotectant solutions;	2002–2003
	Air dehydration before cryoprotection;	
	Unloading	2002–2003
	Recovery (the effect of growth regulators)	2003
	Cooling and rewarming methods[a]	2005–2006
2. Investigating the impact of cryoprotection	Thermal behavior of explants and cryoprotectant solutions during cooling and rewarming	2002–2003
	Analyzing the dynamics of cryoprotectant influx/efflux (HPLC)	2002–2003
	Investigation of genetic stability of recovered plants (field performance, AFLP)	2002–2003
3. Large scale implementation of the standard protocol	Application protocol to diverse genotypes; Genotype-specific adjustments of protocol for sensitive varieties	2005–2010
	Revealing duplicate accessions and filling the gaps in the collection	2005–2010
4. Cryotherapy	Elimination of garlic virus A, onion yellow dwarf virus (OYDV) and leek yellow stripe virus (LYSV) due to cryotherapy	2007–2010

Source: Modified from Kim et al. (2012a)

[a]In 2001–2005, the majority of samples were cryopreserved using a vitrification protocol. Beginning 2005, a newly-developed droplet vitrification procedure was adopted as it generally produced higher recovery after cryopreservation

This approach was successfully tested with embryogenic cultures of the Korean forest species *Kalopanax septemobilis* (Shin et al. 2012) and hairy root cultures of six medicinal plants, which showed high sensitivity to both osmotic and chemical toxicity of loading and vitrification solutions (Kim et al. 2012b). The same principle was later applied to formulate an efficient cryopreservation protocol for chrysanthemum shoot tips which are also very sensitive to chemical toxicity but relatively tolerant to osmotic action of cryoprotectant mixtures (Lee et al. 2011).

Table 3.2 summarizes cryopreservation strategies and optimized pretreatment conditions for the selected species as revealed by applying this rationalized approach to the droplet vitrification method. The following recommendations emerged from the modification of the initial protocol developed for highly tolerant species while transferring it to more sensitive plant materials:

Table 3.2 Strategies and optimized procedures applied for cryopreservation of various species with different explant properties and different sensitivity/tolerance to chemical and osmotic stress of cryoprotectant solutions

Strategies or procedures	Plant species			
	Allium sativum	*Kalopanax septemlobus*	*Chrysanthemum morifolium*	*Rubia akane*
Size and type of the material	Bulbil primordia, clove apices 1.5×3.0 mm	Embryogenic callus, 40 mg fr. wt.	Shoot tips, axillary buds 1.2×1.5 mm	Root apices, 7–10 mm
Tolerance or sensitivity to osmotic stress	Very tolerant	Tolerant	Tolerant	Very sensitive
Tolerance or sensitivity to chemical stress	Very tolerant	Tolerant	Sensitive	Very sensitive
Strategy for developing cryopreservation protocol	Sufficient dehydration through increasing duration of exposure to highly concentrated VS at room temperature	1. Loading is important for regrowth	1. Improve dehydration tolerance through preculture with step-wise increasing sucrose concentrations	1. Use prolonged preculture with moderate sucrose concentrations to induce desiccation tolerance
		2. Avoid DMSO and EG or Reduce their concentration in VS and incubation temperature (ICE)	2. Loading is crucial for regrowth and composition of LS is important	2. Loading is crucial for regrowth and composition of LS is important
				3. Use very short (up to 20 min) incubation with VS
			3. Avoid DMSO and EG or Reduce temperature of VS incubation (ICE)	4. Avoid DMSO and EG or Reduce their concentration in VS and incubation temperature (ICE)
Optimized cryopreservation protocols				
1. Preculture	0.3 M suc, 2–3 days, 10 °C	No	0.3 M suc (31 h) → 0.5 M (17 h) → 0.7 M (7 h)	0.3 M suc (48 h) → 0.5 M (5 h)
2. Loading	No (optional)	Any LS, 20 min, 24 °C	C4, 40 min, 24 °C	C4, 30 min, 24 °C
3. Dehydration	PVS3, 150–180 min, 24 °C	PVS3, 40 min, 24 °C or A3-80 %, 40 min, 0 °C	PVS3, 60 min, 24 °C or A3, 55 min, 0 °C	B5, 15 min, 24 °C or A3-70 %, 20 min, 0 °C
4. Freezing	Foil (droplet)	Foil (droplet) or ampoules	Foil (droplet)	Foil (droplet)

(continued)

Table 3.2 (continued)

Strategies or procedures	Plant species			
	Allium sativum	*Kalopanax septemlobus*	*Chrysanthemum morifolium*	*Rubia akane*
5. Rewarming	37–40 °C, 30 s	37 °C, 60 s	37 °C, 30 s	37 °C, 30 s
6. Unloading	0.8 M suc, 40 min	0.8 M suc, 30 min	0.8 M suc, 40 min	0.8 M suc, 30 min
Regrowth after cryopreservation	77–100 %, depending on genotype	95–100 %	82–85 %	84–89 %

According to Kim et al. (2012a), Lee et al. (2001), Shin et al. (2012)
A3: 37.5 % glycerol + 15 % DMSO + 15 % EG + 22.5 % sucrose, w/v; A3-80 %: 33.3 % glycerol + 13.3 % DMSO + 13.3 % EG + 20.1 % sucrose, w/v; A3-70 %: 29.2 % glycerol + 11.7 % DMSO + 11.7 % EG + 17.4 % sucrose, w/v; B5: 40 % glycerol + 40 % sucrose, w/v; C4: 17.5 % glycerol + 17.5 % sucrose, w/v; DMSO: dimethylsulfoxide; EG: ethylene glycol; PVS3: 50 % glycerol + 50 % sucrose, w/v; Suc: sucrose; VS: vitrification solution

(a) Step-wise preculture with gradually increasing sucrose concentration can improve explant tolerance to further dehydration caused by VS. However, highly sensitive tissues (hairy roots) may not survive exposure to sucrose beyond 0.5 M even for a few hours.
(b) Loading step and composition of loading solution are essential for better survival of sensitive materials after cryopreservation.
(c) The use of DMSO and EG in cryoprotectant solutions should be avoided or minimized, and the cryoprotection treatment performed on ice to reduce chemical toxicity.
(d) Total concentration of cryoprotectants should be reduced as compared to those in standard PVS2 and PVS3.

Although these findings may not be universally applicable, they can be taken into consideration when developing cryopreservation protocols for new genotypes and plant materials.

3.3 Cryopreservation of Horticultural Plants: Chrysanthemum and Lily as Model Plants

Among horticultural plants, chrysanthemum and lily are often favored by breeders due to the ease of the selection and a broad variety of genotypes (Teixeira da Silva 2003). They are commonly propagated vegetatively by bulb division, stem cuttings or by in vitro culture of adventitious shoots and nodal segments (Martín and González-Benito 2009). These plants represent successful models of the application of biotechnological tools in horticultural industry (Teixeira da Silva 2004). Horticultural markets worldwide consistently demand the introduction of new

varieties replacing the existing popular genotypes. This transition renders *old-fashioned* genotypes redundant and the maintenance of their nuclear stock economically unviable. Despite the loss of commercial interest, these genotypes represent a source of genetic diversity, which is essential for plant improvement by conventional and molecular breeding (Martín and González-Benito 2009). In search of a cost-effective option for long-term storage, cryopreservation of lilies and chrysanthemum has been frequently attempted during the past decades. Collectively, these studies have laid the foundation for the development of simple and effective protocols for cryopreserving horticultural species.

Meristematic organs are the most favorable material for cryopreservation due to their high genetic stability and potential use in cryotherapy (Wang et al. 2014a). Cryopreservation of chrysanthemum shoot tips excised from greenhouse (Fukai 1990; Fukai and Oe 1990; Fukai et al. 1991) and in vitro (Martín and González-Benito 2005; Martín et al. 2011; Sakai et al. 2000; Zalewska and Kulus 2013) plants has been reported using programmed freezing and, later, vitrification and encapsulation-dehydration techniques with post-cryogenic regrowth ranging from 35 % to 85 %. All these methods, however, have certain disadvantages that hinder their routine implementation for cryopreserving diverse cultivars. Abnormal development was frequently observed for apices cryopreserved by programmed freezing, presumably as a result of a pretreatment with DMSO (Fukai and Oe 1990; Fukai et al. 1991). Cold hardening of in vitro donor plants for a minimum of 3 weeks was essential for shoot tip recovery after exposure to LN regardless of the cryopreservation procedure used.

A significant step towards simplifying the cryopreservation procedure for chrysanthemum was made by Halmagyi et al. (2004) who replaced cold-hardening of donor plants by 24 h incubation of the excised apices in the medium supplemented with 0.5 M sucrose. Precultured apices showed 80 % regrowth when cryopreserved in 4-µl drops of full-strength or diluted PVS2 stuck to aluminum foil strips. By contrast, programmed freezing, ultra-rapid freezing on top of the hypodermic needle and encapsulation-dehydration tested in the same study resulted in regrowth below 50 %. The only restriction of the new protocol was a narrow time range of explant exposure to vitrification solutions: 5–10 min in full-strength PVS2 or 10–20 min in 60 % PVS2. Exceeding the exposure time by 5 min caused nearly 20 % loss of post-cryopreservation regrowth (Halmagyi et al. 2004).

By expanding the scope of this work, we achieved cryopreservation of chrysanthemum shoot tips using the droplet-vitrification method and rationalized approach described in a previous section. A series of 14 vitrification solutions with varied concentrations of penetrating and non-penetrating cryoprotectants were tested with *Chrysanthemum morifolium* apices excised from 4-week-old in vitro plants (Fig. 3.3). As Fig. 3.3 shows, chrysanthemum apices were moderately sensitive to osmotic stress produced by glycerol and sucrose and very sensitive to toxic action of permeating cryoprotectants DMSO and EG. Increasing DMSO and EG concentration by only 10 % caused a 60 % decrease in regrowth of apices after 30 min exposure with or without cryopreservation while reducing toxic chemicals by 10 % slightly increased the regrowth. Further studies revealed the crucial importance of the load-

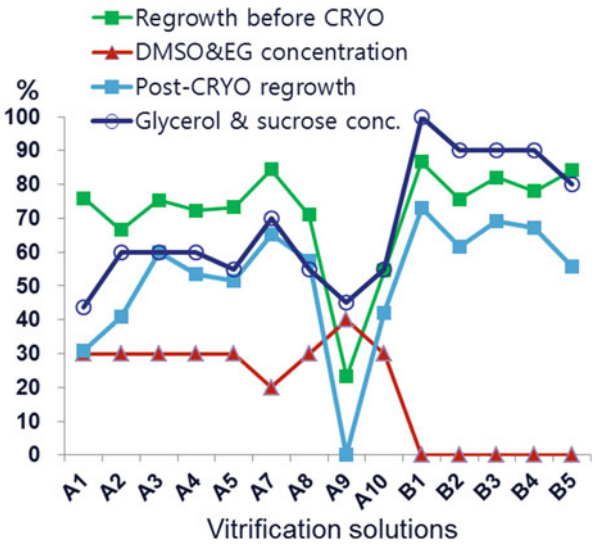

Fig. 3.3 Effect of glycerol and sucrose versus DMSO and EG concentrations in vitrification solutions on regrowth of *Chrysanthemum morifolium* shoot apices before and after cryopreservation (CRYO). Apices excised from 4-week-old in vitro plants were step-wise precultured with 0.3, 0.5 and 0.7 M sucrose for 31, 17 and 7 h, respectively, loaded for 40 min then dehydrated in vitrification solutions A1-A10 for 30 min or B1-B5 for 150 min. Solutions of A series were modified from commonly employed PVS2 (A1) and contained glycerol, sucrose, DMSO and EG. Solutions of B series were based on PVS3 (B1) and contained only glycerol and sucrose (see Table 3.3 for the composition of the solutions)

ing step and composition of the loading solution for post-cryopreservation regrowth of chrysanthemum apices. Thermal analysis of cryoprotected apices during freezing and rewarming suggested that loading had little effect on thermal behavior of the explants. Most likely, it primarily served to mitigate the osmotic stress caused by further exposure to VS. Step-wise preculture of apices in liquid medium supplemented with 0.3, 0.5 and 0.7 M sucrose to induce dehydration tolerance also improved regrowth (Table 3.2). The resulting protocol employing step-wise preculture, loading, exposure to PVS3 or A3 vitrification solution, freezing using aluminum foil strips, rapid rewarming and recovery on GA_3 containing medium resulted in 80–90 % regrowth of two chrysanthemum cultivars (Table 3.2, Lee et al. 2011).

Wang et al. (2014b) developed an alternate procedure for cryopreservation of *Chrysanthemum morifolium*. This new protocol employed shoot tips dissected from 12-day-old nodal segments, one-step preculture with 0.5 M sucrose for 24 h, loading for 20 min and cryoprotection with PVS2 for 30 min at 0 °C followed by freezing in LN using aluminum foil strips. This procedure is highly effective, user-friendly and has been tested in six chrysanthemum genotypes resulting in 43–83 % regrowth. In comparison, programmed freezing method suggested by Fukai et al. (1991) was successful with 3 chrysanthemum cultivars, 12 species and 2 interspecific hybrids, with high variation of regrowth rates of 9.4–100 %.

Table 3.3 Composition and total concentration of the alternative vitrification solutions used in systematic protocol development for *Allium*, chrysanthemum, lily and other species mentioned in the study

Vitrification solution	Composition (%, w/v)	Total concentration (%)	Remarks
A1	Glycerol 30.0 + DMSO 15 + EG 15 + sucrose 13.7	73.7	Classical PVS2
A2	Glycerol 42.5 + DMSO 15 + EG 15 + sucrose 17.5	90.0	
A3	Glycerol 37.5 + DMSO 15 + EG 15 + sucrose 22.5	90.0	
A4	Glycerol 32.5 + DMSO 15 + EG 15 + sucrose 27.5	90.0	
A5	Glycerol 35.0 + DMSO 15 + EG 15 + sucrose 20.0	85.0	
A7	Glycerol 37.5 + DMSO 10 + EG 10 + sucrose 32.5	90.0	
A8	Glycerol 30.0 + DMSO 15 + EG 15 + sucrose 25.0	85.0	
A9	Glycerol 30.0 + DMSO 20 + EG 20 + sucrose 15.0	85.0	
A10	Glycerol 40.0 + DMSO 15 + EG 15 + sucrose 15.0	85.0	
B1	Glycerol 50 + sucrose 50	100.0	Classical PVS3
B2	Glycerol 50 + sucrose 40	90.0	
B3	Glycerol 45 + sucrose 45	90.0	
B4	Glycerol 40 + sucrose 50	90.0	
B5	Glycerol 40 + sucrose 40	80.0	

Source: Modified from Kim et al. (2009)

The majority of studies on cryopreservation of lily germplasm utilized meristems or adventitious shoot primordia regenerated on the scales of in vitro or field grown bulbs (Bouman and de Klerk 1990; Chen et al. 2011; Matsumoto and Sakai 1995). First experiments involved preculture of explants with 10 % sucrose at low temperature (5 °C) followed by direct freezing in LN in cryovials or by programmed freezing (Bouman and de Klerk 1990), but resulted in poor regrowth of <8 %. Vitrification method employed later by the same authors was more successful: 10 min dehydration of precultured and loaded explants with PVS2 was sufficient to achieve 80 % regrowth after cryopreservation. Duration of PVS2 exposure could be extended to 1 h without significant effect on the regrowth (Bouman et al. 2003). Cryopreservation of roots was also achieved effectively though conversion of regenerated roots into plants was not high (Bouman et al. 2003). These results were in agreement with those reported earlier by Matsumoto et al. (1995) who reported the highest post-cryopreservation recovery of lily apices after exposure to PVS2 for 20 min at 25 °C or for 110 min at 0 °C. Chen et al. (2011) obtained highest regrowth of cryopreserved lily apical meristems excised from adventitious buds after 90 min exposure to PVS2.

Despite a noticeable tolerance of lily apices to osmotic and chemical action of cryoprotecting solutions, optimization of preculture procedure, loading duration

and composition of loading solution were found to be important for an efficient cryopreservation (Chen et al. 2011; Matsumoto et al. 1995). Droplet vitrification improved regrowth for two of three cultivars as compared to classical vitrification in cryovials (Chen et al. 2011). Another effective cryopreservation protocol employed two-step pretreatment of excised juvenile corms followed by incubation in alternative loading and vitrification solutions (Fig. 3.4). This method resulted in vigorous regrowth of approximately 60–70 % explants after cryopreservation. In all cases, cold hardening of excised meristems or bulb scales with adventitious buds improved regrowth after cryopreservation.

Yin et al. (2014) used shoot tips of adventitious shoots to test different preculture regimes, loading solutions and PVS2 exposure durations for cryopreservation of lily. The optimized protocol implied preculture with 0.5 sucrose for 24 h and incubation in loading solution containing 0.4 M sucrose and 2 M glycerol for 20 min followed by a PVS2 treatment for 4 h at 0 °C.

When these protocols were tested with different genotypes, they resulted in comparable regrowth. Bouman et al. (2003) reported maximal regrowth from 19 % to 93 % for 10 genotypes using vitrification. Meanwhile, Chen et al. (2011) showed regrowth of 43 %, 65 % and 84 % for three lily cultivars cryopreserved by droplet vitrification. Yin et al. (2014) recorded 42.5–87.5 % regrowth for six genotypes tested with the same method.

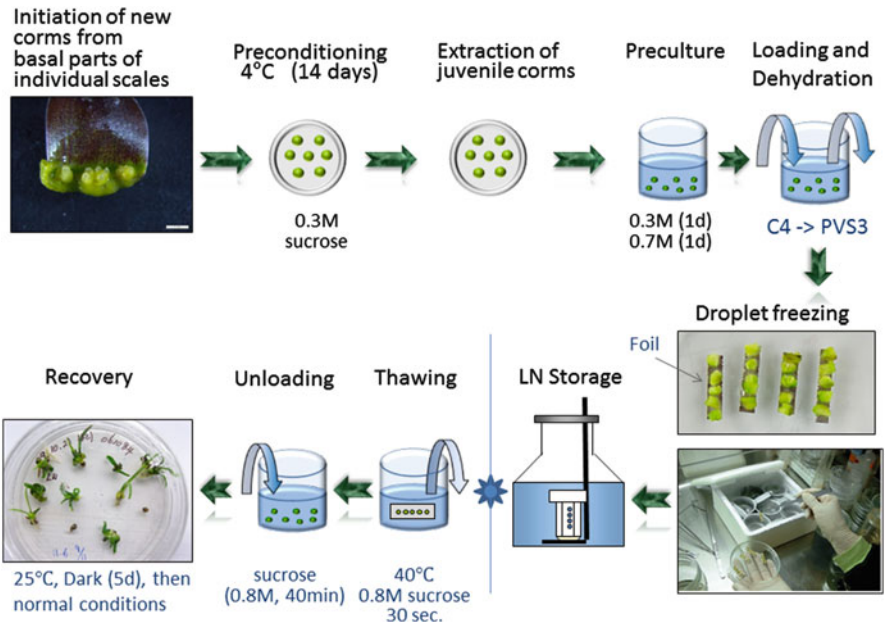

Fig. 3.4 Cryopreservation of lily by droplet vitrification using new corms developed from in vitro growth bulb scales. C4: 17.5 % glycerol + 17.5 % sucrose, w/v; PVS3: 50 % glycerol + 50 % sucrose, w/v

Despite a prominent difference in their physiology, both chrysanthemum and lily remain representative examples of how the cryopreservation methodology has evolved for horticultural plants. First attempts relied on programmed freezing with a number of randomly tested cryoprotective pretreatments. These experiments were rarely successful, mostly due to the lack of knowledge of the physiology of explant response to cryoprotection and low temperature and the experience in resolving these issues. Regrowth was improved by applying PVS-based vitrification and later its modified version, droplet-vitrification methods. These procedures were also simpler and affordable compared to programmed freezing due to exclusion of expensive equipment. Cold hardening of donor plants, once considered essential for regrowth, was recently replaced by using the explants at certain physiological state achieved by preculture with sucrose or manipulations with growth regulators in the medium. In all cases, sequential exposure to increased concentration of cryoprotectants improved regrowth of cryopreserved explants. In vitro cultured plant material remains the best source of explants for cryopreservation due to its proven sterility and availability throughout the year.

Apart from high regrowth rate obtained with number of cultivars, routine implementation of cryopreservation for conserving chrysanthemum and lily germplasm is hampered by genotype-dependent response to standard treatments (Bouman et al. 2003; Fukai et al. 1991; Yin et al. 2014). These difficulties may be overcome in the future by better understanding of mechanisms underlying the response of explants to cryoprotectant treatment and dehydration, and water behavior in cryoprotected samples. Droplet-vitrification method reported in most recent studies (Chen et al. 2011; Lee et al. 2011; Yin et al. 2014) appears to be the most promising for systematic improvement of cryopreservation of lily germplasm.

3.3.1 Genetic Stability

Cryopreservation procedures induce severe stress in plant materials and commonly results in lethal damages to the treated explant. Therefore, monitoring of genetic stability of plants regenerated from cryopreserved explants has been considered very important (Harding 2004). Fukai et al. (1991) found no difference in morphological parameters such as date of flowering, length of flower stem, number of leaves, flower weight and number of florets in 18 regenerants of *Chrysanthemum morifolium* cv. Shuhounitikara after cryopreservation using programmed freezing. Later, however, the authors reported that nearly 70 % of chimeric *Dendranthema grandiflorum* plants derived from cryopreserved shoot tips had altered flower color (Fukai et al. 1994).

Ploidy levels determined by flow cytometry analysis (FCM) remained stable in plants of three chrysanthemum genotypes regenerated from shoot tips cryopreserved by droplet vitrification as compared to control stock cultures (Lee et al. 2011; Wang et al. 2014b). Assessment of genetic stability by RAPD markers did not detect any polymorphic bands in 21 regenerants from shoot tips cryopreserved by vitrification, while one of 25 regenerants developed after encapsulation-dehydration showed a different band pattern (Martín and González-Benito 2005). More detailed investi-

gation into cryopreservation using encapsulation-dehydration techniques with both RAPD and AFLP markers showed that one of the initial steps in explant preparation, 3-day preculture with 0.3 M sucrose at 5 °C, caused modification in the RAPD profile (Martín et al. 2011). Even higher polymorphism was detected with AFLP markers (40.1 % compared to 5.78 % with RAPD). Based on these findings, the authors suggested that genetic variation in the material may be associated with pretreatment steps and not necessarily with the damage caused by the freezing-rewarming process (Martín et al. 2011). Most recently, no polymorphic bands were detected by simple sequence repeats (SSR) in regenerants of two chrysanthemum genotypes cryopreserved by droplet vitrification (Wang et al. 2014b).

Compared to chrysanthemum, analysis of the genetic stability of cryopreserved lily is scarce. In one study (Yin et al. 2014), genetic stability of the regenerants was assessed using inter-simple sequence repeat (ISSR) markers and no differences were found between plants developed in control culture and from cryopreserved adventitious shoot tips of two genotypes within 12 months after cryogenic treatment.

3.4 Cryopreservation of Cell Cultures for Secondary Metabolite Production

In vitro cultures of undifferentiated somatic plant cells are an attractive source of phytochemicals for food and pharmaceutical industries (Nosov 2012; Paek et al. 2005; Smetanska 2008). These cultures are normally heterogenic and prone to genetic instability (Nosov 1999). In the majority of cell lines, the frequency of genetic and epigenetic variations increases in the course of repetitive subcultures causing the loss of regeneration potential and changes in secondary metabolite profile (Heine-Dobbernack et al. 2008). Cryopreservation is often addressed as the only effective option to prevent culture ageing and reduce the risks of culture loss caused by contamination or technical errors (Engelmann 2004; Reinhoud et al. 2000).

The methodology of cryopreserving cultured somatic plant cells has been summarized in a number of reviews (Heine-Dobbernack et al. 2008; Nosov et al. 2014; Popova et al. 2011; Reinhoud et al. 2000; Sakai and Engelmann 2007; Withers 1985). Here, we focus mainly on the cell cultures that have been tested for industrial or semi-industrial production of secondary metabolites of potential economic importance.

Cell cultures derived from *Taxus* spp. have been largely utilizes for the commercial production of paclitaxel, a complex diterpenoid with high anti-tumor activity (Malik et al. 2011). Cryopreservation of *T. chinensis* cell culture in a cryoprotective mixture of 0.5 M DMSO and 0.5 M glycerol by programmed freezing has been reported with 40 % maximum viability after rewarming (Kim et al. 2001). The cryopreserved cell line showed lower accumulation of dry biomass as compared to control culture in the course of 40-day cultivation, however, paclitaxel production was similar to that of the control (Kim et al. 2001).

Programmed freezing was also applied for cryopreserving *Polyscias filicifolia* cell culture which is currently used for the production of bioactive food additive Vitagamal

(Titova et al. 2011). A combination of 15 % glycerol and 10 % sucrose was optimum for cryoprotection and resulted in 45 % post-cryogenic regrowth measured soon after freezing (Krivokharchenko et al. 1999). Same culture was successfully regenerated after 5 years of cryogenic storage with 25–40 % survival rates, proliferated in flasks and scaled up to 20, 75 and 630 L bioreactors for semi-industrial biomass production (Titova et al. 2007). The regenerated culture not only fully retained its specific growth traits to the level recorded before cryopreservation, but showed significantly higher productivity and growth rates compared to the same cell line maintained for 5 years under standard conditions by repetitive subcultures (Table 3.4).

The levels of diosgenine, sitosterol and stigmasterol in *Dioscorea deltoidea* cell culture were unchanged after cryopreservation by programmed freezing following preculture of cells with amino acids asparagines and alanine (Butenko et al. 1984). Cryopreservation of *Panax ginseng* and *P. quinquefolius* cell cultures was successful using programmed and direct freezing respectively (Fedorovskii et al. 1993; Joshi and Teng 2000; Mannonen et al. 1990; Seitz and Reinhard 1987). Glycerol and sucrose in moderate concentrations (up to 20 %) were efficient for cell pretreatment and cryoprotection of ginseng and *Dioscorea* similar to cell cultures of other species. Higher growth, biomass productivity and yield have been recorded for *Panax ginseng* cell suspension regenerated after two-step freezing to cryogenic temperature as compared to unfrozen culture (Joshi and Teng 2000). The ginsenoside pattern was not affected by cryopreservation (Mannonen et al. 1990; Seitz and Reinhard 1987).

Table 3.4 Growth characteristics of *Polyscias filicifolia* cell cultures in flasks and in 20 L bioreactors

Culture	M_{max}* (g/l)	V (%)	μ (day^{-1})	T (day)	P (g/l day)
Cell cultures in 250 ml flasks					
Initial cell culture (before cryopreservation)	14.1 ± 2.0	90 ± 4	0.19 ± 0.03	3.6 ± 0.2	0.86 ± 0.14
Cell culture regenerated after 5 years cryogenic storage	13.2 ± 2.2	91 ± 3	0.17 ± 0.02	4.1 ± 0.6	0.80 ± 0.16
Cell culture maintained for 5 years by repetitive subcultures at 24 ± 3 °C	9.0 ± 0.8	84 ± 2	0.14 ± 0.01	4.9 ± 0.3	0.54 ± 0.06
Cell cultures in 20 L bioreactors					
Initial cell culture (before cryopreservation)	13.9 ± 2.0	88 ± 4	0.17 ± 0.02	4.1 ± 0.4	0.95 ± 0.18
Cell culture regenerated after 5 years cryogenic storage	13.6 ± 1.8	89 ± 5	0.16 ± 0.02	4.3 ± 0.4	1.10 ± 0.15
Cell culture maintained for 5 years by repetitive subcultures at 24 ± 3 °C	8.8 ± 0.8	82 ± 2	0.12 ± 0.04	6.3 ± 2.1	0.58 ± 0.19

Source: Modified from Titova et al. (2011)
M_{max}: maximal biomass concentration (dry weight bases); *V*: cell viability; μ: specific growth rate; *T*: generation (doubling) time of cell culture; *P*: culture productivity on dry weight basis

Thus, on the basis of the review of aforementioned studies, one may conclude that programmed freezing is most applicable to cryopreservation of undifferentiated cell cultures. This may have been due to the decline in interest to use cell cultures as a source of bioactive metabolites in the 1990s when vitrification-based methods started to develop. Meanwhile, more recent studies demonstrated high effectiveness of two-step freezing, vitrification and encapsulation approaches for cryopreservation of embryogenic and non-embryogenic cell cultures of *Gentiana tibetica* (Mikula 2006), *Catharanthus roseus* (Samar et al. 2009), *Sapindus mukorossi* (Kim et al. 2006b) and a few other species (Heine-Dobbernack et al. 2008).

3.5 Cryopreservation of Isolated Root Cultures

In vitro cultures of adventitious or transformed (hairy) roots are widely used for the production of bioactive compounds and constitute a convenient model to study plant morphogenesis, growth, and development (Hahn et al. 2003). Conservation of hairy and adventitious roots is valued for research and development because of the high levels of their morphological uniformity, genetic stability and high regeneration ability.

Cold storage of isolated root cultures at 4 °C has been reported with high survival rate but this method is useful only for short-term conservation. For example, Yoshimatsu et al. (1996) successfully stored genetically transformed hairy root segments and tips (1 mm) on growth regulator-free medium at 4 °C for nearly 4 months. Through cryopreservation of a competent culture, it may be possible to store root material permanently and to reintroduce it into culture without any loss of morphogenic and biosynthetic capacity (Benson and Hamill 1991). Several factors such as culture age, pre-growth, cryoprotection, freezing rate and post-freeze culture conditions play important roles in recovery after cryopreservation. Primary and secondary roots showed similar post-cryopreservation regeneration (88–95 %), while root sections without apices displayed significantly lower regeneration (65 %) (Salma et al. 2014). Park et al. (2014) suggested an optimized cryopreservation process including preculture, osmoprotection, cryoprotection and unloading for high recovery of *Rubia akane* hairy roots which are very sensitive to cytotoxicity of cryoprotectant solutions.

Benson and Hamill (1991) were the first to perform cryopreservation of hairy roots of *Beta vulgaris* and *Nicotiana rustica*, using programmed and ultra-rapid freezing. Since then, cryopreservation of isolated roots of various plant species has been reported using different techniques (Table 3.5) such as programmed freezing (Teoh et al. 1996), vitrification (Jung et al. 2001), encapsulation–dehydration (Hirata et al. 2002; Lambert et al. 2009) and encapsulation– vitrification (Xue et al. 2008). Earlier few study reported cryopreservation of root cultures by conventional programmed freezing or vitrification method, however encapsulation-dehydration technique is used more frequently. In this method encapsulation provides protection against physical damage and does not require toxic cryoprotectants (Xue et al. 2008).

Table 3.5 Cryopreservation of isolated root cultures

Species	Conservation method	Survival[a]	Post-storage stability	References
Adventitious root cultures				
Hyoscyamus niger	Vitrification	93.3 %	Tropane alkaloid content retained at control level	Jung et al. (2001)
Panax ginseng	Vitrification	60 % (32.5 % actual root formation)	Production of 11 ginsenosides retained at control level or above it	Oh et al. (2009), Popova et al., unpublished data
Hairy root cultures				
Astragalus membranaceus	Enc-Vitrif.	5.6 %	Not given	Xue et al. (2008)
Ajuga reptance	Enc.-Deh.	20 %	Not given	Hirata et al. (2002)
Armoracia rusticana	Enc.-Deh.	60 %	Not given	Hirata et al. (1998)
Artemisia annua	Prog. freezing	65 %	Artemisinin content retained at control level	Teoh et al. (1996)
Atropa belladonna	Vitrification	83 %	Tropane alkaloid content retained at control level	Touno et al. (2006)
Beta vulgaris	Prog. freezing to 0 °C then directly to LN	>80 % (6 % actual root conversion)	Betacyanin and betaxanthin content retained at control level	Benson and Hamill (1991)
Eruca sativa	Ultra-Rapid	28.2 % (3.1 % actual root formation)	Not given	Xue et al. (2008)
	Enc.-Vitrif.	73.3 %	Not given	Xue et al. (2008)
Gentiana macrophylla	Enc.-Vitrif.	No regrowth	–	Xue et al. (2008)
Maesa lanceolata	Vitrification	No regrowth	–	Lambert et al. (2009)
	Enc-Deh.	90 %	Not given	Lambert et al. (2009)
Medicago truncatula	Vitrification	No regrowth	–	Lambert et al. (2009)
	Enc.-Deh.	53 %	Not given	Lambert et al. (2009)
Nicotiana rustica	Prog. freezing to 0 °C then directly to LN	23 %	Alkaloid production retained at control level	Benson and Hamill (1991)
Panax ginseng	Vitrification	60 %	Content of five ginsenosides retained at control level	Yoshimatsu et al. (1996)
Rubia akane	Droplet vitrif.	89 %	Not given	Kim et al. (2012b, 2014)
Vinca minor	Enc.-Deh.	>70 %	Vincamine content retained at control level	Hirata et al. (2002)

Modified from Popova et al. (2011)

[a]Mean survival after cryopreservation using an optimized protocol; *Enc.-Vitrif.*: encapsulation-vitrification; *Enc.-Deh.*: encapsulation-dehydration; *Prog. freezing*: programmed freezing

The droplet-vitrification technique developed by Panis et al. (2005) was proven to be very efficient with hairy roots of several species. An efficient protocol for the cryopreservation of madder (*Rubia akane* Nakai) hairy root cultures was developed using droplet-vitrification method (Kim et al. 2012b, 2014). In this method, also known as *foil vitrification*, the explants are dehydrated with highly concentrated vitrification solutions (VSs) and cryopreserved in tiny drops of VS attached onto aluminum foil strips which ensures higher cooling/rewarming rates compared to cryovials (Towill and Bonnart 2003). There is a possibility of combining different procedures. For *Dianthus caryophyllus* a combined encapsulation and droplet-vitrification method was applied with the use of V-Cryo-plate technique (Sekizawa et al. 2011). So far, the combined techniques remain less popular with root cultures but may be used widely in the future with the development of optimized procedures.

Regrowth assessment is critical to optimize the protocol for cryopreservation. The use of staining method to detect the viability without considering proliferation and elongation of explant may be misleading. Benson and Hamill (1991) reported a low percentage (3.1 %) of root development after ultra-rapid freezing of *Beta vulgaris*, even with higher (73.3 %) viability with FDA stain.

3.6 Cryopreservation of Pollen for Breeding

Pollen storage is of key importance to plant breeders and horticulturists involved in fruit and forest tree improvement. Pollen culture techniques have been used for decades to obtain haploids or homozygous diploid plants from various plant species including wheat, maize, rubber tree, apple, poplar and medicinal plants (Maheshwari et al. 1982). A consistent supply of viable pollen provided by pollen banks removes seasonal, geographical or physiological limitations of hybridization programs and supports hybrid development between genera and species. In addition, pollen storage could be critical in saving endangered or threatened plant species without harvesting plants from their natural habitats. There is a vast pool of studies providing strong evidence of a dramatic decline of pollinator species in Europe and North America which in turn leads to a significant reduction in seed production and decrement in heterozygosity of plant populations (Kearns and Inouye 1997). This negatively affects fruit production and may ultimately contribute to the complete loss of economically important genetic diversity (Hanna and Towill 1995).

Traditional pollen bank requires large field areas for maintaining plants at different stages to synchronize flowering, which is both costly and laborious. Other well-known methods of pollen banking for breeding purpose include refrigeration at 3–5 °C for short-term storage, freeze-drying and freeze-storage, vacuum-drying and cold storage in organic solvents (Bajaj 1987). However, frequently observed sharp reduction in pollen viability over the storage period is of major concern (Bajaj 1987). The majority of pollen cryopreservation studies were accomplished by conventional programmed freezing and freeze-drying methods, and were mainly limited to mature pollen. However vitrification has also been tested successfully for pollen cryopreservation in *Brassica* (Xu et al. 1997).

Generally low temperature (−20 °C) is not effective for the long-term storage of pollen whereas storage at −196 °C seems to be more reliable.

Cryopreservation of pollen has been considered the most efficient option for long-term conservation (Parton et al. 2002). Moreover, cryopreserved pollen can be easily transported across the countries as they are usually subjected to less stringent quarantine restrictions (Ganeshan et al. 2008). One of the major advantages of cryopreserving pollen is that it is much easier to handle compared to seeds or other plant tissue due to low water content, highly resistant exine and nonvacuolated cells (Bajaj 1987). Pollen cryopreservation normally does not require any expensive cryostats as pollen grains can be directly immersed in liquid nitrogen for long-term storage. They naturally contain many soluble antioxidants and enzymes. The outer layer of pollen is coated with sporopollenin, which can effectively withstand the oxidative stress during cryopreservation (Smirnova et al. 2009; 2012). Xu et al. (2014) reported that increasing generation of reactive oxygen species (ROS) during cryopreservation may help in improving pollen viability as recent study have shown that ROS may stimulate pollen germination and tube growth (Speranza et al. 2012).

Pollen viability after freezing at −180 °C was first reported by Konwlton in 1922. Since then cryopreservation of pollen has been successful with different plant species, including *Clianthus formosus* (Hughes et al. 1991), *Carica papaya* (Ganeshan 1986), *Juglans regia* (Luza and Polito 1985), *Panax ginseng* (Zhang et al. 1993), *Solanum tuberosum* (Weatherhead et al. 1978), Rose (Rajasekharan and Ganeshan 1994), *Citrus limon* (Rajasekharan et al. 1995), *Gladiolus* (Rajasekharan et al. 1994) and native orchid species (Marks et al. 2014). However, there are limited studies showing that pollen viability can be maintained beyond 1 year of storage in liquid nitrogen (Xu et al. 2014). Visser (1955) successfully germinated tomato pollen stored at −190 °C for 2 years and at −20 °C for 3 years. Successful storage of pollen obtained from *Lycopersicon hirsutum* and *L. peruvianum* at −80 °C for 10 months suggests that *Solanum* species may also be amenable to cryopreservation (Sacks and Clair 1997). However, the cryobiology of pollen is not well understood. Viability of pollen subjected to cryopreservation depends on a number of factors such as pollen moisture content, freezing and thawing procedure, physiological stage of the mother plant, flowering stage, plant vigor and genotypic differences (Crips and Grout 1984; Ganeshan et al. 2008). Aside from genotype-dependent variations in pollen storage behavior and longevity, the stage of pollen development and the method of collection may be important (Stanley and Linskens 1974). Dry-mature pollen grains freshly dehisced from anthers of healthy plants are in an ideal physiological condition for cryopreservation. Lowering the moisture content and ensuring low relative humidity at the time of collection can increase viability of cryopreserved pollen. Generally, pollen moisture content below 35 % is suitable for cryopreservation (Ichikawa et al. 1970); however, relative humidity below 50 % can be harmful and may reduce viability (Daniel 1955) although these parameters may vary from species to species. Binucleate pollen has been shown to be more tolerant to desiccation and storage due to thicker walls as compared to trinucleate pollen (Bajaj 1987).

Considerable interest exists for cryopreservation of orchid pollen as a tool to support both breeding practices and conservation of endangered species (Bernhardt and

Edens-Meier 2010; Vendrame et al. 2008). It is particularly important for the orchid industry, which is known to consistently demand novelty. In orchids, pollen grains are fused into a structure called the *pollinium*. Orchid pollen storage may resolve the issue with temporal and spatial separation between periods of sexual reproduction and can also be used for haploid plant development. Earlier conventional pollen storage methods involving low temperature and low humidity have been developed for several species (Ito 1965; Niimi and Shiokawa 1992; Towill 1985; Yates et al. 1991). Successful cryopreservation of orchid pollen using different methods has also been reported (Marks et al. 2014; Meeyot and Kamemoto 1969; Pritchard and Prendergast 1989; Seaton 1994; Shijun 1984; Vendrame et al. 2008).

3.6.1 General Steps for Pollen Cryopreservation

The first important step in the cryopreservation protocol is the collection of dry mature pollen pretested to verify the viability prior to cryopreservation. Generally pollen does not require any pre-treatment with cryoprotectant before cryopreservation. Pollen samples are kept either in gelatin capsules or butter paper pouches, sealed in airtight poly aluminum pouches, loaded to canisters and cryopreserved. The rewarming procedure of pollen samples is carried out at ambient temperature for 30–60 min followed by viability test and field pollination. Depending upon the purpose of storage, a variety of tissues such as microspore, pollen embryo and whole anthers can be stored in cryopreservation facilities.

Viability of stored pollen can be determined by using acetocarmine-fuchsin/triphenyl tetrazolium chloride, hanging drop technique, PEG technique or in vitro germination. However, fertilization and seeds setting are more effective and reliable methods of viability testing.

In addition to seeds and tissue cultures, cryopreservation of pollen is another area of considerable importance for the conservation of genetic resources. Thus, the establishment of pollen cryobanks will facilitate a regular supply of pollen, which may reduce the maintenance of orchards and nurseries and allow germplasm storage and international exchange with minimal restrictions. User-friendly database software Polbase is available for accessions stored in pollen cryobank.

3.7 Conclusion and Prospects

Climate change is a new threat for conserving global biodiversity and in combination with other human activity could lead to global mass extinctions. Current approaches to in situ conservation are not likely to be sufficient to address anticipated changes, creating an urgent need to develop new models and technologies. Long-term conservation of plants out of their natural habitat represents a number of challenges as maintaining them in the field requires extensive resources and leaves them prone to natural calamities.

Cryopreservation is regarded as the safest and most cost-effective strategy for long-term conservation of the germplasm of economically-important plant species as well as endangered plants. Development of cryopreservation methods for asexually propagated plants is of crucial importance as seeds have a limited shelf life and need to be periodically replenished. Many species produce seeds that cannot be stored for more than a couple of weeks, and some species rarely or never produce seed. During recent years, a significant progress has been made with the development of new cryopreservation techniques for non-orthodox and vegetatively propagated species. Currently available technologies offer the possibility of cryopreservation for many plants. Properly pretreated plant tissue such as dormant buds, pollen, roots can be stored in LN for an indefinite period of time. However, the efficiency of cryopreservation protocols depends on the genotypes of the plant and may vary not only for the individual plant species but also for cultivars and genetic lines. In addition, the physiological stage of plant material, its tolerance to osmotic, chemical, desiccation and cold stress, and the behavior during cooling and warming significantly affects survival and regeneration capacity of cryopreserved tissues. Thus, much more additional research is needed to further improve the amenability of cryopreservation methodologies as well as the rates of recovery of healthy and normal plants from cryopreserved tissues. In this regard, a systematic and rationalized approach based on sensitivity/tolerance of explants to various steps of the procedure may help in developing efficient commonly applicable cryopreservation protocols for new plant genotypes and species that are at high risk of extinction.

References

Atmakuri AR, Chaudhury R, Malik SK et al (2009) Mulberry biodiversity conservation through cryopreservation. In Vitro Cell Dev Biol Plant 45:639–649

Bachiri Y, Bajon C, Sauvanet A et al (2000) Effect of osmotic stress on tolerance of air-drying and cryopreservation of *Arabidopsis thaliana* suspension cells. Protoplasma 214:227–243

Bajaj YPS (1987) Cryopreservation of pollen and pollen embryos, and the establishment of pollen banks. Int Rev Cytol 107:397–420

Bajaj YPS (1995) Cryopreservation of germplasm of medicinal and aromatic plants. In: Bajaj YPS (ed) Biotechnology in agriculture and forestry, vol 32, Cryopreservation of plant germplasm I. Springer, Berlin, pp 419–434

Baskakova OY, Voinkova NM, Nikishina TV et al (2003) Freezing resistance and cryopreservation of cell strains of *Rhaponticum cartlzamoides* and *Thalictrum minus*. Russ J Plant Physiol 50:666–671

Benson EE (1990) Free radical damage in stored plant germplasm. Int Board Plant Genet Res, Rome

Benson EE (2004) Cryoconserving algal and plant diversity: historical perspectives and future challenges. In: Fuller BJ, Lane N, Benson E (eds) Life in the frozen state. CRC Press, Boca Raton, pp 299–328

Benson EE (2008) Cryopreservation of phytodiversity: a critical appraisal of theory & practice. Crit Rev Plant Sci 27:141–219

Benson EE, Bremner D (2004) Oxidative stress in the frozen plant: a free radical point of view. In: Fuller BJ, Lane N, Benson EE (eds) Life in the frozen state. CRC Press, Boca Raton, pp 205–241

Benson EE, Hamill JD (1991) Cryopreservation and post freeze molecular and biosynthetic stability in transformed roots of *Beta vulgaris* and *Nicotiana rustica*. Plant Cell Tiss Org Cult 24:163–171

Bernhardt P, Edens-Meier R (2010) What we think we know vs. what we need to know about orchid pollination and conservation: *Cypripedium* L. as a model lineage. Bot Rev 76:204–219
Bouman H, De Klerk GJ (1990) Cryopreservation of lily meristems. Acta Hort 266:331–337
Bouman H, Tiekstra A, Petutschnig E et al (2003) Cryopreservation of *Lilium* species and cultivars. Acta Hort 612:147–154
Burchett S, Niven S, Fuller MP (2006) The effect of cold-acclimation on the water relations and freezing tolerance of *Hordeum vulgare* L. Cryo Lett 27:295–303
Butenko RG, Popov AS, Volkova LA et al (1984) Recovery of cell cultures and their biosynthetic capacity after storage of *Dioscorea deltoidea* and *Panax ginseng* cells in liquid nitrogen. Plant Sci Lett 33:285–292
Carpentier SC, Witters E, Laukens K et al (2007) Banana *(Musa spp.)* as a model to study the meristem proteome: acclimation to osmotic stress. Proteomics 7:92–105
Chen XL, Li JH, Xin X et al (2011) Cryopreservation of in vitro-grown apical meristems of *Lilium* by droplet-vitrification. S Afr J Bot 77:397–403
Chetverikova EP (1999) Role of abscisic acid in frost-tolerance of plants and cryopreservation of cultured tissues. Russ J Plant Physiol 46:823–829
Crips P, Grout BWW (1984) Storage of broccoli pollen in liquid nitrogen. Euphytica 33:819–823
Cruz-Cruz CA, González-Arnao MT, Engelmann F (2013) Biotechnology and conservation of plant biodiversity. Resources 2:73–95
Cyr DR (2000) Cryopreservation: roles in clonal propagation and germplasm conservation of conifers. In: Engelmann F, Takagi H (eds) Cryopreservation of tropical plant germplasm – current research progress and applications. JIRCAS/IPGRI, Tsukuba/Rome, pp 261–268
Cyr D, Attree SM, El-Kassaby YA et al (2001) Application of somatic embryogenesis to tree improvement in conifers. In: Morohoshi N, Komamine A (eds) Molecular breeding of woody plants. Proceedings of the international wood biotechnology symposium (IWBS), Narita, Chiba, Japan, 14–17 Mar 2001. Elsevier Science B.V, pp 305–312
Dixit S, Ahuja S, Narula A, Srivastava PS (2004) Cryopreservation: a potential tool for long-term conservation of medicinal plants. In: Srivastava PS, Narula A, Srivastava S (eds) Plant biotechnology and molecular markers. Anamaya Publ, New Delhi, pp 278–288
Ellis D, Skogerboe D, Andre C et al (2006) Implementation of garlic cryopreservation techniques in the national plant germplasm system. Cryo Lett 27:99–106
Engelmann F (1997) In vitro conservation methods. In: Callow JA, Ford-Lloyd BV, Newbury HJ (eds) Biotechnology and plant genetic resources. CAB International, Oxford, UK, pp 119–161
Engelmann F (2004) Plant cryopreservation: progress and prospects. In Vitro Cell Dev Biol Plant 40:427–433
Engelmann F (2014) Cryopreservation of clonal crops: a review of key parameters. Acta Hort 1039:31–39
Engelmann F, Takagi H (eds) (2000) Cryopreservation of tropical plant germplasm: current research progress and applications. JIRCAS/IPGRI, Rome
Escobar-Pérez RH (2005) Aspectos logísticos de manejo y determinación de la estabilidad genética de materiales crioconservados de yuca (*Manihot esculenta* Crantz). MSc. thesis. Facultad de Ciencias Agropecuarias, Universidad Nacional de Colombia-Sede Palmira, Colombia
Fabre J, Dereuddre J (1990) Encapsulation-dehydration: a new approach to cryopreservation of *Solanum* shoot-tips. Cryo Lett 11:413–426
Fahy GM, Macfarlane DR, Angell CA, Meryman HT (1984) Vitrification as an approach to cryopreservation. Cryobiology 21:407–426
Fahy GM, Lilley TH, Linsdell H et al (1990) Cryoprotectant toxicity and cryoprotectant toxicity reduction: in search of molecular mechanisms. Cryobiology 27:247–268
Fang JY, Sacandé M, Pritchard H, Wetten A (2009) Influence of freezable/non-freezable water and sucrose on the viability of *Theobroma cacao* somatic embryos following desiccation and freezing. Plant Cell Rep 28:883–889
FAO (2014) Genebank standards for plant genetic resources for food and agriculture. Rev. ed. Rome. http://www.fao.org/docrep/019/i3704e/i3704e.pdf

Fedorovskii DN, Chernyak ND, Popov AS (1993) The disturbing effects of cryopreservation on the plasma membranes of ginseng cells. Sov Plant Physiol 40:94–98
Fukai S (1990) Cryopreservation of chrysanthemum shoot tips. Sci Hort 45:167–174
Fukai S, Oe M (1990) Morphological observations of chrysanthemum shoot cultured after cryoprotection and freezing. J Jap Soc Hort Sci 59:383–387
Fukai S, Goi M, Tanaka M (1991) Cryopreservation of shoot tips of *Chrysanthemum morifolium* and related species native to Japan. Euphytica 54:201–204
Fukai S, Goi M, Tanaka M (1994) The chimeric structure of the apical dome of chrysanthemum (*Dendranthema grandiflorum* (Ramat.) Kitam.) is affected by cryopreservation. Sci Hort 57:347–351
Fukui K, Shirata K, Niino T, Kashif IM (2011) Cryopreservation of mulberry winter buds in Japan. Acta Hort 908:483–488
Fuller BJ, Lane N, Benson EE (eds) (2004) Life in the frozen state. CRC Press, Boca Raton
Ganeshan S (1986) Viability and fertilising capacity of onion pollen (*Allium cepa* L.) stored in liquid nitrogen. Trop Agric (Trinidad) 63:46–48
Ganeshan S, Rajasekharan PE, Shashikumar S, Decruze W (2008) Cryopreservation of pollen. In: Reed BM (ed) Plant cryopreservation: a practical guide. Springer, New York, pp 443–464
Gupta S (2014) Cryopreservation of germplasm through encapsulation-dehydration technique. Acta Hort 1039:147–153
Hahn EJ, Kim YS, Yu KW et al (2003) Adventitious root cultures of *Panax ginseng* C.A. Meyer and ginsenoside production through large-scale bioreactor system. J Plant Biotech 5:1–6
Halmagyi A, Fischer-Klüver G, Mix-Wagner G, Schumacher HM (2004) Cryopreservation of *Chrysanthemum morifolium* (*Dendranthema grandiflora* Ramat.) using different approaches. Plant Cell Rep 22:371–375
Hanna WW, Towill LE (1995) Long term storage of pollen. In: Janick J (ed) Plant breeding reviews, vol 13. John Wiley & Sons, Inc, Oxford, UK, pp 179–207
Harding K (2004) Genetic integrity of cryopreserved plant cells: a review. Cryo Lett 25:3–22
Heine-Dobbernack E, Kiesecker H, Schumacher HM (2008) Cryopreservation of dedifferentiated cell cultures. In: Reed BM (ed) Plant cryopreservation: a practical guide. Springer, New York, pp 141–165
Hirata K, Goda S, Phunchindawan M et al (1998) Cryopreservation of horseradish hairy root culture by encapsulation-dehydration. J Ferment Bioeng 86:418–420
Hirata K, Mukai M, Goda S et al (2002) Cryopreservation of hairy root cultures of *Vinca minor* (L.) by encapsulation-dehydration. Biotech Lett 24:371–376
Hughes HG, Lee CW, Towill LE (1991) Low-temperature preservation of *Clianthus formosus* pollen. Hort Sci 26:1411–1412
Ito I (1965) Ultra-low temperature storage of pollinia and seeds of orchids. Jpn Orchid Soc Bull 11:4–15
Joshi A, Teng WL (2000) Cryopreservation of *Panax ginseng* cells. Plant Cell Rep 19:971–977
Jung D-W, Sung CK, Touno K et al (2001) Cryopreservation of *Hyoscyamus niger* adventitious roots by vitrification. J Plant Physiol 158:801–805
Kaczmarczyk A, Rokka VM, Keller ERJ (2011) Potato shoot tip cryopreservation. A review. Potato Res 54:45–79
Kartha KK (ed) (1985) Cryopreservation of plant cells and organs. CRC Press, Boca Raton
Kearns CA, Inouye DW (1997) Pollinators, flowering plants, and conservation biology: much remains to be learned about pollinators and plants. Bio Sci 47:297–307
Kim SI, Choi HK, Son JS et al (2001) Cryopreservation of *Taxus chinensis* suspension cell cultures. Cryo Lett 22:43–50
Kim JB, Kim HH, Cho EG et al (2004a) Evolution of sucrose and glycerol concentration in garlic shoot tips during a vitrification procedure. Cryo Lett 25:91–100
Kim JB, Kim HH, Cho EG et al (2004b) Evolution of DMSO concentration in garlic shoot tips during a vitrification procedure. Cryo Lett 25:91–100
Kim HH, Lee JK, Yoon JW et al (2006a) Cryopreservation of garlic bulbil primordia by the droplet-vitrification procedure. Cryo Lett 27:143–153

Kim HT, Yang BH, Park YG (2006b) Cryopreservation of somatic embryos of soapberry (*Sapindus mukorossi* Gaertn.) by vitrification. Korean J Plant Res 19:665–669
Kim HH, Lee YG, Shin DJ et al (2009) Development of alternative plant vitrification solutions in dropletvitrification procedures. Cryo Lett 30:320–324
Kim HH, Popova E, Shin DJ et al (2012a) Cryobanking of Korean *Allium* germplasm collections: results from a 10 year experience. Cryo Lett 33:45–57
Kim HH, Popova EV, Shin DJ et al (2012b) Development of a droplet-vitrification protocol for cryopreservation of *Rubia akane* (Nakai) hairy roots using a systematic approach. Cryo Lett 33:506–517
Kim HH, Kong HJ, Shin DJ et al (2014) Development of droplet-vitrification protocol for madder hairy roots: a systematic approach using alternative cryoprotectant solution. Acta Hort 1039:107–112
Knowlton HE (1922) Studies in pollen, with special reference to longevity. Mem Cornell Univ Agric Exp Stn 52:747–793
Krivokharchenko AS, Chernyak ND, Nosov AM (1999) Cryopreservation of suspension cultures of plant cells by the freezing technique elaborated for mammalian embryos. Rus J Plant Physiol 46:831–834
Lambert E, Goossens A, Panis B et al (2009) Cryopreservation of hairy root cultures of *Maesa lanceolata* and *Medicago truncatula*. Plant Cell Tiss Org Cult 96:289–296
Lee YG, Popova E, Cui HY et al (2001) Improved cryopreservation of chrysanthemum (*Chrysanthemum morifolium*) using droplet-vitrification. Cryo Lett 32:487–497
Luza JG, Polito VS (1985) In vitro germination and storage of English walnut pollen. Sci Hort 27:303–316
Lyons JM, Raison JK, Steponkus PL (1979) The plant membrane in response to low temperature: an overview. In: Lyons LM, Graham D, Raison JK (eds) Low temperature stress in crop plants. The role of the membrane. Acad Press Inc, London, pp 1–24
Maheshwari SC, Rashid A, Tyagi AK (1982) Haploids from pollen grains – retrospect and prospect. Am J Bot 69:865–879
Malik S, Cusidó RM, Mirjalili MH et al (2011) Production of the anticancer drug taxol in *Taxus baccata* suspension cultures: a review. Process Biochem 46:23–34
Mannonen L, Toivonen L, Kauppinen V (1990) Effects of long-term preservation on growth and productivity of *Panax ginseng* and *Catharanthus roseus* cell cultures. Plant Cell Rep 9:173–177
Marks TR, Seaton PT, Pritchard HW (2014) Desiccation tolerance, longevity and seed-siring ability of entomophilous pollen from UK native orchid species. Ann Bot 114:561–569
Martín C, González-Benito ME (2005) Survival and genetic stability of *Dendranthema grandiflora* Tzvelev shoot apices after cryopreservation by vitrification and encapsulation-dehydration. Cryobiology 51:281–289
Martín C, González-Benito ME (2009) Cryopreservation and genetic stability of *Dendranthema grandiflora* Tzvelev in vitro cultures. Agric Food Sci 18:129–135
Martín C, Cervera MT, González-Benito ME (2011) Genetic stability analysis of chrysanthemum (*Chrysanthemum* × *morifolium* Ramat) after different stages of an encapsulation–dehydration cryopreservation protocol. J Plant Physiol 168:158–166
Matsumoto T, Niino T (2014) The development of plant vitrification solution 2 and recent PVS2-based vitrification protocols. Acta Hort 1039:21–28
Matsumoto T, Sakai A (1995) An approach to enhance dehydration tolerance of alginate-coated dried meristems cooled to −196°C. Cryo Lett 16:299–306
Matsumoto T, Sakai A, Yamada K (1995) Cryopreservation of in vitro-*grown* apical meristems of lily by vitrification. Plant Cell Tiss Org Cult 41:237–241
Mazur P (1984) Freezing of living cells: mechanisms and implications. Am J Phys (Cell Physiol 16) 247:125–142
Meeyot W, Kamemoto H (1969) Studies on storage of orchid pollen. Am Orchid Soc Bull 38:388–393

Meryman YT (1974) Freezing injury and its prevention in living cells. Annu Rev Biophys Bioeng 3:341–363
Miao NH, Kaneko Y, Sugawara Y (2005) Ultrastructural implications of pretreatment for successful cryopreservation of *Oncidium* protocorm-like body. Cryo Lett 26:333–340
Micuła A, Tykarska T, Kuraś M (2005) Ultrastructure of *Gentiana tibetica* proembryogenic cells before and after cooling treatments. Cryo Lett 26:367–378
Mikula A (2006) Comparison of three techniques for cryopreservation and reestablishment of long-term *Centiana tibetica* suspension culture. Cryo Lett 27:269–282
Niimi Y, Shiokawa Y (1992) A study on the storage of *Lilium* pollen. J Japan Soc Hort Sci 61:399–403
Nishizawa S, Sakai A, Amano Y, Matsuzawa T (1993) Cryopreservation of asparagus *(Asparagus officinalis* L.) embryogenic suspension cells and subsequent plant regeneration by vitrification. Plant Sci 91:67–73
Nosov AM (1999) Plant cell culture: unique system, model, and tool. Russ J Plant Physiol 46:731–738
Nosov AM (2012) Application of cell technologies for production of plant-derived bioactive substances of plant origin. Appl Biochem Microbiol 48:609–624
Nosov AM, Popova EV, Kochkin DV (2014) Isoprenoid production via plant cell cultures: biosynthesis, accumulation and scaling-up to bioreactors. In: Paek KY, Murthy HN, Zhong JJ (eds) Production of biomass and bioactive compounds using bioreactor technology. Springer Science + Business Media, Dordrecht
Oh SY, Wu CH, Popova E et al (2009) Cryopreservation of *Panax ginseng* adventitious roots. J Plant Biol 52:348–354
Paek KY, Chakrabarty D, Hahn EJ (2005) Application of bioreactor systems for large scale production of horticultural and medicinal plants. Plant Cell Tiss Org Cult 81:287–300
Paek KY, Murthy HN, Hahn EJ, Zhong JJ (2009) Large scale culture of ginseng adventitious roots for production of ginsenosides. Adv Biochem Eng Biotechnol 113:151–176
Panis B, Lambardi M (2006) Status of cryopreservation technologies in plants (crops and forest trees). In: Ruane J, Sonnino A (eds) The role of biotechnology in exploring and protecting agricultural genetic resources. Food and Agriculture Organization of the United Nations, Rome, pp 61–78
Panis B, Thinh NT (2001) Cryopreservation of *Musa* germplasm. In: Escalant JV, Sharrock S (eds) INIBAP technical guideline 5. International Network for the Improvement of Banana and Plantain, Montpellier
Panis B, Piette B, Swennen R (2005) Droplet-vitrification of apical meristems: a cryopreservation protocol applicable to all *Musaceae*. Plant Sci 168:45–55
Park SU, Kong H, Shin DJ et al (2014) Development of vitrification protocol in *Rubia akane* (Nakai) hairy roots using a systematic approach. Cryo Lett 35:138–144
Parton E, Vervaeke I, Delen R et al (2002) Viability and storage of bromeliad pollen. Euphytica 125:155–161
Pence VC (1995) Cryopreservation of recalcitrant seeds. In: Bajaj YPS (ed) Biotechnology in agriculture and forestry, vol 32, Cryopreservation of plant germplasm I. Springer Verlag, Berlin, pp 29–52
Pence VC (2011) Evaluating costs for the in vitro propagation and preservation of endangered plants. In Vitro Cell Dev Biol Plant 47:176–187
Popova EV, Lee EJ, Wu CH et al (2009) A simple method for cryopreservation of *Ginkgo biloba* callus. Plant Cell Tiss Org Cult 97:337–343
Popova E, Kim HH, Paek KY (2010) Cryopreservation of coriander (*Coriandrum sativum* L.) somatic embryos using sucrose preculture and air desiccation. Sci Hort 124:522–528
Popova EV, Paek KY, Kim HH (2011) Cryopreservation of medicinal plants: the case of in vitro cultures. In: Kumar A, Roy S (eds) Plant tissue culture and applied plant biotechnology. Pointer publishers, Jaipur, pp 153–196
Pritchard HW (2007) Cryopreservation of desiccation-tolerant seeds. In: Day JG, Stacey GN (eds) Methods in molecular biology, vol 368, 2nd edn, Cryopreservation and freeze-drying protocols. Humana Press Inc, Totowa, pp 185–201

Pritchard HW, Nadarajan J (2008) Cryopreservation of orthodox (desiccation tolerant) seeds. In: Reed BM (ed) Plant cryopreservation: a practical guide. Springer, Berlin, pp 485–501
Pritchard HW, Prendergast FG (1989) Factors influencing the germination and storage characteristics of orchid pollen. In: Pritchard HW (ed) Modern methods in orchid conservation: the role of physiology, ecology and management. Cambridge University Press, Cambridge, pp 1–16
Rajasekharan PE, Ganeshan S (1994) Freeze preservation of rose pollen in liquid nitrogen: feasibility, viability and fertility status after long-term storage. J Hort Sci 69:565–569
Rajasekharan PET, Rao M, Janakiram T, Ganeshan S (1994) Freeze preservation of *Gladiolus* pollen. Euphytica 80:105–109
Rajasekharan PE, Ganeshan S, Thamizharasu V (1995) Expression of trifoliate leaf character in *Citrus limonia* × *Poncirus trifoliata* hybrids through cryostored pollen. J Hort Sci 70:485–490
Ramon M, Geuns JMC, Swennen R, Panis B (2002) Polyamines and fatty acids in sucrose precultured banana meristems and correlation with survival rate after cryopreservation. Cryo Lett 23:345–352
Reed BM (2001) Implementing cryogenic storage of clonally propagated plants. Cryo Lett 22:97–104
Reed BM (ed) (2008) Plant cryopreservation: a practical guide. Springer, New York
Reinhoud PJ, Van Iren FV, Kijne JW (2000) Cryopreservation of undifferentiated plant cells. In: Engelmann F, Takagi H (eds) Cryopreservation of tropical plant germplasm: current research progress and applications. JIRCAS/IPGRI, Rome, pp 91–102
Ren L, Zhang D, Jiang XN et al (2013) Peroxidation due to cryoprotectant treatment is a vital factor for cell survival in *Arabidopsis* cryopreservation. Plant Sci 212:37–47
Ren L, Zhang D, Shen XH, Reed BM (2014) Antioxidants and anti-stress compounds improve the survival of cryopreserved *Arabidopsis* seedlings. Acta Hort 1039:57–60
Sacks EJ, St Clair DA (1997) Cryogenic storage of tomato pollen: effect on fecundity. Hort Sci 31:447–448
Sakai A, Engelmann F (2007) Vitrification, encapsulation-vitrification and droplet-vitrification: a review. Cryo Lett 28:151–172
Sakai A, Kobayashi S, Oiyama I (1990) Cryopreservation of nucellar cells of navel orange *(Citrus sinensis* Osb. var. *brasiliensis* Tanaka) by vitrification. Plant Cell Rep 9:30–33
Sakai A, Matsumoto T, Hirai D, Niino T (2000) Newly developed encapsulation-dehydration protocol for plant cryopreservation. Cryo Lett 21:53–62
Salma M, Engelmann-Sylvestre I, Collin M et al (2014) Effect of the successive steps of a cryopreservation protocol on the structural integrity of *Rubia akane* Nakai hairy roots. Protoplasma 251:649–659
Samar F, Mujib A, Nasim SA, Siddiqui ZH (2009) Cryopreservation of embryogenic cell suspension of *Catharanthus roseus* L. (G) Don. Plant Cell Tiss Org Cult 98:1–9
Seaton P (1994) Orchid seed and pollen storage. Am Orchid Soc Bull 63:918–922
Seitz U, Reinhard E (1987) Growth and ginsenoside patterns of cryopreserved *Panax ginseng* cell cultures. J Plant Physiol 131:215–223
Sekizawa K, Yamamoto S, Rafique T et al (2011) Cryopreservation of in vitro–grown shoot tips of carnation (*Dianthus caryophyllus* L.) by vitrification method using aluminium cryo-plates. Plant Biotech 28:401–405
Shijun C (1984) The study of keeping freshness of orchid pollinia. Acta Hort Sin 11:279–280
Shin DJ, Kong HJ, Popova EV et al (2012) Cryopreservation of *Kalopanax septemlobus* embryogenic callus using vitrification and droplet-vitrification. Cryo Lett 33:402–410
Smetanska I (2008) Production of secondary metabolites using plant cell cultures. In: Stahl U, Donalies UB, Nevoigt E (eds) Food biotechnology, vol 111, Advances in biochemical engineering/biotechnology. Springer, Berlin/Heidelberg/, pp 187–228
Smirnova AV, Matveyeva NP, Polesskaya OG, Yermakov IP (2009) Generation of reactive oxygen species during pollen grain germination. Russ J Dev Biol 40:345–353
Smirnova AV, Timofeyev KN, Breygina MA et al (2012) Antioxidant properties of the pollen exine polymer matrix. Biophysics 57:174–178
Speranza A, Crinelli R, Scoccianti V, Geitmann A (2012) Reactive oxygen species are involved in pollen tube initiation in kiwifruit. Plant Biol 14:64–76

Stanley RG, Linskens HF (1974) Pollen: biology, biochemistry and management. Springer, Berlin
Stanwood PC, Bass LN (1981) Seed germplasm preservation using liquid nitrogen. Seed Sci Tech 9:423–437
Steponkus PL (ed) (1992) Advances in low temperature biology, vol 1. JAI Press Ltd, London, pp 1–61
Suzuki M, Tandon P, Ishikawa M, Toyomasu T (2008) Development of a new vitrification solution, VSL, and its application to the cryopreservation of gentian axillary buds. Plant Biotech Rep 2:123–131
Teixeira da Silva JAT (2003) Chrysanthemum: advances in tissue culture, cryopreservation, post-harvest technology genetics and transgenic biotechnology. Biotech Adv 21:715–766
Teixeira da Silva JAT (2004) Ornamental chrysanthemums: improvement by biotechnology. Plant Cell Tiss Org Cult 79:1–18
Teoh KH, Weathers PJ, Cheetham RD, Walcerz DB (1996) Cryopreservation of transformed (hairy) roots of *Artemisia annua*. Cryobiology 33:106–117
Titova MV, Shumilo NA, Chernyak ND et al (2007) Use of cryoconservation to maintaining strain stability upon outfit culturing of *Polyscias filicifolia* Bailey cell suspension I. Growth characteristics of restored culture. Biotekhnologiya 5:60–65 (in Russian)
Titova MV, Berkovich EA, Reshetnyak OV et al (2011) Respiration activity of suspension cell culture of *Polyscias filicifolia* Bailey, *Stephania glabra* (Roxb.) miers, and *Dioscorea deltoidea* wall. Appl Biochem Microbiol 47:87–92
Touchell DH, Dixon KW (1994) Cryopreservation for seedbanking of Australian species. Ann Bot 40:541–546
Touno K, Yoshimatsu K, Shimomura K (2006) Characteristics of *Atropa belladonna* hairy roots cryopreserved by vitrification method. Cryo Lett 27:65–72
Towill LE (1985) Low temperature and freeze-vacuum-drying preservation of pollen. In: Kartha KK (ed) Cryopreservation of plant cells and organs. CRC Press, Boca Raton, pp 171–198
Towill LE, Bajaj YPS (2002) Cryopreservation of plant germplasm II. Biotechnology in agriculture and forestry, vol 50. Springer, Berlin
Towill LE, Bonnart R (2003) Cracking in a vitrification solution during cooling or warming does not affect growth of cryopreserved mint shoot tips. Cryo Lett 24:341–346
Towill LE, Forsline PL, Walters C et al (2004) Cryopreservation of *Malus* germplasm using a winter vegetative bud method: results from 1915 accessions. Cryo Lett 25:323–334
Turner SR, Senaratna T, Bunn E et al (2001) Cryopreservation of shoot tips from six endangered Australian species using a modified vitrification protocol. Ann Bot 87:371–378
Uchendu EE, Leonard SW, Traber MG, Reed BM (2010a) Vitamins C and E improve regrowth and reduce lipid peroxidation of blackberry shoot tips following cryopreservation. Plant Cell Rep 29:25–35
Uchendu EE, Muminova M, Gupta S, Reed BM (2010b) Antioxidant and anti-stress compounds improve regrowth of cryopreserved *Rubus* shoot tips. In Vitro Cell Dev Biol Plant 46:386–393
Uchendu EE, Shukla MR, Reed BM, Saxena PK (2013) Melatonin enhances the recovery of cryopreserved shoot tips of American elm (*Ulmus americana* L.). J Pineal Res 55:435–442
Uemura M, Steponkus PL (1999) Cold acclimation in plants: relationship between the lipid composition and the cryostability of the plasma membrane. J Plant Res 112:245–254
Vendrame WA, Carvalho VS, Dias JMM, Maguire I (2008) Pollination of *Dendrobium* hybrids using cryopreserved pollen. Hort Sci 43:264–267
Verpoorte R, van der Heijden R, Memelink J (2000) Engineering the plant cell factory for secondary metabolite production. Transgen Res 9:323–343
Visser T (1955) Germination and storage of pollen. Meded Landbouwhogesch Wageningen 55:1–68
Volk GM, Walters C (2006) Plant vitrification solution 2 lowers water content and alters freezing behavior in shoot tips during cryoprotection. Cryobiology 52:48–61
Volk GM, Henk AD, Bonnart RM et al (2014a) Plant shoot tip response to treatment with plant vitrification solution #2. Acta Hort 1039:81–84

Volk GM, Shepherd A, Borrart RM (2014b) Strategies for improved efficiency when implementing plant vitrification techniques. Acta Hort 1039:85–89
Wang QC, Panis B, Engelmann F et al (2009) Cryotherapy of shoot tips: a technique for pathogen eradication to produce healthy planting materials and prepare healthy plant genetic resources for cryopreservation. Ann Appl Biol 154:351–363
Wang QC, Wang RR, Li BQ, Cui ZH (2012) Cryopreservation: a strategy technique for safe preservation of genetically transformed plant materials. Adv Genet Eng Biotechnol 1:1–2
Wang B, Wang RR, Cui ZH et al (2014a) Potential applications of cryogenic technologies to plant genetic improvement and pathogen eradication. Biotech Adv 32:583–595
Wang RR, Gao XX, Chen L et al (2014b) Shoot recovery and genetic integrity of *Chrysanthemum morifolium* shoot tips following cryopreservation by droplet-vitrification. Sci Hort 176:330–339
Weatherhead MA, Grout BWW, Henshaw GG (1978) Advantages of storage of potato pollen in liquid nitrogen. Potato Res 21:331–334
Wink M, Alfermann AW, Franke R et al (2005) Sustainable bioproduction of phytochemicals by plant in vitro cultures: anticancer agents. Plant Genet Res 3:90–100
Withers LA (1985) Cryopreservation of cultured plant cells and protoplasts. In: Kartha KK (ed) Cryopreservation of plant cells and organs. CRC Press, Boca Raton, pp 243–267
Withers LA (1987) The low temperature preservation of plant cell, tissue and organ cultures and seed for genetic conservation and improved agriculture. In: Grout BWW, Morris GJ (eds) The effects of low temperatures on biological systems. Edward Arnold, London, pp 389–409
Xu B, Han H, Zheng C, Sun M (1997) Cryopreservation of pollen by vitrification in *Brassica*. Wuhan Univ J Nat Sci 2:120–123
Xu J, Liu Q, Jia M et al (2014) Generation of reactive oxygen species during cryopreservation may improve *Lilium* × *siberia* pollen viability. In Vitro Cell Dev Biol Plant 50:369–375
Xue SH, Luo XJ, Wu ZH et al (2008) Cold storage and cryopreservation of hairy root cultures of medicinal plant *Eruca sativa* Mill. *Astragalus membranaceus* and *Gentiana macrophylla* Pall. Plant Cell Tiss Org Cult 92:251–260
Yates IE, Sparks D, Connor K, Towill L (1991) Reducing pollen moisture simplifies long-term storage of pecan pollen. J Amer Soc Hort Sci 116:430–434
Yin ZF, Bi WL, Chen L et al (2014) An efficient, widely applicable cryopreservation of *Lilium* shoot tips by droplet vitrification. Acta Phys Plant 36:1683–1692
Yoshimatsu K, Yamaguchi H, Shimomura K (1996) Traits of *Panax ginseng* hairy roots after cold storage and cryopreservation. Plant Cell Rep 15:555–560
Zalewska M, Kulus D (2013) Cryopreservation of in vitro-grown shoot tips of chrysanthemum by encapsulation-dehydration. Folia Hort 25:133–140
Zhang LX, Chang WC, Wei YJ et al (1993) Cryopreservation of ginseng pollen. Hort Sci 28:742–743
Zhao Y, Qi LW, Wang WM et al (2011) Melatonin improves the survival of cryopreserved callus of *Rhodiola crenulata*. J Pineal Res 50:83–88
Zhu GY, Geuns JMC, Dussert S et al (2006) Change in sugar, sterol and fatty acid composition in banana meristems caused by sucrose-induced acclimation and its effects on cryopreservation. Physiol Plant 128:80–94

Chapter 4
Mutants as a Genetic Resource for Future Crop Improvement

Mohd Gulfishan, Tariq Ahmad Bhat, and Mohammad Oves

Abstract Enhanced crop production is seriously needed to meet the challenges of food security imposed by the rapidly increasing human population around the globe. Unpredictable climatic conditions, depleting water resources and finite arable land limit crop production. Thus the development of new crop varieties with improved yield and resistant to biotic and abiotic stresses will make a vital contribution to food security. Induced mutagenesis has played a pivotal role in ensuring food security by creating 3218 mutant varieties around the world. Mutagenesis combined with advanced molecular biology techniques and in vitro culture methods have resulted in enhanced food production. Mutant germplasm resources have been developed for different crop plants and are freely available to speed up crop improvement programs. These mutant resources are also being used for functional genomics studies, molecular breeding and a greater understanding of the molecular basis of other biological process.

Keywords Crop improvement • Food security • Genomics • Induced mutagenesis • Mutant • Mutation breeding

M. Gulfishan (✉)
Assistant Professor, Faculty of Life and Allied Health Sciences, Glocal University, Mirzapur Pole, Saharanpur 247121, Uttar Pradesh, India
e-mail: fishan.amu@gmail.com

T.A. Bhat
Department of Education, Government of Jammu and Kashmir, Jammu and Kashmir, India
e-mail: bhattariq110@gmail.com

M. Oves
Center of Excellence in Environmental Studies, King Abdul Aziz University, Jeddah, Saudi Arabia
e-mail: owais01m@gmail.com

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_4

4.1 Introduction

Globally, the current human population is increasing day by day and expected to reach 9 billion by 2050 and that will lead to food scarcity on earth since 70 % more food will be required to meet this challenge. To overcome this increasing demand for food and proper nourishment, an improvement in food production is urgently needed (Ronald 2014). According to the Food and Agriculture Organization (FAO) of the United Nations, food security *exists when all people, at all times, have physical, social and economic access to sufficient safe and nutritious food that meets their dietary needs and food preferences for an active and good life* (FAO 2014a).

The envisaged increase in food production is daunting because of limited available arable land, depleting water resource and varying climatic conditions. The difficulties are also compounded by urbanization, salinization, biotic stress, drought and desertification that result in a reduction of arable land. Moreover, changing climatic conditions and subsequent variations also limit food production (UNEP 2002). Agrochemicals cannot be used excessively to meet the challenge of food shortage due to their deleterious impacts on health and environment (Mba 2013). FAO (2011) recommended high-yielding varieties to meet the food shortages, as well as efficient use of input, under the projected climate change and adapted to a wide range of agroecosystems conditions. Food production must be increased through fewer agricultural inputs with maximum environmentally-friendly methods (Schlenker and Roberts 2009). Keeping in view the limited arable land, water resources and variable climatic conditions, new crop varieties and cropping system that use more efficiently the limited resources have to be developed to meet the challenges of food security.

There are different mechanisms for harnessing the heritable variations encoded in the genetic make-up of existing crop plants so as to use them in crop improvement programs. The incorporation of desired traits from non-adapted landraces/crop wild resource can speed up crop improvement. Putative parental material can also be induced to mutate so as to obtain new genes that control desired traits for new crop variety development (Suprasanna et al. 2014).

Among the different strategies to enhance crop improvement programs, induced mutagenesis has contributed immensely by creating mutant varieties with improved and desirable genetic changes in agronomically-important traits of the crop plants. Mutagenesis has become more efficient in combination with advanced molecular biology techniques and in vitro culture methods that result in enhancement of crop improvement/breeding programs (Jain 2010a) particularly under the global climate change (Jain 2010b). Such induced mutagenesis also helps in the mining of new gene alleles that do not occur in the germplasm (Roychowdhury and Tah 2013).

4.2 Induced Mutagenesis

Mutation is a term coined by De Varies (1901) upon the appearance of a new phenotype he noted in the common evening primrose, *Oenothera lamarkiana*, to describe the sudden heritable change in the genotype of an organism; the organisms

with such heritable changes are known as *mutants* (Mba 2013). Mutation is the ultimate source of all genetic changes which provide the raw material for evolution and it is a valuable approach for improvement of economic characters of plants. Such genetic changes can occur spontaneously naturally at a very low rate or experimentally induced by physical and chemical mutagens (Jain 2002, 2010a; Mba et al. 2007). Physical mutagens include radiations such as α-rays, β-rays, fast neutrons, thermal neutrons, x-rays, γ-rays and ultraviolet (UV) radiation. Most common chemical mutagens include alkylating agents, such as ethyl methane sulphonate (EMS); methyl methane sulphonate (MMS); ethylene imines (EI); diethyl sulphate (DES) etc. Also, azides e.g. sodium azide; acridine dyes e.g. acriflavine, acridine orange, proflavin etc.; base analogues e.g. 5 bromouracil, 2 aminopurine, 5 chlorouracil, etc.; and other direct-acting chemicals such as nitrous acid, mustard gas etc.

4.3 Spontaneous Mutation

Spontaneous mutations in crop plants occurs naturally during adaptations and evolutionary processes at an extremely low rate i.e. 10^{-5}–10^{-8}. This frequency is inadequate for creating variations in the genetic architecture of a crop for improvement of desirable traits (Zhong-hua et al. 2014). Wheat, peas and barley are the notable example of mutants derived through heritable permanent change i.e. spontaneous mutations during the course of domestication (Table 4.1). Spontaneous mutations in these plants resulted in eradicated pod or head shattering and the reductions in seed dormancy periods. Other examples of spontaneous mutants include those found in almond, lima bean, watermelon, potato, eggplant, cabbage and several types of nuts (Mba 2013).

High yielding and lodging resistance in wheat varieties were developed by the incorporation of spontaneously-mutated alleles of the genes that resulted in the Green Revolution and subsequently secured food for millions of people around the world. Other examples include utilization of dwarf germplasm Dee-geo-woo-gen from China and the release of rice variety IR8 developed in the Philippines by the International Rice Research Institute (IRRI) from a dwarf line (Mba et al. 2012b). Application of cytoplasmic male sterile (CMS) and photoperiod-sensitive genic male sterile rice lines were utilized to develop hybrid rice seeds for commercial

Table 4.1 Spontaneous mutations and domestication of crops

Mutation that facilitated domestication	Examples
Abolishment of bitterness and toxicity	Almond, lima bean, watermelon, potato, eggplant, cabbage, nuts
Abolishment of the need for sexual reproduction (seedlessness or parthenocarpy)	Banana, grape, orange, pineapple
Loss of natural seed dispersal mechanism (shattering of pods and heads)	Pea, wheat, barley
Loss of the hard seed coat and other germination inhibitors (dormancy)	Wheat, barley, pea
Facility for self-compatible hermaphroditism	Grape, papaya

Source: Mba (2013)

release (Mba and Shu 2012). Therefore, induction of mutations in existing crop plant germplasm will be helpful to create maximum genetic variability, to identify new genes/alleles of desired interest and lead to a scaling-up of crop improvement programs.

4.4 Induced Mutation

Over the last six decades, thousands of new mutant crop varieties with improved agronomic characters have been developed by induced mutations throughout the world (Jain and Maluszynski 2004). After the epoch-making discoveries made by Muller (1927) and Stadler (1928a, b), a large amount of genetic variability has been induced by various mutagens/radiation in different crop plant species that have resulted in the development of more than 3218 officially-released mutant varieties (Figs. 4.1 and 4.2) worldwide in about 224 plant species during the past 60 years (FAO 2014b). Among the mutant varieties the majority are of food crops contributing to environmentally-sustainable food security in the world. In India 343 mutant varieties, by using different mutagens, have been released for cultivation (Tables 4.2 and 4.3).

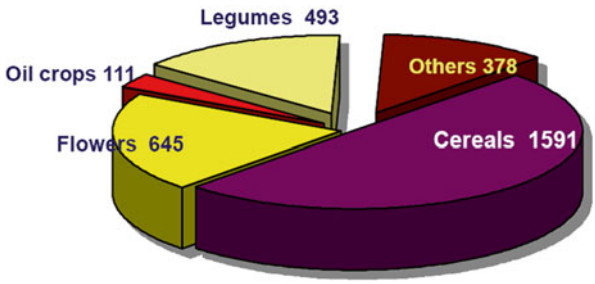

Fig. 4.1 Registered mutant crop varieties (FAO 2014b)

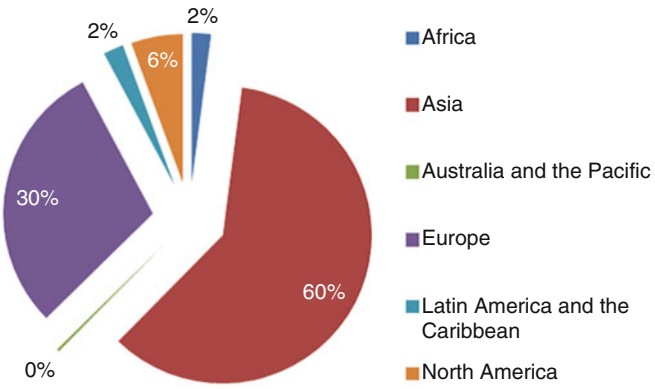

Fig. 4.2 Distributions of mutant crop varieties by continents of official release (Source: FAO 2014b)

Table 4.2 Mutant varieties of different crops released for cultivation in India

Crop	No. of varieties	Specific crop and no. of varieties
Cereals	74	Rice (42), barely (13), pearl millet (5), finger millet (7), foxtail millet (1), wheat (4), sorghum (2)
Pulses	57	Mung bean (15), black gram (9), chickpea (8), cowpea (10), moth bean (5), pea (1), pigeon pea (5), French bean (1) lentil (3)
Oilseeds	44	Groundnut (18), mustard (9), castor bean (4), sesame (5), soybean (7) sunflower (1)
Fiber crops	14	American cotton (8), tossa jute (3), white jute (2), desi cotton (1)
Vegetables	14	Tomato (4), turmeric (2), bitter gourd (1), brinjal (1), green pepper (1), okra (2), ridge gourd (1), snake gourd (1), cluster bean (1)
Cash crops	10	Sugarcane (9), tobacco (1)
Medicinal crops	17	Citronella (9), German chamomile (1), Indian henbane (1), isabgol (2), khasianum (1), opium poppy (2), Spearmint (1)
Fruit trees	2	Mulberry (1), papaya (1)
Forage crops	1	Egyptian clover (1)
Ornamentals	110	Chrysanthemum (49), rose (16), dahlia (11), portulaca (11), bougainvillea (13), wild sage (3), gladiolus (2) Hibiscus sp. (2), tuberose (2) Coleus (1)
Total	343	

Source: Kharkwal and Shu (2009)

Table 4.3 Mutagens used and trait improved in mutant cultivars released in India

Mutagens	No. of mutants	Main attribute	Number of occurrence
Gamma rays	169	High yielding	86
X-rays	26	Early maturity	65
Neutrons	07	Disease resistance	57
EMS	15	Quality character	39
DMS	04	Grain quality	67
Ethylene imine (EI)	02	Abiotic stress resistance	65
Sodium azide (NaN3)	02	Improved plant type	181
Other mutagens	29	Others	9
Cross breeding	47		
Natural mutants	12		

Source: Chopra (2005)

For enhanced food production, there is an urgent need to create genetic variability in the desired crop plant species. This genetic variability could be used in the development of new varieties with increased yield, disease and lodging resistant, modified protein, oil and starch content, increased nutrient uptake, deeper rooting system and resistant to abiotic stresses such a drought, heat and salinity etc. to counter the erratic climate conditions that limit crop production (Jain 2002, 2010b).

4.4.1 Physical Mutagenesis

After the pioneering work using X-rays, by Muller (1927) and Stadler (1928a, b) on fruit fly and maize, scientist began to create genetic variations by using radiation and it became a very prominent method (Ahloowalia et al. 2004). Among physical mutagens, ionizing radiations i.e. gamma rays and X-rays are the most commonly used methods (Mba and Shu 2012; Mba et al. 2012a). Other physical mutagens include alpha (α) and beta (β) particles and fast neutrons to induce genetic changes in crop plants (Table 4.4). These mutagens cause deletion or addition of nucleotides and by substitution of one or more nucleotides for the new combinations of genes (Mba 2013).

With pioneering work in China and Japan (Mei et al. 1994a; Wu et al. 2001), an ion beam, generated by particle accelerators, i.e. cyclotron using 20Ne, 14N, 12C, 7Li, 40Ar, or 56Fe as radioisotope sources, has gained popularity in the induction of mutations. Cosmic radiations have also been used to induce mutations in crop plants (Liu et al. 2007; Mei et al. 1998; Ren et al. 2010). Other researcher found fast neutron bombardment most efficient in mutation induction (Koornneef et al. 1982). Among these, gamma rays are less destructive and result in large deletions (Yuan et al. 2014) or small deletions while translocations, chromosome losses and large deletions have been induced after fast neutron bombardment (Sikora et al. 2011). Li et al. (2001) generated 51,840 lines of *Arabidopsis* population by fast neutron mutagenesis used for screening of mutants, while in rice around 10,000 rice mutant lines have been generated by fast neutron bombardment and around 20,000 lines by γ-ray. (Wu et al. 2005).

4.4.2 Chemical Mutagenesis

Chemical mutagenesis is the use of chemical compounds that can induce mutations (Table 4.5). The demonstration that nitrogen mustard gas caused mutations in the cell (Auerbach 1940, 1947; Auerbach and Robson 1944, 1946) opened the way for identification of chemicals which induce mutations. Plant breeders and geneticists applied chemical mutagens to induce mutations in various crop plants because they are easy to use and no special equipment is required to induce high mutation frequency (Gulfishan et al. 2012, 2013; Henikoff and Comai 2003; Koornneef et al. 1982; Zhu et al. 1995). The chemical mutagen reacts with DNA of the treated seed/cell/tissue or organ and induces somatic and genetic changes and only unrepaired damage to the DNA in initial cells of the sporogenic layer (germ line cells) are transferred as mutations to the next generation. Other mutations in somatic cells of the embryo, including mitotic chromosomal aberrations together with toxic action of mutagen on all components of cytosol, affect plant growth and development, and are called the *somatic effect* of the mutagen (Roychowdhury and Tah 2013). Compared to radiation, chemical mutagens tend to cause single base-pair (bp) changes, or single-nucleotide polymorphisms (SNPs) rather than deletions and translocations of nucleotide (Sikora et al. 2011). Chemicals include base analogues, alkylating agents, azides and others that modify genetic makeup of crop plants. Among chemical mutagens ethyl methane sulphonate (EMS) has been used frequently. It causes point mutations such as nonsense, missense and silent mutations

by chemical modification of nucleotides within the DNA (Jiang and Ramachandran 2010). EMS mainly induces C to T changes that results in substitution of C/G to T/A (Kim et al. 2006; Krieg 1963) and at lower doses it induces G/C to A/T transversions through 7-ethylguanine hydrolysis or A/T to G/C transitions through 3-ethyladenine pairing errors (Greene et al. 2003).

Table 4.4 Different physical mutagens and their properties

Mutagen	Typical frequency (s^{-1})	Typical energy (kJ/m)	Typical photon energy (eV)
Particles			
α-particles		4.1×10^8	
β-particles		1.5×10^7	
Electromagnetic radiation			
Cosmic rays	$6 \times 1021\ s^{-1}$	2.4×10^9	
γ-rays	$3 \times 102\ s^{-1}$	1.2×10^8	1 MeV
x-rays	$3 \times 1017\ s^{-1}$	1.2×10^5	100 keV
Ultraviolet	$3 \times 1015\ s^{-1}$	1.2×10^3	4 eV

Source: Mba (2013)

Table 4.5 Frequently used chemical mutagens and their properties

Chemical mutagenic agent	Mode of action
Base analogues e.g. 5-bromouracil (BU), 5 bromodeoxyuridine, 2-aminopurine (2AP)	Incorporates into DNA in place of the normal bases during DNA replication thereby causing transitions (purine to purine or pyrimidine to pyrimidine)
Nitrous acid	Acts through deamination, the replacement of cytosine byuracil which can pair with adenine and thus from subsequent cycles of replication lead to transitions
Alkylating agents such as: sulfonates e.g. ethylmethanesulfonate (EMS), diethyl sulfonate (DES); Sulphur mustards e.g. ethyl-2-chloroethyl sulphide; Nitrogen mustards e.g. 2-chloroethyl-dimethylamine; and epoxides e.g. ethylene oxide Others are ethyleneimine, hydroxylamine (NH2OH), *N*-methyl-*N*'-nitro-*N* nitrosoguanidine (MNNG), sodium azide and diazomethane	They react with bases and add methyl or ethyl groups and, depending on the affected atom, the alkylated base may then degrade to yield a baseless site, which ismutagenic and recombinogenic or mispair to result in mutations upon DNA replication
Intercalating agents such as acridine orange, proflavin, ethidium bromide	They insert between bases of DNA thereby causing a *stretching* of the DNA duplex and the DNA polymerase in turn recognizes this stretch as an additional base and inserts an extra base opposite this stretched (intercalated) molecule. This results in frameshifts i.e. an alteration of the reading frame since codons are groups of three nucleotides
Miscellaneous group of agents; large molecules referred to as *bulky* lesions (e.g. *N*-acetoxy-*N*-2-acetyl-aminofluorine-NAAAF)	They bind to bases in DNA and cause them to benon coding thereby preventing transcription and DNA replication; they cause intra- and inter-strand crosslinks (e.g. psoralens); they also cause DNA strand breaks (e.g. peroxides)

Source: Mba (2013)

4.5 Other Methods for Induction of Mutation

In recent years, heavy ion irradiation has also attracting attention as an effective method of induction of mutation in plants. These heavy ion beams possess high linear energy transfer (LTE) and relative biological effectiveness (RBE) which is supposed to enhance mutation frequency and also induces wide spectrum of mutations at low doses and short duration of treatments (Abe et al. 2012; Tanaka et al. 2010). Hirano et al. (2012) in *Arabidopsis* and Mei et al. (1994b) in rice investigated the effectiveness of ion beams and concluded that they were more effective as compared to gamma rays and X-rays. In Japan, mutations have been induced by using heavy ion beam irradiations (Hase et al. 2010; Kondo et al. 2009; Tanaka et al. 2010) and in China low energy ion beam have also been used to create mutations (Feng et al. 2009). Bacterial leaf blight and blast resistance in rice (Xiao et al. 2008), yellow mosaic virus in barley (Tanaka et al. 2010) and potato virus Y in tobacco (Hamada et al. 1999) have also been induced by using ion beam radiations. Resistance to sigatoka disease in banana has been induced by carbon ion-beam irradiation (Reyes-Borja et al. 2007). The UVB resistant rice mutant, utr319, was isolated after carbon ion irradiation which was more tolerant to UV than the wild type (Takano et al. 2013). Chundet et al. (2012) created rice mutant expressing short stature and photoperiod insensitivity by using a low-energy N^{+}/N_2^{+} ion beam (Fig. 4.3). A chlorophyll deficient mutant was also isolated in *Arabidopsis thaliana* by carbon ion irradiation (Shikazono et al. 2003). Electron beams induced mutations of leaf shape and color, seed size and shape, trailing, more branching, dwarfing, early or late flowering time and high yield in *Vigna angularisi* (Luo et al. 2012). Tanaka et al. (2010) reviewed the work on ion beam irradiation on different plant and found that the spectrum of mutations induced by ion beams was different than gamma rays. In rice, the mutation spectrum of different ion beams was similar to that of gamma rays but the frequency of mutations was higher (Yamaguchi et al. 2009). Luo et al. (2012) isolated leaf shape mutants in *Vigna angularisi* by electron beam irradiations (Fig. 4.4). In *Chrysanthemum* ray floret mutants has been induced by $^{12}C^{5+}$ ion beam irradiation as shown in Fig. 4.5 (Matsumura et al. 2010).

In recent years, cosmic radiations have also been recognized as a source of induction of mutations in plants and the term *space breeding* has been coined for such breeding programs. Mutations were induced in seeds on space flights (Kranz 1986; Mei et al. 1994b, 1998). Scientist used space craft, satellites and high altitude balloons to carry seeds into space for mutation induction (Li et al.1999), and in the subsequent generation of plants genetic changes have been recorded (Guo et al. 2010; Hu et al. 2010; Ou et al. 2010). Space flight cauliflower experienced changes in plant size and flower head and also resistance to black rot (Wua et al. 2010). Blast-resistant rice mutants have also been developed by using space flight mutagenesis (Xiao et al. 2008). Since 1987, China has developed more than 40 mutant varieties in different crop plant such as rice, wheat, cotton, pepper and tomato by space mutagenesis (Chengzhi 2011).

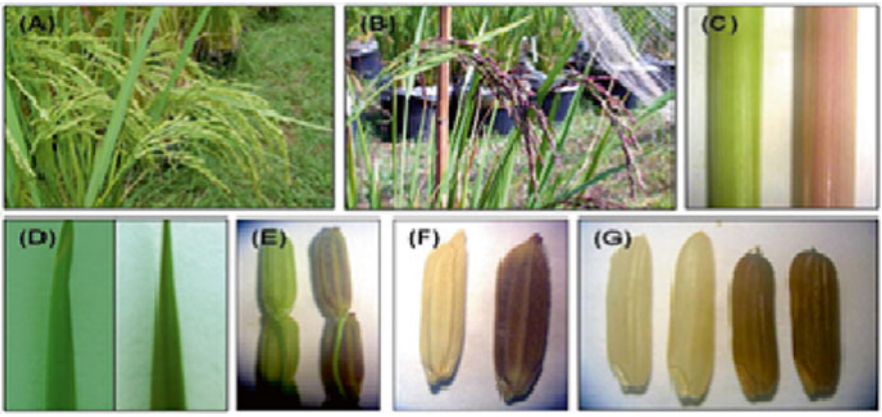

Fig. 4.3 Morphological mutants in KDML105 jasmine rice induced with a low-energy N+/N2+ion-beam. (**a**) and (**b**) illustrate the wild-type (WT) plant versus the mutant (BKOS) phenotype. The mutant has a purple color in the plant stem (**c**), Leaf (**d**), immature-seeds (**e**), husked (**f**) and dehusked mature seeds (**g**) (Source: Chundet et al. 2012)

Fig. 4.4 Mutants of leaf shape induced by electron beam. (**a**) Jingnong 6 control. (**b**) Sword leaf (600 Gy from Jingnong 6). (**c**) Lanceolate leaf (600 Gy from Jingnong 6) (Source: Luo et al. 2012)

Fig. 4.5 Ray floret mutants of Shiroyamate obtained by 12C5+ ion beam irradiation. (**a**) Original flower of Shiroyamate; (**b**) yellow mutant; (**c**) pale-yellow mutant (Source: Matsumura et al. 2010)

Plant cell and tissue culture have helped plant breeders in creating genetic variation in plants. Efficiency of mutation induction has been improved by the development of in vitro techniques (Ahloowalia 1998; Jain 2001, 2006a, 2007). Easy handling of the regenerated population and the wide range of plant material for mutagenic treatment are the main advantage of in vitro mutagenesis (Jain 2000; Predieri 2001) In vitro culture is useful in creating variation specially in vegetatively

propagated plants such banana (Chai et al. 2004; Roux et al. 2009) cassava (Jain 2005) sweet potato (Ahloowalia 1997; Ahloowalia and Maluszynski 2001) sugarcane (Kenganal et al. 2008; Patade et al. 2008; Suprasanna et al. 2009) citrus species, (Predieri 2001; Somsri et al. 2008) indiangrass (Stephens 2009) and ornamental species (Ahloowalia 1997; Jain et al. 2006a, b). Jain (2010a) isolated different types of mutants in banana such as large fruit size, reduced height, early flowering and resistant to *Fusarium* wilt and black sigatoka by in vitro mutagenesis. Bayoud disease resistant mutants were developed in date palm by in vitro mutagenesis (Jain 2012).

4.6 Potential Use of Mutants in Crop Improvement

4.6.1 Mutants as a Raw Material for Crop Improvement

Mutants have long been used by scientists directly or indirectly (cross breeding) for the development of new varieties. Efforts have been made by plant scientists to set up experimental protocols to create heritable genetic variations among crop plants and their subsequent use in crop improvement programs. A mutant varieties database (MVD) is maintained by FAO/IAEA (http://mvgs.iaea.org/AboutMutantVarities.aspx), containing newly-introduced rice varieties of modified zinc and starch content and reduced grain size from China, and an early-maturing, flood- and disease-resistance variety of rice, BINA DHAN-7, from Bangladesh. Most of the mutant varieties (77 %) are seed propagated and approximately (48 %) varieties recorded in the MVD are cereals, with rice at the highest constituting 53 %, followed by barley at 20 %. Induced mutants are freely available for crop improvement programs without restrictions on their use in contrast to protected plant varieties or germplasm. Many mutants have been released directly as new varieties and many others used as parents to create varieties with improved traits such as improved yield, quality of seed propagated crops, modified oil, protein and starch quality, enhanced uptake of specific metals, deeper rooting system and resistance to biotic and abiotic stresses. Outstanding mutant varieties such as rice in Australia, India, China, Pakistan and Thailand; cotton and wheat in Pakistan; pear in Japan; grapefruit, sunflower and peppermint in the USA; barley in several countries of Europe; durum wheat in Italy; sorghum in Mali; groundnut and pulses and several ornamental plants in India, the Netherlands and Germany have been grown successfully and are playing pivotal roles to keep food scarcity at bay (Ahloowalia et al. 2004). Mutant germplasm stock for different plants has been maintained by several countries round the world (Table 4.6) for their use in future crop improvement. The contribution of mutant crop varieties, cultivated worldwide, in food security were reviewed by Kharkwal and Shu (2009) and they concluded that mutant varieties contribute significantly to ensuring food and nutritional security by enhanced resistance to biotic and abiotic factors, higher yield, improved nutrient use efficiencies and less agricultural inputs. All these qualities have made mutant varieties an integral part of daily diet around the world. Promising characteristics of the mutant varieties have been shown in Fig. 4.6.

Table 4.6 Mutants/germplasm stocks of crops and the host institutions

Crop	Host institution
Maize	The Maize Genetics Cooperation Stock Center, University of Illinois, Urbana/Champaign, IL USA
Arabidopsis	European Arabidopsis Stock Centre (or Nottingham Arabidopsis Stock Centre, NASC), University of Nottingham, Sutton Bonington Campus, UK; Arabidopsis Biological Resource Centre, (ABRC) Ohio State University, OH, USA
Tomato	CM Rick Tomato Genetics Resource Centre, University of California at Davis, CA, USA; AVRDC, Taiwan, Tomato Breeding Program Germplasm, University of Florida, USA
Rice	The Oryza base of the National Bio Resource Project-Rice National Institute of Genetics, Japan; IR64 Rice Mutant Database of the International Rice Functional Genomics, International Rice Research Institute, Manila, Philippines; Plant Functional Genomics Lab., Postech Biotech Center, San 31 Hyoja-dong, Nam-gu Pohang, Kyoungbuk, Korea; Institute of Crop Germplasm Resources, CAAS, Bejieng, China; Africa Rice Center (Africa Rice) Cotonou, Benin; China National Rice Research Institute, China; Genetic Resources Center IRRI, Philippines; The International Rice Gene bank Philippines, NBPGR India; Shanghai Agrobiological Gene Center, China
Barley and wheat	Barley mutants, Scottish Crop Research Institute, Dundee, Scotland; Wheat Genetics Resource Center (WGRC) KU, USA. Barley and Wheat Genetic Stock of the USDA-ARS. USDA-ARS Cereal Crops Research Unit, Fargo, ND, USA; Wheat Genetics Resource Center, Kansas State University
	Manhattan, KS, USA; Wheat Genetic Resources Database of the Japanese National BioResource Project; Genetic Stock Centre, Bulgaria; Cereal Genebank, Martonvásár, Hungary
Peas	Pea mutants, John Innes Centre, Norwich, UK; G.A. Marx *Pisum* Genetic Stock Collection

Source: Updated from Mba (2013)

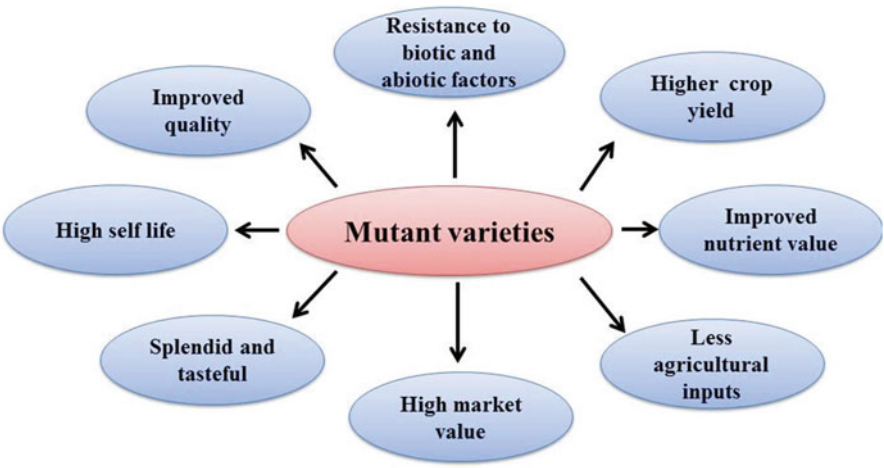

Fig. 4.6 Characteristics of mutant varieties

4.6.2 Resource for Genomics andMolecular Breeding

Induced mutagenesis coupled with molecular biology techniques have made it possible to generate large mutant populations in different crops (Caldwell et al. 2004; Krishnan et al. 2009; Wang et al. 2013; Xin et al. 2008). DNA sequencing, genome wide association analysis, transcriptomics and proteomics data of such mutant populations have been used for functional genomics studies. These mutant populations have also been used for gene discovery by forward and reverse genetic approaches. In rice, 64 genes have been discovered so far that are responsible for mutant phenotypes in plant development, photosynthesis, signaling transduction and disease resistance (Morrell et al. 2012). Knowledge of functions, expression and regulations of these genes responsible for agronomically-important phenotypes will benefit crop improvement.

Induced mutants play pivotal roles in plant breeding. Mutation breeding, by using EMS, fast neutron and gamma irradiation, created 443 rice cultivars (Kharkwal and Shu 2009). In 1976, Calrose 76, the first semi-dwarf rice cultivar, was developed in the USA and afterward numerous semi-dwarf cultivars and other useful mutants with improved characters, such as early maturity, endosperm quality, elongated uppermost internode and genetic male sterile, improved nutritional quality because of low phytic acid, giant embryo mutants of potential interest to the rice oil industry and adapted basmati and jasmine germplasm were developed by using Calrose 76 (Rutger 2009). In maize, a model biological system and also an important agronomic crop, shrunken2 (sh2) mutant kernels were sweet (Laughnan 1953). By using this mutant as a parent in maize improvement programs sh2-based sweet varieties have been developed (Tracy 1997). Amylose extender1, Leafy1, Sugary1, sugary enhancer1, waxy1, opaque2, floury2, and brown midrib3 are other maize mutants that have also been used directly for specialty maize production (Cox and Cherney 2001; Hallauer 2001). Mutant germplasm collections, for a specific character, have also been maintained such as in barley; more than 700 anthocyanin and proanthocyanidin mutants were used for identifying genes involved in the flavonoid metabolic pathway in barley (Jende-Strid 1993). One more barley mutant cultivar, Diamant, has been used for breeding more than 150 leading barley cultivars all over the world (Ahloowalia et al. 2004). High oleic acid mutants of sunflower were widely used in the USA and Europe, low linolenic acid and high oleic acid mutants of rapeseed induced in Canada have been used for breeding rapeseed in Australia, Canada and Europe (Ahloowalia et al. 2004).

4.7 Impact of Mutant Varieties

Over the past six decades, more than 3000 mutant varieties have been developed by mutation induction in more than 200 plant species. Rice, wheat, cotton, rapeseed, sunflower, sesame, grapefruit and banana are economically-important crops with a large number of mutant varieties. Among these varieties, rice in Australia, China

and Thailand; rice and cotton in Pakistan; Japanese pear in Japan; barley varieties in Europe; durum wheat in Italy; sunflower, grapefruit and peppermint in USA; sorghum in Mali; rice, groundnut, pulse crops and ornamentals in India, the Netherlands and Germany, have all had a positive impact on the economy of the individual countries (Ahloowalia et al. 2004). Rice varieties resistant to salinity, and with early-maturing and high-quality traits, have been developed in Vietnam providing extra income to the farmers and generating USD 300 million per year (Jain and Suprasanna 2011). Biotic and abiotic factors hinder crop production and these can be overcome by changing the crop genetic architecture that neutralizes the effect of biotic and abiotic stresses. Reduced height in cereals (Kharkwal and Shu 2009); bushy mutant in chickpea (Gaur et al. 2008) cotton (Ahloowalia et al. 2004) and sunflower (Jambhulkar and Shitre 2009) etc. are examples of changed genetic architecture of these plants. A flax mutant with low linola content and a sunflower mutant with high oleic acid, developed by induced mutagenesis, improve the quality of the product. NIAB Karishma a cotton leaf curl virus resistant cultivar, generates USD 294.4 million in Pakistan (Haq 2009). Calrose 76, a semi-dwarf rice mutant developed through gamma irradiation, has 14 % more yield than the wild type and an added USD 20 million per year to farmers (Rutger 2006). Golden Promise and Diamant barley mutants brought millions of dollars to the malting and brewing industry in Europe.

4.8 Conclusions and Prospects

Currently available staple crops are insufficient to meet the challenges of the twenty-first century and provide total world food security. Enhanced food production, together with reduced postharvest loses are primary approaches to feed the expected 9 billion population by 2050. Induced mutagenesis has played an important role by creating several mutants in different crop plants. These mutant varieties with specific character/trait such as high yield, resistance to biotic and abiotic stresses, have been grown globally bringing a significant positive economic impact and contribute to global food and nutritional security and improved livelihoods. Despite the available mutant resources, challenges still lie ahead to feed an ever-increasing population. To speed up crop production, mutant resources for different crop plants have to be established which can be used to create new mutant cultivars which are high yielding, resistant to biotic and abiotic stresses, enhanced uptake of specific metal, deeper rooting systems and modified oil, starch and protein content that can boost industrial processing. Now in the nanotechnology era, scientists develop can new tools and techniques for crop improvement. Recently, nanoparticles, nanocapsules and nanofibers have been utilized in gene manipulation for crop productivity enhancement. Numbers of biological and chemicals materials applied as a functionalized nanoform to regulate gene expression in plant systems. Nanofiber-based delivery of genetic materials is the best alternative process of a microinjection gene delivery method. Carbon nanofiber arrays are being used for fast and efficient

delivery of genetic material in plant cells in crop engineering. With the innovation of molecular biological and nano-level techniques, it is now possible to study more agronomically-important traits at the molecular level that will help in creating envisaged *smart* crop varieties that overcome the constraints threatening twenty-first century global food security.

References

Abe T, Ryuto H, Fukunishi N et al (2012) Ion beam radiation mutagenesis. In: Shu QY, Forster BP, Nakagawa H (eds) Plant mutation breeding and biotechnology. CABI, Oxford, pp 99–106

AhlALuo WX, Li YS, Wu BM et al (2012) Effects of electron beam radiation on trait mutation in azuki bean (*Vigna angularisi*). Afr J Biotechnol 11:12939–12950

Ahloowalia BS (1997) Improvement of horticultural plants through in vitro culture and induced mutations. In: Altman A, Ziv M (eds) Horticultural biotechnology, in vitro culture and breeding, 3rd International ISHS symposium on in vitro culture and horticultural breeding, Jerusalem, Israel, pp 545–549

Ahloowalia BS (1998) In vitro techniques and mutagenesis for the improvement of vegetatively propagated plants. In: Jain SM, Brar DS, Ahloowalia BS (eds) Somaclonal variation and induced mutations in crop improvement. Kluwer, London, pp 293–309

Ahloowalia BS, Maluszynski M (2001) Induced mutations – a new paradigm in plant breeding. Euphytica 118:167–173

Ahloowalia BS, Maluszynski M, Nichterlein M et al (2004) Global impact of mutation-derived varieties. Euphytica 135:187–204

Auerbach C (1940) Tests of carcinogenic substances in relation to the production of mutation in *Drosophila melanogaster*. Proc R Soc Edinb B 60:164–173

Auerbach C (1947) The induction by mustard gas of chromosomal instabilities in *Drosophila melanogaster*. Proc R Soc Edinb B 62:307–320

Auerbach C, Robson JM (1944) Production of mutations by allylisothiocyanate. Nature 154:81

Auerbach C, Robson JM (1946) Chemical production of mutations. Nature 157:302

Auerbach C, Robson JM (1947) Tests for chemical substances for mutagenic action. Proc R Soc Edinb B 62:284–291

Caldwell DG, McCallum N, Shaw P et al (2004) A structured mutant population for forward and reverse genetics in barley (*Hordeum vulgare* L.). Plant J 40:143–150

Chai M, Ho YW, Liew KW, Asif JM (2004) Biotechnology and in vitro mutagenesis for banana improvement. In: Jain SM, Swennen R (eds) Banana improvement: cellular, molecular biology, and induced mutations. Leuven, Belgium, pp 59–77

Chengzhi L (2011) Agronomy in space – China's crop breeding program. Space Policy 27:157–164

Chopra VL (2005) Mutagenesis: investigating the process and processing the outcome for crop improvement. Curr Sci 89:353–359

Chundet R, Cutler RW, Anuntalabhochai S et al (2012) Induction of anthocyanin accumulation in a Thai jasmine rice mutant by low-energy ion beam. IRJPS 3:120–126

Cox WJ, Cherney DJR (2001) Influence of brown midrib, leafy, and transgenic hybrids on corn forage production. Agron J 93:790–796

De Varies H (1901) Die mutations theorie, Leipzig. (Steite V)

FAO (2011) Save and grow – a policy maker's guide to the sustainable intensification of smallholder crop production. FAO, Rome

FAO (2014a) Committee on World Food Security (CFS). http://www.fao.org/cfs/cfshome/en/. Accessed 2 Nov 2014

FAO (2014b) Mutant varieties database of the joint FAO/IAEA division of nuclear techniques in food and agriculture. http://mvgs.iaea.org/AboutMutantVarities.aspx. Accessed 2 Nov 2014
FAO/IAEA (2014) Mutant variety data base. http://www-naweb.iaea.org/nafa/pbg/index.html. Accessed 5 Nov 2014
Feng HY, Yang G, Yu ZL (2009) Mutagenic mechanisms of ion implantation in plants. In: Induced plant mutations in the genomics era. Proceedings of an international joint FAO/IAEA symposium. International Atomic Energy Agency, Vienna, Austria, pp 220–222
Gaur PM, Gour VK, Srinivasan S et al (2008) An induced brachytic mutant of chickpea and its possible use in ideotype breeding. Euphytica 159:35–41
Greene EA, Codomo CA, Taylor NE, Henikoff JG et al (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. Genetics 164(2):731–740
Gulfishan M, Khan AH, Jafri IF et al (2012) Assessment of mutagenicity induced by MMS and DES in *Capsicum annuum* L. Saudi J Biol Sci 19:251–255
Gulfishan MJ, Afri IF, Khan AH et al (2013) Methyl methane sulphonate induced desynaptic mutants in *Capsicum annuum* L. Chrom Bot 8:59–62
Guo H, Jin W, Zhao L et al (2010) Mutagenic effects of different factors in spaceflight environment of Shijian-8 satellite in wheat. Acta Agron Sin 36:764–770
Hallauer AR (2001) Specialty corns, 2nd edn. CRC Press, Boca Raton
Hamada K, Inoue M, Tanaka A et al (1999) Potato virus Y-resistant mutation induced by the combination treatment of ion beam exposure and anther culture in *Nicotiana tabacum* L. Plant Biotechnol J 16:285–289
Haq MA (2009) Development of mutant varieties of crop plants at NIAB and the impact on agricultural production in Pakistan. In: Induced plant mutations in the genomics era. Proceedings of an international joint FAO/IAEA symposium. International Atomic Energy Agency, Vienna, Austria, pp 61–64
Hase Y, Okamura M, Takeshita D et al (2010) Efficient induction of flower-color mutants by ion beam irradiation in petunia seedlings treated with high sucrose concentration. Plant Biotechnol J 27:99–103
Henikoff S, Comai L (2003) Single-nucleotide mutations for plant functional genomics. Annu Rev Plant Physiol Plant Mol Biol 54:375–401
Hirano T, Kazama Y, Ohbu S et al (2012) Molecular nature of mutations induced by high-LET irradiation with argon and carbon ions in *Arabidopsis thaliana*. Mutat Res 735:19–31
Hu X, Li Y, Gao Y et al (2010) Review and prospect of space mutation application in pepper breeding. China Veg 24:14–18
Jain SM (2001) Tissue culture-derived variation in crop improvement. Euphytica 118:153–166
Jain SM (2002) A review of induction of mutations in fruits of tropical and subtropical regions. Acta Horticult 575:295–302
Jain SM (2005) Major mutation-assisted plant breeding programs supported by FAO/IAEA. Plant Cell Tiss Org Cult 82:113–123
Jain SM (2006a) Mutation-assisted breeding for improving ornamental plants. Proc Int Eucarpia Symp Sect Ornamentals Breed Beauty 714:85–98
Jain SM (2006b) Mutation-assisted breeding in ornamental plant improvement. Acta Horticult 714:85–98
Jain SM (2007) Recent advances in plant tissue culture and mutagenesis. Acta Horticult 736:205–211
Jain SM (2010a) In vitro mutagenesis in banana (*Musa* spp.) improvement. Acta Horticult 879:605–614
Jain SM (2010b) Mutagenesis in crop improvement under the climate change. Romania Biotechnol Lett 15(2):88–106
Jain SM (2012) In vitro mutagenesis for improving date palm (*Phoenix dactylifera* L.). Emir J Food Agric 24:400–407
Jain SM, Maluszynski M (2004) Induced mutations and biotechnology on improving crops. In: Mujib A, Cho M, Predieri S, Banerjee S (eds) In vitro applications in crop improvement: recent progress. IBH-Oxford, India, pp 169–202

Jain SM, Suprasanna P (2011) Induced mutations for enhancing nutrition and food production. Gene Conserv 40:201–215

Jambhulkar SJ, Shitre AS (2009) Development and utilization of genetic variability through induced mutagenesis in sunflower (*Helianthus annuus* L.). In: Induced plant mutations in the genomics era. Proceedings of an international joint FAO/IAEA symposium. International Atomic Energy Agency, Vienna, Austria, pp 104–105

Jende-Strid B (1993) Genetic control of flavonoid biosynthesis in barley. Heredity 119:187–204

Jiang SY, Ramachandran Y (2010) Natural and artificial mutants as valuable resources for functional genomics and molecular breeding. Int J Biol Sci 6(3):228–251

Kenganal M, Hanchinal RR, Nadaf HL (2008) Ethyl methane sulphonate (EMS) induced mutation and selection for salt tolerance in sugarcane in vitro. Indian J Plant Physiol 13:405–410

Kharkwal MC, Shu QY (2009) The role of induced mutations in world food security. In: Shu QY (ed) Induced plant mutations in the genomics era. FAO, Rome, pp 33–38

Kim Y, Schumaker KS, Zhu JK et al (2006) EMS mutagenesis of *Arabidopsis*. Methods Mol Biol 323:101–103

Kondo E, Nakayama M, Kameari N et al (2009) Red-purple flower due to delphinidin 3,5-diglucoside, a novel pigment for *Cyclamen* spp., generated by ion-beam irradiation. Plant Biotechnol J 26:565–569

Koornneef M, Dellaert LW, Van Der Veen JH et al (1982) EMS-and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana*. (L.) Heynh. Mutat Res 93:109–123

Kranz AR (1986) Genetic and physiological damage induced by cosmic radiation on dry plant seeds during space flight. Adv Space Res 6:135–138

Krieg DR (1963) Ethyl methane sulfonate-induced reversion of bacteriophage T4rII mutants. Genetics 48:561–580

Krishnan A, Guiderdoni E, An G et al (2009) Mutant resources in rice for functional genomics of the grasses. Plant Physiol 149:165–170

Laughnan JR (1953) The effect of the sh2 factor on carbohydrate reserves in the mature endosperm of maize. Genetics 38:485–499

Li JG, Wang PS, Zhang J, Jiang XC (1999) Development and prospect of plant mutation breeding induced by aviation and spaceflight in China. Space Med 12:464–468

Li X, Song Y, Century K et al (2001) A fast neutron deletion mutagenesis-based reverse genetics system for plants. Plant J 27:235–242

Liu L, Zhao L, Guo H et al (2007) Current status and perspective outlook of space induced mutation breeding in crop plants. Rev China Agric Sci Technol 9:26–29

Matsumura A, Nomizu T, Furutani N et al (2010) Ray florets color and shape mutants induced by $^{12}C^{5}$+ ion beam irradiation in chrysanthemum. Sci Hortic 123:558–561

Mba C (2013) Induced mutations unleash the potentials of plant genetic resources for food and agriculture. Agronomie 3:200–231

Mba C, Shu QY (2012) Gamma irradiation. In: Shu Q, Forster BP, Nakagawa H (eds) Plant mutation breeding and biotechnology. CABI, Oxfordshire, pp 91–98

Mba C, Afza R, Jain SM et al (2007) Induced mutations for enhancing salinity tolerance in rice. In: Jenks MA, Hasegawa PM, Jain SM (eds) Advances in molecular breeding towards drought and salt tolerant crops. Springer, Berlin, pp 413–454

Mba C, Afza R, Shu QY et al (2012a) Mutagenic radiations: X-rays, ionizing particles and ultraviolet. In: Shu QY, Forster BP, Nakagawa H (eds) Plant mutation breeding and biotechnology. CABI, Oxfordshire, pp 83–90

Mba C, Guimaraes PE, Ghosh K (2012b) Re-orienting crop improvement for the changing climatic conditions of the 21st century. Agric Food Sect 1:7

Mei M, Deng H, Lu Y et al (1994a) Mutagenic effects of heavy ion radiation in plants. Adv Space Res 14:363–372

Mei M, Qiu Y, He Y et al (1994b) Mutational effects of space flight on *Zea mays* seeds. Adv Space Res 14:33–39

Mei MT, Oiu YL, Sun YQ et al (1998) Morphological and molecular changes of maize plant after seeds been flown on recoverable satellite. Adv Space Res 22:1691–1697

Morrell PL, Buckler ES, Ross-Ibarra J et al (2012) Crop genomics: advances and applications. Nat Rev Genet 13:85–96
Muller HJ (1927) Artificial transmutation of the gene. Science 66:84
Ou XF, Long LK, Wu Y et al (2010) Spaceflight-induced genetic and epigenetic changes in the rice (*Oryza sativa* L.) genome are independent of each other. Genome 53:524–532
Patade VY, Suprasanna P, Bapat VA (2008) Gamma irradiation of embryogenic callus cultures and in vitro selection for salt tolerance in sugarcane (*Saccharum officinarum* L.). Agric Sci China 7:1147–1152
Predieri S (2001) Mutation induction and tissue culture in improving fruits. Plant Cell Tiss Org Cult 64:185–210
Ren W, Zhang Y, Deng B et al (2010) Effect of space flight factors on alfalfa seeds. Afr J Biotechnol 9:7273–7279
Reyes-Borja WO, Sotomayor I, Garzon I et al (2007) Alteration of resistance to black sigatoka (*Mycosphaerella fijiensis* Morelet) in banana by in vitro irradiation using carbon ion-beam. Plant Biotechnol J 24:349–353
Ronald PC (2014) Lab to farm: applying research on plant genetics and genomics to crop improvement. PLoS Biol 12(6):e1001878. doi:10.1371/journal.pbio.1001878
Roux NS, Toloza A, Strosse H et al (2009) Induction and selection of potentially useful mutants in banana. Acta Horticult 828:315–322
Roychowdhury R, Tah J (2013) Mutagenesis – a potential approach for crop improvement. In: Hakeen KR, Ahmad P, Öztürk O (eds) Crop improvement. Springer, New York, pp 149–187
Rutger JN (2006) Thirty years of induction, evaluation, and integration of useful mutants in rice genetics and breeding. Plant Mutat Rep 1(2):4–13
Rutger JN (2009) The induced sd-1 mutant and other useful mutant genes in modern rice varieties. In: Shu QY (ed) Induced plant mutations in the genomics era. Proceeding of an international joint FAO/IAEA symposium, Rome, pp 44–47
Schlenker W, Roberts MJ (2009) Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. Proc Natl Acad Sci U S A 106:15594–15598
Shikazono N, Yokota Y, Kitamura S et al (2003) Mutation rate and novel tt mutants of *Arabidopsis thaliana* induced by carbon ions. Genetics 163:1449–1455
Sikora P, Chawade A, Larsson M et al (2011) Mutagenesis as a tool in plant genetics, functional genomics, and breeding. Int J Plant Genom. doi:10.1155/2011/314829
Somsri S, Putivoranat M, Kanhom P et al (2008) Improvement of tropical and subtropical fruit trees: tangerine, pummelo in Thailand through induced mutation and biotechnology. Acta Horticult 787:127–139
Stadler LJ (1928a) Genetic effects of X-rays in maize. Proc Natl Acad Sci U S A 14:69–72
Stadler LJ (1928b) Mutations in barley induced by X-rays and radium. Science 68:186–187
Stephens LC (2009) Ethidium bromide-induced mutations from inflorescence cultures of Indiangrass. Hortic Sci 44:1215–1218
Suprasanna P, Patade VY, Vaidya ER, Patil VD (2009) Radiation induced in vitro mutagenesis, selection for salt tolerance and characterization in sugarcane. In: Induced plant mutations in the genomics era. Proceedings of an international joint FAO/IAEA symposium, International Atomic Energy Agency, Vienna, Austria, pp 145–147
Suprasanna P, Mirajkar SJ, Patade YV, Jain SM (2014) Induced mutagenesis for improving plant abiotic stress tolerance. In: Tomlekova NB, Kozgar MI, Wani MR (eds) Mutagenesis: exploring genetic diversity of crops. Wageningen Academic Publishers, Netherlands, pp 349–378
Takano N, Takahashi Y, Yamamoto M et al (2013) Isolation of a novel UVB-tolerant rice mutant obtained by exposure to carbon-ion beams. J Radiat Res 54:637–648
Tanaka A, Shikazono N, Hase Y et al (2010) Studies on biological effects of ion beams on lethality, molecular nature of mutation, mutation rate, and spectrum of mutation phenotype for mutation breeding in higher plants. J Radiat Res 51:223–233
Tracy WF (1997) History, genetics, and breeding of supersweet (shrunken2) sweet corn. Plant Breed Rev 14:189–236

UNEP Global Environment Outlook 3 (2002) Global Environment Outlook 3. United Nations, New York

Wang N, Long T, Yao W, Xiong L, Zhang Q, Wu C et al (2013) Mutant resources for the functional analysis of the rice genome. Mol Plant 6:596–604

Wu L, Yu Z (2001) Radiobiological effects of a low-energy ion beam on wheat. Radiat Envir Biophys 40:53–57

Wu JL, Wu C, Lei C et al (2005) Chemical- and irradiation-induced mutants of *indica* rice IR64 for forward and reverse genetics. Plant Mol Biol 59:85–97

Wua H, Huang C, Zhang K, Sun Y (2010) Mutations in cauliflower and sprout broccoli grown from seeds flown in space. Adv Space Res 46:1245–1248

Xiao W, Yang Q, Chen Z et al (2008) Blast-resistance inheritance of space-induced rice lines and their genomic polymorphism by microsatellite markers. Sci Sin Agric 4:3952–3958

Xin Z, Wang ML, Barkley NA et al (2008) Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. BMC Plant Biol 8:103. doi:10.1186/1471-2229-8-103

Yamaguchi H, Hase Y, Tanaka A et al (2009) Mutagenic effects of ion beam irradiation on rice. Breed Sci 59:169–177

Yuan L, Dou Y, Kianian SF et al (2014) Deletion mutagenesis identifies a haploinsufficient role for g-Zein in opaque2 endosperm modification. Plant Physiol 164:119–130

Zhong-hua W, Xin-chen Z, Yu-lin J (2014) Development and characterization of rice mutants for functional genomics studies and breeding. Rice Sci 21(4). doi:10.1016/S1672-6308(13)60188-2

Zhu BG, Gu AQ, Deng XD et al (1995) Effects of caffeine or EDTA post treatment on EMS mutagenesis in soybean. Mutat Res 334:157–159

Part II
Breeding Strategies

Chapter 5
Conventional Plant Breeding Principles and Techniques

George Acquaah

Abstract Conventional plant breeding is the development or improvement of cultivars using conservative tools for manipulating plant genome within the natural genetic boundaries of the species. Mendel's work in genetics ushered in the scientific age of plant breeding. The number of genes that control the trait of interest is important to breeders. Qualitative traits (controlled by one or a few genes) are easier to breed than quantitative traits (controlled by numerous genes). General steps in breeding are: objectives, creation/assembly of variability, selection, evaluation and cultivar release. Breeders use methods and techniques that are based on the mode of reproduction of the species self-pollinating, cross-pollinating, or clonally propagated. The general strategy is to breed a cultivar whose genetic purity and productivity can be sustained by its natural mating system. There are six basic types of cultivars: pure line, open-pollinated, hybrid, clonal, apomictic and multilines. The common methods for breeding self-pollinated species include mass selection, pure line selection, pedigree, bulk population, single seed descent, backcrossing, multiline and composite. Methods for breeding cross-pollinated species include mass selection, recurrent selection, family selection and synthetics. Hybrid cultivar breeding exploits the phenomenon of heterosis, and is applicable to both self- and cross-pollinated species. Polyploids have complex genetics. Hybridization of parents is often accompanied by infertility of the hybrid. Mutation breeding may be resorted to when the gene of interest is non-existent in nature and may be induced. Also, sometimes, the desired trait is found in wild relatives of the species and may be introgressed into cultivated species through pre-breeding.

Keywords Breeders equation • Breeding methods • Conventional breeding • Cultivar • Selection • Polyploidy

G. Acquaah (✉)
Department of Natural Sciences, College of Arts and Sciences, Bowie State University, Bowie, MD, USA
e-mail: gacquaah@bowiestate.edu

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_5

5.1 Introduction

Modern plant breeding is both an art and science. It has been practiced for thousands of years, as humans continue to use various methods to change the traits of plants in order to produce desired characteristics (Sleper and Poehlman 1999). With time, the role of science and technology in plant breeding continues to increase (Acquaah 2012). Selection, the discrimination among variability, is the oldest method of plant improvement (Chahal and Gosal 2002). The science of genetics revolutionized the selection process, taking the guesswork out of it, facilitating it and making it more efficient (Acquaah 2012; Briggs and Knowles 1967). Modern plant breeders follow standard methods to create variability, discriminate among the variability, and develop cultivars for release to farmers (Acquaah 2012; Allard 1960; Fehr 1987a).

The objective of this chapter is to present and describe the methods of plant breeding and their genetic underpinnings. Plant breeders follow several general steps in their breeding programs, starting with objectives and ending with cultivar release. These steps are described in detail and illustrated. The chapter opens with a brief definition of conventional breeding and follows up the science that has been brought into it, introducing the advances in conventional breeding technologies and techniques. The key genetic principles and concepts, including the number of genes controlling traits, gene action, heritability, and the breeders' equation are discussed in depth. Decision-making in plant breeding based on biometrical genetics is discussed next, followed by an explanation of the genetic structure and types of plant cultivars that breeders develop. The role of the method of reproduction (mating system) in plant breeding is discussed, leading to a detailed discussion of the methods used to breed self-pollinated, cross-pollinated, and clonally propagated species.

5.1.1 Definition of Plant Breeding

Plant breeding is a deliberate effort by humans to improve certain aspects of plants to perform new roles or enhance existing ones (Acquaah 2012). Consequently, the term *plant breeding* is often used synonymously with *plant improvement* in modern society. Plant breeding is often likened to the process of evolution (Zohary and Hopf 1988). A key difference between the two is that plant breeding is a relatively quick process that is orchestrated by humans (artificial process), while evolution is a very slow and natural process (Gepts 2002). Further, the goal of evolution is to increase the fitness of the species, whereas plant breeders aim to nudge the population towards specific and predetermined goals (i.e. they are not necessarily concerned with fittest because modern growers can produce plants under artificial conditions) (Borojevic 1990).

5.1.2 Definition of Conventional Breeding

The practice of plant breeding has advanced from the cynical view of "crossing the best with the best and hoping for the best," to the now carefully planned and thought-out strategies for developing high performing cultivars with high predictability (Acquaah 2012). Modern plant breeding is both a science and an art. The methods and tools employed, which keep changing with advances in science and technology, provide the basis for categorizing plant breeding approaches into two basic types – *conventional* and *unconventional (molecular).* (Acquaah 2012; Chahal and Gosal 2002). Molecular techniques are used primarily to create new variability, after which the breeder uses conventional breeding methods for cultivar development (Jain and Kharkwal 2004).

Conventional breeding (classical breeding or traditional breeding), is the development of new varieties (cultivars) of plants by using older tools and natural processes, as opposed to the newer, more sophisticated and sometimes radical tools of molecular plant breeding (Jain and Kharkwal 2004). Molecular breeders may violate natural biological boundaries in the manipulation of the genetics of plants. They are able to introduce into the new cultivar desirable genes that are alien to its species. In conventional breeding, breeders assemble desirable traits from different but usually closely related plants into the new cultivar using the techniques of crossing (hybridization) (Acquaah 2012). Consequently, the product of conventional breeding only emphasizes target traits which preexist in the genetic potential of the species, without introducing new genes (Jakowitsch et al. 1999).

5.1.3 Primitive Origins of Conventional Plant Breeding

To feed themselves, early humans gathered whatever they could find in the wild that was edible. As their lifestyle changed from wandering to sedentary, they brought and planted desirable plant types at their abode. This was the beginning of the era of domestication of plants (Gepts 2002). Over many years, and through discrimination (selection) among the available natural biological diversity, humans gradually but systematically favored characteristics that increased the utility of the plants. With time, domesticated plants become markedly different from their wild progenitors in specific ways that were valuable and advantageous to humans (Zohary and Hopf 1988). Selection is the oldest technique of crop improvement and remains the most widely used technique to date (Acquaah 2012). It is accomplished largely on the basis of visual appearance. Armed with a mental picture of the target trait, the person conducting the operation visually discriminates among the available variability based on appearance, to identify and select desirable plants.

Farmers using this technique of crop improvement save seed from the best looking (superior) plants from their current crop for planting the next year's crop (Acquaah 2012). The *improved seed* developed this way is called a *land race* (Allard

1960). Because selection is based on phenotype, the field planted to a land race may look similar in appearance, but is genetically highly heterogeneous. Land races provide a rich source of variability for modern plant breeding (Harlan 1975).

5.2 Bringing Science into Conventional Plant Breeding

As knowledge abounds and technology advances, modern breeders are increasingly depending on science to take the guesswork out of the selection process, or at least reduce it (Chahal and Gosal 2002). However, being an applied science, conventional breeding still relies to some extent on intuition, skill, and judgment of the breeder *(breeder's eye),* in conjunction with scientific knowledge (Allard 1960). A good breeder should have a keen sense of observation. One of the most widely grown potato cultivars in the USA, the Burbank potato, was discovered through the keen observation of the grower (Fehr 1987b).

5.2.1 The Concept of Genetic Manipulation of Plant Traits

The scientific age of plant breeding was ushered in with the ground-breaking work of Gregor Mendel in the nineteenth century which, along with further advances in science that followed his discovery (*genetic theory*), established that plant characteristics are controlled by hereditary factors (*genes*) that consist of DNA (the hereditary material) (Borojevic 1990). The invisible genes (genotype; G) are expressed in an environment (E) to produce a visible trait (phenotype; P). It follows then that in order to change a trait (or its expression), one may change the *nature* (genotype), and/or modify the (environment in which it is expressed). This concept is mathematically expressed as $P = G + E$ (Falconer 1981). Changing the environment essentially entails modifying the growing or production conditions (agronomy). This may be achieved through the application of production inputs (e.g. fertilizers, irrigation) (Acquaah 2012).

This approach is effective in enhancing certain traits only temporarily (i.e. no tempering with the genetics of the plant). Plant breeders on the other hand seek to alter plants in a permanent and heritable way (by modifying the genotype; transmissible from one generation to the next) (Poehlman and Sleper 1995). Genetic theory, coupled with other scientific knowledge (statistics, biometrics, quantitative genetics, etc.) provides the foundational principles for developing modern plant breeding methodologies and techniques (Acquaah 2012; Griffiths 1999). Some of the key genetic and biometrical concepts are discussed further later in this chapter.

5.2.2 *Advances in Conventional Breeding Technologies and Techniques*

In developing cultivars, breeders undertake two basic activities – create or assemble variability, and discriminate among (selection) variability to identify and advance desirable individuals that meet the breeding objectives (Fehr 1987a). The effectiveness and efficiency of a breeding program rest to a large extent on these two activities. Consequently, breeders seek new or refine old technologies and techniques that facilitate these activities. The following are some of the key ones, each of which may have additional associated techniques:

(a) Selection: This is the most fundamental technique used for crop improvement by both trained scientists and untrained farmers (Briggs and Knowles 1967). It is basically the act of discriminating among variability to pick and advance desirable plants. Sometimes, the unit of selection is the individual plant; other times, a number of plants is chosen and advanced in a breeding program (Acquaah 2012). Modern breeders use a variety of *selection* or *breeding schemes* or methods in their programs.
(b) Artificial pollination: This is an ancient and yet very critical technique in modern plant breeding. It is a controlled pollination technique for genetic studies, creating targeted variability, targeted gene transfer, developing breeding stocks, improving fruit set and seed production, among other uses (Acquaah 2012). The success of artificial pollination is beset with issues such as *genetic incompatibility* (genetically-induced prevention of fertilization) (Chahal and Gosal 2002). Various techniques are available for overcoming these issues.
(c) Hybridization: This is the planned crossing of parents used to create new variability to initiate a breeding program, or to create new cultivars *(hybrid breeding*) (Agrawal 1998). It involves the use of controlled pollination that may be effected by artificial means. Depending upon the program, breeders may establish hybridization blocks where controlled pollination occurs. Sometimes, sophisticated hybridization schemes (e.g. diallele crosses) are used to create certain genetic combinations (Brown and Caligari 2008). To facilitate hybridization for developing hybrid cultivars, breeders use various techniques to control pollination.
(d) Wide crosses: Commonly, breeders cross parents within the same species. However, sometimes the gene(s) of interest may be located in other species. Wild germplasm is a rich source of genes for modern crop improvement (Zamir 2001). *Wide crosses* are crosses that involve materials from outside the gene pool for cultivated species (Harlan 1975). They can involve two species (*interspecific cross*) or even genera (*intergeneric cross*). The more distant genetically the parents are, the higher the incidence of genetic complications resulting in infertility and low success (Harlan 1975).
(e) Embryo culture: Because of infertility issues, the embryos resulting from especially wide crosses do not develop normally and need to be extracted (rescued)

prematurely and nurtured in vitro into full plants using tissue culture techniques (Acquaah 2012; Bridgen 1994; Sage et al. 2010). The procedure was successfully used following interspecific crosses in the genus *Hylocereus* (Cactaceae) (Cisneros and Tel-zur 2010).

(f) Chromosome doubling: Interspecific wide crosses involve parents with different numbers of chromosomes. The hybrid resulting from such crosses are reproductively sterile, due to meiotic incompatibility (Acquaah 2004a). Breeders use *chromosome doubling* (effected by application of colchicine) technique to create paring partners for successful meiosis and restoration of fertility (Chahal and Gosal 2002).

(g) Doubled haploids: A doubled haploid genotype results when a haploid cell (e.g. pollen, egg) undergoes chromosome doubling, an event that can produced in vivo or in vitro, and be spontaneous or artificially induced in plant breeding (Maluszynski et al. 2003). Haploids may be generated via wide crossing, as was done in barley (Kasha and Kao 1970). Chromosome doubling is achieved through the application of the chemical colchicine (Winzeler et al. 1987; Acquaah 2012). Whereas doubled haploids can be a cultivar development technique (via elite crossing, backcrossing, or direct releases), other applications in plant breeding include mapping quantitative trait loci (QTLs), bulk segregant analysis (BSA), genetic maps and genomic studies (Maluszynski et al. 2003). Currently, several hundreds of plants are amenable to doubled haploid protocols, but the technology has advantages and disadvantages. The key disadvantage is that selection cannot be imposed on the population. Doubled haploid technology has been successfully applied in hybrid maize breeding (Geiger and Gordillo 2009) as well as wheat improvement (Liu et al. 2002).

(h) Bridge crossing: *Bridge crossing* is the technique used in wide crosses or for creating a transitional or intermediate cross as an indirect way of crossing two parents with different ploidy levels (different numbers of chromosomes) (Zamir 2001). The intermediate hybrid is reproductively sterile and is subjected to chromosome doubling to restore fertility.

(i) Protoplast fusion: This is also a technique that is helpful to wide crossing. Protoplasts (cells without cell wall) may be fused in vitro to achieve hybridization in cases where crossing via pollination and normal fertilization is challenging or impossible (Acquaah 2012). In potato, Helgeson et al. (1986) used somatic fusion to achieve potato plants that were resistant to potato leaf roll disease.

(j) Seedlessness: Fertility for seed formation is desired in seed-bearing species. However, certain fruits are best enjoyed when they are seedless (e.g. watermelon, grapes) (Acquaah 2004). Odd chromosome number set (e.g. triploidy) results in reproductive sterility. Consequently, crossing a diploid ($2n$) with a tetraploid ($4n$) yields a triploid ($3n$) fruit that is reproductively sterile and seedless (Andrus et al. 1971).

(k) Genetic marker: Visual selection in breeding may be facilitated by using techniques such as *genetic markers.* A genetic marker is a simply inherited and readily identified trait that is linked to a target trait that is difficult to identify or

observe (Acquaah 2004). Conventional breeders use morphological markers in their work. In modern plant breeding molecular or biochemical makers, which are more versatile and efficient, are used to facilitate selection (marker assisted selection – MAS) (Betrand et al. 2008). MAS is not genetic modification (genetic engineering) as alien genes are not incorporated into the product (cultivar) developed. Molecular markers may be enzyme-based or DNA based. Isozyme (or *isoenzymes*; multiple forms of certain enzymes) technology is the first generation biochemical marker technology that was first used in conventional plant breeding (Acquaah 1992). It is an obsolete technology. Modern molecular markers include RFLPs (*restriction fragment length polymorphisms),* SSRs (*minisatellites),* and PCR-based markers (e.g. *random amplified polymorphic DNA* – RAPD and *single nucleotide polymorphism* – SNPs) (Collard et al. 2005).

(l) Mutagenesis: Heritable variation is the lifeblood of plant breeding. Hybridization reorganizes existing variability into new matrices or combinations without creating novel genes. Plant breeders can create non-existing or novel genes through the application of mutagens (agents of mutation – chemical or physical) to create novel sources of variation for breeding (Broetjes and Harten 1998). Mutation breeding continues to be used in modern conventional plant breeding vis-a-vis more precise techniques provided by modern biotechnology (Bado et al. 2013; Matijevic et al. 2013; Shu et al. 2012). Some recent specific applications include the modification of vernalization, growth and development of winter wheat (Czyczyło-Mysza et al. 2013), cultivar development in banana (Jain et al. 2011) and nutritional quality improvement in sorghum (Mehlo et al. 2013). The International Atomic Energy Agency (IAEA) supports research in the use of mutagens for plant breeding (Anonymous 1991).

5.3 Key Genetic Principles and Concepts Pertinent to Conventional Breeding

With time, Mendel's basic principles of genetics were advanced and extended to provide a deeper understanding of heredity, and how to apply it to facilitate plant breeding.

5.3.1 Number of Genes Controlling a Qualitative Trait

The methods used for breeding, and the ease and success of breeding depend on the number of genes that control the trait of interest (Agrawal 1998). Mendel established his laws of inheritance from work conducted on traits that were controlled by one or a few genes (simply inherited or qualitatively inherited; *qualitative traits*).

These traits are easy to manipulate and evaluate in a breeding program because their expressions produce discrete phenotypes that can be distinguished by counting and placing into clear-cut categories (discrete distribution) because the effects of these genes can fall into such non-overlapping groups (Brown and Caligari 2008).

5.3.2 Number of Genes Controlling a Quantitative Trait

Unfortunately, most traits of interest to plant breeders are controlled by many genes (quantitatively inherited; *quantitative traits*), each contributing a small effect to the overall phenotypic expression of a trait (Kearsey and Pooni 1998). In plant breeding, quantitative traits are distinguishable by measuring (metrical traits) the plants (instead of counting).

Quantitative traits are controlled by polygenes (multiple genes) and are also called *polygenic traits* (Lynch and Walsh 1998). The genes have effects that are too small to be individually distinguished. Polygenic inheritance exhibits a phenotypic pattern that is continuous (non-discrete) because segregation occurs at a large number of loci. As a consequence, quantitative traits are more susceptible, than qualitative traits, to modification by the variation in environmental factors to which plants in the population are subjected (Acquaah 2012).

It is challenging to measure the role of the environment on trait expression because it is difficult to measure the environmental effect on plant basis (Mackay et al. 2009). Consequently, it is imperative that breeders evaluate their materials in environments that are similar to those in which the cultivars being developed would be used in crop production (Briggs and Knowles 1967). Another aspect of polygenic trait inheritance is that it is advantageous to breeders if desirable polygenes occur in tight linkages (linkage blocks are transferred together) (Allard 1960).

5.3.3 Gene Action

Biometrical genetics is used to aid breeders in understanding quantitative variation by describing the segregating population in terms of means and variances. Because the effects of quantitative genes do not fall into discrete categories, it is more useful to describe quantitative traits by their gene action rather than by the number of genes by which they are encoded (Mackay et al. 2009). The genetics of a quantitative trait centers on the study of its variation (Acquaah 2012). It is in terms of variation that the primary genetic questions are formulated. Further, the researcher is interested in partitioning variance into its components that are attributed to different causes or sources. The genetic properties of a population are determined by the relative magnitudes of the components of variance (Crow and Kimura 1970). In addition, by knowing the components of variance, one may estimate the relative

importance of the various determinants of phenotype. Total variance of a quantitative trait may be mathematically expressed as follows (Falconer 1981):

$$V_P = V_G + V_E + V_{GE}$$

where V_P = total phenotypic variance of the segregating population; V_G = genetic variance; V_E = *environmental variance*; and V_{GE} = variance associated with the genetic and environmental interaction.

The four types of gene action are additive, dominance, overdominance, and epistatic (Holland 2001). Additive gene action occurs when each additional gene enhances the expression of the trait by equal increments. In a breeding program, selection is most effective for additive variance; it can be fixed by breeding (i.e. a cultivar that is homozygous can be developed) (Acquaah 2012). Dominance gene action pertains to allelic relationship at the same locus. Dominance gene effects are deviations from additivity that make the heterozygote resemble one parent more than the other (Lamkey and Edwards 1999). When dominance is complete, the heterozygote is equal to the homozygote in effects making it impossible for the breeder to distinguish between the two phenotypes. Because both kinds of phenotypes will be selected in a breeding program, fixing superior genes will be less effective where dominance gene action prevails.

Overdominance gene action exists when each allele at a locus produces a separate effect on the phenotype, and their combined effect exceeds the independent effect of the alleles (i.e., $aa = 1$, $AA = 1$, $Aa = 2$) (Lamkey and Edwards 1999). Plant breeders can fix overdominance effects only in the first generation (i.e. F_1 hybrid cultivars) through apomixis (asexual reproduction via seed), or through chromosome doubling of the product of a wide cross). Epistasis is the non-allelic interaction of genes. Epistatic gene action, when it occurs in qualitative traits, manifests as the masking of one gene expression by another (Holland 2001). In quantitative traits, epistasis is described simply as non-allelic gene interactions which can result in an effect where none existed (e.g., $Aabb = 0$, $aaBB = 0$, but $A\text{-}B\text{-} = 4$).

5.3.4 Heritability and Plant Breeding

Plant breeders practice selection in the field on the basis of phenotype (the product of the interaction of genotype and the environment). If a trait is significantly influenced by the environment (e.g. a quantitative trait), a less desirable genotype may be accidentally selected to advance the breeding program. The expected breeding progress or gain will not be realized. *Heritability* is the concept of the reliability of the phenotypic value of a plant as a guide to the breeding value (additive genotype) of a metrical trait (Falconer 1981). It does not measure genetic control, but rather how this control can vary (broad sense, $H = V_G/V_P$ or narrow sense, $h^2 = V_A/V_P$) (Falconer and Mackay 1996). It is also the property of the trait, the population, and the environment. Quantitative traits tend to have lower heritabilities versus

qualitative traits, and hence selection for the former is challenging in breeding. A high heritability (especially narrow sense heritability) is desired in breeding for rapid progress (Lynch and Walsh 1998). In other words, high heritability indicates that a trait can be readily improved by plant breeding.

5.3.5 *The Breeders' Equation*

After the plant breeder has determined that a trait can be improved through breeding, genetic variability is assembled or generated to initiate the breeding program. As the program continues, rounds of selection are implemented with the goal that the trait of interest would be progressively improved with each round. The difference between the mean phenotypic value (value of an individual judged by the mean value of its progeny; it is the value that is transferred from an individual to its progeny) of the offspring of the selected parents and the whole of the parental generation before selection is called the *response to selection (R)* or the *genetic gain* or *genetic advance* (Falconer 1981) Stated differently, the response to selection is simply the change of population mean between generations following selection (Fig. 5.1). The response to selection depends on three factors (Acquaah 2012): (a) The total (phenotypic) variation in the population in which selection will be conducted; (b) Heritability of the target trait; and (c) Selection pressure (i.e. the proportion of the population that is selected for the next generation) to be imposed by the plant breeder.

The mean phenotypic value of the individuals selected as parents for the next generation expressed as a deviation of the population mean is called *selection differential (S).* Response to selection is related to heritability by the following equation (Falconer and Mackay 1996):

$$R = h^2 S$$

Qualitative traits are easier to select (breed) than quantitative traits (low heritability). When heritability is high, a small number of top performers may be selected (high selection pressure). The reverse is true when heritability is low. Further, when heritability is unity ($V_A = V_P$; no environmental variance), progress with selection should be perfect. On the other hand if heritability is zero, ($R = 0$); in theory, the breeder can predict the response to selection in one generation (heritability estimate is valid for one generation).

The response to selection in one generation may be mathematically expressed following (Falconer and Mackay 1996):

$$\bar{X}_0 - \bar{X}_p = ih^2\sigma = R\left(\text{or } \Delta G = ih^2\sigma\right)$$

where $\bar{X}_0$ = mean phenotype of the offspring of selected parents, $\bar{X}_p$ = mean phenotype of the whole parental generation, R = the advance in one generation of

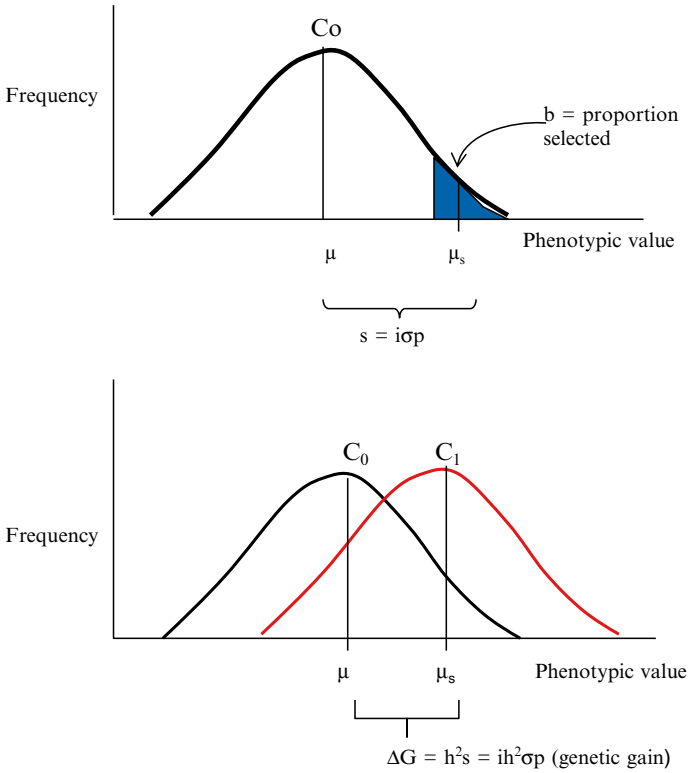

Fig. 5.1 The response to one generation of selection is the genetic gain. It depends on the heritability of the trait and the selection differential

selection, h^2 = heritability, σ = phenotypic standard deviation of the parental population, i = intensity of selection, ΔG = genetic gain or genetic advance.

This equation has been suggested by some to be one of the fundamental equations of plant breeding that must be understood by all breeders, hence called the *breeder's equation* (Acquaah 2012). However, its utility is largely academic because the parameters are seldom available (Bernardo 2010).

5.3.6 Genetic Linkage and Markers

Plant breeders generate large amounts of variability, from which they select individuals to advance their programs. The success of a breeding program depends on selecting the right plants at each stage in the program (Dudley et al. 1991). Field selection is done on phenotypic basis, whereby the breeder selects on visual examination of the variability. If the trait is difficult to visually observe, the gene of interest can be missed during selection, and hence genetic advance during breeding can

be negated (Allard 1960). This can prolong the time for cultivar development. Breeders benefit from knowing of the existence of linkage between the gene(s) of interest and a genetic marker (easy-to-observe trait). This way, genes of interest that are difficult to observe can be assumed to be present when genetic markers to which they are strongly associated are observed (Jain and Kharkwal 2004; Pillen et al. 2003).

5.4 General Steps in Plant Breeding

Certain general steps (activities) are associated with conventional plant breeding described by Acquaah (2012) are: (a) Setting objective(s); (b) Creation/assembly of variation; (c) Selection; (d) Evaluation; (e) Release; (f) Multiplication and (g) Distribution of the new cultivar.

Plant breeders are not responsible for seed multiplication and distribution. However, a seed company usually conducts all these seven activities. A breeder should have a comprehensive plan for a breeding project that addresses these steps.

5.4.1 Objectives

Breeding objectives depend on the species and the intended use of the cultivar to be developed. The breeder should first define a clear objective (or set of objectives) for initiating the breeding program, taking the needs of end-users into account as follows (Acquaah 2012):

(a) The *producer* (grower) – Consider from the point of view of growing the cultivar profitably (e.g. need for high yield, disease resistance, early maturity, lodging resistance).
(b) The *processor* (industrial user) – Consider as it relates to efficiently and economically using the cultivar as raw material for producing new products (e.g. canning qualities, fiber strength, wood quality, mechanized production).
(c) The *consumer* (household user) – Consider their preference (e.g. taste, high nutritional quality, shelf life, aesthetics).

5.4.2 Creation/Assembly of Variation

Genetic variation or variability is indispensable to plant breeding (Allard 1960). After determining the breeding objective(s), the next step is to assemble the requisite germplasm for initiating the breeding program. If, for example, breeding for disease resistance is the objective, the gene conferring resistance to the disease must be available for initiating the base population (Acquaah 2012). The most common

method of introducing the desired gene into the base population is via artificial crossing of appropriate parents (Gur and Zamir 2004). If the gene does not exist, the breeder may attempt to create (induce) it. The common conventional method of creating a nonexistent gene is via mutagenesis (use of mutation agents to induce variation) (Micke 1992).

Variability exists in many plant *gene banks* (repositories for plant germplasm) to which scientists have access. Gene banks may be operated on a small scales by national governments. However, comprehensive operations are undertaken by international entities such as the international research centers (Acquaah 2012). These centers focus on mandate crops (e.g., CYMMIT in Mexico for maize and ICRISAT in India for sorghum). The desired variability may already exist in the breeder's local collection and in remnants from previous breeding projects (Harland and de Wet 1971).

5.4.3 Selection and Evaluation

Selection is simply discriminating among the available or created variability to identify individuals with the desired combination of genes (genotype) or expressed trait(s) (Dudley and Lambert 1992). With the introduction of genetics and statistics into modern plant breeding, breeders have developed standard breeding methods for the species, the genetics of the trait of interest and the type of product desired. There are selection or breeding methods for species based on their modes of reproduction, genetics or whether the product should be uniform or variable (Briggs and Knowles 1967).

The final selection cycle in breeding results in a small number of genotypes that are potential candidates for advancing as cultivars for release to producers. These genotypes are subjected to rigorous *evaluation* under conditions which must include those under which the cultivars will be commercially grown (Borojevic 1990). Evaluations may be conducted at multiple locations and over multiple years (Fehr 1987a). Included in such evaluations are standard cultivars of known performance (for comparison), which could be replaced should superior performers emerge from the trials.

5.4.4 Certification and Cultivar Release

In countries with more advanced agricultural operations, there exist standardized and approved protocols for releasing new cultivars to growers. There may be national crop certifying agencies that oversee the seed certification process for various crops (Agrawal 1998). The ultimate purpose of seed certification is to ensure that the seed produced by the plant breeder reaches the public (consumer) in its highest quality, original genetic identity and highest genetic purity.

5.4.5 *Multiplication and Distribution*

Certified seed is multiplied by certified seed growers contracted by independent breeders and seed companies to mass-produce released cultivars for sale to growers (Allard 1960). New cultivars are sold to consumers via a variety of outlets. Commercial seed companies have elaborate sales mechanisms.

5.5 Mode of Reproduction (Mating System) and Plant Breeding

Modern plant breeding is a highly planned and managed enterprise. A breeding program follows a carefully planned scheme or method. Methods of breeding are classified into these categories based on their modes of reproduction of the species as those for self-pollinated, cross-pollinated or clonally propagated (those that do not bear flowers and are vegetatively propagated) (Allard 1960). Modes of reproduction (sexual or asexual) impact the genetic structure of plants and their populations. Methods of breeding are selected by breeders such that the natural genetic structure of the species is retained in the new cultivar (Acquaah 2012).

Some breeding methods require artificial hybridization or crossing, which entails effective control of pollination to allow only the desired pollen to be included in the cross. To be successful, the breeder needs to understand the reproductive behavior of the species. After a cultivar has been developed, the procedure for its multiplication and maintenance depends on its mode of reproduction. Breeders need to also know about the lifecycle of the species – annual, biennial, or perennial (Borojevic 1990). The methods and strategies used in breeding differ according to these three lifecycles.

5.6 Decision-Making in Breeding

Biometrical genetics can be used to aid plant breeders in making fundamental and critical decisions in their work (Acquaah 2012). Two of these decisions are the cultivar to breed and the selection method that would facilitate the breeding program.

5.6.1 *Best Cultivar to Breed*

The type of cultivar that is desirable is closely related to the breeding or mating system of the species, but more importantly on the genetic control of the traits targeted for manipulation (Acquaah 2012). The breeding system can be temporally

and artificially altered (e.g. cross-pollinated species can be forced to self-pollinate). However, the genetic control of the trait of interest cannot be changed. Breeders should make decisions regarding the type of cultivar to breed based on the genetic architecture of the trait, especially, the nature and extent of dominance and gene interaction, more so than the breeding system of the species (Falconer 1981).

Generally, where additive variance and additive x additive interaction predominate, pure lines and inbred cultivars are appropriate to develop (Falconer 1981). However, where dominance variance and dominance x dominance interaction suggest that overdominance predominate, hybrids would be successful cultivars. Open-pollinated cultivars are suitable where a mixture of the above genetic architecture occurs (Fehr 1987a).

5.6.2 Effective Selection Method for Traits

The genetic control of the trait of interest determines the most effective selection method to use. The breeder should pay attention to the relative contribution of the components of genetic variance (additive, dominance, epistasis) and environmental variance in choosing the best selection method, as previously stated (Kearsey and Pooni 1998). Additive genetic variance can be exploited for long-term genetic gains by concentrating desirable genes in the homozygous state in a genotype. The breeder can make rapid progress where heritability is high by using selection methods that are dependent solely on phenotype (e.g. mass selection) (Acquaah 2012). However, where heritability is low, the method of selection based on families and progeny testing are more effective and efficient. When overdominance predominates, the breeder can exploit short-term genetic gain very quickly by developing hybrid cultivars for the crop (Falconer 1981).

It should be pointed out that as self-fertilizing species attain homozygosity following a cross, they become less responsive to selection (Briggs and Knowles 1967). However, additive genetic variance can be exploited for a longer time in open-pollinated populations because relatively more genetic variation is regularly being generated through ongoing intermating (Eberhart and Gardner 1966).

5.6.3 Selection of Single Traits or Multiple Traits

Plant breeders are often interested in more than one trait that they seek to improve. Even though a breeding program may focus on disease resistance, for example, the breeder is not interested in achieving disease resistance only, but in addition, high yield and other agronomic traits are important (Poehlman and Sleper 1995). It makes little sense to develop a new cultivar that is resistant to a biotic or abiotic stress, and be so low-yielding that it cannot be economically produced. It is fair,

therefore, to say that high or reasonable yield is part of every breeding program (Acquaah 2012).

If traits are negatively correlated, improving one diminishes the other. Selection schemes (e.g. recurrent selection) increase the chance for linkages to be broken to allow negatively linked traits to be simultaneously selected for in a breeding program (Paterniani and Vencovsky 1977).

5.6.4 Pre-breeding

Plant breeding may be categorized into two broad phases or groups of activities – *pre-breeding* and *cultivar development* (Acquaah 2012). Also called germplasm enhancement, this preliminary breeding activity is widely used to introgress or transfer genes (e.g. disease resistance genes) from the wild into breeding materials (Harlan 1976; Zamir 2001) (Fig. 5.2). It is time consuming and resource intensive, making it unattractive to breeders in general. Consequently, such undertakings are often funded by governments or international agencies. The products are then available to the general public.

Plant genetic resources (germplasm) used in plant breeding were categorized into three gene pools (Harlan and de Wet 1971). Gene pool 1 (GP1) consists of natural biological species, domesticated, and wild and weedy, which are able to intercross without genetic consequences (infertility). Crosses involving GP2 and more so GP3 species are problematic. A newly added category, GP4, consists of synthetic strains (do not occur in nature) created by scientists (Harlan and de Wet 1971). To be accessible to breeders for cultivar development, traits existing in distant gene pools must first be transferred into domesticated breeding materials by pre-breeding

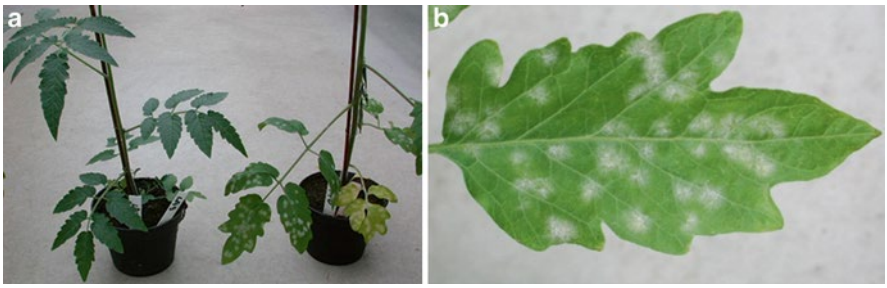

Fig. 5.2 Tomato plants inoculated with tomato powdery mildew. (**a**) The *left* plant is from tomato wild species *Solanum pervianum* LA2172, showing no powdery mildew infection; the *right* plant is from *S. lycoerpsicum* cv. Moneymaker (MM), showing fungal colonies growing on infected leaves. (**b**) A closer look at the colonization of tomato powdery mildew (*Oidium neolycopersici*) growing on the upper-side of MM leaf. Pictures were taken 15 days post inoculation (Source: Courtesy of Yurling Bai, Wageningen UR Plant Breeding, Netherlands)

(Gepts 2002). Consequently, the duration of a breeding program depends on whether the breeder has a cultivar development-ready germplasm or not.

5.6.5 Duration of Plant Breeding

Plant breeding, conventional or molecular, is a lengthy activity that is influenced by numerous factors. Duration of breeding can be anywhere from a few years (4–6 years) to over 10 years (Poehlman and Sleper 1995). The lifecycle of the species is important, annuals taking a shorter time to breed than tree crops. If the gene of interest exists in the wild gene pool, the time for pre-breeding to introduce it into domesticated pool, thereby lengthening the breeding program (Chahal and Gosal 2002). Qualitative traits are easier to breed than quantitative traits. If the trait is recessive, an additional step of selfing would be required during each cycle for selection purposes. Some breeding methods are more involved than others (Acquaah 2012).

The breeding objective also determines the duration of the program. Breeding for abiotic or biotic stress resistance may require ideal environmental conditions for optimal selection for rapid genetic gains. Breeding for resistance to airborne diseases and improved aboveground traits are much easier than addressing soil-borne diseases or underground traits. If high uniformity of the final product is needed, breeders may need additional cycles of selfing, thus prolonging the program (Brown and Caligari 2008).

The location of the breeding program is also important in the duration of cultivar development It helps if the breeder can have two growing seasons in which to work. That failing, some breeders in the temperate regions have *winter nurseries* in tropical areas for work during the winter season. Sometimes, it may be possible to do some work in the greenhouse. When putative cultivars have been identified for the evaluation stage, it helps if field trials can be conducted more than once a year (Allard 1960).

5.6.6 Plant Breeding as a Numbers Game

Plant breeders encounter many situations in their breeding work which involves numbers (Acquaah 2012). They need to know if a trait is qualitative or quantitative (number of genes – few or numerous). They also need to know the number of parents to initiate the program or advance it at each stage in the selection process (selection intensity) (Brown and Caligari 2008). After crossing, breeders must decide the number of cycles of selfing needed, number of times to backcross to the adapted parent, or number of recurrent cycles to conduct (Borojevic 1990).

At the evaluation stage, the breeder decides the number of locations at which to conduct the trials, the number of years of testing and the number of cultivars to release.

5.7 Genetic Structure of Plant Cultivars

Based on the mating system of the species and the method of breeding, the cultivars developed vary in genetic structure and phenotypic uniformity. This condition determines how the farmer maintains the cultivars, with regard to planting in subsequent seasons. To this end, cultivars may be classified into several categories (see Sect. 5.14) (Agrawal 1998):

(a) Homozygous and homogeneous: This breeding strategy is applied to self-pollinated species. The product of economic importance is uniform, and famers may, if legally permissible, save seed from the current season for planting the next season without genetic consequences. (Acquaah 2012).
(b) Heterozygous and homogeneous: This breeding strategy is widely applied to cross-pollinated species. The cultivar may be genetically heterozygous and remain phenotypically homogeneous, as in hybrid cultivars (Crow 1988). The genetic structure is maintained only for one season. Consequently, saving seed from the current crop to plant in the next season's crop has severe genetic and yield consequences (the hybrid cultivar is F_1 and uniform, but the next seed is F_2 and heterogeneous) (Bauman and Crane 1992).
(c) Heterozygous and heterogeneous: Some cultivars are deliberate blends or mixtures, giving them broad adaptation for success in production systems such as obtains in developing countries (Allard 1960).
(d) Homozygous and heterogeneous: An example of such a cultivar is a landrace. The component genotypes may each be homogeneous but there is such a large amount of diversity in the composite genotypes such that the cultivar is heterogeneous (Borojevic 1990).
(e) Clonal: Clones are identical to the parent material. The uniqueness about this cultivar is that heterozygosity can be propagated without change or genetic consequence (hybrid vigor is fixed) through subsequent generations as would occur in a hybrid of an open-pollinated species. The offspring of a clonal population is homogeneous, even though the parents are heterozygous (Acquaah 2004).

5.8 Types of Cultivars

Plant breeders develop six basic types of cultivars from four basic populations: inbred pure lines, open-pollinated populations, hybrids and clones (Acquaah 2012).

5.8.1 Pure-Line Cultivars

Pure-line cultivars are developed for species that are highly self-pollinated (inbred). These cultivars are homogeneous and homozygous in genetic structure, a condition attained through a series of self-pollination (Briggs and Knowles 1967). These cultivars are often used as parents in the production of other kinds of cultivars. Pure line cultivars have a narrow genetic base (Brown and Caligari 2008). They are desired in regions where uniformity of the market product has high premium.

5.8.2 Open-Pollinated Cultivars

Contrary to pure-lines, *open-pollinated cultivars* are developed for species that are naturally cross-pollinated. The cultivars are genetically heterogeneous and heterozygous (Burton and Brim 1981). Two basic types of open-pollinated cultivars are developed by breeders. One type is developed by improving the general population by *recurrent* (or repeated) *selection* or bulking and increasing the material from selected superior inbred lines (Burton and Brim 1981). The other type, called a *synthetic cultivar,* is derived from planned inter-matings involving selected genotypes. Open pollinated cultivars have a broad genetic base (Borojevic 1990).

5.8.3 Hybrid Cultivars

Hybrid cultivars are produced by crossing inbred lines that have been evaluated for their ability to produce hybrids with superior vigor (performance) over and above those of the parents used in the cross (Betrán and Hallauer 1996). Hybrid production exploits the phenomenon of *hybrid vigor* (or *heterosis*) to produce superior yields (Lamkey and Edwards 1999). Heterosis is usually less in crosses involving self-pollinated species than those involving cross-pollinated species.

Hybrid cultivars are homogeneous but highly heterozygous. Pollination is highly controlled and restricted in hybrid breeding to only the designated pollen source (Stuber et al. 1992). In the past, physical (human) intervention was required to enforce this strict pollination requirement, making hybrid seed expensive. However, with time, various techniques have been developed to capitalize on natural reproductive control systems (e.g. *male sterility*) to facilitate hybrid production (Norskog 1995). Hybrid production is more widespread in cross-pollinated species (e.g. maize, sorghum), because the natural reproductive mechanisms (e.g. cross fertilization, *cytoplasmic male sterility*) are more readily economically exploitable than in self-pollinated species (Betrán and Hallauer 1996).

5.8.4 Clonal Cultivars

Seeds are used to produce most commercial crop plants. However, a significant number of species are propagated by using plant parts (e.g. stems and roots) other than seed (Fehr 1987a). By using vegetative parts, the cultivar produced consists of plants with identical genotypes and is homogeneous. However, the cultivar is genetically highly heterozygous. Some plant species are sexually reproducing but are propagated clonally by choice (Fehr 1987a). Such species are improved through hybridization, so that when hybrid vigor exists, it can be fixed (i.e. vigor is retained from one generation to another), and then the improved cultivar propagated asexually (Acquaah 2012).

In seed propagated hybrids, hybrid vigor is highest in the F_1, but is reduced by 50 % in each subsequent generation (Briggs and Knowles 1967). In other words, whereas clonally propagated hybrid cultivars may be harvested and used for planting the next season's crop without adverse genetic effects, producers of sexually reproducing species using hybrid seed, must obtain a new supply of seed as previously indicated (Allard 1960).

5.8.5 Apomictic Cultivars

Apomixis is the phenomenon of producing seed without the benefit of the union of sperm and egg cells (i.e. without fertilization) (Savidan 2000). The seed harvested is hence genetically identical to the mother plant (in much the same way as clonal cultivars) (Hanna and Bashaw 1987). Hence, apomictic cultivars have the same benefits of clonally propagated ones, as previously discussed. In addition, they have the convenience of vegetative propagation through seed (versus propagation through cuttings or vegetative plant parts). Apomixis is common in perennial forage grasses and tree fruits (e.g. mango, citrus) (Savidan 2000).

5.8.6 Multilines

Multilines are developed for self-pollinating species (Allard 1987). These cultivars consist of a mixture of specially developed genotypes called *isolines* (or *near isogenic lines*) because they differ only in a single gene (or a defined set of genes). Isolines are developed primarily for disease control, even though these cultivars, potentially, could be developed to address other environmental stresses (Chahal and Gosal 2002). Isolines are developed by using the techniques of *backcrossing* in which the F_1 is repeatedly crossed to one of the parents that lacked the gene of interest (e.g. disease resistance) (Acquaah 2012).

5.9 Developing Self-Pollinated Cultivars

Self-pollinated cultivars may be developed from either a single plant or a mixture of plants, each with implications for the genetic structure of the cultivar (Briggs and Knowles 1967). The methods used for breeding self-pollinated species may be classified into two – one that is preceded by a planned cross (hybridization) and one that is not (Acquaah 2012). Further, the final product (cultivar) may be derived from a single plant (after the cross, if there is one), or a mixture. The methods used for cultivars deriving from a single plant include the following basic ones – pure line, pedigree, single seed descent, bulk, backcross. Methods that use a mixture of plants include the following basic ones: mass selection, composites or blends, and recurrent selection (Borojevic 1990).

5.9.1 Mass Selection

Mass selection is applicable to both self- and cross-pollinated species, but with different genetic consequences (Allard 1960). The persistence of inbreeding alters population gene frequencies reducing heterozygosity from one generation to the next in self-pollinated species (Acquaah 2012). It is a population improvement strategy whose purpose is to increase the gene frequencies of desirable genes and therefore increase the average performance of the base population. It acts on existing variability without creating a new one. Procedurally, off-types or undesirable plants are rogued out of the population; alternatively, desirable plants may be selected and bulked from the existing population to form the improved cultivar (Fig. 5.3). Single plants, pods, or heads as applicable, may be the selection unit (Baezinger et al. 2006).

Mass selection may be used to maintain the purity of an existing cultivar after it has been contaminated (e.g. by natural outcrossing, spontaneous mutation, mechanical mixing). It is used in breeding horizontal (durable) resistance into a cultivar, and for adapting a new cultivar to a production region. Mass selection is based on plant phenotype and is hence most effective if the trait of interest is highly heritable (Brown and Caligari 2008).

The resulting cultivar is fairly uniform (for the trait of interest), but genotypically, it could comprise a large number of pure lines, thus making is genetically broad-based, adaptable and stable. It is inexpensive, simple and rapid to conduct, requiring only one generation per cycle. Large populations can be handled at one time.

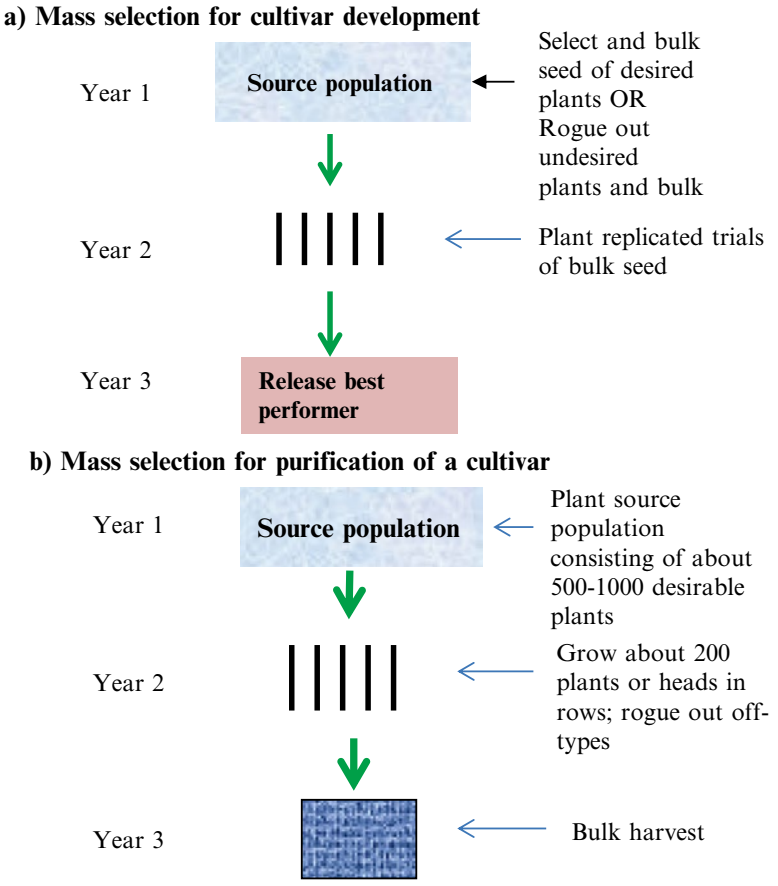

Fig. 5.3 Mass selection may be conducted for cultivar development (**a**) or cultivar purification (**b**), by roguing out off-types or selecting and advancing desirable plants in a bulk

5.9.2 Pure Line Selection

In theory, a *pure line* cultivar is one in which the plants have identical alleles at all loci, something that is impractical to achieve (Allard 1960). It has a narrow genetic base, and the traits of interest tend to be phenotypically uniform (Acquaah 2012). Such cultivars are suited to uses that demand uniformity of traits (e.g. uniform maturity or ripening for mechanized farming, uniform shape and size for processing or a discriminating market). It is rapid and easy to conduct, for once a genotype has been selected from a variable population, it is repeatedly selfed until there is no noticeable segregation in the progeny (Poehlman and Sleper 1995) (Fig. 5.4).

Selection based on progeny performance makes it suitable for improving traits with low heritabilities. Pure line selection can also be part of other breeding methods (e.g. bulk breeding). Pure line cultivars lack yield stability over a wide range of

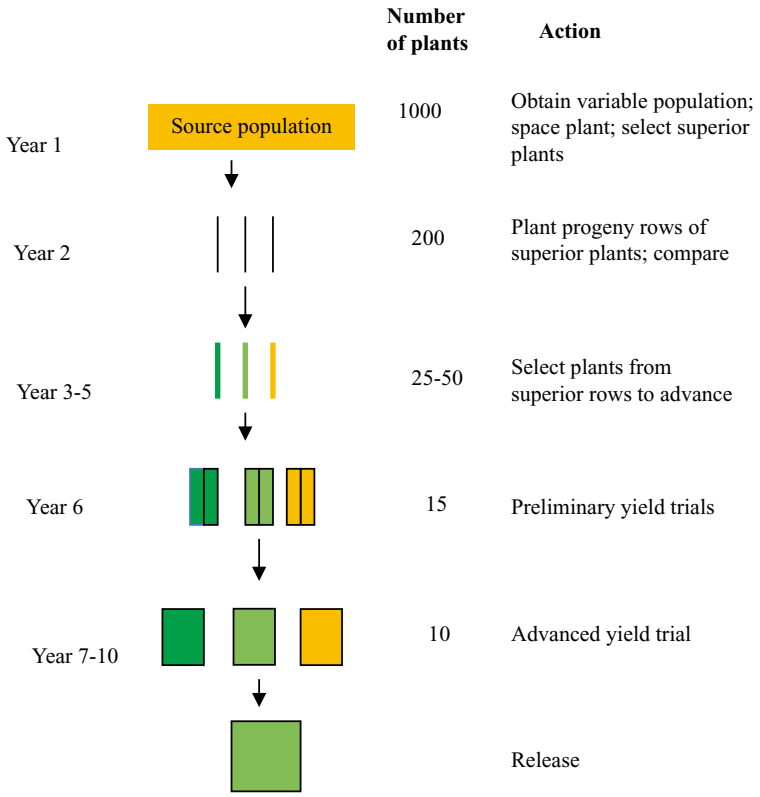

Fig. 5.4 General steps for breeding by pure line selections

environments and are susceptible to devastation by pathogenic outbreaks (Allard 1960).

5.9.3 Pedigree Selection

A key difference between *pedigree* and mass or pure line selection methods is that, as the name implies, pedigrees are descended from known parents (Poehlman and Slepper 1995). Consequently, hybridization is used to establish the base population for the breeding program. To maintain ancestry of the cultivar, the breeder must keep meticulous records such that the parent-progeny can be traced back to an individual F_2 plant from any subsequent generation (Acquaah 2012). The method is suited to species that allow individual plants to be observed, selected and harvested separately (e.g. peanut, tomato, tobacco).

Starting, usually with the F_2, plants are reselected in subsequent generations until a desirable level of homozygosity is attained (Allard 1960) (Fig. 5.5). Lines begin to form by the F_4 and as such selection should be on the basis of progenies rather

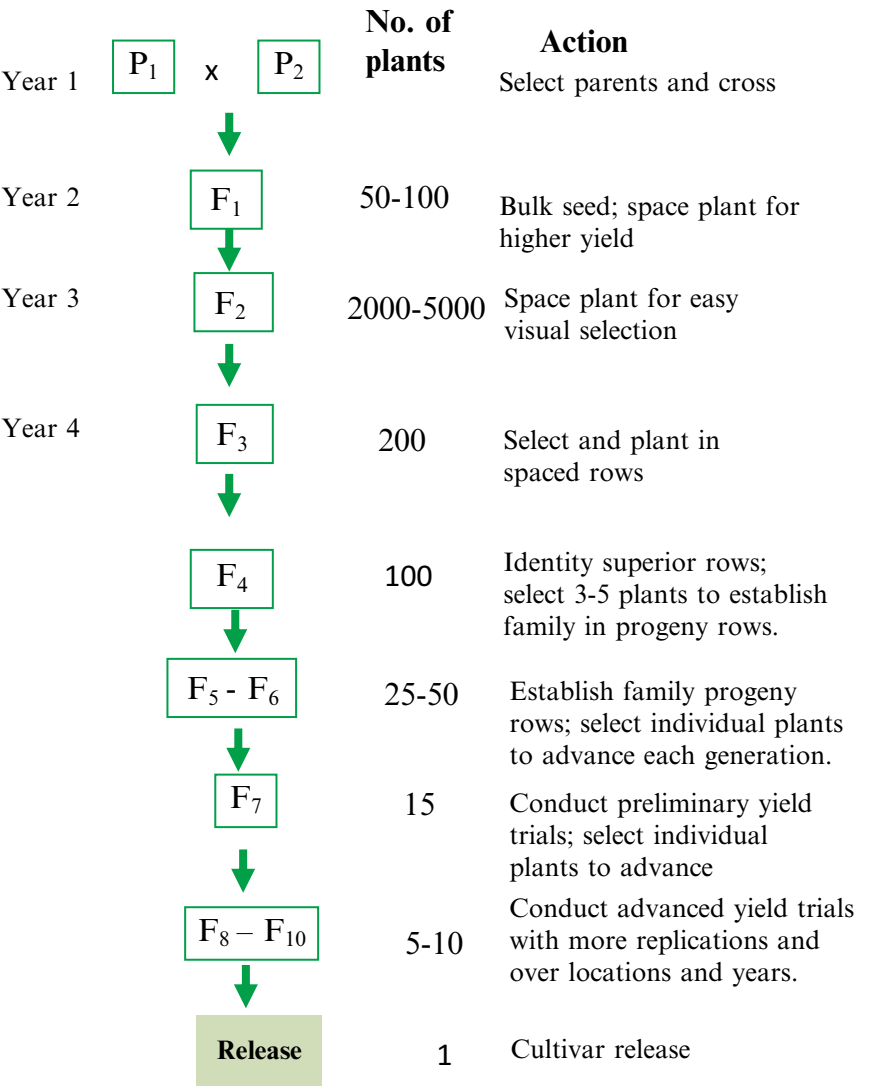

Fig. 5.5 Pedigree selection starts with a cross of known parents whose identity is maintained through meticulous record-keeping throughout the breeding process

than individual plants. The method is tedious and time-consuming, prolonging the breeding program.

5.9.4 Bulk Population Breeding

A key distinguishing feature of the bulk population breeding method is that artificial selection is delayed, thereby allowing natural selection pressure to bear on the initial variability (Briggs and Knowles 1967). The rationale is that natural selection would weed out individuals that are susceptible to abiotic factors in the production region (e.g. drought, cold, photoperiodic response), before emphasis is placed on the trait of interest (Acquaah 2012). The cultivar that emerges would already be adapted to the production region. It is suitable for breeding self-pollinating species that are closely spaced in production (e.g. small grains like wheat, barley; also field beans, soybean).

After initiating the program with a cross, the F_2 is planted and bulk-harvested (Fig. 5.6). A sample of the harvest is used to initiate the next bulk planting cycle until F_5 when selected plants are space-planted for individual observation (Allard 1960). It is easy to conduct and less labor intensive in the early stages, allowing large amounts of segregating material to be handled. A negative aspect of this scheme is that, under natural selection, a desirable genotype (even possibly the target trait) that is not competitive can be lost in the early generations, while an aggressive but undesirable genotype may persist to later generations (Acquaah 2012).

5.9.5 Single-Seed Descent

Single-seed descent is a method for speeding up a breeding program by randomly selecting a single seed from each of the desirable F_2 plants for planting the next generation (Allard 1960). This strategy is continued through F_3. This way, the breeder is able to advance a larger than normal number of F_2 plants through several generations, thereby attaining homozygosity rapidly, while delaying selection. It is best suited to breeding self-pollinated small grains and legumes like soybean (Tigchelaat and Casali 1976). The early generations require a small space (can be conducted in a greenhouse).

5.9.6 Backcross Breeding

Backcross breeding is undertaken to transfer one or a few specific genes of interest from a source (donor parent) to an adapted cultivar (breeding line), while preserving all other qualities (Acquaah 2012). After initiating the transfer via a cross, the

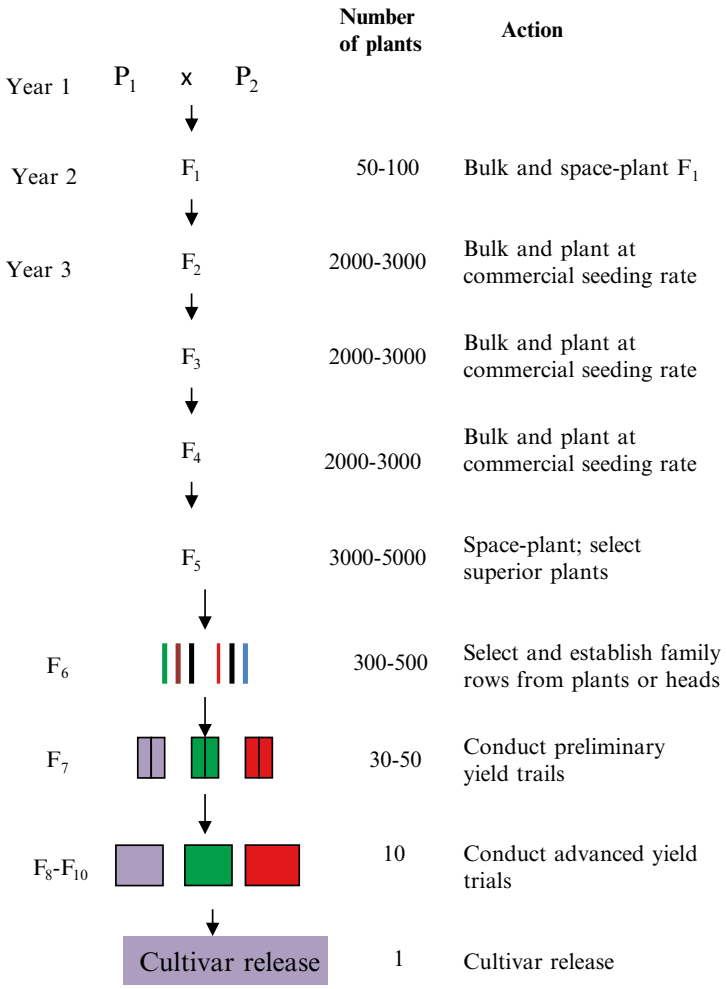

Fig. 5.6 Bulk selection advances the products of a cross for several generations before space planting for single plant selection

desirable (adapted cultivar; called the *recurrent parent*) is repeatedly crossed (backcrossed) to the F_1 to retrieve all the desirable genes of the adapted cultivar (Fig. 5.7). Backcrossing may be done for 2–5 cycles (BC_1 – BC_5), depending on how much of the recurrent parent the breeder wants to recover, and how easy it is to observe the trait of interest.

After adequate backcrossing, the breeder selfs the product (e.g. BC_5F_1, and then BC_5F_2) to stabilize the gene in a homozygous state (Chahal and Gosal 2000).

The method works best for qualitative traits. The procedure is used to introgress genes from wild gene pools into domesticated ones. It is also used to develop

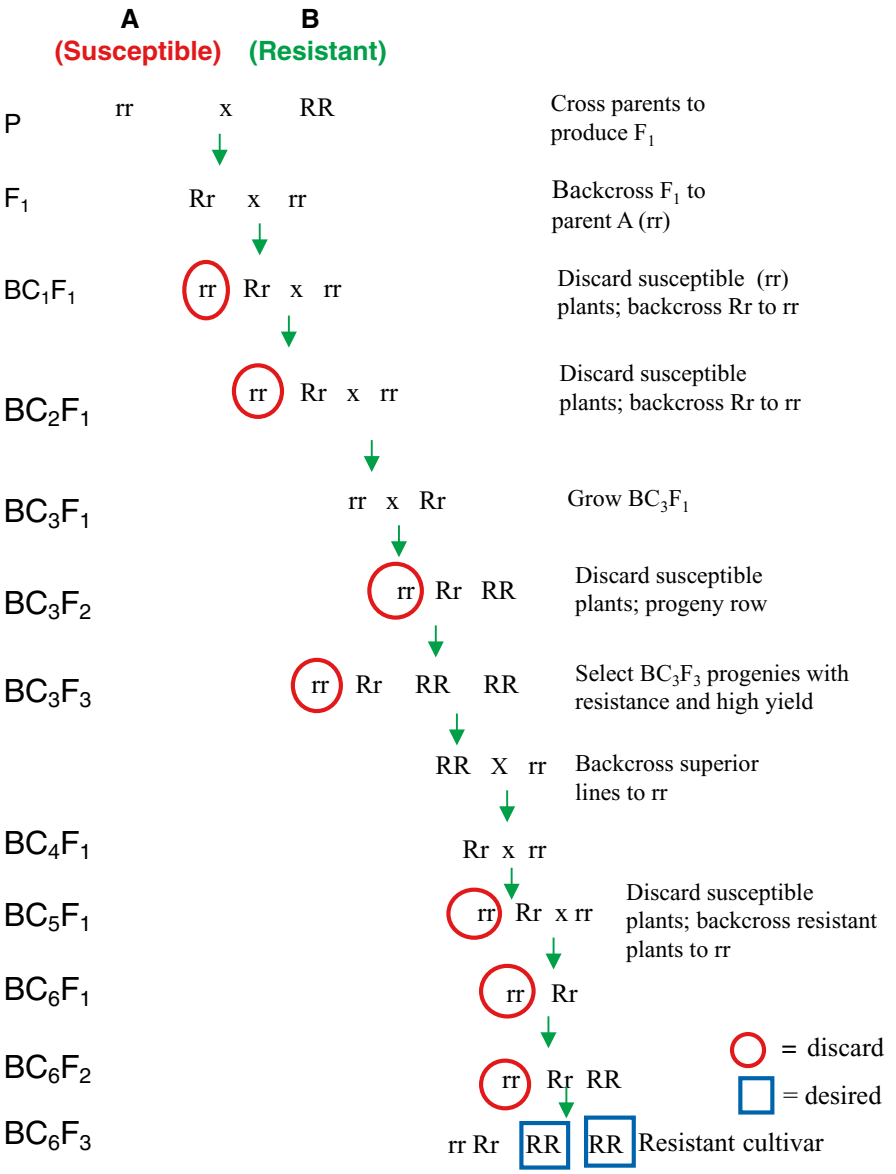

Fig. 5.7 General steps for backcross breeding for improving a dominant trait

isogenic lines (Borojevic 1990). When the gene of interest is recessive, an additional step is needed in each cycle to allow the breeder to distinguish between the homozygote (*RR)* from the heterozygote (*Rr*) which carries the desired recessive gene (Acquaah 2012). By selfing, the breeder will be able to identify the desired *rr* genotype for subsequent backcrossing cycles to the recurrent parent. It is important

what parent is used as female in a cross, because of cytoplasmic inheritance (especially, when using CMS in breeding). If the gene of interest is tightly linked with undesirable genes, the product of backcross breeding could suffer from the consequences of linkage drag (Allard 1960).

5.9.7 Multiline Breeding and Cultivar Blends

A *multiline* or *blend* is a planned seed mixture of cultivars or lines (multiple pure lines) composted in a predetermined ratio (Fehr 1987a). Each component line (isoline) is developed by a separate backcross, making the method expensive to conduct. The strategy of multiline is to increase heterogeneity of the self-pollinated cultivar so that it would have greater environmental buffering to reduce the risk of total loss from biotic or abiotic factors (Agrawal 1998). Technically, a multiline is spatially differentiated, plant-to-plant, such that the cultivar creates a mosaic of genotypes in the field. This physical layout buffers against rapid spread of disease. The multiline cultivar appears phenotypically uniform and provides yield stability (Acquaah 2012). For example, the component lines may provide resistance to different races of a pathogen.

The cultivar is widely used in the turfgrass industry. A blend or multiline can be reconstituted, should a component line fail to perform as planned. As disease patterns change the component lines are periodically changed as well, through reconstituting (Fehr 1987a).

5.9.8 Composites

A *composite* cultivar is similar to a multiline, in that it consists of a mixture of different genotypes (Jensen 1978). However, the component genotypes are not closely related (as in isolines), but may be inbred lines, hybrids, populations and other dissimilar genotypes, which are carefully selected to have some similarity in agronomic benefit (e.g. similar growth habit, disease resistance) (Acquaah 2012).

5.10 Developing Cross-Pollinated Cultivars

Breeders adopt one of three general approaches in developing cross-pollinated cultivars – population improvement, synthetic cultivars, or hybrid cultivars. Each of these have consequences on the genetic structure of the cultivar develop and hence how it is maintained and propagated (Acquaah 2012).

5.10.1 Population Improvement and Recurrent Selection

Self-pollinated species utilize mostly their own pollen for pollination, whereas cross-pollinated species may receive pollen from other plants. Consequently, plant breeding methods for self-pollinated species tend to focus on improving individual plants, whereas those for cross-pollinated species tend to focus on improving the population. The material may be from one specific population (*intrapopulation improvement*) or based on a cross between two populations (*interpopulation improvement;* for hybrid cultivar development). Intrapopulation improvement may be used to produce a synthetic cultivar or develop pure lines for hybrid cultivar development (Acquaah 2012).

Recurrent selection is a widely used method for population improvement. In this method of breeding, a number of plants are advanced to the next generation in which another cycle of intermating occurs (Burton and Brim 1981). This process is repeated for several cycles (hence, recurrent selection), the outcome being an improved population (superior to original population in mean performance as pertains to trait(s) of interest) with high genetic variability (Menz and Hallauer 1997). Repeated crossing provides opportunity for genetic recombination (also, increased opportunity for linkages to be broken) to occur to increase genetic diversity in the population.

Recurrent selection comprises three key steps – parents are crossed in all possible combinations; plants or families are evaluated and a new set of parents selected; and the selected parents are intermated to produce material for the next selection cycle (Menz and Hallauer 1997) (Fig. 5.8). There are several recurrent selection schemes, some of which involve the use of testers. Testers are useful especially when the traits of interest are of low heritability.

The four basic recurrent selection schemes are (a) Simple recurrent selection (no tester used), (b) recurrent selection for general combining ability (is a half-sib progeny test involving a wide genetic base tester; only one parent in the cross is known), (c) recurrent selection for specific combining ability (uses narrow genetic base tester like an inbred line) and (d) reciprocal recurrent selection (involves two heterozy-

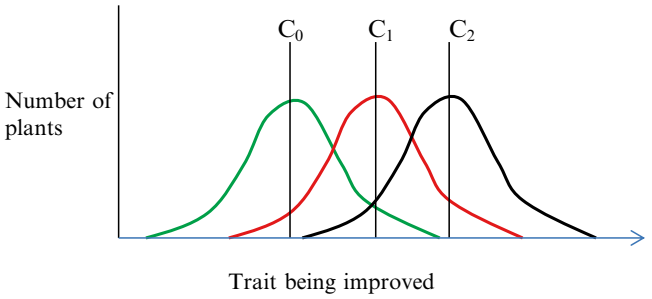

Fig. 5.8 Recurrent selection progressively increases the phenotypic expression of the target trait

gous populations, each serving as a tester for the other) (Acquaah 2012; Briggs and Knowles 1967).

5.10.2 *Intrapopulation and Interpopulation Improvement Methods*

Intra-population methods include mass selection and family selection (half- or full-sib). Just like intrapopulation methods, interpopulation schemes include both half- and full-sib methods.

(a) Mass selection: Mass selection for population improvement differs from that for self-pollinating cultivar development in that it aims to improve the general population performance by selecting and bulking superior genotypes already present in the base population, following a progeny test (Allard 1960).
(b) Family selection: Family selection methods may involve half-sib selection (e.g. ear-to-row selection) in which half-sib families are created for evaluation and recombination, both steps occurring in just one generation (Acquaah 2012; Feng et al. 2004). The method is applicable to breeding perennial forage grasses and legumes (Fig. 5.9). Full-sib methods start with biparental crosses using parents from the base population, which are then evaluated in replicated trials to identify superior full-sib families for initiating the next cycle (Paterniani and Vencovsky 1977, 1978). A third family selection, selfed-families (S_1 or S_2), is best suited to breeding self-pollinated species, even though it has been applied to breeding maize (Fig. 5.10).

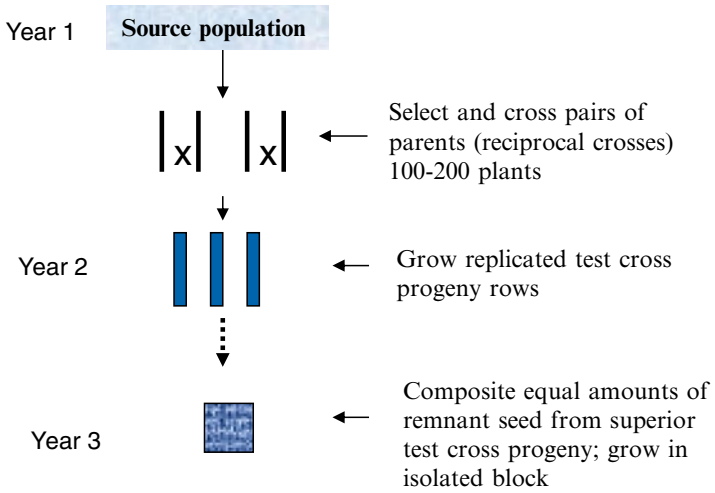

Fig. 5.9 General steps for conducting breeding using the full-sib method

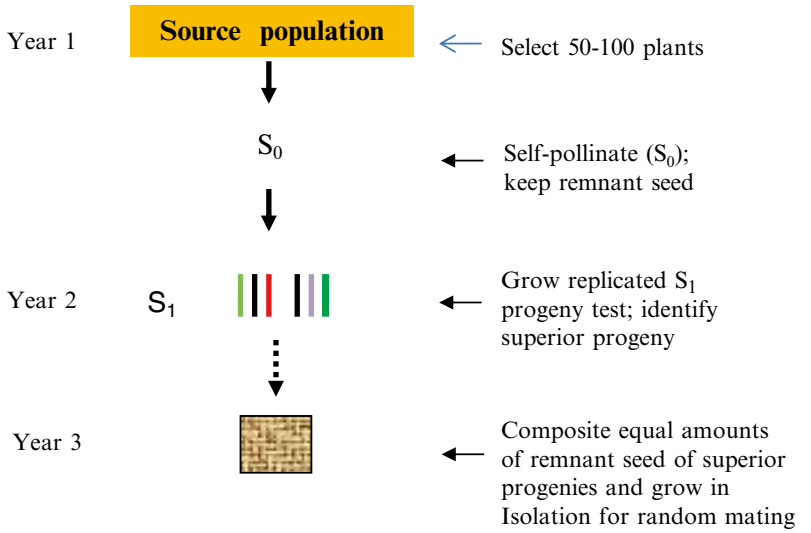

Fig. 5.10 General steps for conducting breeding using the S_1/S_2 progeny performance method

5.11 Development of Synthetic Cultivars

Population improvement methods may involve just purely phenotypic selection (mass selection) and those that involve progeny testing (Comstock et al. 1949). Whereas composites are basically mixes of different materials on the basis of selected agronomic traits (e.g. maturity, disease resistance), a *synthetic cultivar* is created from a cross-fertilized (random mating in all combinations) mixture of parents (e.g. clones, hybrids) selected on the basis of progeny testing or their GCA analysis using a test cross or top cross, but commonly a polycross, for evaluation (Acquaah 2012). Even though *syn-1* (equivalent of F_1 for hybrid cultivars) has the highest yield productivity, there is often insufficient seed produced for distribution to farmers. Practically, breeders open pollinate the *syn-1* to create *syn-2*. However, synthetics of autotetraploids (e.g. alfalfa) suffer significant decline in vigor (inbreeding depression) when the *syn-2* generation is created (Briggs and Knowles 1967).

A synthetic is usually reconstituted after use in cultivation for a number of generations. The method is well suited for breeding forage species, but has been successfully used for other crops (e.g. maize – Iowa stiff-stalk synthetic, sugar beets). Synthetics are suited for production regions where farmers commonly save seed for planting the next year's crop (Allard 1960).

5.12 Development of Hybrid Cultivars

In breeding *hybrid cultivars*, plant breeders exploit the phenomenon of *hybrid vigor* (*heterosis*), whereby the performance (e.g. vigor, fertility, overall productivity) of the hybrid (F_1) exceeds the average performance of the parents (P_1 and P_2). Though widespread in the plant kingdom, heterosis is more pronounced in cross-pollinated species than self-pollinated species. It increases as the genetic distance between the parental inbreds increases (Springer and Stupar 2007).

Heterosis is calculated as the difference between the crossbred and inbred means (Falconer 1981):

$$\text{Hybrid vigor} = \left\{ \left[F_1 \quad \left(P_1 + P_2 \right) / 2 \right] / \left[\left(P_1 + P_2 \right) / 2 \right] \right\}$$

It is manifest when the parents are genetically diverse. Typically, breeders start the program by developing inbred lines (repeatedly crossed or selfed parent) of outcrossed species (Comstock et al. 1949). Selfing (inbreeding) plant produces the opposite effect of heterosis (*inbreeding depression*). Crossing two inbred lines restores the lost vigor via genetic mechanisms that remain debatable. Because heterosis is subject to genotype x environment interaction, breeders prefer to determine it for a particular hybrid line and for a specific trait and specified environmental conditions (Nassimi et al. 2006).

The most costly and time consuming phase in a hybrid program is the identification of parental lines that would produce superior hybrids when crossed (Hallauer 1967). Breeders utilize various techniques to find parents that will *nick* in a program to optimize heterosis, one of which is *combining ability* (Griffing 1956). Two parental lines have high combining ability when their cross produces high heterosis in their progeny. Similarly, a group of related or unrelated genotypes from the same or different populations, which display combining ability when crossed with other genotypes from other germplasm groups may be identified as a *heterotic group*. Such grouping or classification helps breeders develop hybrid cultivars with high predictability of performance (Melchinger and Gumbler 1998).

5.12.1 Types of Commercial Hybrids

The earliest commercial hybrids were single crosses (AxB). Later double crosses, which involved four parents ([*AxB*] x [*CxD*]), were thought to provide more yield stability and more economical. Other parental combinations have been proposed (Fehr 1987b). However, modern hybrids are mostly single crosses, largely because breeders are able to use techniques to identify superior parents with outstanding combining abilities (Crow 1998).

5.12.2 Inbred Lines

An *inbred line* is a breeding material that is homozygous. The rationale of developing inbred lines for hybrid seed production is that heterosis would be highest when one allele is fixed in one parent to be used in a cross and the other allele is fixed in the other parent (Lamkey and Edwards 1999). Inbred lines may be developed for both self- and cross-pollinated species. Because species vary in tolerance to inbreeding, the number of cycles of inbreeding used in inbred line development is variable, but usually 5–7 generations would suffice for self-fertilized species (Acquaah 2012).

Developing inbred lines for cross-pollinated species is more challenging, requiring the need for pollen control to be successful (Norskog 1995). They may be initiated with materials from natural populations or from F_2. Two strategies, conventional (normal) and non-conventional, may be used. Normal inbreds are developed by repeatedly self-pollinating selected plants, from S_o-S_n (when using natural populations) or from F_1 to F_n (when using materials obtained from crossing of F_2) (Allard 1960). In cross-pollinated species, pollen control is critical and may be effected via physical of genetic methods. Physical methods are laborious and include the covering of floral parts (with paper bags) or emasculation (Fehr 1987b).

Genetic control is commonly achieved by using male sterility genes in lieu of emasculation (Poehlman and Sleper 1995). To achieve this, breeders incorporate *cytoplasmic male steritly (CMS)* into inbred lines. Two different female lines and one male line are needed. The female lines are *A-line* (male sterile, sterile cytoplasm (s), nonrestorer genes (*rfrf*) in the nucleus), and a *B-line* (male fertile, fertile cytoplasm (N), with nonrestorer genes *(rfrf*) in the nucleus). The *A-line* is the seed-producing parent. The male line, called the *R-line*, has the genotype *Rf Rf* and is the pollen parent in the hybrid program. These three unique lines are maintained by special techniques (Bauman and Crane 1992). The male sterility system may also be exploited for hybrid seed production (Fig. 5.11).

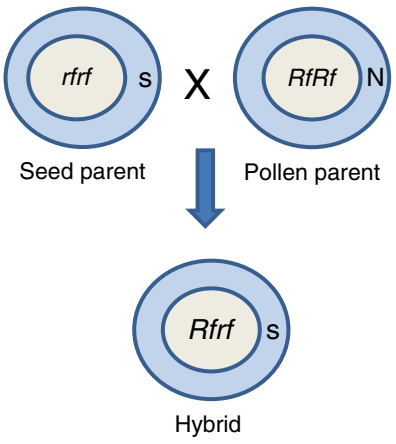

Fig. 5.11 Using cytoplasmic male sterility (CMS) genetic system in single cross hybrid breeding

It is important to evaluate the inbred lines for performance and the presence of general agronomic qualities (e.g. drought resistance, disease resistance, lodging resistance, etc.) that are of importance to the regions where the cultivar will be used, and the end-user needs (Acquaah 2012). Once superior inbreds have been identified for hybrid seed development, the breeder plants them in the field according to a design that would maximize hybrid seed yield (Kempe and Gils 2011). Factors to consider in field planting include planting time, synchronization of flowering, spacing or plant density.

5.13 Developing Clonal Cultivars

Clones are identical copies of a genotype. The economic parts of some clonal cultivars are vegetative while others are fruits. Some clones produce flowers (with poor or good seed set); others are non-flowering. Breeding methods differ for these categories (Fehr 1987b).

Species with normal flowers and good seed set are usually commercially produced vegetatively (e.g. sugarcane). They are bred via hybridization methods to transfer desirable genes into improved cultivars (Fig. 5.12) Heterosis in clones is fixed in the hybrid product and maintained indefinitely (Melchinger and Gumber 1998). The hybrid is ready to be released immediately following a cross without the lengthy process of additional steps as obtains in other species. In species with poor seed set, crossing can nonetheless be used to transfer genes into adapted cultivars, albeit with more difficulty than where seed is copious.

Non-flowering species are essentially obligate and asexually propagated species. Mutation breeding technique may be used to generate genetic variability for crop improvement. Somatic mutations are characterized by chimerism (tissues in the same plant with different genetic characteristics) (Micke 1992).

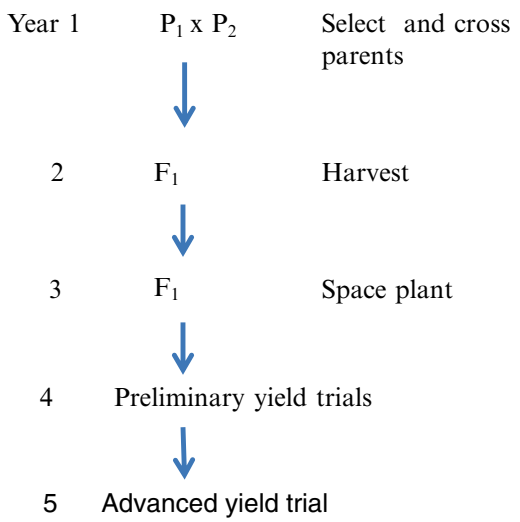

Fig. 5.12 General steps for breeding flowering apomictic species

5.14 Apomixis

The widespread phenomenon whereby certain plant species produce seed without fertilization is called *apomixis* (asexual production of seed) (Acquaah 2012). Apomictically produced seed are essentially clones of the mother plant (natural mechanism for cloning plants through seed) (Hanna and Bashaw 1987). Some species (e.g. bluegrass, *Poa pratensis*) have the capacity to produce seeds via both sexual and apomictic modes (partial apomicts), versus complete apomicts like bahiagrass (*Paspalum notatum*).

If apomixis is effectively exploited as a breeding tool, hybrid seed breeders would not need to make crosses each year, and vigor would be retained and duplicated year after year without decline (Fehr 1987a). The commercial hybrid production method could be simplified (no need for fertility control and field isolation). For the producer, if legally permissible, seed could be saved for planting in subsequent years without repurchase from the seed company (hybrid vigor is fixed). It should be noted that apomixis arise by several mechanism, which impact how the phenomenon can be effectively exploited in breeding (Hanna and Bashaw 1987).

5.15 Polyploidy in Plant Breeding

Polyploids are individuals whose cells possess multiples of the basic set of chromosomes in the somatic cell in excess of the diploid number for the species (Comai 2005). The variation may be multiples of the complete set of chromosomes (euploidy) or incomplete set (aneuploidy). They may also be duplicates of the chromosomes of one species (autoploidy; e.g. banana – AAA) or different species (alloploidy; e.g. wheat – AABBDD) (Poehlman and Sleper 1995). The genetics of polyploids is hence very complex.

5.15.1 Breeding Autoploids and Alloploids

Important natural autotetraploids (AAAA) include alfalfa, peanut, potato, and coffee (Katepa-Mupondwa et al. 2002). The primary effect of polyploidy is gigas effect, which increase size of vegetative plant parts as well as the incidence of sterility. For example, tetraploid rye ($2n = 4x = 28$) has about 2 % more protein and superior baking qualities than diploid cultivars, but also has higher incidence of spike sterility, leading to lower yields than diploid rye. Autopolyploidy is suited to crops whose economic part is vegetative (Comai 2005).

Autotriploidy is associated with sterility that arises from meotic disorders due to odd chromosome numbers. It is desirable in species whose economic parts are vegetative (e.g. sugar beet, banana, and grasses). Seedlessness is desired in banana

(AAA), while triploid monogerm types (3x = 27) of sugar beet are important in the sugar industry (Borojevic 1990).

Triploids are developed by crossing diploids with tetraploids. In seedless watermelon breeding, breeders develop tetraploids ($2n = 4x = 44$) by using colchicine to double the diploid ($2n = 2x = 22$). Induced autoploidy is not common in plant breeding (ryegrass is one of a few success stories; rye (*Secale cereale*) is a synthetic autoploid) (Fehr 1987a).

Important natural alloploids include wheat, oats, tobacco, cotton, sugarcane and strawberry (Acquaah 2012). They are reproductively fertile because most alloploids have evolved certain genetic systems that ensure that pairing occurs between chromosomes of the same genome. Though not commonly done by breeders, alloploids may be induced by crossing two species with different genomes, followed by chromosome doubling of the hybrid (Poehlman and Slepper 1995).

5.15.2 Using Aneuploidy in Plant Breeding

Aneuploidy involves a gain or a loss of one or a few chromosomes that make up the ploidy of the species (Borojevic 1990). Just like polyploidy, aneuploidy has its own unique genetics. A successful application of aneuploidy in breeding entails replacing normal chromosomes with alien ones (chromosome substitution). This was achieved by first developing monosomic lines for the species (e.g. in wheat, cotton, tobacco). In wheat, the technique resulted in substitutions that conferred disease resistance to the crop (Fehr 1987b).

5.16 Mutation Breeding as a Breeding Scheme

Mutations may be induced by chemical agents (e.g. colchicine, EMS) or physical agents (e.g. gamma rays, x-rays), the latter especially gamma rays, being the most successfully used in breeding to date (Anonymous 1991). In conventional breeding, mutations remain random and rare events. Products of mutagenesis are commonly recessive in gene action (Maluszynski et al. 2000). Consequently, organisms can carry a number of them, even deleterious ones (*genetic load*), without adverse effects. However, upon selfing, recessive alleles become homozygous and are expressed, the deleterious effects causing a reduction in vigor (inbreeding depression). Gametic mutations (occur in gametic cells) are heritable; somatic mutations (occur in somatic cells) are not, and result in chimeras. Chimeras are desirable in breeding certain ornamental species such as African violet (*Saintpaulia ionantha*) (Broertjes and van Harten 1988).

When used in a breeding program, recessive mutations prolong cultivar development because additional steps are required at each cycle or generation for selfing for gene expression for selection to be made (see backcross breeding) (Micke 1992). Whereas genes may mutate, meiotic disorders associated with the spindle mechanism

may lead to abnormal distribution of chromosomes to daughter cells. These genomic mutations may involve duplication of the basic chromosome set (euploidy), or duplication or deletion of certain chromosomes in the set (aneuploidy). Various parts of the plant (e.g. seed, pollen, cuttings) may be used in mutagenesis (Broertjes and van Harten 1988). The mutagen may be administered in an acute dose (high dose over a short period) or chronic dose (continuous or prolonged exposure at low dose). The plant material used, breeding objective, screening method, and dose rate are some of the key factors to consider in mutation breeding (Micke 1992). Because mutations are random and rare events, the breeder must plant a large number of segregating population (M_2) is needed to give a high chance of finding the desired induced allele (Fig. 5.13). Besides, most mutations are deleterious. Morphological mutations are easier to spot than subtle mutations that are hidden and consequently, a large number of plants would not survive.

As a breeding method, mutation breeding is applicable to both self- and cross-pollinated species (Ahloowalia 2004). It is commonly used to correct a deficiency in an adapted cultivar. When using mutagenesis in breeding, it is important to note that other mutations besides that for which selection is conducted, may occur. Backcrossing the mutant to its original parent helps to reselect a purer version of the mutation of interest (Broertjes and van Harten 1988).

5.17 Performance Trials and Release

Breeders conduct performance or field trials of advanced generations of the materials developed in a breeding program (Acquaah 2012). Performance trials are often called *yield trials* because yield is usually the fundamental trait interest in a breeding program. The performance trials may be conducted in two stages – *preliminary*

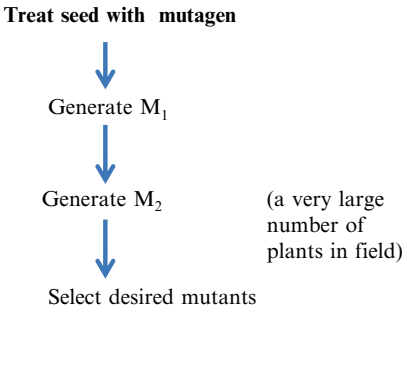

Fig. 5.13 General steps for mutation breeding of plants

yield trials (involves a few plots and replications), followed by a more detailed *advanced yield trial* (Fehr 1987a).

Performance trials are usually conducted at multiple locations (including especially those in which the cultivar is going to be used) and over several years. If the analysis shows a lack of significant G x E interactions for genotype x location or location x years, the breeder may be able to select a superior genotype for release as a cultivar for use throughout a specified production region (Briggs and Knowles 1967). On the other hand, if G x E interaction is significant and the environmental variations are unpredictable, the breeder should conduct a *stability analysis* to find out which genotypes have stable performance (i.e. perform consistently, whether at a high or low level, across a wide range of environments) (Falconer 1981).

Following performance trials, the breeder may identify a superior genotype for release as a new cultivar. The new cultivar may be registered (given a name, full description of its attributes, etc. for the sake of the scientific community). There are states or national guidelines for registration for various crops (Acquaah 2012). Certification is a legal process that is conducted by a certifying agency. The goal is to ensure that the seed (cultivar) reaches the user (producer) in its highest quality, original genetic identity, and highest purity. Certified seed multiplication may be contracted to professional seed growers (Allard 1960).

Plant breeders are among the most prolific inventors in the world, developing new cultivars to meet the needs of producers in ever changing biotic and abiotic production environments, coupled with changes in consumer preferences. Consequently, the matter of intellectual property is very important to plant breeders, the most common protection sought being the breeder's rights (Acquaah 2012). Laws protecting plant varieties vary among countries. The first international effort for plant variety protection was enacted in 1961 (Barton 1982). Called the International Union for the Protection of New Varieties of Plants (UPOV), this intergovernmental organization headquartered in Geneva, Switzerland, has guidelines for securing protection (Moore 2005).

The plant variety for which a breeder seeks protection must meet the following criteria: distinctness (clearly distinguishable from any other variety known to exist at the time of filing), uniformity (not a mixture of several varieties) stability (relevant characteristics remain unchanged after repeated propagation (the three criteria form the so-called DUS criteria); other criteria are, novelty (new and not sold in that country for a certain period), acceptable denomination (variety name), and rightful person (breeder's name) (Barton 1982; Moore 2005).

5.18 Germplasm Conservation, Utilization and Exchange

As previously stated, germplasm is the lifeblood of plant breeding without which breeding is impossible to conduct (Fig. 5.14). Plant breeders derive their breeding materials from five major types: advanced (elite) germplasm, improved germplasm, landraces, wild or weedy relatives, which can be generally categorized into

Fig. 5.14 Diversified fruit colors and shapes of modern tomato cultivars (Source: Courtesy of Yurling Bai, Wageningen UR Plant Breeding, The Netherlands)

three-domesticated plants, undomesticated plants, and plants from other species or genera (Acquaah 2012). Collection, evaluation, utilization, conservation and exchange of genetic resources assume considerable significance, especially in view of the rapid degradation and exploitation of the available biodiversity all over the world (Mehra and Arora 1982; Mengesha 1984).

Conservation of biodiversity is urgent because of the problem of *genetic erosion* (decline in genetic variation in cultivated or natural populations) due to natural causes (e.g. wild fires, prolonged droughts) but largely human actions (indiscriminate land clearing, settlement of new lands, action of breeders in types of cultivars developed – narrow genetic base) (Acquaah 2012).

Biodiversity is deliberately sought and collected by researchers who undertake planned trips to centers of plant diversity. Biodiversity can be maintained away from their natural habitats and ecosystems (ex situ) such as gene banks, or within them (in situ conservation) (Maxted 2013). Germplasm banks are national or international institutions for preservation and sustainable use of biodiversity (AMCOST 2007).

Conservation of biodiversity efforts facilitate availability and access to materials for breeding. However, there is a significant gap between availability and utilization of germplasm, due in part to the lack of information on a large number of accessions in gene banks (Upadhyaya and Laxmipathi Gowda 2009). When available, the ease of utilization of germplasm depends on the gene pool to which it belongs. As discussed in Sect. 5.10, gene transfer (crossing) between domesticated plants (GP1 or primary gene pool) (commercial cultivars, remnants of breeding materials, landraces) are usually without fertility problems (Harlan and de Wet 1971). However, when wild germplasm and more distant plants are used in crosses, pre-breeding is needed for such introgression.

The germplasm is considered as a pan of human biological heritage without whose free exchange and availability, present day farm productivity would not have been possible (Jain 1982). Related to germplasm utilization is the subject of

germplasm exchange that also has implications in intellectual property. The breeder is protected from unauthorized use of the cultivar developed from a breeding program, but they must ensure that they have the right to use germplasm in their breeding program in the first place. But nothing can be achieved, even by genetic engineering, unless those desirable genes are at our disposal (Mengesha 1984).

In 1983, the International Treaty on Plant Genetic Resources for Food and Agriculture (PGRFA) was signed by a number of countries (Moore 2005). All international germplasm exchanges (except the return of germplasm to the country of its origin) are governed by a Material Transfer Agreement (MTA) appropriate to the material (Moore 2005). International Agricultural Research Centers (IARC) are strategically located in regions of rich crop diversity to facilitate the collection, conservation, and exchange of germplasm for use by breeders (Mengesha 1984). These centers have protocols that govern their own germplasm exchange programs.

5.19 Conclusion and Prospects

Conventional breeding is an age-old approach to plant improvement that is increasingly becoming sophisticated as it becomes infused with science and technology. It uses variability that already exists primarily within the species and occasionally within close relatives, provided they are not reproductively isolated. Modern plant breeding is a methodical and systematic approach to plant improvement. Science and technology has made it more efficient. However, its chief limitation is the fact that it depends on variability that can be exchanged within natural boundaries. Advances in knowledge and DNA technology currently allow genes to be exchanged across natural biological boundaries. Essentially, there is one universal gene pool from which breeders may obtain variability for crop improvement. Called molecular plant breeding, plant breeders may now access genes from the animal kingdom for plant improvement, but not without controversy. Whereas genetically modified (GM) technology remains controversial, conventional plant breeding can benefit from the use of molecular technologies for various tasks to facilitate their programs without incorporating alien genes into the final product (the cultivar that is bred). Advances in molecular breeding without genetic modification continue to be made.

References

Acquaah G (1992) Practical protein electrophoresis for genetic research. Dioscoredes Press, Portland

Acquaah G (2004a) Horticulture: principles and practices, 3rd edn. Prentice Hall, Upper Saddle River

Acquaah G (2004b) Understanding biotechnology: an integrated and cyber-based approach. Prentice Hall/Pearson, Upper Saddle River

Acquaah G (2012) Principles of plant genetics and breeding, 2nd edn. Wiley-Blackwell, Oxford
Agrawal RL (1998) Fundamentals of plant breeding and hybrid seed production. Science Publishers, Inc, Enfield
Ahloowalia BS (2004) Global impact of mutation-derived varieties. Euphytica 135:187–204
Allard RW (1960) Principles of plant breeding. John Wiley and Sons, New York
AMCOST (2007) Conservation and sustainable use of biodiversity. http://nepadst.org/platform/biodiv.shtml
Andrus C, Seshadri V, Grimball P (1971) Production of seedless watermelons. US Agricultural Research Service, USDA, Washington, DC
Anonymous (1991) Plant mutation breeding for crop improvement. In: Proceedings of the FAO/IAEA symposium, Vienna. IAEA, Vienna
Bado S, Kozak K, Sekander H et al (2013) Resurgence of X-rays in mutation breeding. In: Plant genetics and breeding technologies; plant diseases and resistance mechanisms: Proceedings, 18–20 Feb 2013, Vienna, Austria. Medimond-Monduzzi Editore International Proceedings Division, Pianoro, 2013, pp 13–16
Baezinger PS, Russel WK, Graef GL, Campbell BT (2006) 50 years of crop breeding, genetics, and cytology. Crop Sci 46:2230–2244
Barton JH (1982) The international breeder's rights system and crop plant innovation. Science 216:1071–1075
Bauman F, Crane PL (1992) Hybrid corn – history, development and selection considerations, National corn handbook. Purdue University, West Lafayette
Bernardo R (2010) Breeding for quantitative traits in plants, 2nd edn. Stemma Press, Woodbury
Betrán FJ, Hallauer AR (1996) Hybrid improvement after reciprocal recurrent selection in BSSS and BSCB1 maize populations. Maydica 41:25–33
Betrand C, Collard Y, Mackill DJ (2008) Marker-assisted selection: an approach for precision breeding in the 21st century. Philos Trans Roy Soc Lond Ser B Biol Sci 363:557–572
Borojevic S (1990) Principles and methods of plant breeding. Elsevier, Amsterdam
Bridgen MP (1994) A review of plant embryo culture. Hort Sci 29:1243–1246
Briggs FN, Knowles PF (1967) Introduction to plant breeding. Reinhold Publishing Corporation, New York
Broertjes C, van Harten AM (1988) Applied mutation breeding for vegetatively propagated crops. Elsevier, Amsterdam
Brown J, Caligari P (2008) An introduction to plant breeding. Blackwell Publishing, Ltd, Oxford/Ames
Burton JW, Brim CA (1981) Recurrent selection in soybeans III. Selection for increased percent oil in seeds. Crop Sci 21:31–34
Chahal GS, Gosal SS (2000) Principles and procedures of plant breeding: biotechnological and conventional approaches. CRC Press, New York
Chahal GS, Gosal SS (2002) Principles and procedures of plant breeding. Alpha Science International
Cisneros A, Tel-zur N (2010) Embryo rescue and plant regeneration following interspecific crosses in the genus *Hylocereus* (Cactaceae). Euphytica 174:73–82
Collard BCY, Jahurfer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci, mapping, and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169–196
Comai L (2005) The advantages and disadvantages of being polyploid. Nat Rev Genet 6:836–846
Comstock RE, Robinson HF, Harvey PH (1949) A breeding procedure designed to make maximum use of both general and specific combining ability. Agron J 41:360–367
Coors IG, Pandey S (eds) (1997) In: Proceedings of the international symposium on the exploitation of heterosis in crops. CIMMIT, Mexico City, 17–22 Aug 1997; ASA, Madison
Crow JF (1998) 90 years ago: the beginning of hybrid maize. Genetics 148:923–928

Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper and Row, New York
Czyczyło-Mysza I, Marcińska I, Jankowicz-Cieślak J, Dubert F (2013) The effect of ionizing radiation on vernalization, growth and development of winter wheat. Acta Biol Cracov Ser Bot 55(1):1–6
Dudley JW, Lambert RJ (1992) Ninety generations of selection for oil and protein in maize. Maydica 37:81–87
Dudley JW, Saghai-Maroof MA, Rufener GK (1991) Molecular markers and grouping of parents in maize breeding programs. Crop Sci 31:718–723
Eberhart SA, Gardner CO (1966) A general model for genetic effects. Biometrics 22(4):864–881
Falconer DS (1981) Introduction to quantitative genetics. Longman Group, Ltd, New York
Falconer DS, Mackay TFC (1996) Introduction to quantitative genetic, 4th edn. Longman, Harlow
Fehr WR (1987a) Principles of cultivar development, vol 1, Theory and technique. Macmillan, New York
Fehr WR (1987b) Principles of cultivar development, vol 2, Crops species. Macmillan, New York
Feng L, Burton JW, Carter TE Jr, Pantalone VR (2004) Recurrent half-sib selection with testcross evaluation for increased oil content in soybean. Crop Sci 44:63–69
Geiger HH, Gordillo GA (2009) Doubled haploids in hybrid maize breeding. Maydica 54:485–499
Gepts P (2002) A comparison between crop domestication, classical plant breeding, and genetic engineering. Crop Sci 42:1780–1790
Griffing JB (1956) Concept of general and specific combining ability in relation to diallel systems. Aust J Biol Sci 9:463–493
Griffiths JF (1999) An introduction to genetic analysis. WH Freeman Ltd, San Francisco
Gur A, Zamir D (2004) Unused natural variation can lift yield barriers in plant breeding. PLoS Biol 2:e245
Hallauer AJ (1967) Development of single-cross hybrids from two-eared maize populations. Crop Sci 7:192–195
Hanna WW, Bashaw EC (1987) Apomixis: its identification and use in plant breeding. Crop Sci 27:1136–1139
Harlan JR (1975) Our vanishing genetic resources. Science 188:618–621
Harlan JR (1976) Genetic resources in wild relatives of crops. Crop Sci 16:329–333
Harlan JR, de Wet JMT (1971) Toward a rational classification of cultivated plants. Taxon 20:509–517
Helgeson JP, Hunt GJ, Haberlach GT, Austin S (1986) Somatic hybrids between *Solanum brevidens* and *Solanum tuberosum*: expression of a late blight resistance gene and potato leaf roll resistance. Plant Cell Rep 5(3):212–214
Holland JB (2001) Epistasis and plant breeding. Plant Breed Rev 21:27–92
Jain HK (1982) Plant breeders' rights and genetic resources. Indian J Plant Breed 42:121–128
Jain HK, Kharkwal MC (2004) Plant breeding: Mendelian to molecular approaches. Springer, Dordrecht
Jain SH, Till BJ, Suprasanna P, Roux N (2011) Mutations and cultivar development in banana. In: Banana breeding: progress and challenges. CRC Press, pp 203–218
Jakowitsch J, Mette MF, van der Winden J et al (1999) Integrated pararetroviral sequences define a unique class of dispersed repetitive DNA in plants. Proc Natl Acad Sci 96(23):13241–13246
Jensen NF (1978) Composite breeding methods and the DSM system in cereals. Crop Sci 18:622–626
Kasha KJ, Kao KN (1970) High frequency haploid production in barley (*Hordeum vulgare* L.). Nature 225:874–876
Katepa-Mupondwa FM, Christie BR, Michaels TE (2002) An improved breeding strategy for autotetraploid alfalfa (*Medicago sativa* L.). Euphytica 123:139–146

Kearsey MJ, Pooni HS (1998) The genetical analysis of quantitative traits. Stanley Thornes Publishers, Cheltenham
Kempe K, Gils M (2011) Pollination control technologies for hybrid breeding. Molec Breed 27:417–437
Lamkey K, Edwards J (1999) Quantitative genetics of heterosis. The genetics and exploitation of heterosis in crops. pp 31–48
Liu W, Zheng MY, Polle EA, Konzak CF (2002) Highly efficient doubled-haploid production in wheat (*Triticum aestivum* L.) via induced microspore embryogenesis. Crop Sci 42:686–692
Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Inc, Sunderland
Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nat Rev Genet 10:565–577
Maluszynski MK, Nichterlein K, van Zanten L, Ahloowalia BS (2000) Officially released mutant varieties – the FAO/IAEA database. Mutat Breed Rev 12:1–84
Maluszynski M, Kasha KJ, Forster BP, Szarejko I (2003) Doubled haploid production in crop plants: a manual. Kluwer Academic Publ, Dordrecht/Boston/London
Matijevic M, Bado S, Lagoda PJL, Forster BP (2013) Impact of induced mutations in plant breeding. In: Plant genetics and breeding technologies; plant diseases and resistance mechanisms: Proceedings, 18–20 Feb 2013, Vienna, Austria. Medimond-Monduzzi Editore international Proceedings Division, Pianoro, 2013, pp 45–47
Maxted N (2013) In situ and ex situ conservation. Elsevier, Amsterdam
Mehlo L, Mbambo Z, Bado S et al (2013) Induced protein polymorphisms and nutritional quality of gamma irradiation mutants of sorghum. Mutat Res 749:66–72
Mehra KL, Arora RK (1982) Plant genetic resources of India, their diversity and conservation, NBPGR scientific monograph 4:60. National Bureau of Plant Genetic Resources, New Delhi
Melchinger AE, Gumber RK (1998) Overview of heterosis and heterotic groups in agronomic crops. In: Lamkey KR, Staub JE (eds) Concept and breeding of heterosis in crop plants, SP. Pub. No. 25. CCSSA, Madison, pp 29–44
Mengesha MH (1984) International germplasm collection, conservation, and exchange at ICRISAT. In: Conservation of crop germplasm-international perspective. Crop Science Society of America, Madison, pp 47–54
Menz MA, Hallauer AR (1997) Reciprocal recurrent selection of two tropical corn populations adapted to Iowa. Maydica 42(3):239–246
Micke A (1992) 50 years induced mutations for improving disease resistance of crop plants. Mutat Breed Newsl 39:2–4
Moore G, Tymowski W (2005) Explanatory guide to the international treaty on plant genetic resources for food and agriculture. IUCN, Gland/Cambridge
Nassimi AW, Raziuddin S, Ali G et al (2006) Combining ability analysis for maturity and other traits in rapeseed (*Brassica napus* L.). Agron J 5(3):523–526
Norskog C (1995) Hybrid seed corn enterprises. A brief history. Curtis Norskog, Willmar
Paterniani E, Vencovsky R (1977) Reciprocal recurrent selection in maize (*Zea mays* L.) based on testcrosses of half-sib families. Maydica 22:141–152
Paterniani E, Vencovsky R (1978) Reciprocal recurrent selection based on half-sib progenies and prolific plants in maize (*Zea mays* L.). Maydica 23:209–219
Pillen K, Zacharias A, Leon J (2003) Advanced backcross QTL analysis in barley (*Hordeum vulgare* L). Theor Appl Genet 107:340–352
Poehlman JM, Sleper DA (1995) Breeding field crops. Iowa State Univ Press, Iowa
Sage TL, Strumas F, Cole WW, Barret S (2010) Embryo rescue and plant regeneration following interspecific crosses in the genus *Hylocereus* (Cactaceae). Euphytica 174:73–82
Savidan YH (2000) Apomixis: genetics and breeding. Plant Breed Rev 18:13–86
Shu QY, Forster BP, Nakagawa H (2012) Plant mutation breeding and biotechnology. CABI International, Wallingford/Cambridge
Sleper DA, Poehlman JM (1999) Breeding field crops. Wiley and Sons, New Jersey

Springer NM, Stupar RM (2007) Allelic variation and heterosis in maize; How do two halves make more than a whole? Genome Res 17(3):264–275

Stuber CW, Lincoln SE, Wolff DW et al (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genet 132:823–839

Tigchelaat EC, Casali VWD (1976) Single seed descent: applications and merits in breeding self-pollinated crops. Acta Hort 63:85–90

Upadhyaya HD, Laxmipathi Gowda CL (2009) Managing and enhancing the use of germplasm – strategies and methodologies, vol 10, Technical manual. International Crops Research Institute for the Semi-Arid Tropics, Patancheru, p 236

Winzeler H, Schmid J, Fried PM (1987) Field performance of androgenetic doubled haploid spring wheat line in comparison with line selected by the pedigree system. Plant Breed 99:41–48

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nat Rev Genet 2:983–989

Zohary D, Hopf M (1988) Domestication of plants in the old world. Clarendon, Oxford

Chapter 6
Applying Epigenetics in Plant Breeding: Balancing Genome Stability and Phenotypic Plasticity

Estelle Jaligot and Alain Rival

Abstract The correct implementation of epigenetic mechanisms is often a prerequisite for the timely regulation of genome expression and structure and ultimately for the development of higher plants. Developmental regulation is thus playing a paramount role in the elaboration of yields in agricultural crops. However, numerous studies have shown that this tight control includes a certain degree of freedom as epigenetic regulations can be loosened in the course of the reproductive development, after hybridization or as part of the response to environmental constraints – both in vitro and in vivo – whereas genome stability is globally maintained. As a result, several modified epigenetic marks and associated altered gene or transposable element expression can eventually give rise to qualitative or quantitative phenotypic changes on the long term. The present chapter is intended to present the main concepts governing epigenetic regulation of gene expression in higher plants and to review its potential applications for the selection of heritable phenotypes. It illustrates how epigenetic variations can be smartly used in breeding schemes and which questions remain to be addressed in order to make such integration successful. The Next Generation Sequencing revolution has also impacted our approach of plant epigenetics as more genomes, epigenomes and transcriptomes and are made available for crop plants and the simplistic Arabidopsis model is being questioned.

Keywords Epigenetics • Chromatin • DNA methylation • Paramutation • Transposable elements

E. Jaligot • A. Rival (✉)
CIRAD – UMR DIADE, Montpellier 34398 Cedex 5, France
e-mail: estelle.jaligot@cirad.fr; alain.rival@cirad.fr

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_6

6.1 Introduction

Plant breeding relies on the selection of individuals with desired traits from populations bearing phenotypically variable traits. The amount of trait variation which can be caused by heritable differences in chromatin states is far lower than variation due to changes in the primary sequence of DNA. In addition, epigenetic states are often unstable, and selection on only a small number of epigenetic states could lead to consistent plant improvement (House and Lukens 2014).

In plants, epigenetic variation contributes to phenotypic differences in developmental traits (Turck and Coupland 2014). The originality of plant developmental programs relies on the absence of dedicated germ line set up, segregated and maintained from early development onwards. Instead, cells giving rise to the germ line are generated de novo from somatic tissues (Feng et al. 2010a, b). There is, therefore, a theoretical window of opportunity for any epigenetic information acquired by any somatic tissue to be transmitted to the next generation, provided two conditions are fulfilled.

First is that this information escapes the erasing events which take place in between generations. Indeed, although the removal of certain epigenetic marks has been described in plants, the global reprogramming of the epigenome appears to be less thorough than that of other living organisms (Baroux et al. 2011). Second is that active maintenance mechanisms allow for a particular piece of epigenetic information to be faithfully transmitted throughout gametogenesis, then fertilization and early embryogenesis (Heard and Martienssen 2014).

Such requirements actually differentiate heritable from non-heritable epigenetic states and the mechanisms governing the resetting and the maintenance of epigenetic marks work in close association and share many effectors. The heritability of epigenetic states (or transgenerational epigenetic inheritance) has been studied in several recent articles and some of them have tackled the question of its potential implications for plant breeding (Paszkowski and Grossniklaus 2011).

The major concepts governing epigenetic regulation of gene expression and its potential applications for the selection of heritable phenotypes rely on the in-depth understanding of differentiation and development in plants. In such phenomena the expression of specific set of genes and the concomitant decrease of the organogenic potential is mirrored by dynamic changes in DNA methylation patterns and chromatin conformation (Pikaard and Mittelsten-Scheid 2014).

Environmental constraints that naturally occur during the plant's life cycle induce epigenetic alterations and vernalization provides a well documented example of such epigenetic footprints. Mechanisms underlying the erasing of epigenetic marks during meiosis clearly impacts heritability of traits and epigenetic stability. Thus the application of epigenetics in breeding relies on a clearer understanding of intricate mechanisms governing the transmission of phenotypes through generations.

Somaclonal variation is hampering the development of large-scale micropropagation process and the development of genetic engineering in several plant species

(Us-Camas et al. 2014). At the same time, this epigenetic phenomenon can be considered as an interesting source of variability which can be exploited by breeders (Schlichting and Wund 2014). New options for breeding strategies are offered by the stable transmission of epialleles; indeed, Transgenerational Epigenetic Inheritance is now changing our way of considering heredity as it opens a new field of research (Heard and Martienssen 2014). Interestingly, several basic concepts in genetics such as hybrid vigor are now revisited in the light of epigenetics, as the epigenetic regulations and gene expression patterns that are implemented by the hybrid genome are not additive with respect to those of each parent.

Plant epigenetics is still a young field of research and there is still a lot of work required in order to integrate epigenetics concepts into plant breeding. As more plant epigenomes are now sequenced and made available to researchers, it is likely that rules governing the relatively simple epigenome of Arabidopsis will not apply to larger polyploid plant genomes which are made of over 80 % of Transposable Elements (TEs).

The present chapter aims at providing updated information on the latest developments in basic research on plant epigenetics and their present and future applications for plant breeding. It covers several aspects of epigenetics which are closely intricate and presents an updated review of recent breakthrough in the fields of plant cell differentiation and development, environment and genetic footprints, somaclonal variation and epigenetic inheritance.

6.2 Differentiation and Development: Harnessing and Unleashing Gene Expression

In both animals and plants the newly formed zygote is totipotent, which means that it is theoretically capable of modulating the expression of any part of its genome and to generate a complete organism. The zygote genome has a very low level of DNA methylation and is wrapped in chromatin in an open, accessible configuration (euchromatin) (Messerschmidt et al. 2014). After a few symmetrical non-orientated cell divisions the future body pattern of the embryo emerges from a succession of asymmetrical cell divisions. In plants the orientation of the division plan is under tight genetic control and is dependent on the proper functioning of both gene- and TE-methylation mechanisms (Xiao et al. 2006) and auxin response pathway (Yoshida et al. 2014). The pluripotent cells that form the different parts of the embryo display more limited properties with respect to the initial totipotent cells, and can only generate cell lineages corresponding to a determined range of organs. As embryo development progresses, so does cell differentiation within each tissue and the increasing specialization/commitment of each cell type is mirrored by the expression of specific set of genes and the concomitant decrease of the organogenic potential (De Vega-Bartol et al. 2013; Hemberger et al. 2009). Concomitantly, DNA methylation patterns and chromatin structure are dynamically redistributed

throughout the genome under the control of DNA- and histone-modifying enzymes. The inactivated parts of the genome undergo a global increase of both their DNA methylation rate and chromatin compaction, leading to the formation of: (a) inducible/facultative heterochromatin (mostly inactivated genes) and (b) constitutive heterochromatin (corresponding to repetitive sequences and TEs) domains (Cantone and Fisher 2013; Exner and Hennig 2008; Li et al. 2011). Polycomb Repressive Complexes (PRCs) are chromatin modifiers that are crucial in both animals and plants to ensure the mitotically stable repression of key developmental genes until their expression is required and their switching from an inducible heterochromatic state to a euchromatic state in a timely and tissue-appropriate manner. The balance between their action and that of activating trithorax-group proteins determine cell identity by orientating it towards either the maintenance of a proliferation-competent state (dominated by trxG activity, which promotes the deposition of the euchromatic mark H3K4me3) or the orientation towards a differentiated state, which depends on the coordinated inactivation of specific gene sets by PcG activity (Cantone and Fisher 2013; Köhler and Hennig 2010; Springer 2013). In other terms, cell identity is determined through the gradual repression of gene expression whereas the maintenance of an undifferentiated, stem-like state relies on the exclusion of targets genes from PRC-mediated inactivation (Sun et al. 2014).

In animals, small groups of stem cells remain poorly differentiated and therefore retain the capacity to regenerate the tissue they contribute to although this ability decreases during development and is eventually lost. By contrast, plant meristems present the hallmarks of the stem cell status, such as the prevalence of euchromatin in the nucleus (Lafon-Placette et al. 2013) and retain the ability to produce new organs throughout the life cycle of the plant. It has recently been shown that both embryo development and the maintenance of meristematic stem cell populations are strongly dependent on chromatin-remodeling ATPases (Sang et al. 2012). Also, the germinal lineage separates from the other cell lineages very early in animal development, whereas plant germ cells are formed late in development within fully differentiated floral organs (Bourc'his and Voinnet 2010; Feng et al. 2010b; Jacob and Martienssen 2011).

In plants, most of the knowledge available on PRCs comes from studies focusing on the PRC2, which is responsible for the deposition of H3K27me3, whereas evidence regarding the occurrence in plants of the other two PRC complexes known in animals, PRC1 (catalyzing mono-ubiquitination of H2K119) and PhoRC, is still scarce. Genes encoding putative members of these complexes have been isolated and they govern similar functions as in Drosophila. Nevertheless complexes themselves have not been detected and it could be that the plant complexes are structurally distinct from the animal ones (Wang et al. 2014). Compared to Drosophila and mammals, plants were found to have a higher number of PRC2 complexes (based on the combinatorial diversity offered by the multigenic families generating the different components of the complex) with functions in distinct developmental phase changes. Such functions include notably the repression of the seed development program in the absence of fertilization (FIS2 complex) and the transition between vegetative and reproductive phase through the impairment of precocious flowering

and the integration of developmental and environmental cues (EMF2 and VRN2 complexes) (Bemer and Grossniklaus 2012; Köhler and Hennig 2010; Tang et al. 2012). It is therefore clear that the notion of cell fate is much more straightforward and much more difficult to circumvent in animals (when the notions of commitment and that of cell lineage are prevalent) than it is in plants, in which the relative positions of the cells seem to be essential to their future specification and for which the nature of the co-factors recruited with the PRC2 can promote different specifications and, ultimately, different cell fates (Baroux et al. 2011; Lee et al. 2014).

Once the cell is fully differentiated and specialized, the life cycle of the plant unfolds with a continuous integration of both external (climatic, pathogenic) and internal (developmental, hormonal, nutritional) signals in its regulation (Bender 2004; Boyko and Kovalchuk 2011; Exner and Hennig 2008; Finnegan 2002). Seed germination is paralleled by light-dependent chromatin relaxation and decrease of genomic DNA methylation (Cho et al. 2012), and photomorphogenesis in general is associated with large-scale histone modifications (Charron et al. 2009). In another example, Yaish et al. (2011) have underlined the interplay between floral induction and stress response pathways. The interactions occurring between epigenetic mechanisms based on either DNA or histone modifications are represented in Fig. 6.1.

For the most part, epigenetic patterns are faithfully transmitted among somatic cells belonging to the same tissue through mitosis (with the exception of the occasional stress response, see below). This transmission is ensured through the coupling of mechanisms aimed at maintaining methylation or histone modification patterns with DNA replication processes. The so-called maintenance methylation of DNA, a function commanded by the enzymes MET1 (CG sites in both genes and repetitive sequences) and CMT3 (CHG in repetitive sequences) in plants, relies on the symmetry of their respective target sites so that methylation patterns of the mother strand can be duplicated on the daughter strand (Cao and Jacobsen 2002; Cao et al. 2003; Finnegan and Kovac 2000; Finnegan et al. 1996; Henikoff and Comai 1998; Kato et al. 2003; Law and Jacobsen 2010). In the case of repetitive sequences, CHG methylation is consecutive to histone methylation (H3K9me2) by KRYPTONITE (Jackson et al. 2002).

Meanwhile, the histone proteins forming the nucleosome are temporarily dissociated from the DNA molecule during the passage of the replication fork. Then the reassembly of chromatin around the daughter strands involves both the incorporation of recycled parental histones (thereby leaving room for some mitotic transmission of histone marks) and the deposition of newly synthesized histones. Histone-modifying enzymes are then recruited to homogenize histone marks between the parental histones and the new one (Rivera et al. 2014).

It has been recently demonstrated in Arabidopsis (Blevins et al. 2014) that this mitotic inheritance of epigenetic repressive marks was promoted through the sequential *tagging* of the silenced loci during DNA replication through a combination of HDA6-mediated histone deacetylation and MET1-dependent maintenance of CG methylation, followed by a second step involving the recruitment of the RNA-polymerase variants responsible for the production of 24-nucleotide siRNAs (RNA PolIV and PolV) (Herr et al. 2005; Kanno et al. 2005; Onodera et al. 2005).

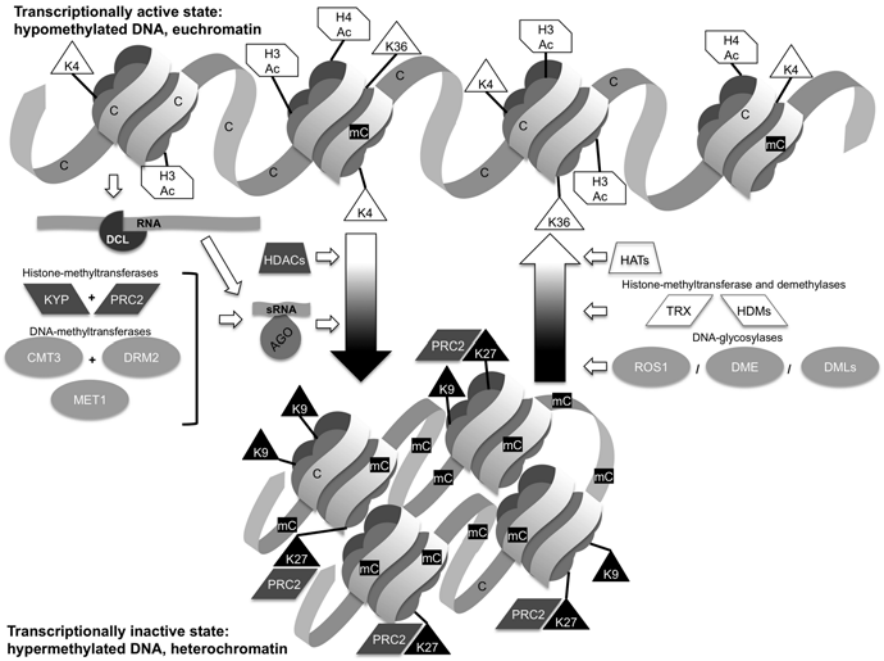

Fig. 6.1 Epigenetic mechanisms regulating gene expression. The transition from transcriptional activity to transcriptional repression is controlled through variable combinations of epigenetic mechanisms involving DNA methylation changes, histone modifications and/or small RNA-mediated targeting (note that, for the sake of clarity, only some of the main actors from different pathways have been displayed). *Light grey ribbon*: DNA; *dark grey circles*: histones. C: cytosine; mC: 5-methylcytosine. H3Ac, H4Ac: acetylated H3 histone and acetylated H4 histone, respectively; K4, K36, K9, K27: methylated lysines 4, 36, 9 and 27 on histone 3, respectively. HDACs: histone deacetylases; KYP: KRYPTONITE; PRC2: Polycomb Repressive Complex 2. sRNA: small non-coding RNA; DCL: Dicer-like enzyme; AGO: Argonaute protein. MET1/CMT3/DRM2: DNA-methyltransferases belonging to the DNMT1/MET1 class, the CHROMOMETHYLASE class or the DOMAIN-REARRANGED METHYLTRANSFERASE class, respectively. HATs: histone acetyltransferases; TRX: Trithorax Complex; HDMs: histone demethylases; DME/DMLs: DEMETER/DEMETER-like glycosylases (Source: Miguel and Marum 2011, with modifications)

These siRNAs then trigger the subsequent induction of further methylation of these loci through RdDM, a pathway involving the de novo DNA methyltransferase DRM2, responsible for the methylation of CHH sites, i.e. cytosines in asymmetrical sequence contexts (Cao et al. 2000).

Ultimately, in both animals and plants the layering of different epigenetic marks on DNA and chromatin determines a small number of states corresponding either to transcriptional activity, conditional inducibility or constitutive repression, with specific combinations of marks which depend both on the nature of the target sequences (genic, intergenic or TE) and on the tissue (Roudier et al. 2011; Rye et al. 2014).

6.3 Stress Leaves Epigenetic Footprints

6.3.1 Chromatin Condensation at the Crossroad

The environmental constraints that naturally occur during the plant's life cycle also induce epigenetic alterations. Some, if not most, of these changes are non-specific and are thought to result from the transient, random misregulation of epigenetic processes upon stress application; for instance, through the temporary impairment of DNA methylation maintenance mechanisms during mitosis (Finnegan 2002; Springer 2013). Another part of these changes are targeted and believed to participate to stress response and short-term acclimation through their effect on gene expression (for instance, the activation of genes involved in stress tolerance) and, ultimately, phenotypic plasticity (Angers et al. 2010; Bossdorf et al. 2008; Boyko and Kovalchuk 2011; Lukens and Zhan 2007; Nicotra et al. 2010; Wada et al. 2004). Moreover, Grafi et al. (2011) have suggested that proper stress response necessitates the temporary return of the cell to a pluripotent-like state via a partial de-differentiation (loss of DNA methylation, chromatin decondensation, up- or down-regulation of TFs…). In support to this hypothesis, Van Zanten et al. (2012) have noted that both biotic and abiotic stresses resulted in genome-wide decrease of chromatin condensation in both euchromatic and heterochromatic domains, similar to what is observed during phase changes as if this transient structural relaxation was a prerequisite before large-scale changes in gene expression could take place. The impact of environmental changes on epigenetic mechanisms is well documented although not all of these changes have phenotypic consequences (Fig. 6.2).

6.3.2 Common Memories of Multiple Environmental Changes

The most thoroughly studied example is, by far, the vernalization process in temperate plants. Indeed, the *FLC* floral repressor of *Arabidopsis thaliana* is gradually repressed as a result of the prolonged application of winter-like conditions (short days, low temperatures) and flowering is induced upon the return to warmer, spring-like conditions. This repression is correlated with the gradual transition of *FLC* from a transcriptionally active to a stably silenced state through successive steps of histone deacetylation then PRC2-mediated histone methylation (H3K27me3) (Bastow et al. 2004; Sheldon et al. 1999, 2002; Sung and Amasino 2004). Also, Choi and Sano (2007) showed that both loss of CG methylation and transcriptional upregulation of a glycerophosphodiesterase gene were induced in response to different abiotic stresses. Epigenetic mechanisms underlying drought tolerance have also been intensively studied. Indeed, Chen et al. (2012) have shown that the adaptation of root growth to drought stress depends on miRNA-mediated cleavage of the transcripts for two auxin receptors. Studies from both Kim et al. (2012) and Zhao et al. (2013) have established that promoters of genes activated after osmotic or

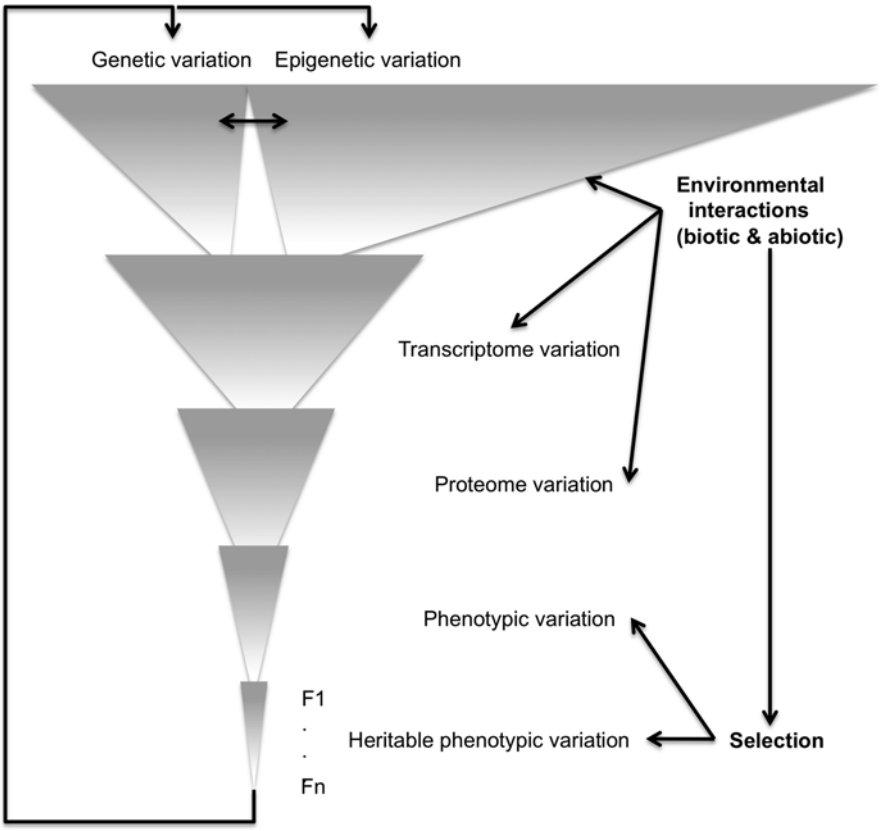

Fig. 6.2 Schematic representation of the variation *funnel*. Most qualitative or quantitative genetic and epigenetic variations do not result in variations in gene expression visible at the transcriptome level, either because they are reverted through editing/repair mechanisms or because they do not affect functionally relevant sequences. Similarly, not all variations at the transcriptome level have an impact on protein production because of the occurrence of post-transcriptional regulation mechanisms aimed at modifying or degrading transcripts (a similar reasoning is applicable to variations at the peptide level vs. post-translational modifications and turn-over). Environmental conditions can affect all upper levels (epigenetic, transcriptomic and peptidic variations). Only a small fraction of the initial genetic/epigenetic variation pool results in phenotypic alterations, an even smaller fraction of which is susceptible to be stably transmitted to successive generations (under the influence of selection). In turn, these heritable variations can then trigger further genetic and epigenetic variations (Source: Bossdorf et al. (2008) and Delker and Quint (2011), with modifications)

drought stress underwent changes in histone marks in conjunction with increased transcriptional activity.

In many other studies, however, only correlations between stress and changes in epigenetic patterns have been established, therefore the occurrence of a causal relationship with specific consequences on either gene expression or phenotype remains to be demonstrated. Both quantitative and qualitative changes in light trigger either the activation or the ubiquitin-dependent degradation of transcription factors

involved in light signaling. Both high and low light intensities promote genome-wide decondensation of chromatin, and histone marks are widely re-dispatched upon de-etiolation (Charron et al. 2009; Van Zanten et al. 2012). A mechanical stress mimicking herbivory attack on the leaves of yellow monkey flower results in both enhanced expression of a MYB transcription factor and increased trichome density on the leaves (Scoville et al. 2011). In soybean, Kulcheski et al. (2011) have identified miRNAs displaying differential accumulation in either water or pathogen stress conditions, with distinct patterns for sensitive and tolerant phenotypes.

A similar correlation between miRNA expression and abiotic stress response was found in rice inflorescences (Barrera-Figueroa et al. 2012). Similarly, in the promoter sequences of known stress-responsive genes of the rubber tree, Uthup et al. (2011) have identified genotype- and planting location-specific methylation polymorphisms. The main genes involved in the PTGS pathway (siRNA-dependent gene silencing) in tomato are overexpressed as a result of viral infection or abiotic stresses (Bai et al. 2012). Mastan et al. (2012) have detected both changes in methylation rates and methylation patterns arising in the genome of *Jatropha curcas* after salt stress.

Through mutant-based studies it has been possible to unveil functional connection between epigenetic regulatory pathways and stress responses. Two histone deacetylase genes have been shown to be involved in ABA and salt stress response (Chen et al. 2010; Luo et al. 2012). Also, it has been demonstrated that components of the RdDM pathway of transcriptional silencing are essential in the regulation of stress-responsive genes (Khan et al. 2014; Popova et al. 2013).

6.3.3 Transposable Elements: From Evolution to Adaptation

Transposable Elements (TEs) are enriched in both methylated cytosines and repressive chromatin marks in their default repressed state. In addition to their direct impact on genes, the temporary alleviation of repressive epigenetic mechanisms by stress also allows the transcription and, sometimes, remobilization of TEs (Grandbastien 1998). Exposure to high temperatures or UV light (Lang-Mladek et al. 2010; Molinier et al. 2006; Pecinka et al. 2010; Tittel-Elmer et al. 2010) result in TE reactivation through histone hyper-acetylation without loss of repressive marks. Not all of such reactivation mechanisms are detrimental to the organism: indeed, TEs for which *cis*-element methylation and/or small RNA production is altered by stress seem to facilitate the occurrence of epigenetic changes in neighboring stress-responsive genes, thereby contributing to the acquisition of stress tolerance through increased phenotypic plasticity. Such a mechanism has been proposed for the response of *Arabidopsis thaliana* to biotic stress and it has been suggested that siRNAs produced by the de-repressed TEs could contribute to the acquisition of Systemic Acquired Resistance (SAR) (Dowen et al. 2012; Yu et al. 2013). Alternatively, the insertion polymorphisms generated through TE remobilization could allow a quick modulation of neighboring genes expression, as shown in salt

and light stress conditions in wheat by Woodrow et al. (2010). On a more global level, it has been proposed that the epigenetic pathways regulating gene and TE expression, respectively, communicate through their respective small RNA intermediates: in a study from McCue et al. (2012), the transcriptional and translational activity of stress-responsive gene is functionally connected with the epigenetic status of an *Athila* retroelement.

6.3.4 *Environment and Development Govern Alternative Splicing*

As part of the phenotypic plasticity induced by environmental constraints, an increased occurrence of alternative splicing (AS) events is also observed in response to biotic and abiotic stresses (Kwon et al. 2014; Seo et al. 2013; Staiger and Brown 2013). Until recently it was believed that most of these alternative transcripts were degraded through the Nonsense-Mediated Decay (NMD) pathway, so that AS events had only indirect and limited consequences on the phenotype through the decreased availability of the functional transcript. However, the exact proportion of alternate transcripts that are actually degraded through the NMD pathway remains largely unknown (Reddy et al. 2013) and recent evidence suggests that even transcripts with known hallmarks of NMD targeting might be able to evade it (Leviatan et al. 2013). Moreover, the increasing amount of high-throughput transcriptome sequence data available is changing the perception of AS. Indeed, it shows that these events (which affect over 60 % of intron-containing genes in both Arabidopsis and soybean) are far more ubiquitous than previously thought (Marquez et al. 2012; Shen et al. 2014). The expression of splicing factors themselves is regulated by AS and is responsive to both environmental and developmental cues with a high degree of tissue specificity (Guerra-Peraza et al. 2009; Petrillo et al. 2014; Quesada et al. 2003, 2005; Simpson et al. 2003; Staiger and Brown 2013; Vitulo et al. 2014). The majority of the MADS-box genes involved in either flowering time or in the regulation of flower development produce transcript isoforms (Severing et al. 2012) and flower morphogenesis itself coincides with a surge of AS events (Jiao and Meyerowitz 2010) thus raising the possibility that AS is required to increase the plasticity of genome expression in the course of development. Moreover, the fact that some of these alternative transcripts encode putative peptides with altered biological properties, such as those produced by R genes during pathogen infection, supports the idea that AS can also trigger phenotypic variations directly through increased proteome plasticity (Jaligot et al. 2014; Mastrangelo et al. 2012; Reddy et al. 2013; Seo et al. 2012; Syed et al. 2012). Taken together, these results suggest that AS represents a quick solution for modulating the production of the functional protein and/or for promoting the emergence of new biological functions without the need to involve either complex regulatory cascades or changes in transcription levels (Leviatan et al. 2013; Mastrangelo et al. 2012; Mazzucotelli et al. 2008).

6.3.5 *Border Controls Along Chromatin*

The recent demonstrations in both animals and plants that changes in chromatin condensation over exon-intron borders contribute to shifting the balance between different transcript isoforms (Luco et al. 2010, 2011) could provide a molecular basis for stress-induced changes in pre-mRNA splicing. Indeed, genome-wide mapping of DNA methylation and histone modifications have demonstrated the occurrence of high amounts of CG methylation and of usually repressive chromatin marks, such as H3K27me3, in the body of actively transcribed genes (by contrast with both 5′ and 3′ gene extremities) in species belonging to different kingdoms (Elling and Deng 2009; Feng et al. 2010a; Li et al. 2008; Lister et al. 2008; Roudier et al. 2011; Zemach et al. 2010b; Zhang et al. 2008; Zilberman et al. 2006). This result could suggest that a high background level of repressive marks, possibly correlated with abundant intragenic TE-derived sequences and contributing to exon definition (Chodavarapu et al. 2010; Tilgner et al. 2009), is a requirement for gene transcription (Saze and Kakutani 2011).

6.4 Meiosis: Wiping the Slate Clean

By contrast with mitosis, most epigenetic marks targeting genes are not maintained throughout meiosis, although plant germ cells do not undergo as complete a resetting of their epigenome as the one that is observed in animal organisms (Feng et al. 2010b). However, the repressive marks aimed at inactivating TEs, involving high levels of DNA methylation at CG and CHG and dimethylated H3K9-enriched chromatin (Kasschau et al. 2007; Lippman et al. 2003, 2004; Zhang et al. 2008) need to be faithfully transmitted to the next generation so that genome integrity is not compromised. Indeed, massively reactivated TE populations that are inherited from a parent defective in either the main maintenance DNA methyltransferase (*met1* mutants) or an important chromatin-remodeling factor (*ddm1* mutants) will only return to their initial silent state several generations after segregation from the defective alleles: it therefore seems that once silencing marks are lost at these loci, they are extremely difficult to re-establish (Teixeira et al. 2009; Tsukahara et al. 2009). Nevertheless, the exact modalities and kinetics of both TE reactivation and their re-silencing may differ depending on the family of elements (Mirouze et al. 2009). Moreover, because of the large TE content of plant genomes the silencing of TEs is always a trade-off between the necessity to silence them in order to preserve genome stability, and the risk to induce position-dependent gene silencing because of the spreading of repressive marks from nearby TEs (Woodhouse et al. 2014).

The initiation and subsequent maintenance of repressive marks in TE sequences is largely dependent on their targeting by the so-called heterochromatic 24-nucleotide siRNAs which promote their initial methylation at asymmetrical sites (CHH) through the RdDM pathway. RdDM in turn triggers the application of further silenc-

ing marks through other pathways involving KRYPTONITE-mediated H3K9 dimethylation and symmetrical CHG methylation by the CHROMOMETHYLASE3 DNA-methyltransferase (Lippman et al. 2004) and thus, their complete transcriptional silencing (Law and Jacobsen 2010; Saze and Kakutani 2011; Teixeira et al. 2009). Therefore, the inheritance of TE epigenetic repression paradoxically implies that the silencing mechanisms targeting these deleterious elements must be alleviated at some stage; this will enable TEs to be transcribed again into mRNAs further processed into siRNAs and silencing epigenetic marks to be re-applied. This partial TE demethylation-dependent reactivation takes place in *accessory* tissues that do not contribute to zygote formation (vegetative nucleus in the male germ line, central cell in the female germ line, possibly companion somatic cells surrounding the female gametophyte) and preferentially targets located within euchromatic, gene-rich regions (Ibarra et al. 2012). It is hypothesized that the siRNAs that are synthesized as a result could be transferred into the gametes (sperm cells or egg cell) to establish TE silencing in a cooperative fashion. After fertilization has occurred, it is hypothesized that TE silencing is reinforced in the embryo through the RdDM pathway mediated by the flux of siRNAs produced by the reactivated endosperm TEs (Calarco et al. 2012; Feng et al. 2010b; Hsieh et al. 2009; Ibarra et al. 2012; Slotkin et al. 2009). So far this phenomenon of TE reactivation in companion tissues has been shown to occur in Arabidopsis and rice (Zemach et al. 2010a) but also in other species with larger TE contents. Jaaskelainen et al. (2013) have shown the presence of the *BARE* TE (accounting for over 10 % of the barley genome) proteins in both ovaries and embryo. Sun et al. (2013) have demonstrated that TEs with the highest level of transcriptional activity in the endosperm of the developing wheat grain were also the most efficient at producing 24-nucleotide siRNAs for the induction of RdDM. Nevertheless, the occurrence of inter-cellular siRNAs movements has only been indirectly demonstrated so far, although other studies have shown that RdDM could be induced in tissues that are distant from the siRNA source after long-range movement of the small RNAs through the phloem (Dunoyer et al. 2010a, b; Molnar et al. 2010). It has been proposed that for some TE families this mechanism is complemented by the miRNA-dependent processing of TE transcripts into 21 nucleotide siRNAs triggering the post-transcriptional silencing processes (through transcript degradation). This pathway could prevail to silence TE populations that are reactivated during the epigenetic reprogramming of the germ line, in circumstances where 24 nucleotide siRNAs are not produced and DNA methylation cannot operate (Creasey et al. 2014).

In Arabidopsis, both the reactivation of TEs in companion cells of the germline and the resetting of DNA methylation marks in the endosperm dependent on *DEMETER*, a gene encoding a DNA glycosylase that acts by physically removing the methylated cytosines, leaving the DNA repair pathway to fill the gap (base-excision repair pathway). Monocotyledonous species lack *DEMETER* but undergo a similar extensive demethylation of the companion cells and the endosperm. In monocots related genes like *REPRESSOR OF SILENCING1* (*ROS1*) and *DEMETER-LIKE3* (*DML3*) might serve the same purpose (Ibarra et al. 2012). The TEs of the central cell are demethylated both actively through the action of

DEMETER and passively through the inhibition of methylation maintenance during DNA replication but whether both mechanisms take place in the male germ line is unclear since the lower number of cell divisions involved makes it unlikely that passive demethylation alone would be sufficient (Calarco and Martienssen 2011). Besides, the repression of a subpopulation of endosperm TEs by H3K27me3 in the absence of DNA methylation makes it likely that more than one type of epigenetic mark needs to be reset at repressed loci during the reprogramation process (Wollmann and Berger 2012). Methylated cytosines could also be oxidized by Tet-like enzymes and subsequently replaced by unmethylated cytosines through unknown mechanisms, as observed in the gametes and the early embryos of mammals (Cantone and Fisher 2013). How chromatin marks are erased between generations is so far unclear. The PRC2-dependent H3K27 trimethylation associated with the inactivation of the floral repressor *FLC* in the annual *Arabidopsis thaliana* is a prerequisite for vernalization-dependent flowering to occur but it needs to be reset so that the flowering of the following generation can be re-induced by cold treatment (Henderson and Jacobsen 2007; Yaish et al. 2011). All that is known is that this mark is erased during periconception, i.e. somewhere between gametogenesis and early embryo development. In its perennial relative *Arabidopsis halleri*, it has been shown that *FLC* expression is progressively restored on a yearly basis in autumn and winter (Aikawa et al. 2010). In some instances the resetting of chromatin modifications has been hypothesized to occur through the active, non-replication dependent replacement of modified histones H3 by newly synthesized *virgin* histones, such as has been observed in the zygote (Feng et al. 2010a; Ingouff et al. 2010; Wollmann and Berger 2012). While the central role of the chromatin remodelers DDM1 and MOM1 in the resetting of stress-related epigenetic marks between sexual generations has been demonstrated in a recent paper by Iwazaki and Paszkowski (2014), the exact molecular mechanism remains to be determined.

6.5 Applications to Breeding

The qualitative, quantitative and cumulative effects of epigenetic modifications on both genome stability and gene expression form a virtually bottomless reservoir of phenotypic plasticity, even in situations where very limited to no genetic variation is observed (Alonso-Blanco et al. 2009; Rapp and Wendel 2005; Reinders et al. 2009; Teixeira et al. 2009). Adaptation to environmental stress has important differences to the objective of agriculture to maximize crop yields, even if this concept is rapidly changing in the perspective of adapting agriculture to climate changes. Understanding the nature of adaptation in wild populations at the whole genome level may suggest strategies for crop breeding to deliver agricultural production with more resilience to climate variability (Henry and Nevo 2014).

The highly dynamic nature of both DNA methylation and histone modification patterns, the non-Mendelian transmission of traits and the long-held view that all epigenetic marks were erased between generations made it seem unlikely that

breeding science could benefit from such a source of variation. While information regarding the mechanisms governing the sexual transmission of epigenetic changes is still scarce, an increasing number of recent studies have explored this promising area, as well as the potential application of heritable epigenetic variation in breeding (Boyko and Kovalchuk 2011; Daxinger and Whitelaw 2010; Hauser et al. 2011; Mirouze and Paszkowski 2011; Paszkowski and Grossniklaus 2011; Richards 2011; Rival and Jaligot 2011; Skinner 2011).

6.6 Reprogramming and Somaclonal Variation

Waddington's take on epigenetic mechanisms, while interesting, did not consider that differentiation could be reverted. In animals, obstacles to dedifferentiation seem to be difficult to circumvent, and if pluripotent cells can be produced to regenerate tissues and organs, the production of an entire organism through reproductive cloning (nuclear transfer from a somatic cell to an enucleated egg) has a very low success rate and results in severe developmental abnormalities (Cantone and Fisher 2013). It has been found that most of these anomalies could be traced back to improper implementation of the developmental programs, possibly deriving from incomplete erasure of the pre-existing DNA methylation for the activation of pluripotency genes. Indeed, such a *methylation barrier* seems to be the main obstacle impeding the complete and stable reprogramming of mammalian somatic cells. More, it has been observed that in the pre-implantation of embryo in mammals a *de-methylation wave* sweeps throughout the nucleus and re-initializes nearly all methylation marks that were transmitted through the original parental gametes (Cantone and Fisher 2013; Feng et al. 2010a, b). Interestingly, a comparable genome-wide loss of differentiation-associated DNA methylation patterns is also achieved prior to cell tumorization (Cantone and Fisher 2013; Gordon et al. 2014).

To the contrary, plant tissues are generally amenable to dedifferentiation and cloning, even though cells sampled from mature organs and/or from older plants tend to lose their organogenic abilities. Depending on the plant species and the nature and initial differentiation state of the explants, dedifferentiation can be induced after treatment with plant growth regulators. Among them, auxin analogs such as 2,4-D are the most commonly used, with the occasional addition of cytokinins. Both families of molecules are thought to act through the erasing of the pre-existing DNA methylation patterns of the explant tissues and the relaxation of chromatin structures, leading to the resetting of the developmental program and the return to the proliferative state (Exner and Hennig 2008; LoSchiavo et al. 1989; Miguel and Marum 2011). By contrast with animal zygotes, the *de-methylation wave* does not seem to take place upon plant zygote formation (Baroux et al. 2011; Feng et al. 2010b; Hajkova et al. 2008; Ingouff et al. 2010; Pillot et al. 2010; Slotkin et al. 2009), which could indicate that reprogramming in plants either involves a lesser requirement for the erasing of pre-existing epigenetic marks, or that these marks are both more labile and more responsive to de-differentiating stimuli.

During their development, higher plants undergo phase changes towards sexual maturity and DNA methylation has been speculated to play a key role in physiological ageing of tree species, The prevailing hypothesis is that as trees age and develop, genomic DNA could become hypermethylated and induce changes in the expression of genes related to maturation traits (Monteuuis et al. 2008; Smulders and de Klerk 2011).

Aberrant development and somaclonal variations can also result from the cloning of plant tissues (Kaeppler et al. 2000), they are generally not lethal even in cases where fitness is severely impaired and fertility is compromised. This points out to the greater plasticity of plant organisms, which are able to accommodate both a large-scale misregulation of genomic stability and its phenotypic consequences and yet keep on functioning (mostly) normally throughout their life cycle. Today most studies aimed at understanding somaclonal variations focus on detrimental examples of variant phenotypes in a bid to either prevent their emergence or to eliminate them from further commercial exploitation through the assessment of clonal conformity. The potential use of genetic/cytologic/epigenetic markers is expected to ensure that the clonal progeny is identical to the mother plant for all the characteristics and traits that are deemed essential (Miguel and Marum 2011; Smulders and de Klerk 2011). However, theoretically a proportion of somaclonal variations can also translate into the acquisition of advantageous traits and therefore, in enhanced fitness of the clonal progeny (Rival et al. 2013; Wang and Wang 2012). This could be especially true of the epigenetic variations resulting in the enhanced production of stress response factors since it has been observed that these mechanisms are instrumental in the adaptation of tissues to in vitro culture conditions. This could be put to good use through the combined engineering and in vitro selection of variant lines displaying desirable stress tolerance traits (Rai et al. 2011) provided these traits can be stabilized in subsequent generations.

6.7 More Cards in the Deck of Heredity

Most epigenetic alterations are meant to be reversed either within the lifetime of the plant or between sexual generations. However, instances of stable transmission of epialleles (different epigenetic versions of a gene, altering its expression without any permanent DNA polymorphism) across many generations have been described in plants (Hauser et al. 2011; Kakutani 2002; Kalisz and Purugganan 2004). There are several studied examples of Transgenerational Epigenetic Inheritance (TEI) such as the *peloric* variation in *Linaria vulgaris*, where the hypermethylation of a single gene promotes radial instead of bilateral flower symmetry and has been stably inherited over centuries (Cubas et al. 1999). More recent illustrations of spontaneous epiallele inheritance also include the colorless non-ripening (*Cnr*) variant of tomato (Manning et al. 2006), sex determination in melon (Martin et al. 2009) or the *S*-locus based self-incompatibility in Brassica (Tarutani et al. 2010). In each instance, the inheritance of the phenotype is correlated with the faithful

transmission of a highly stable repressed state, through DNA methylation (some of it mediated by small non-coding RNAs), imposition of repressive histone marks leading to heterochromatinization, or both. Bond and Baulcombe (2014) recently reviewed how RNA-based system in plants involving small (s)RNAs can influence heritable phenotypes in plants through the de novo establishment and maintenance of DNA methylation at many sites in the genomes. Paszkowski and Grossniklaus (2011) observed that the ability to maintain a basal level of DNA methylation at a given locus could be a prerequisite for TEI to occur, since inheritance of epigenetic states was generally limited to a few generations in methylation-impaired backgrounds (Johannes et al. 2009; Reinders et al. 2009; Teixeira et al. 2009), possibly as a result of enhanced genomic instability through homologous recombination (Colomé-Tatché et al. 2012; Mirouze et al. 2012; Pecinka et al. 2012). Small RNAs could be important in recruiting DNA- and histone- modifying enzymes and targeting them to the appropriate loci, thereby ensuring the maintenance of their epigenetic status (Boyko and Kovalchuk 2011; Boyko et al. 2010; Brosnan and Voinnet 2011; Teixeira et al. 2009).

The phenomenon of paramutation, which has been extensively studied in maize, might provide clues towards the identification of the molecular mechanisms underlying TEI. In paramutation, one paramutagenic epiallele of a gene is transmitting both its transcriptional activity and its epigenetic status to a paramutable epiallele. The newly acquired state can be faithfully maintained through many generations, regardless of further exposure to the original paramutagenic allele (Arteaga-Vazquez and Chandler 2010; Arteaga-Vazquez et al. 2010; Chandler 2010; Stam 2009; Suter and Martin 2010). Several authors have indicated that proteins with unusual structural features and unknown functions might contribute to stabilize further the repression of the paramutable allele (Barbour et al. 2012; Brzeska et al. 2010). This phenomenon might complement other well-documented and more general epigenetic mechanisms such as RdDM and H3K9me2-mediated chromatin condensation (Arteaga-Vazquez et al. 2010; Erhard Jr and Hollick 2011; Erhard et al. 2009; Sekhon et al. 2012; Stonaker et al. 2009). Interestingly, the *cis* motifs and *trans* factors that are necessary to induce the epiallelic interactions underlying the paramutation phenomenon seem to be shared by evolutionary distant organisms such as drosophila and maize (McEachern and Lloyd 2012). Finally, results from drosophila and human studies have prompted the hypothesis that meiotic transmission of Polycomb Repressive Complexes could provide a molecular scaffold enabling the inheritance of epigenetic patterns as a form of *molecular memory* (Francis 2009; Hansen and Helin 2009; Köhler and Grossniklaus 2002).

In addition to the inheritance of natural epigenetic variation, stress-induced epigenetic changes can persist beyond the duration of said stress and promote within-generation stress tolerance, inducing a quicker/stronger response to future occurrences of the same stress through a phenotypic change (Ding et al. 2014). Such modifications could set ground for longer-term *stress memory* in the progeny of the stressed plants through the fixation of specific epialleles or epigenetic states in a population and their transmission, together with the corresponding stress tolerance trait, through several sexual generations thus resulting in an enhanced fitness of the

next generation in similar conditions (Transgenerational Epigenetic Inheritance). Attempts to assess such a hypothesis have been undertaken (Bilichak et al. 2012; Boyko and Kovalchuk 2011; Lang-Mladek et al. 2010; Molinier et al. 2006; Pecinka et al. 2009; Scoville et al. 2011; Verhoeven et al. 2010) with, so far, mixed results regarding the meiotic stability of the stress-induced variations. It must be kept in mind that actual TEI is difficult to demonstrate since partial carry-over of the epigenetic marks from the *stressed* plants (F0) to the first *un-stressed* generation (F1) by parental transmission (of transcripts, non-coding RNAs and/or proteins) can occur stochastically without involving true inheritance mechanisms (Mirouze and Paszkowski 2011; Youngson and Whitelaw 2008). As a consequence TEI can only be characterized through the analysis of epigenetic marks and the phenotyping of the successive progenies deriving from the original stressed plants so that the parallel transmission of an epigenetic alteration and of the corresponding phenotype can be established up to at least generation F2. However, most studies published to date used an experimental design that was limited to the stressed parents and their F1 progeny, and for this reason they are not conclusive as to the generalization of the TEI phenomenon. The study from Kou et al. (2011), on the other hand, successfully demonstrated the combined inheritance of a tolerance to N deficit and stress-induced changes in DNA methylation patterns over three selfed generations of rice. Ou et al. (2012) provided another good example, with the effects of heavy metal stress on the inheritance of both altered DNA methylation patterns and heavy metal tolerance followed in two successive selfed generations of rice. Clearly, such an experimental design is not easily manageable for all plant models (such as perennials because of their long life cycle and bulkiness) and can require costly analyses depending on the nature of the phenotype. Also, constraints applied in a controlled environment will be easier to study than stresses occurring either in an open-field environment or in the wild, since those will necessarily involve highly fluctuating conditions and complex epigenetic and transcriptional changes resulting from the interaction between several parameters (Richards 2011; Staiger and Brown 2013). Most importantly the genetic variation between parental genotypes impacts both the epigenetic response and the resulting stress-induced phenotype. In most instances, it will therefore be difficult to demonstrate any association between epigenetic variation and phenotypic alteration without interference from the genetic background, especially in natural populations of strictly outcrossing species (Richards 2006, 2008, 2011). The mutagenic effects of the stress might also prove a problem when trying to tear apart genetic and epigenetic effects, such as the UV light stress used by Molinier et al. (2006). Asexually reproducing species such as the apomictic dandelion (Verhoeven and Preite 2014; Verhoeven and van Gurp 2012; Verhoeven et al. 2010) are a material of choice when studying the inheritance of stress-induced epigenetic modifications because of the genetic homogeneity between parents and clonal progenies, which allows the study of *pure* epigenetic inheritance. However, it might not prove possible to generalize to sexually propagated species the mechanisms underlying TEI in asexual populations since the latter *skip* the epigenetic reprogramming steps (Verhoeven and Preite 2014).

When studying phenotypic variation of 134 epigenetic recombinant inbred lines (epiRILs) and 23 control lines (Ctr lines) of *Arabidopsis thaliana* across two environments, Zhang et al. (2013) suggested that variation in DNA methylation can cause substantial heritable variation of ecologically important plant traits, including root allocation, drought tolerance and nutrient plasticity, and that rapid evolution based on epigenetic variation alone should thus be possible.

It has been hypothesized that stress-induced epigenetic marks that are preserved through TEI could be one of the driving forces of speciation; thereby somewhat fulfilling Lamarck's prophecy that environment could influence heritable traits. This could occur as a two-track system, with high-frequency, labile epigenetic variations leading to the rapid emergence of new phenotypes, some of which (the most advantageous) being subsequently stabilized or fixated in the variant population through mutations (Bossdorf et al. 2008; Richards et al. 2010). Indeed, methylated cytosines (mC) have long been assumed to be prone to spontaneous deamination, opening the way for the emergence of transitional point mutations (Goll and Bestor 2005). Alternatively, numerous studies have underlined the abundance of TE-derived sequences in the vicinity of many genes, some of them essential for plant development. The presence of TE or TE remnants can cause mutations with visible phenotypic impact, such as is the case with the Pinot blanc phenotype of grapevine (Vezzulli et al. 2012; Walker et al. 2007) and the nectarine phenotype of peach (Vendramin et al. 2014). Moreover these insertions, even when they are no longer transcriptionally or transpositionally active, can interfere with the expression of nearby genes (Bennetzen 2000; Feschotte 2008; Maumus and Quesneville 2014a; Weil and Martienssen 2008).

A number of recent articles have illustrated how the intrinsic properties of TE sequences to be targeted by the host's epigenetic silencing mechanisms and to be re-activated in stress conditions have been integrated in the functioning of endogenous genes and have resulted in distinctive, and in some cases highly stable, phenotypes. Among the abovementioned examples of natural epialleles, sexual differentiation in melon is determined by changes in the epigenetic status of a TE inserted in the vicinity of the *CmWIP1* transcription factor gene (Martin et al. 2009). Ripening in tomato is controlled by an SBP-box transcription factor which expression is downregulated through promoter DNA methylation in the *Colorless non-ripening* (*Cnr*) epimutant (Manning et al. 2006). It was later hypothesized by other authors that this repression could be dependent on the presence of a neighboring TE insertion in *Cnr*-prone cultivars (Paszkowski and Grossniklaus 2011). The variable methylation of TE-derived sequences in the promoter of the *FWA* gene is an important determinant of flowering time in Arabidopsis (Soppe et al. 2000). Li et al. (2010) demonstrated that the altered floral phenotype of *Hose in hose* variants of primrose is associated with a tissue-specific TE insertion in the promoter of a flower developmental gene, triggering changes in both its methylation status and its expression.

The blood orange phenotype is dependent on the cold-dependent upregulation of a TE insertion upstream of the *Ruby* MYB transcription factor that enhances anthocyanin biosynthesis (Butelli et al. 2012). Changes in the chromatin status of an

intronic retrotransposon insertion directly affect the production of functional transcripts by the *RPP7* disease resistance gene of Arabidopsis (Tsuchiya and Eulgem 2013). This process of *TE domestication* could be at the origin of the acquisition of environmental responsiveness by many plant genes. It could also help explaining why TE silencing and gene silencing rely on similar mechanisms and partially share some regulatory pathways (Lisch and Bennetzen 2011; Saze and Kakutani 2011). In accordance with this hypothesis of inter-connected regulations, the activation of the *ONSEN* retrotransposon in heat stress conditions is dependent on its ability to bind some of the host's heat shock factors through a *captured cis*-element (Cavrak et al. 2014). In this context, it would not be surprising to find that TE-derived sequences have also been co-opted to ensure the heritability of changes in gene expression (Erhard Jr and Hollick 2011; Erhard et al. 2013; Suter and Martin 2010).

6.8 Hybrids: Being More than the Sum of Its Parts

Among other functions, epigenetic mechanisms are aimed at ensuring that the appropriate balance is achieved in terms of expressed copies among paralogous and homeologous genes, but also within TE families. Nevertheless, some proportion of TE reactivation is *authorized* between sexual generations so that the newly formed zygote can implement repressive epigenetic mechanisms targeting sequences that are contributed by each parental genome. This mechanism not only participates to the maintenance of genome integrity during the reproductive phase of the plant's life cycle, but it could also contribute to a small RNA-dependent compatibility assessment between the parental genomes, similarly to what is observed in Drosophila and in ciliates (Bourc'his and Voinnet 2010). According to this scenario, the labile silencing signal mediated by the siRNA production from each parental TE set would have to be *confirmed* through chromatin-based mechanisms so that the epigenetic repression that has been initiated is maintained subsequently throughout development. Hybrid incompatibility could then result from qualitative (sequence divergence) or quantitative (dosage imbalance) differences in TE population and TE-derived siRNAs between the parental genomes.

Imprinted genes, which are expressed in a mono-allelic manner in either the endosperm or the embryo as a result of parent-of-origin-specific DNA methylation, could constitute a sort of genomic *gauge* that would help ensuring that the appropriate genome dosage has been achieved after double fertilization (Jullien and Berger 2010; Wollmann and Berger 2012). The observation that TEs are inserted in the vicinity of many recently identified candidate imprinted genes suggests that the differential parental methylation of these genes could have derived from these insertions.

Even when parental genomes are compatible, evidence show that epigenetic marks are not always additive, nor are they subjected to simple relationships of dominance vs. recessivity. Rather, some of the epigenetic patterns found in the progeny result from complex rearrangements with respect to the patterns character-

ized in each parental genome, in order to manage the joint regulation of both parental gene sets and TE populations (some of them common to both parents, others specific to one parent). This point is all the more important when polyploidization is involved and results in a dramatic increase in genome size. Illustration to this are provided by studies focusing on hybrid formation in Arabidopsis, maize (Barber et al. 2012), wheat (Kenan-Eichler et al. 2011), rice (He et al. 2010), and tomato (Shivaprasad et al. 2012). Hollister et al. (2009, 2011) have demonstrated how the production of TE-derived 24 and 21 nucleotides siRNAs is decreased in different F1 allotetraploid hybrids of Arabidopsis compared to the parental diploid species, resulting in the re-expression of several TE families. As an explanation, it has been proposed that siRNAs contributed through each parental germline fail to recognize the TE populations from the other parental genome, resulting in their transient reactivation in the early F1 embryo (Calarco and Martienssen 2011). Alternatively, it is possible that the merger of two different genomes is perceived as a genomic stress that triggers stress response under the form of epigenetic changes resulting in altered gene expression and TE reactivation in the F1 hybrid (He et al. 2010). This transient alleviation of TE repression would in turn allow their transcription and the production of new siRNAs, which would contribute in the course of the F1 generation to the establishment of *fresh* silencing mechanisms based on the new *borders* of this hybrid genome.

Interestingly, similar mechanisms between homeologous genes contributed by each parents (with differences in both levels of expression and epigenetic patterns) could provide a molecular basis to the phenomenon of heterosis. The difficulty, in these inter-specific hybrids, is to discriminate the effects contributed by the genetic variations from those contributed by the epigenetic variations. Over time, ancient polyploid genomes have progressively lost some of the duplicated genes and returned to a diploid behavior. It has been noticed that this phenomenon has affected subgenomes asymmetrically: one subgenome has lost more genes and those that have been retained have a lesser level of expression compared to that of the corresponding homeologous genes in the *dominant* subgenome. This process, which seems to have played a role in genome stabilization after allopolyploidization, is mediated through the 24-nucleotide-based silencing of TE elements inserted upstream of the genes in the *dominated* subgenome (with both negative and, surprisingly, positive effects on gene expression) and is meiotically heritable, in a combination of position effect and epigenetic effect (Woodhouse et al. 2014).

It must be observed that a strong heterosis effect can be observed even in the F1 progeny resulting from intra-specific crosses between different diploid ecotypes of *Arabidopsis thaliana*, implying that the differences between the two parental epigenomes is sufficient to generate heterosis in the absence of significant genetic differences (Groszmann et al. 2011b). In this case, the decrease in 24 nucleotide siRNA level is only observed for a limited number of genes and gene-flanking sequences (also showing hypomethylation), and the levels of 21 nucleotide siRNAs and miRNAs is not altered. The loci with altered DNA methylation and/or reduced siRNA production match loci presenting contrasted methylation and/or siRNAs production in the parents, respectively (Greaves et al. 2012, 2014). Also, *Trans*

Chromosomal Methylation/deMethylation (TCM/TCdM) phenomena, which underlie transmission of the epigenetic state of one of the parental alleles to the other one in the hybrid, in a mechanism similar to paramutation (Erhard and Hollick 2011; Mittelsten Scheid et al. 2003; Suter and Martin 2010) occur essentially in genes and gene-flaking regions and affect one third of those showing differential methylation, with the possibility that these methylation changes are induced by the altered siRNA production. Similar observations (interactions between epialleles inherited from parents with contrasted genomic DNA methylation levels and TE reactivation in the offspring) were made in synthetic epiRIL populations of Arabidopsis (Johannes et al. 2009; Reinders et al. 2009) and in RIL of maize (Regulski et al. 2013). Indeed, 24 nucleotide siRNAs are involved in RdDM, although not all siRNA-producing loci with different methylation patterns between the parents undergo TCM/TCdM, pointing out to the existence of other factors (such as histone modifications?) with a role in this phenomenon (Groszmann et al. 2013). It is however unclear which proportion of the TCM/TCdM events results in actual changes in gene expression. Indeed, a majority of the epigenetic changes do not trigger any visible expression or phenotypic change: they are random, phenotypically neutral, part of the non-specific epigenetic background noise or only serve structural purposes. As a consequence, only a subset of the loci displaying altered epigenetic patterns in the hybrid will actually participate to the heterotic phenotype. Those that do, on the other hand, tend to behave as epigenetic QTLs in both synthetic and natural populations, and could explain as much as 90 % of the variation of a complex trait (Cortijo et al. 2014). They can also be selected, even in the absence of genetic variation (Hauben et al. 2009).

Another mechanism that might contribute to genomic and phenotypic plasticity in hybrids is homologous recombination. Indeed, genome size in plants is positively correlated with both TE content and the amount of nuclear heterochromatin and it is negatively correlated with crossing-over frequency (Henderson 2012). Conversely, it has been shown that both decreased genomic methylation/decreased chromatin condensation and neopolyploidization increased meiotic recombination rates, mainly in euchromatic, gene-rich chromosome arms vs. repeat-rich centromeric regions (Colomé-Tatché et al. 2012; Mirouze et al. 2012; Pecinka et al. 2012).

As for histone modifications, histone methylation at either H3K27 or H3K4 and H3K9 acetylation are mostly transmitted an additive fashion in both Arabidopsis and rice (He et al. 2010; Moghaddam et al. 2011). However, in both species it has been found that loci displaying both TCM/TCdM and changes in expression also had non-additive histone modifications (Greaves et al. 2012; He et al. 2010), and in rice an allelic bias has been observed. The exact function of histone marks in hybrid formation and heterosis therefore remains unclear.

Ultimately, the fact that the epigenetic regulations and gene expression patterns that are implemented by the hybrid genome are not additive with respect to those of each parent could increase the allelic diversity contributed by both parental gene sets and might contribute to the phenomenon of heterosis, in which the F1 hybrids have higher biomass and productivity than either parental lines (Calarco and Martienssen 2011; Druka et al. 2010; Groszmann et al. 2011a; Qi et al. 2010). The

lack of inheritance to F2 and subsequent generations (for most loci) and the associated loss of heterosis are proposed to correspond to the gradual dilution over the F1 of siRNA production from *donor* chromosomal segments with respect to *recipient* ones. Once the ratio of parental *donor* segments is no longer sufficient to insure maintenance of the parental methylation pattern, this could trigger the random segregation of the TCM/TCdM polymorphisms and corresponding expression changes (Greaves et al. 2014).

6.9 Conclusions and Prospects

Fully understanding the rules governing epigenetic inheritance is crucial for the exploitation of epigenetic diversity through progressive selection schemes aimed at optimizing growth and yields for targeted agro-ecological conditions (Boyko and Kovalchuk 2011; King et al. 2010; Kou et al. 2011; Lukens and Zhan 2007). Especially how meta-stable repressed states can be induced and maintained throughout meiosis, involving interactions between epialleles or between heterochromatic sequences. Exploring the epigenetic interactions occurring in trans between different loci within each parental genomes, and how these interactions evolve in the newly formed hybrid (diploid or polyploid) will also be instrumental for designing future breeding schemes (Springer 2013). It must be kept in mind, however, that *Arabidopsis thaliana* has a small, diploid genome harboring an unusually small proportion of TEs compared with crop plants, especially when considering the polyploid genomes of cereals (Tenaillon et al. 2010), although a more recent in silico study estimates that *A. thaliana*'s *repeatome* represents over 30 % of the genome size (Maumus and Quesneville 2014a, b). It is therefore likely that rules governing the relatively simple epigenome of Arabidopsis will not apply to larger, polyploid plant genomes which are made of over 80 % of TEs (with possibly an even larger percentage of the genome being repeat-derived) with a comparatively higher potential for epigenetic interactions in trans between homologous/homeologous sequences to occur. Indeed, Hollister et al. (2011) have hypothesized that large polyploid genomes with a high TE load might have a greater tolerance to TE expression compared to small, TE-poor ones. When considering the longer-term (evolutive) implications of TEI, precautions must be taken when comparing results from studies focusing on wild species and weeds (Arabidopsis, dandelion) vs. crop genotypes resulting from selection (tomato, apple) since selection is unlikely to have the same influence on all of these species.

Also, a better knowledge on the epigenetic events surrounding hybrid genome formation, especially the managing of new TE insertions, will provide insights into how the host genome reacts to the insertion of a foreign sequence in terms of readjusting repressive mechanisms. Because protocols in genetic engineering more often involve the in vitro regeneration of transformed plants, more in-depth studies of stress-responsive epigenetic mechanisms will also give us information on how to optimize, and how to stabilize over time the expression of transgenes.

References

Aikawa S, Kobayashi MJ, Satake A et al (2010) Robust control of the seasonal expression of the Arabidopsis FLC gene in a fluctuating environment. Proc Natl Acad Sci 107:11632–11637. doi:10.1073/pnas.0914293107

Alonso-Blanco C, Aarts MGM, Bentsink L et al (2009) What has natural variation taught us about plant development, physiology, and adaptation? Plant Cell 21:1877–1896. doi:10.1105/tpc.109.068114

Angers B, Castonguay E, Massicote R (2010) Environmentally induced phenotypes and DNA methylation: how to deal with unpredictable conditions until the next generation and after. Mol Ecol 19:1283–1295. doi:10.1111/j.1365-294X.2010.04580.x

Arteaga-Vazquez MA, Chandler VL (2010) Paramutation in maize: RNA mediated trans-generational gene silencing. Curr Opin Genet Dev 20:156–163. doi:10.1016/j.gde.2010.01.008

Arteaga-Vazquez M, Sidorenko L, Rabanal FA et al (2010) RNA-mediated trans-communication can establish paramutation at the b1 locus in maize. Proc Natl Acad Sci 107:12986–12991. doi:10.1073/pnas.1007972107

Bai M, Yang G-S, Chen W-T et al (2012) Genome-wide identification of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analyses in response to viral infection and abiotic stresses in *Solanum lycopersicum*. Gene 501:52–62. doi:10.1016/j.gene.2012.02.009

Barber WT, Zhang W, Win H et al (2012) Repeat associated small RNAs vary among parents and following hybridization in maize. Proc Natl Acad Sci 109:10444–10449. doi:10.1073/pnas.1202073109

Barbour J-ER, Liao IT, Stonaker JL et al (2012) Required to maintain repression2 is a novel protein that facilitates locus-specific paramutation in maize. Plant Cell Online 24:1761–1775. doi:10.1105/tpc.112.097618

Baroux C, Raissig MT, Grossniklaus U (2011) Epigenetic regulation and reprogramming during gamete formation in plants. Curr Opin Genet Dev 21:124–133. doi:10.1016/j.gde.2011.01.017

Barrera-Figueroa BE, Gao L, Wu Z et al (2012) High throughput sequencing reveals novel and abiotic stress-regulated microRNAs in the inflorescences of rice. BMC Plant Biol 12:132. doi:10.1186/1471-2229-12-132

Bastow R, Mylne JS, Lister C et al (2004) Vernalization requires epigenetic silencing of FLC by histone methylation. Nature 427:164–167. doi:10.1038/nature02269

Bemer M, Grossniklaus U (2012) Dynamic regulation of polycomb group activity during plant development. Curr Opin Plant Biol 15:523–529. doi:10.1016/j.pbi.2012.09.006

Bender J (2004) DNA methylation and epigenetics. Annu Rev Plant Biol 55:41–68. doi:10.1146/annurev.arplant.55.031903.141641

Bennetzen JL (2000) Transposable element contributions to plant gene and genome evolution. Plant Mol Biol 42:251–269. doi:10.1023/A:1006344508454

Bilichak A, Ilnystkyy Y, Hollunder J, Kovalchuk I (2012) The progeny of *Arabidopsis thaliana* plants exposed to salt exhibit changes in DNA methylation, histone modifications and gene expression. PLoS One 7, e30515. doi:10.1371/journal.pone.0030515

Blevins T, Pontvianne F, Cocklin R et al (2014) A two-step process for epigenetic inheritance in Arabidopsis. Mol Cell 54:30–42. doi:10.1016/j.molcel.2014.02.019

Bond D, Baulcombe D (2014) Small RNAs and heritable epigenetic variation in plants. Trends Cell Biol 24(2):100–107. doi:http://dx.doi.org/10.1016/j.tcb.2013.08.001

Bossdorf O, Richards CL, Pigliucci M (2008) Epigenetics for ecologists. Ecol Lett 11:106–115. doi:10.1111/j.1461-0248.2007.01130.x

Bourc'his D, Voinnet O (2010) A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. Science 330:617–622. doi:10.1126/science.1194776

Boyko A, Kovalchuk I (2011) Genome instability and epigenetic modification – heritable responses to environmental stress? Curr Opin Plant Biol 14:260–266. doi:10.1016/j.pbi.2011.03.003

Boyko A, Blevins T, Yao Y et al (2010) Transgenerational adaptation of Arabidopsis to stress requires DNA methylation and the function of Dicer-like proteins. PLoS One 5, e9514. doi:10.1371/journal.pone.0009514

Brosnan CA, Voinnet O (2011) Cell-to-cell and long-distance siRNA movement in plants: mechanisms and biological implications. Curr Opin Plant Biol 14:580–587. doi:10.1016/j.pbi.2011.07.011

Brzeska K, Brzeski J, Smith J, Chandler VL (2010) Transgenic expression of CBBP, a CXC domain protein, establishes paramutation in maize. Proc Natl Acad Sci 107:5516–5521. doi:10.1073/pnas.1001576107

Butelli E, Licciardello C, Zhang Y et al (2012) Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. Plant Cell Online 24:1242–1255. doi:10.1105/tpc.111.095232

Calarco JP, Martienssen RA (2011) Genome reprogramming and small interfering RNA in the Arabidopsis germline. Curr Opin Genet Dev 21:134–139. doi:10.1016/j.gde.2011.01.014

Calarco JP, Borges F, Donoghue MTA et al (2012) Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell 151:194–205. doi:10.1016/j.cell.2012.09.001

Cantone I, Fisher AG (2013) Epigenetic programming and reprogramming during development. Nat Struct Mol Biol 20:282–289. doi:10.1038/nsmb.2489

Cao X, Jacobsen SE (2002) Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci 99:16491

Cao X, Springer NM, Muszynski MG et al (2000) Conserved plant genes with similarity to mammalian de novo DNA methyltransferases. Proc Natl Acad Sci U S A 97:4979

Cao X, Aufsatz W, Zilberman D et al (2003) Role of the DRM and CMT3 Methyltransferases in RNA-directed DNA methylation. Curr Biol 13:2212–2217. doi:10.1016/j.cub.2003.11.052

Cavrak VV, Lettner N, Jamge S et al (2014) How a retrotransposon exploits the plant's heat stress response for its activation. PLoS Genet 10, e1004115. doi:10.1371/journal.pgen.1004115

Chandler VL (2010) Paramutation's properties and puzzles. Science 330:628–629. doi:10.1126/science.1191044

Charron J-BF, He H, Elling AA, Deng XW (2009) Dynamic landscapes of four histone modifications during deetiolation in Arabidopsis. Plant Cell. doi:10.1105/tpc.109.066845

Chen L-T, Luo M, Wang Y-Y, Wu K (2010) Involvement of Arabidopsis histone deacetylase HDA6 in ABA and salt stress response. J Exp Bot 61:3345–3353. doi:10.1093/jxb/erq154

Chen H, Li Z, Xiong L (2012) A plant microRNA regulates the adaptation of roots to drought stress. FEBS Lett 586:1742–1747. doi:10.1016/j.febslet.2012.05.013

Cho J-N, Ryu J-Y, Jeong Y-M et al (2012) Control of seed germination by light-induced histone arginine demethylation activity. Dev Cell 22:736–748. doi:10.1016/j.devcel.2012.01.024

Chodavarapu RK, Feng S, Bernatavichute YV et al (2010) Relationship between nucleosome positioning and DNA methylation. Nature 466:388–392. doi:10.1038/nature09147

Choi C-S, Sano H (2007) Abiotic-stress induces demethylation and transcriptional activation of a gene encoding a glycerophosphodiesterase-like protein in tobacco plants. Mol Gen Genomics 277:589–600. doi:10.1007/s00438-007-0209-1

Colomé-Tatché M, Cortijo S, Wardenaar R et al (2012) Features of the Arabidopsis recombination landscape resulting from the combined loss of sequence variation and DNA methylation. Proc Natl Acad Sci 109:16240–16245. doi:10.1073/pnas.1212955109

Cortijo S, Wardenaar R, Colomé-Tatché M et al (2014) Mapping the epigenetic basis of complex traits. Science 343:1145–1148. doi:10.1126/science.1248127

Creasey KM, Zhai J, Borges F et al (2014) miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. Nature. doi:10.1038/nature13069

Cubas P, Vincent C, Coen E (1999) An epigenetic mutation responsible for natural variation in floral symmetry. Nature 401:157–161. doi:10.1038/43657

Daxinger L, Whitelaw E (2010) Transgenerational epigenetic inheritance: more questions than answers. Genome Res 20:1623–1628. doi:10.1101/gr.106138.110

De Vega-Bartol JJ, Simões M, Lorenz WW et al (2013) Transcriptomic analysis highlights epigenetic and transcriptional regulation during zygotic embryo development of *Pinus pinaster*. BMC Plant Biol 13:123. doi:10.1186/1471-2229-13-123

Delker C, Quint M (2011) Expression level polymorphisms: heritable traits shaping natural variation. Trends Plant Sci 16:481–488. doi:10.1016/j.tplants.2011.05.009

Ding Y, Virlouvet L, Liu N et al (2014) Dehydration stress memory genes of *Zea mays*; comparison with *Arabidopsis thaliana*. BMC Plant Biol 14:141. doi:10.1186/1471-2229-14-141

Dowen RH, Pelizzola M, Schmitz RJ et al (2012) Widespread dynamic DNA methylation in response to biotic stress. Proc Natl Acad Sci 109:E2183–E2191. doi:10.1073/pnas.1209329109

Druka A, Potokina E, Luo Z et al (2010) Expression quantitative trait loci analysis in plants. Plant Biotechnol J 8:10–27. doi:10.1111/j.1467-7652.2009.00460.x

Dunoyer P, Brosnan CA, Schott G et al (2010a) An endogenous, systemic RNAi pathway in plants. EMBO J 29:1699–1712. doi:10.1038/emboj.2010.65

Dunoyer P, Schott G, Himber C et al (2010b) Small RNA duplexes function as mobile silencing signals between plant cells. Science 328:912–916. doi:10.1126/science.1185880

Elling AA, Deng XW (2009) Next-generation sequencing reveals complex relationships between the epigenome and transcriptome in maize. Plant Signal Behav 4:760–762. doi:10.1105/tpc.109.065714

Erhard KF Jr, Hollick JB (2011) Paramutation: a process for acquiring trans-generational regulatory states. Curr Opin Plant Biol 14:210–216. doi:10.1016/j.pbi.2011.02.005

Erhard KF, Stonaker JL, Parkinson SE et al (2009) RNA polymerase IV functions in paramutation in *Zea mays*. Science 323:1201–1205. doi:10.1126/science.1164508

Erhard KF, Parkinson SE, Gross SM et al (2013) Maize RNA polymerase IV defines trans-generational epigenetic variation. Plant Cell Online. doi:10.1105/tpc.112.107680

Exner V, Hennig L (2008) Chromatin rearrangements in development. Curr Opin Plant Biol 11:64–69. doi:10.1016/j.pbi.2007.10.004

Feng S, Cokus SJ, Zhang X et al (2010a) Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci 107:8689–8694. doi:10.1073/pnas.1002720107

Feng S, Jacobsen SE, Reik W (2010b) Epigenetic reprogramming in plant and animal development. Science 330:622–627. doi:10.1126/science.1190614

Feschotte C (2008) The contribution of transposable elements to the evolution of regulatory networks. Nat Rev Genet 9:397–405. doi:10.1038/nrg2337

Finnegan EJ (2002) Epialleles – a source of random variation in times of stress. Curr Opin Plant Biol 5:101–106. doi:10.1016/S1369-5266(02)00233-9

Finnegan EJ, Kovac KA (2000) Plant DNA methyltransferases. Plant Mol Biol 43:189–201

Finnegan EJ, Peacock WJ, Dennis ES (1996) Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. Proc Natl Acad Sci U S A 93:8449–8454

Francis NJ (2009) Does maintenance of polycomb group proteins through DNA replication contribute to epigenetic inheritance? Epigen Off J DNA Methyl Soc 4:370–373

Goll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases. Annu Rev Biochem 74:481–514. doi:10.1146/annurev.biochem.74.010904.153721

Gordon K, Clouaire T, Bao XX et al (2014) Immortality, but not oncogenic transformation, of primary human cells leads to epigenetic reprogramming of DNA methylation and gene expression. Nucleic Acids Res 42:3529–3541. doi:10.1093/nar/gkt1351

Grafi G, Chalifa-Caspi V, Nagar T et al (2011) Plant response to stress meets dedifferentiation. Planta 233:433–438. doi:10.1007/s00425-011-1366-3

Grandbastien M-A (1998) Activation of plant retrotransposons under stress conditions. Trends Plant Sci 3:181–187. doi:10.1016/S1360-1385(98)01232-1

Greaves IK, Groszmann M, Ying H et al (2012) Trans chromosomal methylation in Arabidopsis hybrids. Proc Natl Acad Sci 109:3570–3575. doi:10.1073/pnas.1201043109

Greaves IK, Groszmann M, Wang A et al (2014) Inheritance of trans chromosomal methylation patterns from Arabidopsis F1 hybrids. Proc Natl Acad Sci 111:2017–2022. doi:10.1073/pnas.1323656111

Groszmann M, Greaves IK, Albert N et al (2011a) Epigenetics in plants--vernalisation and hybrid vigour. Biochim Biophys Acta BBA Gene Regul Mech 1809:427–437. doi:10.1016/j.bbagrm.2011.03.006

Groszmann M, Greaves IK, Albertyn ZI et al (2011b) Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. Proc Natl Acad Sci 108:2617–2622. doi:10.1073/pnas.1019217108

Groszmann M, Greaves IK, Fujimoto R et al (2013) The role of epigenetics in hybrid vigour. Trends Genet 29:684–690. doi:10.1016/j.tig.2013.07.004

Guerra-Peraza O, Nguyen HT, Stamp P, Leipner J (2009) ZmCOI6.1, a novel, alternatively spliced maize gene, whose transcript level changes under abiotic stress. Plant Sci 176:783–791. doi:10.1016/j.plantsci.2009.03.004

Hajkova P, Ancelin K, Waldmann T et al (2008) Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature 452:877–881. doi:10.1038/nature06714

Hansen KH, Helin K (2009) Epigenetic inheritance through self-recruitment of the polycomb repressive complex 2. Epigenetics Off J DNA Methy Soc 4:133–138

Hauben M, Haesendonckx B, Standaert E et al (2009) Energy use efficiency is characterized by an epigenetic component that can be directed through artificial selection to increase yield. Proc Natl Acad Sci 106:20109–20114. doi:10.1073/pnas.0908755106

Hauser M-T, Aufsatz W, Jonak C, Luschnig C (2011) Transgenerational epigenetic inheritance in plants. Biochim Biophys Acta BBA Gene Regul Mech 1809:459–468. doi:10.1016/j.bbagrm.2011.03.007

He G, Zhu X, Elling AA et al (2010) Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. Plant Cell 22:17–33. doi:10.1105/tpc.109.072041

Heard E, Martienssen R (2014) Transgenerational epigenetic inheritance: myths and mechanisms. Cell 157(1):95–109. doi:10.1016/j.cell.2014.02.045

Hemberger M, Dean W, Reik W (2009) Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. Nat Rev Mol Cell Biol 10:526–537. doi:10.1038/nrm2727

Henderson IR (2012) Control of meiotic recombination frequency in plant genomes. Curr Opin Plant Biol 15:556–561. doi:10.1016/j.pbi.2012.09.002

Henderson IR, Jacobsen SE (2007) Epigenetic inheritance in plants. Nature 447:418–424

Henikoff S, Comai L (1998) A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in Arabidopsis. Genetics 149:307–318

Henry, Nevo (2014) Exploring natural selection to guide breeding for agriculture. Plant Biotechnol J 12:655–662. doi:10.1111/pbi.12215

Herr AJ, Jensen MB, Dalmay T, Baulcombe DC (2005) RNA polymerase IV directs silencing of endogenous DNA. Science 308:118

Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res 19:1419–1428. doi:10.1101/gr.091678.109

Hollister JD, Smith LM, Guo Y-L et al (2011) Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. Proc Natl Acad Sci 108:2322–2327. doi:10.1073/pnas.1018222108

House M, Lukens L (2014) The role of germinally inherited epialleles in plant breeding. In: Epigenetics in plants of agronomic importance: fundamentals and applications. Springer International Publishing, Switzerland, pp 1–11

Hsieh T-F, Ibarra CA, Silva P et al (2009) Genome-wide demethylation of Arabidopsis endosperm. Science 324:1451–1454. doi:10.1126/science.1172417

Ibarra CA, Feng X, Schoft VK et al (2012) Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science 337:1360–1364. doi:10.1126/science.1224839

Ingouff M, Rademacher S, Holec S et al (2010) Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in Arabidopsis. Curr Biol 20:2137–2143. doi:10.1016/j.cub.2010.11.012

Iwasaki M, Paszkowski J (2014) Identification of genes preventing transgenerational transmission of stress-induced epigenetic states. Proc Natl Acad Sci. doi:10.1073/pnas.1402275111, 201402275

Jääskeläinen M, Chang W, Moisy C, Schulman AH (2013) Retrotransposon BARE displays strong tissue-specific differences in expression. New Phytol 200:1000–1008. doi:10.1111/nph.12470

Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature 416:556–560. doi:10.1038/nature731

Jacob Y, Martienssen RA (2011) Chromatin reprogramming: gender equality during Arabidopsis germline differentiation. Curr Biol 21:R20–R22. doi:10.1016/j.cub.2010.11.052

Jaligot E, Hooi WY, Debladis E et al (2014) DNA methylation and expression of the EgDEF1 gene and neighboring retrotransposons in mantled somaclonal variants of oil palm. PLoS One 9, e91896. doi:10.1371/journal.pone.0091896

Jiao Y, Meyerowitz EM (2010) Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. Mol Syst Biol 6:419. doi:10.1038/msb.2010.76

Johannes F, Porcher E, Teixeira FK et al (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. PLoS Genet 5, e1000530. doi:10.1371/journal.pgen.1000530

Jullien PE, Berger F (2010) Parental genome dosage imbalance deregulates imprinting in Arabidopsis. PLoS Genet 6, e1000885. doi:10.1371/journal.pgen.1000885

Kaeppler SM, Kaeppler HF, Rhee Y (2000) Epigenetic aspects of somaclonal variation in plants. Plant Mol Biol 43:179–188

Kakutani T (2002) Epi-alleles in plants: inheritance of epigenetic information over generations. Plant Cell Physiol 43:1106–1111. doi:10.1093/pcp/pcf131

Kalisz S, Purugganan MD (2004) Epialleles via DNA methylation: consequences for plant evolution. Trends Ecol Evol 19:309–314. doi:10.1016/j.tree.2004.03.034

Kanno T, Huettel B, Mette MF et al (2005) Atypical RNA polymerase subunits required for RNA-directed DNA methylation. Nat Genet 37:761–765. doi:10.1038/ng1580

Kasschau KD, Fahlgren N, Chapman EJ et al (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. PLoS Biol 5, e57. doi:10.1371/journal.pbio.0050057

Kato M, Miura A, Bender J et al (2003) Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. Curr Biol 13:421–426

Kenan-Eichler M, Leshkowitz D, Tal L et al (2011) Wheat hybridization and polyploidization results in deregulation of small RNAs. Genetics 188:263–272. doi:10.1534/genetics.111.128348

Khan A, Garbelli A, Grossi S et al (2014) The Arabidopsis STRESS RESPONSE SUPPRESSOR DEAD-box RNA helicases are nucleolar- and chromocenter-localized proteins that undergo stress-mediated relocalization and are involved in epigenetic gene silencing. Plant J. doi:10.1111/tpj.12533

Kim J-M, To TK, Ishida J et al (2012) Transition of chromatin status during the process of recovery from drought stress in *Arabidopsis Thaliana*. Plant Cell Physiol 53:847–856. doi:10.1093/pcp/pcs053

King GJ, Amoah S, Kurup S (2010) Exploring and exploiting epigenetic variation in crops. Genome Natl Res Counc Can Génome Cons Natl Rech Can 53:856–868. doi:10.1139/G10-059

Köhler C, Grossniklaus U (2002) Epigenetic inheritance of expression states in plant development: the role of polycomb group proteins. Curr Opin Cell Biol 14:773–779. doi:10.1016/S0955-0674(02)00394-0

Köhler C, Hennig L (2010) Regulation of cell identity by plant polycomb and trithorax group proteins. Curr Opin Genet Dev 20:541–547. doi:10.1016/j.gde.2010.04.015

Kou HP, Li Y, Song XX et al (2011) Heritable alteration in DNA methylation induced by nitrogen-deficiency stress accompanies enhanced tolerance by progenies to the stress in rice (*Oryza sativa L.*). J Plant Physiol 168:1685–1693. doi:10.1016/j.jplph.2011.03.017

Kulcheski F, de Oliveira L, Molina L et al (2011) Identification of novel soybean microRNAs involved in abiotic and biotic stresses. BMC Genomics 12:307. doi:10.1186/1471-2164-12-307

Kwon Y-J, Park M-J, Kim S-G et al (2014) Alternative splicing and nonsense-mediated decay of circadian clock genes under environmental stress conditions in Arabidopsis. BMC Plant Biol 14:136. doi:10.1186/1471-2229-14-136

Lafon-Placette C, Faivre-Rampant P, Delaunay A et al (2013) Methylome of DNase I sensitive chromatin in *Populus trichocarpa* shoot apical meristematic cells: a simplified approach revealing characteristics of gene-body DNA methylation in open chromatin state. New Phytol 197:416–430. doi:10.1111/nph.12026

Lang-Mladek C, Popova O, Kiok K et al (2010) Transgenerational inheritance and resetting of stress-induced loss of epigenetic gene silencing in Arabidopsis. Mol Plant 3:594–602. doi:10.1093/mp/ssq014

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11:204–220. doi:10.1038/nrg2719

Lee E, Lucas JR, Goodrich J, Sack FD (2014) Arabidopsis guard cell integrity involves the epigenetic stabilization of the FLP and FAMA transcription factor genes. Plant J. doi:10.1111/tpj.12516

Leviatan N, Alkan N, Leshkowitz D, Fluhr R (2013) Genome-wide survey of cold stress regulated alternative splicing in *Arabidopsis thaliana* with tiling microarray. PLoS One 8, e66511. doi:10.1371/journal.pone.0066511

Li X, Wang X, He K et al (2008) High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression. Plant Cell Online 20:259

Li J, Dudas B, Webster MA et al (2010) Hose in hose, an S locus–linked mutant of *Primula vulgaris*, is caused by an unstable mutation at the Globosa locus. Proc Natl Acad Sci 107:5664–5668. doi:10.1073/pnas.0910955107

Li W, Liu H, Cheng ZJ et al (2011) DNA methylation and histone modifications regulate de novo shoot regeneration in Arabidopsis by modulating WUSCHEL expression and auxin signaling. PLoS Genet 7, e1002243. doi:10.1371/journal.pgen.1002243

Lippman Z, May B, Yordan C et al (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS Biol 1, e67. doi:10.1371/journal.pbio.0000067

Lippman Z, Gendrel A-V, Black M et al (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430:471–476. doi:10.1038/nature02651

Lisch D, Bennetzen JL (2011) Transposable element origins of epigenetic gene regulation. Curr Opin Plant Biol 14:156–161. doi:10.1016/j.pbi.2011.01.003

Lister R, O'Malley RC, Tonti-Filippini J et al (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133:523–536. doi:10.1016/j.cell.2008.03.029

LoSchiavo F, Pitto L, Giuliano G et al (1989) DNA methylation of embryogenic carrot cell cultures and its variations as caused by mutation, differentiation, hormones and hypomethylating drugs. Theor Appl Genet 77:325–331. doi:10.1007/BF00305823

Luco RF, Pan Q, Tominaga K et al (2010) Regulation of alternative splicing by histone modifications. Science 327:996–1000. doi:10.1126/science.1184208

Luco RF, Allo M, Schor IE et al (2011) Epigenetics in alternative pre-mRNA splicing. Cell 144:16–26. doi:10.1016/j.cell.2010.11.056

Lukens L, Zhan S (2007) The plant genome's methylation status and response to stress: implications for plant improvement. Curr Opin Plant Biol 10:317–322. doi:10.1016/j.pbi.2007.04.012

Luo M, Wang Y-Y, Liu X et al (2012) HD2C interacts with HDA6 and is involved in ABA and salt stress response in Arabidopsis. J Exp Bot 63:3297–3306. doi:10.1093/jxb/ers059

Manning K, Tor M, Poole M et al (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nat Genet 38:948–952. doi:10.1038/ng1841

Marquez Y, Brown JWS, Simpson C et al (2012) Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. Genome Res 22:1184–1195. doi:10.1101/gr.134106.111

Martin A, Troadec C, Boualem A et al (2009) A transposon-induced epigenetic change leads to sex determination in melon. Nature 461:1135–1138. doi:10.1038/nature08498

Mastan SG, Rathore MS, Bhatt VD et al (2012) Assessment of changes in DNA methylation by methylation-sensitive amplification polymorphism in *Jatropha curcas L.* subjected to salinity stress. Gene 508:125–129. doi:10.1016/j.gene.2012.07.063

Mastrangelo AM, Marone D, Laidò G et al (2012) Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity. Plant Sci 185–186:40–49. doi:10.1016/j.plantsci.2011.09.006

Maumus F, Quesneville H (2014a) Deep investigation of *Arabidopsis thaliana* Junk DNA reveals a continuum between repetitive elements and genomic dark matter. PLoS One 9, e94101. doi:10.1371/journal.pone.0094101

Maumus F, Quesneville H (2014b) Ancestral repeats have shaped epigenome and genome composition for millions of years in *Arabidopsis thaliana*. Nat Commun. doi:10.1038/ncomms5104

Mazzucotelli E, Mastrangelo AM, Crosatti C et al (2008) Abiotic stress response in plants: when post-transcriptional and post-translational regulations control transcription. Plant Sci 174:420–431. doi:10.1016/j.plantsci.2008.02.005

McCue AD, Nuthikattu S, Reeder SH, Slotkin RK (2012) Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. PLoS Genet 8, e1002474. doi:10.1371/journal.pgen.1002474

McEachern L, Lloyd V (2012) The maize paramutation control region causes epigenetic silencing in *Drosophila melanogaster*. Mol Genet Genomics 287:591–606. doi:10.1007/s00438-012-0702-z

Messerschmidt DM, Knowles BB, Solter D (2014) DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes Dev 28:812–828. doi:10.1101/gad.234294.113

Miguel C, Marum L (2011) An epigenetic view of plant cells cultured in vitro: somaclonal variation and beyond. J Exp Bot 62:3713–3725. doi:10.1093/jxb/err155

Mirouze M, Paszkowski J (2011) Epigenetic contribution to stress adaptation in plants. Curr Opin Plant Biol 14:267–274. doi:10.1016/j.pbi.2011.03.004

Mirouze M, Reinders J, Bucher E et al (2009) Selective epigenetic control of retrotransposition in Arabidopsis. Nature 461:427–430. doi:10.1038/nature08328

Mirouze M, Lieberman-Lazarovich M, Aversano R et al (2012) Loss of DNA methylation affects the recombination landscape in Arabidopsis. Proc Natl Acad Sci 109:5880–5885. doi:10.1073/pnas.1120841109

Mittelsten Scheid O, Afsar K, Paszkowski J (2003) Formation of stable epialleles and their paramutation-like interaction in tetraploid *Arabidopsis thaliana*. Nat Genet 34:450–454. doi:10.1038/ng1210

Moghaddam AMB, Roudier F, Seifert M et al (2011) Additive inheritance of histone modifications in *Arabidopsis thaliana* intra-specific hybrids. Plant J 67:691–700. doi:10.1111/j.1365-313X.2011.04628.x

Molinier J, Ries G, Zipfel C, Hohn B (2006) Transgeneration memory of stress in plants. Nature 442:1046–1049. doi:10.1038/nature05022

Molnar A, Melnyk CW, Bassett A et al (2010) Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. Science 328:872–875. doi:10.1126/science.1187959

Monteuuis O, Doulbeau S, Verdeil J-L (2008) DNA methylation in different origin clonal offspring from a mature *Sequoiadendron giganteum* genotype. Trees 22:779–784. doi:10.1007/s00468-008-0238-3

Nicotra AB, Atkin OK, Bonser SP et al (2010) Plant phenotypic plasticity in a changing climate. Trends Plant Sci 15:684–692

Onodera Y, Haag JR, Ream T et al (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell 120:613–622. doi:10.1016/j.cell.2005.02.007

Ou X, Zhang Y, Xu C et al (2012) Transgenerational inheritance of modified DNA methylation patterns and enhanced tolerance induced by heavy metal stress in rice (*Oryza sativa* L.). PLoS One 7, e41143. doi:10.1371/journal.pone.0041143

Paszkowski J, Grossniklaus U (2011) Selected aspects of transgenerational epigenetic inheritance and resetting in plants. Curr Opin Plant Biol. doi:10.1016/j.pbi.2011.01.002

Pecinka A, Rosa M, Schikora A et al (2009) Transgenerational stress memory is not a general response in Arabidopsis. PLoS One 4, e5202. doi:10.1371/journal.pone.0005202

Pecinka A, Dinh HQ, Baubec T et al (2010) Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. Plant Cell 22:3118–3129. doi:10.1105/tpc.110.078493

Pecinka A, Fang W, Rehmsmeier M et al (2012) Polyploidization increases meiotic recombination frequency in Arabidopsis. BMC Biol 9:24. doi:10.1186/1741-7007-10-33

Petrillo E, Herz MAG, Fuchs A et al (2014) A chloroplast retrograde signal regulates nuclear alternative splicing. Science. doi:10.1126/science.1250322

Pikaard CS, Mittelsten-Scheid O (2014) Epigenetic regulation in plants. In: Allis D, Jenuwein T, Reinberg D (eds) Epigenetics. CSHL Press, USA, pp 81–100

Pillot M, Baroux C, Vazquez MA et al (2010) Embryo and endosperm inherit distinct chromatin and transcriptional states from the female gametes in Arabidopsis. Plant Cell 22:307–320. doi:10.1105/tpc.109.071647

Popova OV, Dinh H, Aufsatz W, Jonak C (2013) The RdDM pathway is required for basal heat tolerance in Arabidopsis. Mol Plant. doi:10.1093/mp/sst023

Qi X, Li ZH, Jiang LL et al (2010) Grain-yield heterosis in L. shows positive correlation with parental difference in CHG methylation. Crop Sci 50:2338. doi:10.2135/cropsci2010.01.0023

Quesada V, Macknight R, Dean C, Simpson GG (2003) Autoregulation of FCA pre-mRNA processing controls Arabidopsis flowering time. EMBO J 22:3142–3152. doi:10.1093/emboj/cdg305

Quesada V, Dean C, Simpson GG (2005) Regulated RNA processing in the control of Arabidopsis flowering. Int J Dev Biol 49:773–780. doi:10.1387/ijdb.051995vq

Rai MK, Kalia RK, Singh R et al (2011) Developing stress tolerant plants through in vitro selection-an overview of the recent progress. Environ Exp Bot 71:89–98. doi:10.1016/j.envexpbot.2010.10.021

Rapp RA, Wendel JF (2005) Epigenetics and plant evolution. New Phytol 168:81–91. doi:10.1111/j.1469-8137.2005.01491.x

Reddy ASN, Marquez Y, Kalyna M, Barta A (2013) Complexity of the alternative splicing landscape in plants. Plant Cell Online 25:3657–3683. doi:10.1105/tpc.113.117523

Regulski M, Lu Z, Kendall J et al (2013) The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. Genome Res 23:1651–1662. doi:10.1101/gr.153510.112

Reinders J, Wulff BBH, Mirouze M et al (2009) Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. Genes Dev 23:939–950. doi:10.1101/gad.524609

Richards EJ (2006) Inherited epigenetic variation – revisiting soft inheritance. Nat Rev Genet 7:395–401

Richards EJ (2008) Population epigenetics. Curr Opin Genet Dev 18:221–226. doi:10.1016/j.gde.2008.01.014

Richards EJ (2011) Natural epigenetic variation in plant species: a view from the field. Curr Opin Plant Biol 14:204–209. doi:10.1016/j.pbi.2011.03.009
Richards CL, Bossdorf O, Pigliucci M (2010) What role does heritable epigenetic variation play in phenotypic evolution? BioScience 60:232–237. doi:10.1525/bio.2010.60.3.9
Rival A, Jaligot E (2011) Epigenetics and plant breeding. CAB Rev Perspect Agric Vet Sci Nutr Nat Resour. doi:10.1079/PAVSNNR20116048
Rival A, Ilbert P, Labeyrie A et al (2013) Variations in genomic DNA methylation during the long-term in vitro proliferation of oil palm embryogenic suspension cultures. Plant Cell Rep 32:359–368. doi:10.1007/s00299-012-1369-y
Rivera C, Gurard-Levin ZA, Almouzni G, Loyola A (2014) Histone lysine methylation and chromatin replication. Biochim Biophys Acta BBA Gene Regul Mech. doi:10.1016/j.bbagrm.2014.03.009
Roudier F, Ahmed I, Berard C et al (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. EMBO J 30:1928–1938. doi:10.1038/emboj.2011.103
Rye M, Sandve GK, Daub CO et al (2014) Chromatin states reveal functional associations for globally defined transcription start sites in four human cell lines. BMC Genomics 15:120. doi:10.1186/1471-2164-15-120
Sang Y, Silva-Ortega CO, Wu S et al (2012) Mutations in two non-canonical Arabidopsis SWI2/SNF2 chromatin remodeling ATPases cause embryogenesis and stem cell maintenance defects. Plant J 72:1000–1014. doi:10.1111/tpj.12009
Saze H, Kakutani T (2011) Differentiation of epigenetic modifications between transposons and genes. Curr Opin Plant Biol 14:81–87. doi:10.1016/j.pbi.2010.08.017
Schlichting CD, Wund MA (2014) Phenotypic plasticity and epigenetic marking: an assessment of evidence for genetic accommodation. Evolution 68:656–672. doi:10.1111/evo.12348
Scoville AG, Barnett LL, Bodbyl-Roels S et al (2011) Differential regulation of a MYB transcription factor is correlated with transgenerational epigenetic inheritance of trichome density in *Mimulus guttatus*. New Phytol 191:251–263. doi:10.1111/j.1469-8137.2011.03656.x
Sekhon RS, Wang P-H, Sidorenko L et al (2012) Maize unstable factor for orange1 is required for maintaining silencing associated with paramutation at the pericarp color1 and booster1 Loci. PLoS Genet 8, e1002980. doi:10.1371/journal.pgen.1002980
Seo PJ, Hong S-Y, Ryu JY et al (2012) Targeted inactivation of transcription factors by overexpression of their truncated forms in plants. Plant J 72:162–172. doi:10.1111/j.1365-313X.2012.05069.x
Seo PJ, Park M-J, Park C-M (2013) Alternative splicing of transcription factors in plant responses to low temperature stress: mechanisms and functions. Planta 237:1415–1424. doi:10.1007/s00425-013-1882-4
Severing EI, van Dijk ADJ, Morabito G et al (2012) Predicting the impact of alternative splicing on plant MADS domain protein function. PLoS One 7, e30524. doi:10.1371/journal.pone.0030524
Sheldon CC, Burn JE, Perez PP et al (1999) The FLF MADS box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. Plant Cell 11:445–458. doi:10.1105/tpc.11.3.445
Sheldon CC, Conn AB, Dennis ES, Peacock WJ (2002) Different regulatory regions are required for the vernalization-induced repression of FLOWERING LOCUS C and for the epigenetic maintenance of repression. Plant Cell 14:2527–2537. doi:10.1105/tpc.004564
Shen Y, Zhou Z, Wang Z et al (2014) Global dissection of alternative splicing in paleopolyploid soybean. Plant Cell Online. doi:10.1105/tpc.114.122739
Shivaprasad PV, Dunn RM, Santos BA et al (2012) Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs. EMBO J 31:257–266. doi:10.1038/emboj.2011.458
Simpson GG, Dijkwel PP, Quesada V et al (2003) FY is an RNA 3' end-processing factor that interacts with FCA to control the Arabidopsis floral transition. Cell 113:777–787. doi:10.1016/S0092-8674(03)00425-2

Skinner MK (2011) Environmental epigenetic transgenerational inheritance and somatic epigenetic mitotic stability. Epigenetics 6:838–842. doi:10.4161/epi.6.7.16537
Slotkin RK, Vaughn M, Borges F et al (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136:461–472. doi:10.1016/j.cell.2008.12.038
Smulders M, de Klerk G (2011) Epigenetics in plant tissue culture. Plant Growth Regul 63:137–146. doi:10.1007/s10725-010-9531-4
Soppe WJ, Jacobsen SE, Alonso-Blanco C et al (2000) The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Mol Cell 6:791–802. doi:10.1016/S1097-2765(05)00090-0
Springer NM (2013) Epigenetics and crop improvement. Trends Genet 29:241–247. doi:10.1016/j.tig.2012.10.009
Staiger D, Brown JWS (2013) Alternative splicing at the intersection of biological timing, development, and stress responses. Plant Cell Online 25:3640–3656. doi:10.1105/tpc.113.113803
Stam M (2009) Paramutation: a heritable change in gene expression by allelic interactions in trans. Mol Plant 2:578–588. doi:10.1093/mp/ssp020
Stonaker JL, Lim JP, Erhard KF, Hollick JB (2009) Diversity of Pol IV function is defined by mutations at the maize rmr7 locus. PLoS Genet 5, e1000706. doi:10.1371/journal.pgen.1000706
Sun F, Guo W, Du J et al (2013) Widespread, abundant, and diverse TE-associated siRNAs in developing wheat grain. Gene 522:1–7. doi:10.1016/j.gene.2013.03.101
Sun X, Zhang Y, Zhu X et al (2014) Advances in identification and validation of plant microRNAs and their target genes. Physiol Plant. doi:10.1111/ppl.12191
Sung S, Amasino RM (2004) Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3. Nature 427:159–164. doi:10.1038/nature02195
Suter CM, Martin DIK (2010) Paramutation: the tip of an epigenetic iceberg? Trends Genet 26:9–14. doi:10.1016/j.tig.2009.11.003
Syed NH, Kalyna M, Marquez Y et al (2012) Alternative splicing in plants – coming of age. Trends Plant Sci 17:616–623. doi:10.1016/j.tplants.2012.06.001
Tang X, Lim M-H, Pelletier J et al (2012) Synergistic repression of the embryonic programme by SET DOMAIN GROUP 8 and EMBRYONIC FLOWER 2 in Arabidopsis seedlings. J Exp Bot 63:1391–1404. doi:10.1093/jxb/err383
Tarutani Y, Shiba H, Iwano M et al (2010) Trans-acting small RNA determines dominance relationships in Brassica self-incompatibility. Nature 466:983–986. doi:10.1038/nature09308
Teixeira FK, Heredia F, Sarazin A et al (2009) A role for RNAi in the selective correction of DNA methylation defects. Science 323:1600–1604. doi:10.1126/science.1165313
Tenaillon MI, Hollister JD, Gaut BS (2010) A triptych of the evolution of plant transposable elements. Trends Plant Sci 15:471–478. doi:10.1016/j.tplants.2010.05.003
Tilgner H, Nikolaou C, Althammer S et al (2009) Nucleosome positioning as a determinant of exon recognition. Nat Struct Mol Biol 16:996–1001. doi:10.1038/nsmb.1658
Tittel-Elmer M, Bucher E, Broger L et al (2010) Stress-induced activation of heterochromatic transcription. PLoS Genet 6, e1001175. doi:10.1371/journal.pgen.1001175
Tsuchiya T, Eulgem T (2013) An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication. Proc Natl Acad Sci 110:E3535–E3543. doi:10.1073/pnas.1312545110
Tsukahara S, Kobayashi A, Kawabe A et al (2009) Bursts of retrotransposition reproduced in Arabidopsis. Nature 461:423–426. doi:10.1038/nature08351
Turck F, Coupland G (2014) Natural variation in epigenetic gene regulation and its effects on plant developmental traits. Evolution 68(3):620–631. doi:10.1111/evo.12286
Us-Camas R, Rivera-Solís G, Duarte-Aké F, De-la-Peña C (2014) In vitro culture: an epigenetic challenge for plants. Plant Cell Tiss Organ Cult 118(2):187–201
Uthup TK, Ravindran M, Bini K, Thakurdas S (2011) Divergent DNA methylation patterns associated with abiotic stress in *Hevea brasiliensis*. Mol Plant 4:996–1013. doi:10.1093/mp/ssr039
Van Zanten M, Tessadori F, Peeters AJM, Fransz P (2012) Shedding light on large-scale chromatin reorganization in *Arabidopsis Thaliana*. Mol Plant 5:57–64. doi:10.1093/mp/sss030

Vendramin E, Pea G, Dondini L et al (2014) A unique mutation in a MYB gene cosegregates with the nectarine phenotype in peach. PLoS One 9, e90574. doi:10.1371/journal.pone.0090574
Verhoeven KJF, Preite V (2014) Epigenetic variation in asexually reproducing organisms. Evolution 68:644–655. doi:10.1111/evo.12320
Verhoeven KJF, van Gurp TP (2012) Transgenerational effects of stress exposure on offspring phenotypes in apomictic dandelion. PLoS One 7, e38605. doi:10.1371/journal.pone.0038605
Verhoeven KJF, Jansen JJ, van Dijk PJ, Biere A (2010) Stress-induced DNA methylation changes and their heritability in asexual dandelions. New Phytol 185:1108–1118. doi:10.1111/j.1469-8137.2009.03121.x
Vezzulli S, Leonardelli L, Malossini U et al (2012) Pinot blanc and Pinot gris arose as independent somatic mutations of Pinot noir. J Exp Bot 63:6359–6369. doi:10.1093/jxb/ers290
Vitulo N, Forcato C, Carpinelli EC et al (2014) A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. BMC Plant Biol 14:99. doi:10.1186/1471-2229-14-99
Wada Y, Miyamoto K, Kusano T, Sano H (2004) Association between up-regulation of stress-responsive genes and hypomethylation of genomic DNA in tobacco plants. Mol Genet Genomics. doi:10.1007/s00438-004-1018-4
Walker AR, Lee E, Bogs J et al (2007) White grapes arose through the mutation of two similar and adjacent regulatory genes. Plant J 49:772–785. doi:10.1111/j.1365-313X.2006.02997.x
Wang Q-M, Wang L (2012) An evolutionary view of plant tissue culture: somaclonal variation and selection. Plant Cell Rep 31:1535–1547. doi:10.1007/s00299-012-1281-5
Wang Y, Gu X, Yuan W et al (2014) Photoperiodic control of the floral transition through a distinct polycomb repressive complex. Dev Cell 28:727–736. doi:10.1016/j.devcel.2014.01.029
Weil C, Martienssen R (2008) Epigenetic interactions between transposons and genes: lessons from plants. Curr Opin Genet Dev 18:188–192. doi:10.1016/j.gde.2008.01.015
Wollmann H, Berger F (2012) Epigenetic reprogramming during plant reproduction and seed development. Curr Opin Plant Biol 15:63–69. doi:10.1016/j.pbi.2011.10.001
Woodhouse MR, Cheng F, Pires JC et al (2014) Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. Proc Natl Acad Sci 111:5283–5288. doi:10.1073/pnas.1402475111
Woodrow P, Pontecorvo G, Fantaccione S et al (2010) Polymorphism of a new Ty1-copia retrotransposon in durum wheat under salt and light stresses. Theor Appl Genet 121:311–322. doi:10.1007/s00122-010-1311-z
Xiao W, Custard KD, Brown RC et al (2006) DNA methylation is critical for Arabidopsis embryogenesis and seed viability. Plant Cell Online 18:805
Yaish MW, Colasanti J, Rothstein SJ (2011) The role of epigenetic processes in controlling flowering time in plants exposed to stress. J Exp Bot 62:3727–3735. doi:10.1093/jxb/err177
Yoshida S, Barbier de Reuille P, Lane B et al (2014) Genetic control of plant development by overriding a geometric division rule. Dev Cell 29:75–87. doi:10.1016/j.devcel.2014.02.002
Youngson NA, Whitelaw E (2008) Transgenerational epigenetic effects. Annu Rev Genomics Hum Genet 9:233–257. doi:10.1146/annurev.genom.9.081307.164445
Yu A, Lepère G, Jay F et al (2013) Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense. Proc Natl Acad Sci 110:2389–2394. doi:10.1073/pnas.1211757110
Zemach A, Kim MY, Silva P et al (2010a) Local DNA hypomethylation activates genes in rice endosperm. Proc Natl Acad Sci 107:18729–18734. doi:10.1073/pnas.1009695107
Zemach A, McDaniel IE, Silva P, Zilberman D (2010b) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328:916–919. doi:10.1126/science.1186366
Zhang X, Shiu S, Cal A, Borevitz JO (2008) Global analysis of genetic, epigenetic and transcriptional polymorphisms in *Arabidopsis thaliana* using whole genome tiling arrays. PLoS Genet 4, e1000032. doi:10.1371/journal.pgen.1000032
Zhang Y-Y, Fischer M, Colot V, Bossdorf O (2013) Epigenetic variation creates potential for evolution of plant phenotypic plasticity. New Phytol 197:314–322. doi:10.1111/nph.12010

Zhao L, Wang P, Yan S et al (2013) Promoter-associated histone acetylation is involved in the osmotic stress-induced transcriptional regulation of the maize ZmDREB2A gene. Physiol Plant. doi:10.1111/ppl.12136

Zilberman D, Gehring M, Tran RK et al (2006) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet 39:61–69. doi:10.1038/ng1929

Chapter 7
Progress and Perspectives of Distant Heterosis in Rice

Deming Jin and Tondi Yacouba Nassirou

Abstract Heterosis, or hybrid vigor, represents one of the greatest contributions of genetics to the improvement of major crops such as maize and rice, which is intensively studied and exploited by breeders and seed companies worldwide. The exploitation of distant heterosis is a promising way to further raise the yield potential of crops. This review describes aspects of the genetic basis for heterosis, diversification of rice and its related species, principles and methodologies for exploiting distant heterosis in hybrid rice breeding. Both promises and constraints are discussed for understanding the importance and difficulties of distant heterosis utilization. Progress of distant heterosis is demonstrated by the intersubspecific heterosis between two subspecies of Asian rice (*Oryza sativa* L.), ssp. *indica* and ssp. *japonica*, and the interspecific heterosis between two cultivated species, Asian rice and African rice (*O. glaberrima* Steud.). Strategies of using the doubled haploid method and molecular genetics approaches are suggested for accelerating distant heterosis breeding, eliminating sterility loci to overcome the hybrid sterility of distant crosses, and identifying and pyramiding QTLs of distant heterotic loci and favorable genes to develop a more typical intersubspecific hybrid rice with higher ratio of *indica/japonica* heterozygotic loci, and to create a more adaptive and productive partial interspecific hybrid rice incorporating *sativa/glaberrima* heterozygotic loci and favorable genes from the two cultivated species.

Keywords Doubled haploid • Genetic diversity • Heterosis breeding • Heterotic loci • Hybrid rice • Interspecific heterosis • Intersubspecific heterosis • *Oryza sativa* • *Oryza glaberrima* • QTL

D. Jin (✉) • T.Y. Nassirou
MOA Key Laboratory of Crop Ecophysiology and Farming System in the Middle Reaches of the Yangtze River, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China
e-mail: djin@mail.hzau.edu.cn; nassiroutondi@yahoo.fr

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_7

7.1 Introduction

Hybrid vigor is widespread in plants and animals, and this phenomenon has been observed and utilized purposely or unconsciously ever since the domestication of crops and animals. Plant hybrid vigor was first reported by Kölreuter (1761) who emphasized two aspects of the phenomenon, that hybrid vigor was related to the dissimilarity of parents and that nature favored outcrossing. Darwin (1876) showed that cross-pollination was generally beneficial while self-pollination was disadvantageous in plant evolution.

Crop heterosis breeding started with cross-pollinated maize. Shull (1908) established the concept of heterosis from his study of a maize field, discovering inbreeding depression and hybrid vigor in crosses between inbred lines. He also designed a procedure to obtain maize single hybrids. However, seed production cost of single hybrids was too high for commercial production at the time. Jones (1918) designed the procedure of producing double hybrids from two single hybrids to lower seed costs, which made possible the successful commercial production of first-generation hybrid maize. Relatively easy mechanical or hand emasculation for this diclinous crop facilitated the early success of hybrid maize production.

Heterosis in rice was first reported by Jones (1926) who observed marked increase in tiller number and grain yield in some F1 hybrids in comparison to their parents. Since then, hybrid vigor has been reported for various agronomic traits such as yield, grain weight, spikelets per panicle, panicles per plant, plant height, days to flower, etc. (Cheng et al. 2007; Jaikishan et al. 2010; Jones et al. 1998; Virmani 1994; Virmani et al. 1981). However, hybrid seed production via artificial emasculation for this self-pollinated monoclinous crop was much more difficult than that of the diclinous maize. Therefore, genetic tools such as cytoplasmic male sterility (CMS) were essential for developing commercially hybrid rice.

Katsuo and Mizushima (1958) reported the first attempt to develop CMS rice. The first complete CMS system derived from an intersubspecific cross between an *indica* cv. Chinsurah Boro II and a *japonica* cv. Taichung 65 was developed in 1966. This *japonica* CMS line is gametophytic male sterile. The F1 hybrid has 50 % fertile pollen but almost normal fertile seed setting (Shinjyo 1975). The first *japonica* hybrid rice, however, showed limited heterosis because its two parental lines were derived from the same original cross and thus had limited genetic diversity.

The first successful commercial hybrid rice, released in China in 1973, was an *indica* hybrid variety Nanyou 2. Its CMS line was derived from an interspecific cross between a pollen abortive common wild rice plant and a Chinese local variety Erjiu Nan and its restorer IR24 was released by IRRI (Yuan 2002). Hybrid rice currently occupies more than 16 million ha, making up more than 50 % of total rice area in China (Li et al. 2007), and it is also being planted commercially in several South and Southeast Asian countries (Sheeba et al. 2009). Hybrid rice produced via both a three-line system based on CMS and a two-line system dependent on photoperiod and/or temperature-sensitive genetic male sterility (Jin et al. 1987) is cultivated in China, although the three-line system is still the major method for hybrid

seed production in other countries. Large-scale production of hybrid rice worldwide is still mainly confined to *indica* rice which shows relatively wider genetic diversity in comparison to *japonica* rice.

The magnitude of heterosis of a hybrid mainly depends on the genetic diversity between its parents. Therefore, exploiting heterosis of distant crosses is a promising way for further raising yield potential of crops. Attempts at exploiting distant heterosis have been made in rice heterosis breeding; encouraging progress has been achieved in both intersubspecific heterosis between two subspecies of Asian rice or common rice (*Oryza sativa* L.), ssp. *indica* and ssp. *japonica*, and interspecific heterosis between two cultivated species, Asian rice and African rice (*O. glaberrima* Steud.). This chapter covers aspects of the genetic basis for heterosis, diversification of rice and its related species, potential and constraints of distant heterosis utilization in rice, principles and methodologies for hybrid rice breeding to exploit distant heterosis, and strategies for more effective approach to distant heterosis using doubled haploid and molecular technologies to solve key problems such as hybrid sterility as well as pyramiding QTLs of heterotic loci and other favorable genes.

7.2 Heterosis and Genetic Diversity

Heterosis, or hybrid vigor, refers to the phenomenon that a hybrid organism often exhibits increased size, growth rate and productivity over those of its parents. The fundamental ideas of heterosis (Shull 1908) include: (1) deleterious recessive alleles persist in large random mating populations; (2) inbreeding reduces vigor due to increasing homozygosity of deleterious alleles and (3) vigor is restored by crossing divergent inbred lines as recessive deleterious alleles are complemented in the hybrid. Crop breeders exploit heterosis by mating two divergent inbred lines that have complementary desirable traits. The F1 hybrids generally show heterosis with desired characteristics of both parents.

Classical genetic theories of heterosis are divided into three categories: dominance hypothesis (Bruce 1910; Davenport 1908): overdominance hypothesis (Crow 1948; Hull 1945) and epistatic hypothesis (Powers 1944; Williams 1959). The dominance hypothesis suggests the complementary effects between different desirable dominant alleles from the two parents result in the superiority of hybrids, whereas the overdominance hypothesis focuses on the strong interaction effects between different allelic genes from the parents. The epistatic hypothesis, however, emphasizes the role of non-allelic gene interactions in hybrids. These three classical hypotheses of heterosis have been analyzed by new molecular technologies such as molecular markers and QTL mapping. Molecular evidences supporting the three hypotheses respectively have been found by different researchers (Fu et al. 2014; Hua et al. 2003; Larièpe et al. 2012; Melchinger et al. 2007; Piepho 2009; Schön et al. 2010; Schrag et al. 2009).

Emerging genomic and epigenetic studies have provided new insights into the genetic basis of heterosis (Chen 2013; Groszmann et al. 2013). Heterosis may arise

from allelic and non-allelic interactions between parental genomes, thus changing the programming of genes that promote the growth, development, stress tolerance and fitness of hybrids. Epigenetic modifications of key regulatory genes in hybrids can alter complex regulatory networks of physiology and metabolism, leading to the expression of heterosis in complex traits such as grain yield.

Genetic diversity between parents is the primary cause of heterosis in the hybrid. Breeders have extensively investigated the genetic diversity between parents and its relationship with heterosis of the hybrid for predicting heterotic cross combinations. Most researchers found positive correlations between the genetic diversity and heterosis, i.e. the greater the genetic distances between two parents, the higher the possibility to obtain hybrids with strong heterosis. For example, genetic distances analysis revealed by both traditional methods based on agronomic-morphological traits and molecular markers were positively correlated with heterosis in maize (Larièpe et al. 2012) and rice (Kaw 1995; Luo et al. 1996, 1999; Mahapatra et al. 1995; Phetmanyseng et al. 2010; Saghai et al. 1997; Xu et al. 2002). However, the correlation between the genetic distance and heterosis depends on the materials and traits investigated (Zhou et al. 2012).

Although genetic diversity between parents is often emphasized to be positively correlated to heterosis of hybrids, it is also correlated to hybrid sterility (Moehring 2011). Reproductive isolation can occur when the genetic distance between parents surpasses a critical point. Negative heterosis of traits related to sexual reproduction such as seed set rate and grain yield may happen when crossing two distantly related parents. Reduced vigor or outbreeding depression is sometimes observed in distant crosses. For example, the hybrid sterility problem remains a crucial factor that negatively affects grain yield in distant crosses of rice, resulting in low seed set rate in intersubspecific crosses between *indica* and *japonica* rice (Li et al. 1997; Long et al. 2008), and almost zero seed setting in interspecific crosses between Asian and African rice (Adedze et al. 2012; Heuer and Miezan 2003; Koide et al. 2008; Sano 1986).

7.3 The Diversification of Rice

The genus *Oryza* has 23 species with 10 recognized genome types (AA, BB, CC, BBCC, CCDD, EE, FF, GG, HHJJ and HHKK), including 21 wild species and 2 domesticated species (Gramene.org). The two cultivated species, Asian rice (*O. sativa*) and African rice (*O. glaberrima*), both have the AA genome but they were independently domesticated from different progenitors of wild species in different geographic locations, respectively (Fig. 7.1).

The Asian rice was domesticated from the Asian common wild rice (*Oryza rufipogon*) approximately 10,000 years ago in Southern Asia and most probably in southern China (Jiang and Liu 2006; Molina et al. 2011; Sweeney and McCouch 2007). Asian rice contains two major subspecies: ssp. *indica* and ssp. *japonica*. The *indica* rice is grown throughout tropical and subtropical Asia. The *japonica* rice,

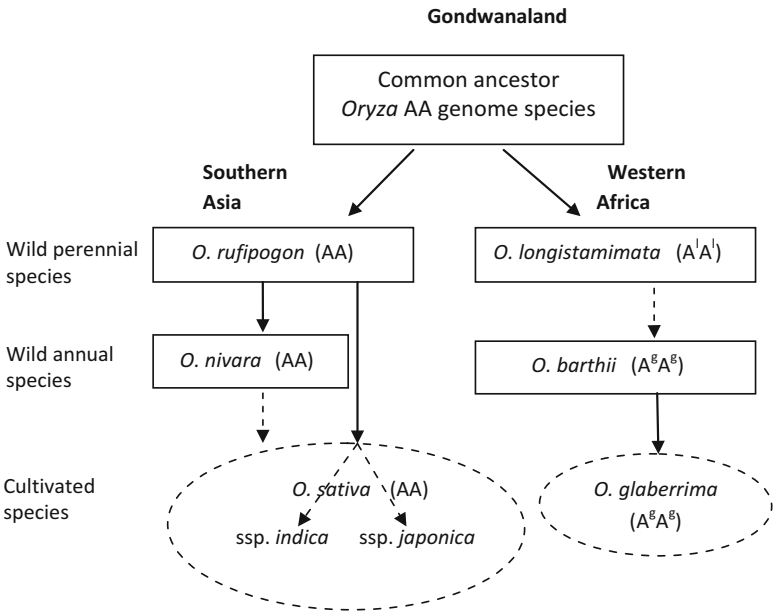

Fig. 7.1 The evolution and domestication of two cultivated rice species

which is more tolerant to low temperatures, is mainly cultivated in temperate East Asia or in high elevation areas in South and Southeast Asia. Previous archaeological and genetic marker studies suggest that *indica* was domesticated from *O. rufipogon* growing in the lowland regions of Southern China; *japonica* rice was later developed in upland regions and selected from *indica* rice (Oka and Morishima 1982). A molecular study demonstrated that the *japonica* subspecies suffered a more severe bottleneck than the *indica* subspecies and thus a greater loss of genetic variation during its domestication, supporting the non-independent domestication of the two rice subspecies (Gao and Innan 2008). However, some molecular evidence suggests the two subspecies might have been domesticated from different *O. rufipogon* progenitors (Cheng et al. 2003; Li et al. 2012; Londo and Schaal 2007; Wei et al. 2012). Although the domestication history of Asian rice remains a controversial issue (Ikehashi 2009), it has been introduced successfully to Europe, Africa, America and Australia and it is now the major cultivated species worldwide.

African rice derived from the wild species *Oryza barthii* originated in West Africa around 1000 BC (Agnoun et al. 2012; Linares 2002; Murray 2004; Portères 1962; Wang et al. 2014). Its productivity is much lower in comparison to Asian rice, but it is well adapted to certain extreme environments and possesses traits resistant or tolerant to biotic and abiotic stresses including drought, iron toxicity, virus disease and weed competitiveness etc. (Barry et al. 2007; Efisue et al. 2009a; Gutierrez et al. 2010; Hiroko et al. 1962; Wang et al. 2014). African rice continues to be cultivated sporadically in West Africa.

The common ancestor AA-genome species of the two cultivated rice species probably occurred in Gondwanaland, and their respective progenitors diverged between one and two million years ago after the breakup of the ancient continent (Ma and Bennetzen 2004; Vaughan et al. 2008). Recent comparative genomic analyses on Asian rice and the other five diploid AA-genome species including African rice revealed rapid diversification of AA genomes with massive levels of genomic structural variation, including segmental duplication and gene family turnover, with particularly high instability in defense-related genes (Zhang et al. 2014).

7.4 Promises and Constraints of Distant Heterosis in Rice

The rich genetic diversification of Asian rice and closely related AA-genome species has provided opportunities for exploiting distant heterosis in rice. Three categories of heterosis in rice have been defined based on the genetic diversity between parental lines:

Intra-subspecific heterosis is resulted from the interaction of different genes from two parental varieties belonging to the same subspecies, either *indica* or *japonica* rice.

Intersubspecific heterosis is resulted from the interaction of different genes between the two subspecies, *indica* and *japonica* rice, respectively.

Interspecific heterosis is resulted from the interaction of different genes from two species respectively, such as Asian rice vs. African rice, or Asian rice vs. a wild rice species.

Since the heterosis of hybrids is generally positively correlated to the genetic diversity of the parents, the heterosis of distant crosses including intersubspecific and interspecific crosses is theoretically higher than that of the intra-subspecific crosses (Fig. 7.2).

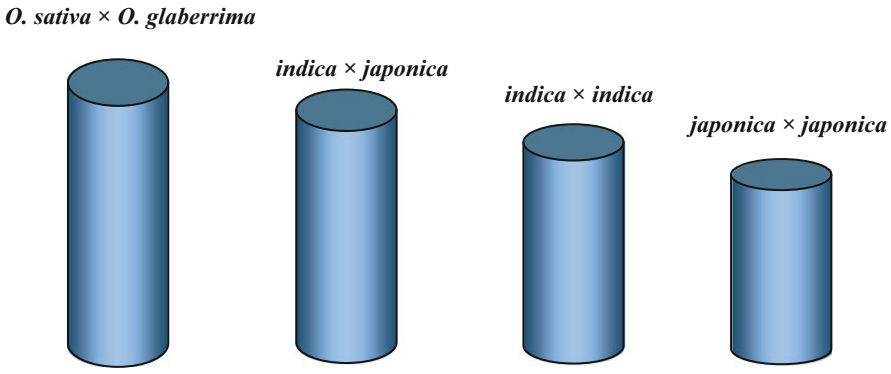

Fig. 7.2 Theoretic heterosis expression of intra-subspecific hybrids of *japonica* or *indica* rice intersubspecific hybrids between *indica* and *japonica* rice and interspecific hybrids between Asian and African rice as estimated according to their genetic diversity

While the observed vegetative vigor of the hybrids of different categories of genetic diversity generally conforms to the theoretical expectations, the actual grain yield of distant crosses, however, is discouraging due to the low seed set rate of the intersubspecific hybrids between *indica* and *japonica* rice and almost no seed setting in the interspecific hybrids between two cultivated species. Apparently, hybrid sterility is a major obstacle for exploiting the distant heterosis in rice.

Reproductive isolation mechanisms restrict gene exchange between diverging species or populations include prezygotic barriers that limit the potential for mating or zygote formation and postzygotic barriers that reduce the fertility of hybrid offspring, which plays a vital role during speciation in evolution (Andrea and Willis 2012). The hybrid sterility of distant crosses is a typical form of postzygotic barrier. The Dobzhansky-Muller model (Dobzhansky 1937; Muller 1942) explains that hybrid sterility can be caused by incompatible gene interactions between two diverging species or populations (Fig. 7.3).

Ikehashi et al. (1984, 1986) studied hybrid sterility between *indica* and *japonica* rice and screened for compatibility varieties to overcome the intersubspecific reproductive barrier and proposed a model which is essentially consistent with the Dobzhansky-Muller model to explain the intersubspecific hybrid incompatibility. The two incompatibility alleles designated as S5i and S5j for *indica* and *japonica*, respectively, and deleterious interactions between S5i and S5j resulted in reduced spikelet fertility of the intersubspecific hybrid. They also reported an additional allelic gene S5n, which could overcome the fertility barrier in the *indica*/*japonica* hybrids and thus it is named a wide-compatible (WC) gene. Molecular studies have revealed the precise location of the S5n gene on chromosome 6 (Ji et al. 2005) and

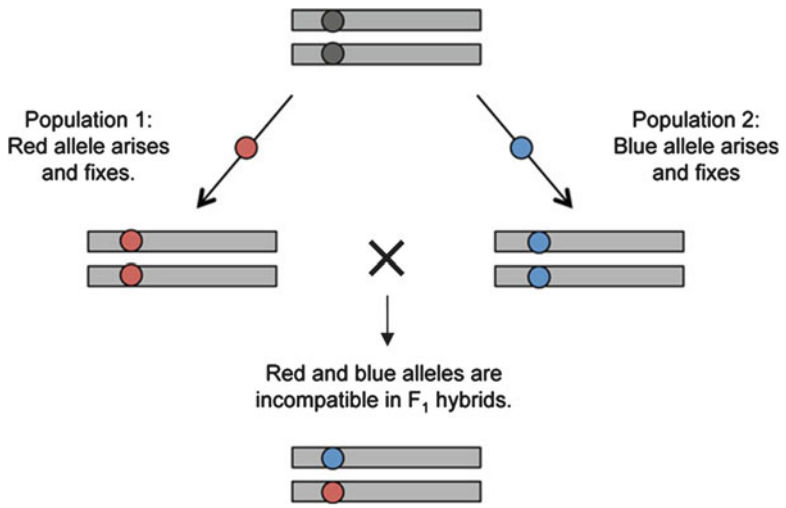

Fig. 7.3 The Dobzhansky-Muller model for a single-locus hybrid incompatibility: an ancestral population splits into two geographically isolated populations that diverge genetically and eventually fix different alleles (*red* or *blue*) at the same locus. In the F1 hybrid, these two derived alleles are incompatible (Source: Adopted from Andrea and Willis (2012))

its genetic mechanism and complex evolution (Du et al. 2011). Further study revealed a killer-protector system at the S5 locus encoded by three tightly linked genes (ORF3, ORF4 and ORF5) regulates fertility in *indica-japonica* hybrids. The combined actions of ORF5^{+} (killer) and ORF4^{+} (partner) cause endoplasmic reticulum (ER) stress. ORF3^{+} (protector) prevents ER stress and produces normal gametes, but ORF3^{-} cannot prevent ER stress, resulting in premature programmed cell death that leads to embryosac abortion (Yang et al. 2012). The discovery of wide-compatible gene S5n greatly encouraged research on intersubspecific heterosis utilization.

A more intensive and complex reproductive isolation including both prezygotic and postzygotic barriers exists in the crosses between *O. sativa* and *O. glaberrima* (Sano et al. 1986). Several major genes responsible for the interspecific hybrid sterility between the two species have been detected from *O. glaberrima*, such as S1 (Guyot et al. 2011; Heuer et al. 2003; Koide et al. 2008), S29 (Hu et al. 2006) and S36 (Li et al. 2011). The interspecific hybrid sterility was mainly caused by an arrest of pollen development at the microspore stage (Bimpong et al. 2011). However, S1 could induce abortion of both male and female gametes possessing its allelic alternative (Koide et al. 2008). No effective wide-compatibility gene has been reported to overcome the interspecific reproductive isolation up to present. The hybrid sterility problem remains a major concern for exploiting the interspecific heterosis between the two cultivated species (Abebrese et al. 2011; Adedze et al. 2012).

7.5 Intersubspecific Heterosis between *indica* and *japonica* Rice

Spontaneous intersubspecific hybrid plants from open pollination between *indica* and *japonica* rice growing in adjacent paddy fields in previous seasons are often seen in regions growing both subspecies. They are called *big green plants* by rice farmers because of their prominent features of longer growth duration, taller and more robust stems, big panicles with fewer filled grains and usually staying green when all other rice plants are turning yellow at maturity in the paddy field. It is apparent that positive heterosis exists in both yield related *source* traits such as leaf area and biomass, and *sink* traits such as spikelets per panicle. Most rice breeders are convinced that there is strong heterosis in the intersubspecific hybrids between *indica* and *japonica* rice (Yang and Liu 1991; Yuan et al. 1994). However, the apparent negative heterosis in grain yield is caused by spikelet sterility which blocks the photosynthetic products from the *source* flow into the *sink*.

Because the postzygotic reproductive barrier of hybrid sterility constitutes a bottleneck for exploiting the intersubspecific heterosis, the genetic mechanism of intersubspecific hybrid sterility has been extensively studied. Both embryosac abortion and pollen sterility are found to be associated with the intersubspecific hybrid sterility. Sterility gene S5 (Ikehashi and Araki 1986), S24 and S35 (Qiu et al. 2005)

have been identified to cause intersubspecific hybrid sterility between *indica* and *japonica* rice. The identification of more than one sterility gene explained why the wide compatibility genes S5n alone could not always effectively overcome the reproductive barrier between the two subspecies of rice.

Remarkable progress has been achieved in the intersubspecific heterosis breeding in China. Chinese rice breeders have adopted a more effective approach to overcome the reproductive barrier by producing partial intersubspecific hybrids, rather than simply depending on the wide compatible gene. For example, testing crosses between *indica* varieties and chromosome segment substitution lines (CSSL) carrying *japonica* genes in the genetic background of *indica* proved to be effective for improving yield potential of hybrid rice (Wang et al. 2012). The super rice program sponsored by the Chinese Ministry of Agriculture has announced a few super high yielding hybrid varieties that incorporated intersubspecific heterosis (Table 7.1), including 4 *indica-japonica* hybrid varieties, i.e. Yongyou6, Yongyou12 and Yongyou15 released by the Ningbo Academy of Agricultural Sciences, and Chunyou84 released by the China National Rice Research Institute (Fig. 7.4c).

Hybrid varieties with a relatively small proportion of intersubspecific heterozygotic loci are usually recognized as intra-subspecific hybrid varieties. For example, a super high yielding hybrid variety Xieyou 9308 carried about 12.5 % *japonica* genes as estimated according to the pedigree of its male parent Zhonghui9308, which was derived from an intersubspecific multi-cross C57(*japonica*)/No300(*indica*)//IR26(*indica*) (Cheng et al. 2007). The initial female parent C57 itself was a *japonica* restorer also derived from an intersubspecific multi-cross IR8(*indica*)/Taida1(*japonica*)//Nonglin34(*japonica*). Xieyou 9308 was regarded as an *indica* hybrid because of its relatively low *japonica* genomic composition. Similarly, almost all three-line *japonica* hybrids contained a small fraction of *indica* genomic genes because the restorer genes of their pollen parents came from *indica* rice.

Recently released *indica-japonica* hybrids Yongyou12 and Yongyou15 have a common restorer line F5032 derived from an intersubspecific cross carrying up to 49 % of *indica* genes as evaluated by molecular markers. Although the two half-brother varieties have different female parents, a *japonica* CMS line Yongjing2A and an *indica* CMS line JingshuangA, respectively, both are named as intersubspecific *indica-japonica* hybrids (Liu 2012; Lu et al. 2007) because they have relatively higher percentages of intersubspecific heterozygous loci.

The new *indica-japonica* hybrid rice is now very popular and has a total area of over a million hectares in China. The hybrid variety Yongyou12 scored a record high yield of 15,214.5 kg/ha in 2012. Still, most so-called *indica-japonica* hybrid varieties released in China up to present are actually partial intersubspecific hybrids with one of their parents derived from *indica*/*japonica* crosses (Table 7.1) instead of a direct intersubspecific hybrids between a typical *indica* rice and a typical *japonica* rice with wide compatibility genes to overcome its hybrid sterility, which is still the ultimate goal.

The two subspecies of rice possess different favorable traits respectively. For example, the *indica* rice generally has more tillers and is more resistant to heat

Table 7.1 Hybrid rice varieties incorporated intersubspecific heterosis released in China

Hybrid variety	Parental combination	Subspecies/source of the parents		Year released
		Female	Male	
[1]Xieyou 9308	Xieqingzao A/R9308	*indica* CMS	*indica* restorer derived from *indica/japonica*	1995
[2]Jinyou207	Jin23A/Xianhui207	*indica* CMS	*indica* restorer derived from *indica/japonica*	1998
[3]Liangyou Peijiu	Pei'ai 64 S/R9311	EGMS derived from *indica/japonica*	*indica* restorer	2001
[4]Liaoyou 5218	Liao 5216A/C418	*japonica* CMS	*japonica* restorer derived from *indica/japonica*	2001
[5]II you 602	II-32A/Luhui 602	*indica* CMS	*indica* restorer derived from *indica/japonica*	2002
[6]II you 7954	II-32A/Zhehui 7954	*indica* CMS	*indica* restorer derived from *indica/japonica*	2002
[1]Guodao 1	Zhong 9A/R8006	*indica* CMS	*indica* restorer derived from *indica/japonica*	2004
[1]Guodao 3	Zhong 8A/R8006	*indica* CMS	*indica* restorer derived from *indica/japonica*	2004
[1]Guodao 6	Neixiang 2A/R8006	*indica* CMS	*indica* restorer derived from *indica/japonica*	2006
[7]Yongyou6	Yongjing2A/K4806	*japonica* CMS	Intermediate restorer derived from *indica/japonica*	2005
[7]Yongyou12	Yongjing2A/F5032	*japonica* CMS	Intermediate restorer derived from *indica/japonica*	2010
[7]Yongyou15	JingshuangA/F5032	*indica* CMS	Intermediate restorer derived from *indica/japonica*	2012
[1]Chunyou84	Chunjiang16A/C84	*japonica* CMS	Intermediate restorer derived from *indica/japonica*	2013

Note: Hybrid varieties were released by
1. The China National Rice Research Institute
2. Hunan Hybrid Rice Research Center
3. The Academy of Agricultural Sciences of Jiangsu Province
4. The Academy of Agricultural Sciences of Liaoning Province
5. The Academy of Agricultural Sciences of Sichuan Province
6. The Academy of Agricultural Sciences of Zhejiang Province
7. The Ningbo Academy of Agricultural Sciences

stress, while the *japonica* rice has more erect and greener leaves and is more tolerant to cold stress. A successful intersubspecific hybrid could have combined favorable traits of both species in addition to strong positive heterosis in yield related agronomic traits (Fig. 7.4).

Fig. 7.4 The development of intersubspecific hybrid rice. (**a**) Panicles of *indica* rice at maturity. (**b**) Panicles of *japonica* rice at maturity. (**c**) Field performance of a partial intersubspecific hybrid rice (Chunyou84, released by the China National Rice Research Institute in 2013), showing an intermediate form of panicle type between *indica* and *japonica* rice (Credit: Image courtesy of the China National Rice Research Institute)

7.6 Interspecific Heterosis Between Asian and African Rice

Written records of African rice (*Oryza glaberrima*) began when the first Portuguese explorers reached the West African coast and witnessed the cultivation of rice in the floodplains and marshes of the Upper Guinea Coast in 1446. African rice was believed to have been domesticated and cultivated by the local inhabitants many centuries before the first Europeans arrived and this cereal played a very important role in the native diet. Asian rice (*O. sativa*) was introduced into West Africa by the Europeans beginning in the sixteenth century, the species spread and was adopted by local peoples who had previous experience growing the African rice species. African rice (*O. glaberrima*) varieties have certain negative features in comparison to the Asian rice (*O. sativa*): the seeds shatter easily, the grains are brittle and difficult to mill, and more importantly, the yields are lower. As a result, African rice *O. glaberrima* has been replaced gradually by the introduced Asian rice (*O. sativa*) as the major cultivated species (Agnoun et al. 2012; Linares 2002; Murray 2004).

The *Oryza glaberrima* varieties also offer distinct advantages: the plants have luxurious wide leaves that shade out weeds, and the species is more resistant to diseases and pests. In addition, African rice is more tolerant of fluctuations in water depth, iron toxicity, infertile soils, severe climates and human neglect. Some *O. glaberrima* varieties also mature faster than *O. sativa* varieties, making them important as emergency food. Some African rice (*O. glaberrima*) varieties have survived for one more reason: the supreme deity in ritual contexts of the traditional religion. It was believed that the rain god gave Diola rice (*O. glaberrima*) to their ancestors (Linares 2002).

African rice (*O. glaberrima*) is a valuable germplasm attractive to rice breeders. It is a unique germplasm with distinct advantages as compared with the wild species of genus *Oryza*. First of all, it is a cultivated species which has undergone about 3000 years of artificial selection, which conserved relatively favorable agronomic traits. Secondly, it has a relatively appropriate genetic distance to Asian rice (*O. sativa*), more distant than the AA-genome common wild rice (*O. rufipogon*) but closer than many other wild species with different genome types, which make it an ideal counterpart for exploiting interspecific heterosis. Last but not least, it possesses rich defense-related genes resistant to various abiotic and biotic stresses, which may facilitate development of more adaptive rice. However, the intensive reproductive isolation including both prezygotic and postzygotic barriers in the interspecific crosses between *O. sativa* and *O. glaberrima* makes rice genetic improvement by crossing the two species extremely difficult (Ikeda et al. 2009).

The NERICA rice varieties derived from interspecific crosses between the two cultivated rice species via in vitro embryo rescue and anther culture techniques in the African Rice Center (WARDA 2000, 2006; africarice.org) has provided a successful example for developing elite inbred varieties with higher productivity and stronger stress tolerance by combining the favorable genes of both species (Aluko 2003; Bimpong et al. 2010; Efisue et al. 2009a, b; He et al. 2010; Ikeda et al. 2009; Ndjiondjop et al. 2012; Semagn et al. 2007). The NERICA rice has demonstrated

the possibility of a favorable recombination of beneficial genes of the two cultivated rice species to create a new type of inbred rice. Is it possible to take one more step forward, to develop a new type of hybrid rice incorporating the interspecific heterosis between the two cultivated rice?

Although almost all direct interspecific hybrids are completely sterile, it could be possible to develop fertile partial interspecific hybrid (PIH) rice by crossing Asian rice male sterile lines to introgression line restorers (ILR) carrying African rice genes. Efforts have been made to develop PIH varieties with higher yield potential in Huazhong Agricultural University since 2004 (Jin et al. 2012). New indica CMS lines with *Oryza glaberrima* cytoplasm (AF-CMS) have been developed (Huang et al. 2012). The AF-CMS lines were completely sterile whereas the fertility of its F1 hybrids was readily restored. The development of AF-CMS system has made available a third species, *O. glaberrima*, in addition to *O. rufipogon* and *O. sativa*, as the cytoplasmic source of CMS for commercial hybrid rice production, which should further help diversify the sterile cytoplasms for the three-line hybrid rice. The *indica* ILRs carrying African rice genes and effective restorer genes (Table 7.2) have also been obtained (Adedze et al. 2012). Some ILRs show certain traits which are intermediate between the two cultivated species *O. sativa* and *O. glaberrima*, such as the middle-sized sterile lemmas on the grains (Fig. 7.5b). Three-line PIHs or two-line PIHs can be produced by crossing CMS lines or PTGMS lines to the ILRs as pollen parents, respectively. A few PIH rice showed outstanding high yield potential (>20 % higher than control varieties of *indica* hybrid, unpublished data) in field experiments (Fig. 7.5c).

Developing introgression line restorers with an appropriate genome composition is crucial to mitigate reproductive barrier and exploit interspecific heterosis between *Oryza glaberrima* and *O. sativa*. A moderate introgression of *O. glaberrima* genes (15–30 %) into ILRs was considered to be favorable for producing high yielding

Table 7.2 Pedigree of introgression line restorer (ILR) carrying genes of *Oryza glaberrima*

ILR	Pedigree
ILR1	RAM3/Paddy//Paddy///Mianhui725
ILR2	RAM152/Paddy//Paddy///Mianhui725
ILR3	RAM3/Jin23B//Jin23B///YuetaiB
ILR4	RAM3/Jin23B//Jin23B///Jin23B
ILR5	RAM3/Jin23B//Wanxian98
ILR6	RAM3/Jin23B//Nongxiang16/// Minghui63
ILR7	RAM152/Paddy//Paddy///Jiachunnian
ILR8	RAM54/Jin23B//Wanxian98///R80
ILR9	RAM3/Jin23B//Nongxiang16// Minghui63////Basmati370

Note: RAM3, RAM54 and RAM152 are accessions of African rice (*O. glaberrima*). All the other varietal names are *indica* varieties (*O. sativa* ssp. *indica*)

Fig. 7.5 The development of partial interspecific hybrid rice. (**a**) Panicles of Asian rice (*left*), African rice (*right*) and interspecific F1 (the two in *middle*) at maturity. The panicle of Asian rice is much bigger than that of the African rice. And the interspecific F1 is completely sterile with long red awns Lingshui, China 2006. (**b**) Grains of Asian rice (*right*), African rice (*left*) and the introgression lines (ILs) (*middle*). African rice grain has two prominent sterile lemmas at two sides; Asian rice grain has tiny degenerated sterile lemmas at bottom, while the size of the sterile lemmas on the grains of some ILs falls in between the two cultivated rice. (**c**) Field performance of a partial interspecific hybrid rice (AF-ST1A/ILR1) (Huazhong Agricultural University, Wuhan, China 2014). The female parental line AF-ST1A is an *indica* CMS line with African rice cytoplasm while the male parent ILR1 is an *indica* introgression line carrying genes of African rice

PIHs. Further improvement of the genomic composition of ILs and the yield performance of the PIHs could be achieved via marker-assisted selection for pyramiding interspecific heterotic QTL loci and eliminating sterility genes.

7.7 Strategies for More Effective Approach to Distant Heterosis

Conventional plant breeding has dramatically increased the productivity and quality of plants grown for food, fodder and industrial use; it has been practiced for hundreds of years, and is still widely used today. The basic methods involve crossing two different parents, followed by pedigree selection dependent on phenotyping through generations of self-pollination, until a set of derived lines that combines the favorable characteristics of both parents is obtained. This traditional methodology, however, is labor intensive, time consuming and rather inefficient. It usually takes 8–10 years or more from start to finish for developing a new variety in annual crops.

This situation can be changed with the advent and dissemination of new technologies such as doubled haploid (DH) methodology and molecular genetics. For example, conventional breeding procedures commonly take eight or nine generations to achieve approximately complete homozygosity inbred lines; more generations are needed to obtain genetically-stable inbred lines derived from distant crosses, whereas DH lines achieve it in just one generation regardless of the types of crosses. Precise genotyping by marker assisted selection (MAS) is much more efficient and labor-saving than traditional visual phenotyping.

Doubled haploid (DH) lines can be produced by androgenesis or in vitro anther culture and parthenogenesis or in vivo induction of maternal haploids. The DH lines based on in vivo techniques have been routinely applied in commercial hybrid maize. The major advantages of DH lines in hybrid maize breeding include the complete homozygosity and maximum genetic variance, as well as short *time to market* etc. (Geiger and Gordillo 2009). However, the DH produced by in vitro anther culture is the more common method used in rice breeding (Alemanno and Guiderdoni 1994; Reiffers and Freire 1990; Yi et al. 2014). Anther culture response or androgenetic potential is largely species and genotype dependent. DH lines are more readily induced via anther culture of *Oryza glaberrima* as well as *O. sativa* ssp. *japonica* genotypes, while *O. sativa* ssp. *indica* genotypes are more recalcitrant (Gueye and Ndir 2010; Reiffers and Freire 1990). DH plants can be produced via spontaneous chromosome doubling of the haploid cells of the callus induced from the microspore, while artificial chromosome doubling with colchicine treatment effectively increases doubled haploid plant regeneration in anther culture of rice (Alemanno and Guiderdoni 1994). Before a more effective in vivo technique is set up, DH lines produced by anther culture can be used to accelerate distant heterosis breeding in rice.

Molecular genetics has provided powerful tools to study the genetic basis of heterosis and made monitoring the transmission of foreign genes into introgression lines possible. Although we are still at the beginning of understanding the complex mechanism of heterosis, advances of molecular genetics, especially genomics, will facilitate the investigation and the understanding of heterotic alleles involved in distant crosses. QTL mapping of heterotic loci of yield related traits is a particularly important foundation work for heterosis breeding in rice (Wang et al. 2013). Fine

mapping of the molecular markers closely linked to distant heterotic loci and other favorable genes as well as sterility genes is essential for pyramiding distant heterotic alleles and favorable genes and eliminating the sterility genes in elite ILs to be used as parents of hybrids incorporated distant heterosis.

The success of *indica-japonica* hybrid rice in China has demonstrated the potential of intersubspecific heterosis. However, these hybrids currently planted in China are partial intersubspecific hybrids with relatively limited *indica/japonica* heterozygotic loci. Developing more typical intersubspecific hybrids with higher ratio of *indica/japonica* heterotic loci and employing more wide compatibility genes is a promising approach to realize the full potential of the intersubspecific heterosis.

The two cultivated rice species *Oryza glaberrima* and *O. sativa* constitute a potential genetic reservoir for breeding hybrid rice incorporated interspecific heterosis. However, a phenomenon known as the reproductive barrier constitutes a bottleneck for exploiting heterosis in this more distant cross. Strategies have to be developed to overcome or circumvent the crossing barriers and F1 hybrid sterility. Introgression lines containing a short alien chromosome fragment could be developed. Since each introgression line has only a fragment of alien chromosome, it can be effectively monitored to eliminate sterility genes. Technologies such as embryo rescue, anther culture and molecular markers can be employed for more effective approach to the interspecific heterosis between the two cultivated rice.

A pragmatic technical route is proposed for effective distant heterosis breeding. Doubled haploid lines via anther culture are used to speed up the development of genetically stable ILs to be used as parental lines of hybrids incorporated distant heterosis. Marker assisted selection is used for pyramiding distant heterotic alleles and favorable genes and eliminating the sterility genes to obtain elite ILs which are used as pollen parents to cross with *indica/japonica* CMS lines or PTGMS lines to produce experimental three line or two line hybrids incorporated distant heterosis (Fig. 7.6).

7.8 Conclusions and Prospects

The ultimate goal of all rice breeding programs is to continuously develop varieties with higher yield potential, superior grain quality and increased tolerance to abiotic and biotic stresses. Exploiting distant heterosis provides a promising approach to this goal.

Remarkable progress has been achieved in the intersubspecific heterosis breeding in China. The intersubspecific *indica-japonica* hybrid rice has already been grown successfully in large hectarage in China, demonstrating super high yield as well as combined favorable traits of both subspecies.

Significant progress has also been achieved in the interspecific heterosis between two cultivated rice. The partial interspecific *sativa-glaberrima* hybrid rice grown in experimental fields shows super high-yielding potential as well as combined favorable traits of both species.

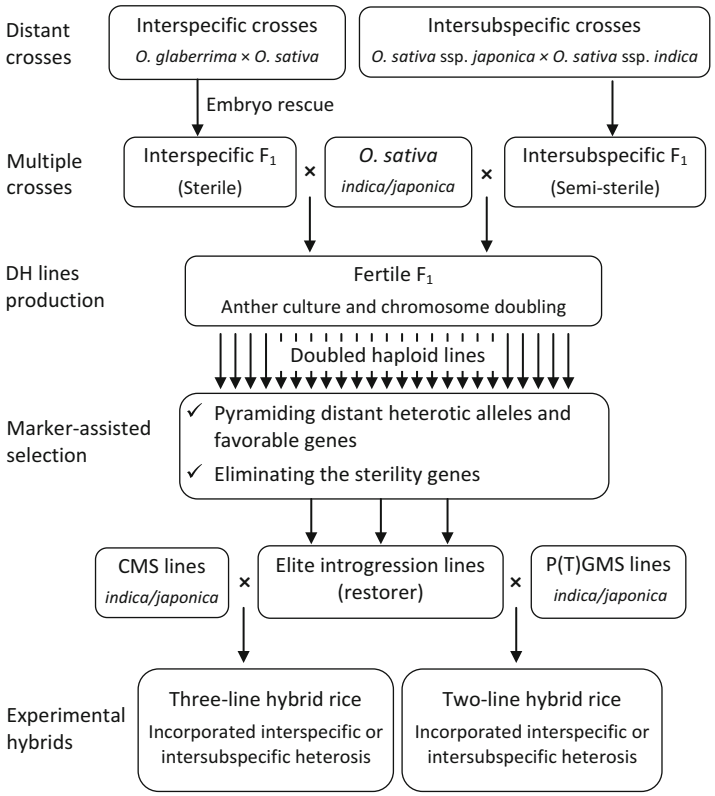

Fig. 7.6 Technical route proposed for effective distant heterosis breeding

Future prospects may be dependent on the more effective use of new technologies such as doubled haploid (DH) methodology and molecular marker technology. For example, the employment of DH may greatly accelerate the breeding for elite ILRs carrying foreign genes for exploiting distant heterosis. Screening and pyramiding intersubspecific and interspecific heterozygous loci could enhance the understanding of the mechanisms of distant heterosis and facilitate its utilization in hybrid rice breeding. Identification and elimination of major sterility genes could help to overcome the hybrid sterility of distant crosses. The combination between parents carrying more heterotic loci and exempting major sterility genes may lead to the development of a more typical intersubspecific hybrid rice with a higher ratio of *indica-japonica* heterozygotic loci and the creation of a more adaptive and productive partial interspecific *sativa-glaberrima* hybrid rice.

Acknowledgements The study of interspecific heterosis was funded by the Natural Science Foundation of China (NSFC: 30871703), the National High Technology Research and Development Program (863 Program) of China (2012AA101101), the Natural Science Foundation of Hubei province (2010CBB01901), and the Fundamental Research Funds for the Central Universities (2009PY025).

References

Abebrese SO, Akromah R, Dartey PKA (2011) Crossability of selected progeny from interspecific crosses between *Oryza sativa* and *Oryza glaberrima* (NERICAs). Afr J Agric Res 6(1):79–83

Adedze Y, Efisue A, Zhang S et al (2012) Identification of interspecific grain yield heterosis between two cultivated rice species *Oryza sativa* L. and *Oryza glaberrima* Steud. Aust J Crop Sci 6(11):1558–1564

Africarice. http://www.africarice.org/. Accessed 8 Jan 2015

Agnoun Y, Biaou S, Sié M, Djedatin G et al (2012) The African rice Oryza glaberrima Steud: knowledge distribution and prospects. Int J Biol 4(3):158–180

Alemanno L, Guiderdoni E (1994) Increased doubled haploid plant regeneration from rice (*Oryza sativa* L.) anthers cultured on colchicine-supplemented media. Plant Cell Rep 13:432–436

Aluko GK (2003) Genetic mapping of agronomic traits from the interspecific cross of *Oryza sativa* (L.) and *Oryza glaberrima* (Steud). PhD dissertation, Louisiana State University and Agricultural and Mechanical College

Andrea LS, Willis JH (2012) Molecular evolution and genetics of postzygotic reproductive isolation in plants. Biol Rep 4:23. doi:10.3410/B4-23

Barry MB, Pham JL, Noyer JL et al (2007) Genetic diversity of the two cultivated rice species (*O. sativa & O. glaberrima*) in Maritime Guinea. Euphytica 154(1–2):127–137

Bimpong IK, Carpena AK, Mendioro MS et al (2010) Evaluation of *Oryza sativa* x *Oryza glaberrima* derived progenies for resistance to rootknot nematode and identification of introgressed alien chromosome segments using SSR markers. Afr J Biotechnol 9(26):3988–3997

Bimpong IK, Chin JH, Ramos J, Koh HJ (2011) Application of subspecies-specific marker system identified from *Oryza sativa* to *Oryza glaberrima* accessions and *Oryza sativa* x *Oryza glaberrima* F1 interspecific progenies. Int J Genet Mol Biol 3(1):7–24

Bruce AB (1910) The Mendelian theory of heredity and the augmentation of vigor. Science 32(827):627–628

Chen ZJ (2013) Genomic and epigenetic insights into the molecular bases of heterosis. Nat Rev Genet 14:471–482

Cheng C, Reiko M, Suguru T et al (2003) Polyphyletic origin of cultivated rice: Based on the interspersion pattern of SINEs. Mol Biol Evol 20:67–75

Cheng SH, Li YC, Jie YZ et al (2007) Super hybrid rice breeding in China: achievements and prospects. J Integr Plant Biol 49(6):805–810

Crow JF (1948) Alternative hypotheses of hybrid vigor. Genetics 33(5):477–487

Darwin CR (1876) The effects of cross and self fertilisation in the vegetable kingdom. John Murray, London

Davenport CB (1908) Degeneration, albinism and inbreeding. Science 28(718):454–455

Dobzhansky TH (1937) Genetics and the origin of species. Columbia University Press, New York

Du H, Ouyang Y, Zhang C, Zhang Q (2011) Complex evolution of S5, a major reproductive barrier regulator, in the cultivated rice *Oryza sativa* and its wild relatives. New Phytol 191(1):275–287

Efisue AA, Tongoona P, Derera J (2009a) Screening early-generation progenies of interspecific rice genotypes for drought-stress tolerance during vegetative phase. J Crop Improv 23(2):174–193

Efisue AA, Tongoona P, Derera J et al (2009b) Genetics of morpho-physiological traits in segregating populations of interspecific hybrid rice under stress and non-stress conditions. J Crop Improv 23(4):383–401

Fu DH, Xiao ML, Hayward A et al (2014) Utilization of crop heterosis: a review. Euphytica 197:161–173

Gao LZ, Innan H (2008) Nonindependent domestication of the two rice subspecies, *Oryza sativa* ssp. *indica* and ssp. *Japonica*, demonstrated by multilocus microsatellites. Genetics 179(2):965–976

Geiger HH, Gordillo GA (2009) Doubled haploids in hybrid maize breeding. Maydica 54:485–499

Gramene. http://archive.gramene.org/species/oryza/rice_intro.html. Accessed 8 Jan 2015

Groszmann M, Greaves IK, Fujimoto R et al (2013) The role of epigenetics in hybrid vigour. Trends Genet 29(12):684–690

Gueye T, Ndir KN (2010) *In vitro* production of double haploid plants from two rice species (*Oryza sativa* L. and *Oryza glaberrima* Steud.) for the rapid development of new breeding material. Sci Res Essay 5(7):709–713

Gutierrez AG, Silvio JC, Olga XG et al (2010) Identification of a rice stripe necrosis virus resistance locus and yield component QTLs using *Oryza sativa* X *O. glaberrima* introgression lines. Plant Bio 10:6. doi: 10.1186/1471-2229-10-6

Guyot R, Garavito A, Gavory F et al (2011) Patterns of sequence divergence and evolution of the S1 orthologous regions between Asian and African cultivated rice species. PLoS ONE 6(3). doi:10.1371/journal.pone.001772

Heuer S, Miezan KM (2003) Assessing hybrid sterility in *Oryza glaberrima* x *O. sativa* hybrid progenies by PCR marker analysis and crossing with wide compatibility varieties. Theor Appl Genet 107(5):902–909

He YX, Zheng TQ, Hao XB et al (2010) Yield performances of japonica introgression lines selected for drought tolerance in a BC breeding program. Plant Breed 129:167–175

Hiroko M, Hinata K, Oka HI (1962) Comparison between two cultivated rice species *Oryza sativa* L. and *Oryza glaberrima* Steud. Jpn J Breed 12:153–165, English with Japanese summary

Hu FY, Xu P, Deng XN et al (2006) Molecular mapping of a pollen killer gene S29 (t) in *Oryza glaberrima* and co-linear analysis with S22 *in O. glumaepatula*. Euphytica 151(3):273–278

Hua JP, Xing YZ, Wu WR et al (2003) Single-locus heterotic effects and dominance-by-dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. Proc Nat Acad Sci 100(5):2574–2579

Huang F, Fu X, Efisue A et al (2012) Genetically characterizing a new *indica* cytoplasmic male sterility with *Oryza glaberrima* cytoplasm for its potential use in hybrid rice production. Crop Sci 53:132–140

Hull FH (1945) Recurrent selection for specific combining ability in corn. J Am Soc Agron 37(2):134–145

Ikeda R, Yoshimi S, Inouss A (2009) Seed fertility of F1 hybrids between upland rice NERICA cultivars and *Oryza sativa* L. or *O. glaberrima* Steud. Breed Sci 59:27–35

Ikehashi H (2009) Why are there indica type and japonica type in rice?--history of the studies and a view for origin of two types. Rice Sci 16(1):1–13

Ikehashi H, Araki H (1984) Varietal screening of compatibility types revealed in F1 fertility of distant crosses in rice. Jpn J Breed 34:304–313

Ikehashi H, Araki H (1986) Genetics of F1 sterility in remote crosses of rice. Rice Genet (International Rice Research Institute, Manila), pp 119–130

Jaikishan I, Rajendrakumar P, Ramesha MS et al (2010) Prediction of heterosis for grain yield in rice using key informative EST-SSR. Plant Breed 129(1):108–111

Ji Q, Lu J, Chao Q et al (2005) Delimiting a rice wide-compatibility gene S5n to a 50 kb region. Theor Appl Genet 111(8):1495–1503

Jiang L, Liu L (2006) New evidence for the origins of sedentism and rice domestication in the lower Yangzi River, China. Antiquity 80(308):355–361

Jin DM, Li ZB, Wan JM (1987) The use and value of photoperiod-sensitive gene in rice breeding. Sci Agric Sin 20(1):6–12

Jin DM, Efisue AA, Zhang SS et al (2012) Developing a new indica CMS system and introgression restorer lines via interspecific crosses between *Oryza glaberrima* Steud and *Oryza sativa* L. Abstract presented at the International Conference on Utilization of Heterosis in Crops, Xi'an, pp 264–265

Jones DF (1918) The effects of inbreeding and crossbreeding on development. PNAS 4(8):246–250

Jones JW (1926) Hybrid vigour in rice. J Am Soc Agr 18(5):423–428
Jones MP, Dingkuhn M, Johnson DE, Sow A (1998) Growth and yield potential of *Oryza sativa* and *O. glaberrima* upland rice cultivars and their interspecific progenies. Field Crop Res 57(1):57–69
Kölreuter JG (1761) Vorläufige nachricht von einigen, das geschlecht der pflanzen betreffenden versuchen und beobachtungen. In der gleditschischen handlung, Leipzig
Kaw RN (1995) Analysis of divergence in some cold tolerant rices. Ind J Genet 55(1):84–89
Katsuo K, Mizushima U (1958) Studies on the cytoplasmic difference among rice varieties, *Oryza sativa* L. 1. On the fertility of hybrids obtained reciprocally between cultivated and wild varieties. Jpn J Breed 8(1):1–5 (in Japanese with English summary)
Koide YA, Onishi K, Nishimoto D (2008) Sex-independent transmission ratio distortion system responsible for reproductive barriers between Asian and African rice species. New Phytol 179(3):888–900
Larièpe A, Mangin B, Jasson S et al (2012) The genetic basis of the heterosis: multiparental quantitative traits loci mapping reveals contrasted levels of apparent overdominance among traits of agronomical interest in maize (*Zea mays* L.). Genetics l90(2):795–811
Li Z, Pinson SRM, Paterson AH et al (1997) Genetics of hybrid sterility and hybrid breakdown in an intersubspecific rice (*Oryza sativa* L.) population. Genetics 145(4):1139–1148
Li SQ, Yang DC, Zhu YG (2007) Characterization and use of male sterility in hybrid rice breeding. J Integr Plant Biol 49(6):791–804
Li F, Liu FH, Morinaga DZ (2011) A new gene for hybrid sterility from a cross between *Oryza sativa* and *O. glaberrima*. Plant Breed 130(2):165–171
Li WJ, Zhang B, Huang GW et al (2012) Chloroplast DNA polymorphism and evolutional relationships between Asian cultivated rice (*Oryza sativa*) and its wild relatives (*O. rufipogon*). Genet Mol Res 11(4):4418–4431
Linares OF (2002) African rice (*Oryza glaberrima*): history and future potential. Proc Nat Acad Sci 99(25):16360–16365
Liu WM (2012) Correlation and path analysis of yield traits on 'Yongyou12' *indica-japonica* intersubspecific hybrid rice. Chin J Agric 2:1–4
Londo JP, Schaal BA (2007) Origins and population genetics of weedy red rice in the USA. Mol Ecol 16:4523–4535. doi:10.1111/j.1365-294X.2007.03489.x
Long Y, Zhao L, Niu B et al (2008) Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. Proc Nat Acad Sci USA 105(48):18871–18876
Lu YF, Ma RR, Wang XY et al (2007) SSLP—based SSR fingerprinting and *indica*/*Japonica* classification of yongyou series hybrid rice. Chin J Rice Sci 21(4):443–446
Luo XJ, Li Yuan J, McCouch SR (1996) Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR-based markers. Theor Appl Genet 92:637–643
Luo LJ, Mei H, Yu X et al (1999) Yield heterosis performance and their parental genetic diversity in rice. Rice Sci 13(1):6–10
Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Nat Acad Sci USA 101(34):12404–12410
Mahapatra KC, Mishra CHP, Acharya B (1995) Clustering of rice mutants by different methods of analysis. Ind J Genet 55(2):138–147
Melchinger A, Utz HF, Piepho HP et al (2007) The role of epistasis in the manifestation of heterosis-a systems-oriented approach. Genetics 177(3):1815–1825
Moehring AJ (2011) Heterozygosity and its unexpected correlations with hybrid sterility. Evolution 65(9):2621–2630
Molina J, Sikora M, Garud N et al (2011) Molecular evidence for a single evolutionary origin of domesticated rice. Proc Nat Acad Sci USA 108(20):8351–8356
Muller HJ (1942) Isolating mechanisms, evolution, and temperature. Biol Sym 6:71–125
Murray SS (2004) Searching for the origins of African rice domestication. Antiquity 78(300):1–3

Ndjiondjop MN, Futakuchi K, Seck PA et al (2012) Morpho-agronomic evaluation of *Oryza glaberrima* accessions and interspecific *Oryza sativa* X *Oryza glaberrima* derived lines under drought conditions. Afr J Agric Res 7(16):2527–2538

Oka HI, Morishima H (1982) Phylogenetic differentiation of cultivated rice, XXIII. Potentiality of wild progenitors to evolve the *indica* and *japonica* types of rice cultivars. Euphytica 31(1):41–50

Phetmanyseng X, Xie F, Hernandez JE, Boirromeo TH (2010) Hybrid rice heterosis and genetic diversity of IRRI and Lao rice. Field Crop Res 117(1):18–23

Piepho HP (2009) Ridge regression and extensions for genome wide selection in maize. Crop Sci 49(4):1165–1176

Portères R (1962) Primary cradles of agriculture in the African continent. J Afr Hist 3:195–210

Powers L (1944) An expansion of Jones's theory for the explanation of heterosis. Am Nat 78:275–280

Qiu SQ, Liu K, Jiang JX et al (2005) Delimitation of the rice wide compatibility gene S5n to a 40-kb DNA fragment. Theor Appl Genet 111(6):1080–1086

Reiffers I, Freire AB (1990) Production of doubled haploid rice plants (*Oryza sativa* L.) by anther culture. Plant Cell Tiss Org Cult 21(2):165–170

Saghai MMA, Yang GP, Qifa Z, Gravois KA (1997) Correlation between molecular marker distance and hybrid performance in US Southern long grain rice. Crop Sci 37(1):145–150

Sano Y (1986) Sterility barriers between Oryza sativa and O. glaberrima. International Rice Research Institute-Rice Genet, IRRI, Manila, pp 109–118

Schön CC, Dhillon BS, Utz HF, Melchinger AE (2010) High congruency of QTL positions for heterosis of grain yield in three crosses of maize. Theor Appl Genet 120(2):321–332

Schrag TA, Frisch M, Dhillon BS, Melchinger AE (2009) Marker-based prediction of hybrid performance in maize single-crosses involving doubled haploids. Maydica 54(2/3):353–362

Semagn K, Ndjiondjop MN, Lorieux M et al (2007) Molecular profiling of an interspecific rice population derived from a cross between WAB 56–104 (*Oryza sativa*) and CG 14 (*Oryza glaberrima*). Afr J Biol 6(17):2014–2022

Sheeba NK, Viraktamath BC et al (2009) Validation of molecular markers linked to fertility restorer gene(s) for WA-CMS lines of rice. Euphytica 167(2):217–227

Shinjyo C (1975) Genetical studies of cytoplasmic male sterility and fertility restoration in rice, *Oryza sativa* L. Sci Bull Coll Agr Univ Ryukyus 22:1–57

Shull GH (1908) The composition of a field of maize. Am Breed Ass Rep 4(1):296–301

Sweeney M, McCouch S (2007) The complex history of the domestication of rice. Ann Bot 100:951–957

Vaughan DA, Lu BR, Tomooka N (2008) The evolving story of rice evolution. Plant Sci 174(4):394–408

Virmani SS (1994) Heterosis and hybrid rice breeding. Springer, Berlin

Virmani SS, Chaudhary RC, Khush GS (1981) Current outlook on hybrid rice. Oryza 18:67–84

Wang MH, Yu YS, Georg H et al (2014) The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. Nat Genet 46:982–988

Wang Z, Yu C, Liu X et al (2012) Identification of *indica* rice chromosome segments for the improvement of *japonica* inbreds and hybrids. Theor Appl Genet 124(7):1351–1364

Wang ZQ, Jiang L, Yin CB et al (2013) QTL mapping of heterotic loci of yield related traits in rice. Chin J Rice Sci 27(6):569–576

West Africa Rice Development Association (2000) Potential for a green revolution in rice in west and central Africa. Report of the second biennial, M'bé, Bouaké, Côte d'Ivoire. (Now renamed Africa Rice Center)

West Africa Rice Development Association (2006) Report of the 5th biennial regional consultative meeting of the national experts committee (NEC V) Cotonou, Republic of Benin, 19–20 Jun 2006 (Now renamed Africa Rice Center)

Wei X, Qiao WH, Chen YT et al (2012) Domestication and geographic origin of *Oryza sativa* in China: insights from multilocus analysis of nucleotide variation of *O. sativa* and *O. rufipogon*. Mol Ecol 21(20):5073–5087

Williams W (1959) Heterosis and the genetics of complex characters. Nature 184:527–530

Xu W, Virmani SS, Hernandez JE, Sebastian LS (2002) Genetic diversity in parental line and heterosis of the tropical rice hybrids. Euphytica 127(1):139–148

Yang ZY, Liu WY (1991) Classification of the F1 intersubspecific hybrids between *Indica* and *Japonica* and its relation to the F1 hybrid vigor. Chin J Rice Sci 5(4):151–156

Yang JY, Zhao XB, Cheng K et al (2012) A killer-protector system regulates both hybrid sterility and segregation distortion in rice. Science 337(6100):1336–1340

Yi G, Lee HS, Kim KM (2014) Improved marker-assisted selection efficiency of multi-resistance in doubled haploid rice plants. Euphytica. doi:10.1007/s10681-014-1303-1

Yuan LP (2002) Hybrid rice. China Agr Press, Beijing, China

Yuan LP, Yang ZY, Yang JB (1994) Hybrid rice in China. In: Virmani SS (ed) Hybrid rice technology: new development and future prospects. International Rice Research Institute, Manila, Philippines, pp 143–147, P.O. Box 933

Zhang QJ, Ting Z, Xia EH et al (2014) Rapid diversification of five Oryza AA genomes associated with rice adaptation. Proc Nat Acad Sci USA 111(46):E4954–E4962

Zhou G, Chen Y, Yao W et al (2012) Genetic composition of yield heterosis in an elite rice hybrid. Proc Nat Acad Sci USA 109(39):15847–15852

Chapter 8
Forward and Reverse Genetics in Crop Breeding

Joanna Jankowicz-Cieslak and Bradley J. Till

Abstract The ability to identify, study, create and utilize genetic variation are arguably fundamental aspects in the development of human beings. Traditional, or *forward*, genetics has its roots in man's first attempts at the domestication of plants and animals. Novel and preferred traits such as the lack of seed shattering were selected from large populations and maintained for future propagation, leading to the first cultivars. The discovery of DNA as the heritable material thousands of years later enabled the development of *reverse* genetics whereby targeted lesions in the genome are recovered to test and utilize functional variation in genes. A major contributor to both forward and reverse genetics was the discovery in the early twentieth century that mutations can be induced in genomes at frequencies in several orders of magnitude higher than typically observable in nature. The ability to produce novel variation has fueled the development of thousands of new crop cultivars. Examples exist of increased disease resistance, higher yields, tolerance to abiotic stresses such as drought and salinity and improved nutritional quality. In an era where global food security is threatened by climate change and variation and growing populations, use of induced mutations is an important method in the breeder's toolbox. This review describes the use of induced mutations for forward and reverse genetics in plants, with a focus on crops. Different mutagens and random versus targeted approaches are described. Additionally, newly emerging methods and technologies are discussed that promise to advance basic and applied plant sciences.

Keywords Chemical mutagen • Forward genetics • Induced mutations • Mutant variety • Mutation breeding • Physical mutagen • Reverse genetics • TILLING

J. Jankowicz-Cieslak (✉) • B.J. Till
Plant Breeding and Genetics Laboratory, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, IAEA Laboratories Seibersdorf, International Atomic Energy Agency, Vienna International Centre, PO Box 100, Vienna 1400, Austria
e-mail: j.jankowicz@iaea.org; b.till@iaea.org

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_8

8.1 Introduction

Spontaneous natural mutation provides the genetic variation that underpins heritable biological diversity. While early man unknowingly utilized genetics to domesticate the first crops, it was Gregor Mendel who formalized what was to become the discipline of genetics and provided us with its first laws (Mendel 1865). Mendel observed heritable phenotypic differences in seed color, seed shape and other characteristics of the common garden pea. Although he did not know the cause of the phenotypic variation, he did observe patterns of inheritance that followed specific rules. Since this seminal work in genetics, studies have been undertaken to determine the variation in DNA sequence causing the traits that Mendel observed. For example, it was found that the round versus wrinkled phenotype of pea is caused by a transposon insertion in a starch branching enzyme gene (Bhattacharyya et al. 1990). The variation driving the tall versus short pea plant phenotype is thought to be the result of a single nucleotide mutation causing an amino acid substitution (Ellis et al. 2011). It is noteworthy, but not surprising, that in these two examples the traits are caused by different mutation types, insertions and point mutations, that would later be widely applied in forward and reverse genetics.

It was not long after the rediscovery of Mendel's laws in the early twentieth century that researchers sought ways to overcome the bottleneck of the relatively low frequency of spontaneous mutation. Hermann Muller found that treating *Drosphila melanogaster* with ionizing radiation resulted in a much higher frequency of aberrant phenotypes (Muller 1927). Concurrently, the work of Stadler in mutagenizing cereals laid the foundation for mutation induction and mutation breeding in crops (Stadler 1928). While Muller recognized the potential value of his discovery for plant breeding, Stadler was more pessimistic owing to the frequency of mutation events that had a negative phenotypic consequence (Sigurbjoernsson and Micke 1974). This initial skepticism may have impacted the rate of acceptance and use of induced mutations into breeding projects. One can note that before 1950 only 3 mutant varieties were officially released. Perhaps equally important to the initial slow uptake of induced mutations is the well-established multi-phasic acceptance of new technologies and the fact that World War II suppressed crop research and breeding in many parts of the world.

The ability to induce novel genetic variation allowed new avenues of research. For example, Stadler's work showed that polyploids were more protected from the effects of ionizing radiation due to the presence of homologous gene copies (Stadler 1929). This was later observed at the DNA level in chemically-mutagenized wheat where TILLING (Targeted Induced Local Lesions IN Genomes) assays showed that polyploids can accumulate a higher density of induced point mutations (Slade et al. 2005). Recombinant DNA and sequencing technologies emerging in the 1970s provided the means to recover induced lesions causative for observed heritable phenotypic trait differences (Sanger et al. 1977). Thus all tools became available for what can be considered the canonical forward genetic method: (1) induce mutations in a population, (2) identify interesting and newly emerged phenotypes and (3) take

appropriate steps to clone the gene(s) causing phenotypic variation (Fig. 8.1). This deductive, cause and effect approach has been a powerhouse for basic biological research across species and has led to fundamental discoveries of genes and gene-function. The list of accomplishments is much too long to cover adequately, but includes the discovery of cell cycle control, segment polarity, flowering time, semi-dwarfism, meristem development and organization, and disease control (Coen and Meyerowitz 1991; Hartwell et al. 1970; Koornneef and van der Veen 1980; Lawrence 1968; Nusslein-Volhard and Wieschaus 1980).

Techniques for sequencing DNA enabled the development of databases containing annotated gene sequences. In plants, the availability of such sequences expanded

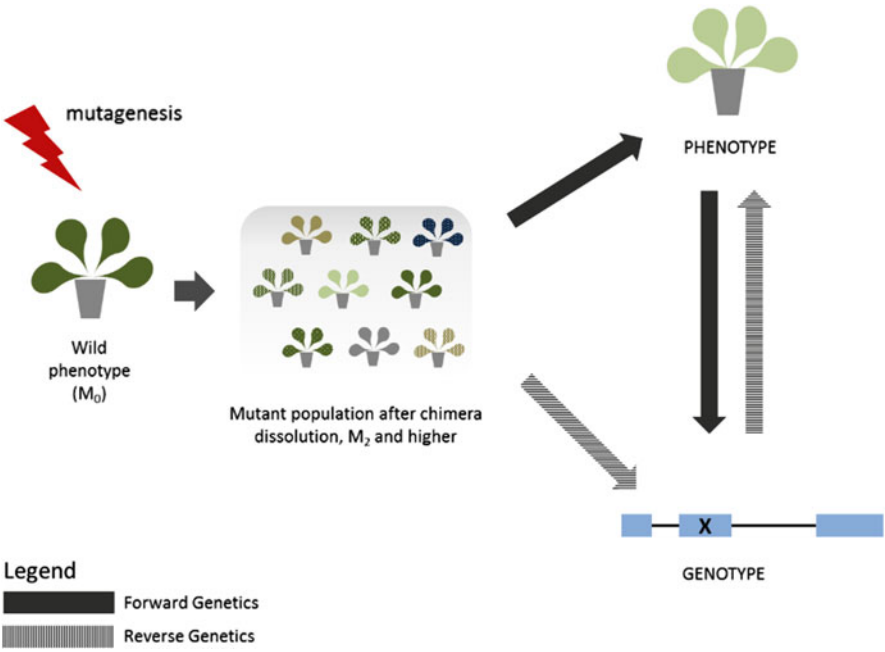

Fig. 8.1 A schematic of forward and reverse genetic approaches using induced mutations. Both approaches begin with the development of a suitably mutagenized population (*left*). Mutagen choice and dosage affect the spectrum and density of induced genomic mutations. Mutagenesis of multi-cellular tissues such as seed result in the first generation (termed M_1) being a genetic mosaic because each cell carries unique mutations. Self-fertilization, when available, is used to make plants non-chimeric (*center*). Traditional or *forward* genetics involves phenotypic evaluation and selection of novel plant phenotypes. Plants with interesting characteristics can be incorporated immediately into breeding programs and, when desired, the mutation causing the phenotype can be cloned to understand the function of genes and to produce a perfect genetic marker (*right, black arrows*). Reverse genetic strategies begin with genotypic screening of the mutant population (typically DNA extracted from M_2 plants) to identify novel induced mutations in candidate genes. This is followed by phenotypic evaluation of those individuals harboring putative deleterious mutations (*right, grey arrows*)

dramatically beginning in the 1990s (Somerville and Somerville 1999). The result was large collections of genes with no validated biological function. Methods emerged that allowed disruption of specific genes to test their in vivo function. Because gene sequences are identified first, the general procedure was coined *reverse genetics* to distinguish it from the canonical forward genetic method. The objective of this chapter is to review forward and genetic approaches in crops while providing examples of successful uses. New and emerging technologies are described as well as key differences between seed and vegetatively-propagated plants.

8.2 Traditional Mutation Breeding of Seed Propagated Crops

The first reported mutant crop variety produced via irradiation was a tobacco Chlorina (pale green) type variety released in the period 1934–1938 (Coolhaas 1952; Konzak 1957; Tollenaar 1934, 1938). X-ray irradiation was applied to young flower buds of Vorstenlanden tobacco by Tollenaar and colleagues. The recovered dominant Chlorina mutant phenotype exhibited better leaf quality for cigar making. Many successful varieties have been developed using induced mutations following this pioneering work. Today, IAEA through its joint program with FAO curates a database of officially released mutant crop varieties known as the Mutant Variety Database (Ref URL: (MVGS)). There are presently over 3200 officially released mutant varieties in over 170 species (Fig. 8.2). Interestingly, species with the highest number of varietal entries are rice, *Oryza sativa* (25 %) followed by barley, *Hordeum vulgare* (10 %), wheat, *Triticum* sp. (9 %) and the ornamental plant *Chrysanthemum* sp. (9 %). The three top entries are cereals collectively important for the food security of much of the world's population. Examples of released mutant cultivars can be found with improvements to major traits including resistance to biotic and abiotic stresses, improved quality and yield, plant architecture and maturity (Table 8.1). These examples suggest that induced mutations can be utilized to affect most characteristics of plants that are the targets of improvement through breeding. While induced mutation has been applied broadly, the approach can be considered especially advantageous in crops with a narrow genetic base where natural allelic diversity is limiting. While difficult to evaluate comprehensively, the economic impact of mutant varieties has been studied and examples of varieties contributing millions of dollars annually to local economies has been documented (Ahloowalia et al. 2004). One example of high economic impact is disease-resistant peppermint, *Mentha piperita*. Starting in the 1960s, commercial peppermint fields were increasingly suffering from verticillium wilt, a fungal disease that reduces oil yields and ultimately kills the plants. Murray, a researcher for the A.M. Todd Company of Kalamazoo, Michigan, developed two verticillium-resistant strains of Mitcham variety, the main commercial variety of peppermint, which were

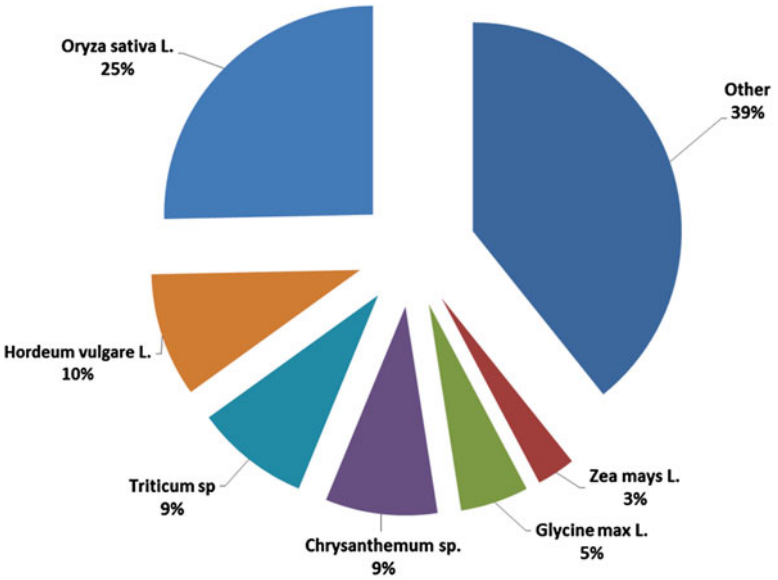

Fig. 8.2 Mutant varieties officially registered in the mutant variety database (Ref URL: (MVGS)) curated by the IAEA. Entries are grouped by genus and species. Approximately one-half of the total 3218 entries represent seed propagated cereals

named Todd's Mitcham and Murray Mitcham (Murray 1969, 1971; Murray and Todd 1972; Todd et al. 1977). It was estimated that by 1989 approximately half of peppermint production in the USA came from these mutants (Todd 1990). There are other documented examples of mutation breeding adding profit to local economies. A study comparing varieties produced from induced mutations estimated that mutant varieties provided about USD 804 million and USD 11.2 billion to the economies of Japan and the US, respectively, in the late 1990s (Kume et al. 2002). There are many ongoing mutation breeding projects for which the economic impact of newly developed crops has yet to be measured. For example, the IAEA has recently led an inter-regional project for the development of mutant wheat with increased resistance to black stem rust race Ug99. The first variety to be released from this project was announced in 2014 in Kenya (Ref URL: (KenyaGazette)). If resistance is durable, the project will potentially produce many novel alleles for introgression into other elite wheat lines in other countries.

Interestingly, approximately 87 % of officially released mutant varieties registered in the IAEA database were produced by treatment with physical mutagens with the majority being developed via gamma irradiation. This may be due to the fact that physical mutagens produce a different spectrum of mutations that may be most suitable for forward genetic breeding approaches (see below). Another possibility for the preponderance of registered varieties produced with physical irradiation may be due to a bias resulting from the IAEA funding projects that utilize what can be termed the nuclear technique of ionizing radiation. Examples also exist of the

Table 8.1 Some examples of mutant varieties derived from forward genetic screens with published references

Species (common name)	Country	Mutation method (dose)	Improved trait(s)	References
Avena sativa (oat), OT 207	Canada	Fast neutron (1150 rads)	Dwarfism (*Dw6* dwarfism gene), lodging shattering resistance	Brown et al. (1980)
Oryza sativa L (rice), Calrose series of cultivars	USA	Introgression of mutant alleles	Aroma, dwarfism (semi-dwarf), early maturing	Rutger and Bryant (2004)
Dahlia sp. (dahlia)	Netherlands	X-ray irradiation of tubers (10–40 Gy)	Improved decorative traits, cadmium-orange flower color with scarlet centre	Broertjes and Ballego (1967)
Arachis hypogaea L. (groundnut), Binachinabadam-3	Bangladesh	Gamma irradiation (200 Gy)	Dark green-obovate leaf pod; increased seed size, higher yield, moderately resistant to diseases, increased oil and protein content	Hamid et al. (2006)
Allium cepa L (onion), Brunette	Netherlands	X-rays (150 Gy)	Very early maturity, high yield and good quality	Sigurbjoernsson and Micke (1974)
Citrus paradisi Macf. (grapefruit), Star Ruby	USA	Thermal neutrons	Red flesh like parent variety, but almost seedless (0–9 seeds instead of 40–60)	Sigurbjoernsson and Micke (1974)
Linum usitatissimum L. (flax/linseed), Dufferin	Canada	Introgression of X-ray induced mutations	Resistant to rust races, higher in oil content	Kenaschuk (1977)
Hordeum vulgare L (barley), Gamma-4	Japan	Gamma rays (150 Gy)	Short stem and early maturity	Hirai et al.(1969)
Triticum aestivum L (wheat), H6765	China	Irradiation of pollen with gamma rays (1.5 Gy)	High yield, drought and salinity tolerance	Liu et al. (2007)
Syringa vulgaris L. (lilac), Prairie Petite	USA	Thermal neutrons	Dwarfness, altered leaf morphology	Lindgren et al. (1996)
Oryza sativa (rice), VND 95–20	Vietnam	Gamma rays (200 Gy)	Large-scale adaptation, good quality	Do (2009)

(continued)

Table 8.1 (continued)

Species (common name)	Country	Mutation method (dose)	Improved trait(s)	References
Mentha piperita (peppermint), Todd's Mitcham	USA	X-ray and neutrons (parental line, Mitcham)	Resistance against Helminthosporium; darker green herbage color, smaller leaves, more erect and less branched, earlier maturity	Murray and Todd (1972), Todd et al. (1977)
Hordeum vulgare (barley)	Denmark	EMS	Improved storage quality of beer	Skadhauge and Doderer (2006)
Iris cv. (iris)	Netherlands	X-ray irradiation of bulbs	Flower color, number of flowers per stalk, flowering dates	Hekstra and Broertjes (1968)

successful use of chemical mutagenesis for forward genetic breeding. A collaborative breeding project between the Carlsberg and Heineken Breweries used sodium azide (NaN_3) to induce random point mutations in barley. Phenotypic screens were used to recover mutants with reduced LOX enzymatic activity for improved flavor stability of beer during storage (Skadhauge and Doderer 2006).

8.3 Development of Mutant Populations of Seed Propagated Crops

The first step in forward and reverse genetics is the development of a suitably mutagenized population. This involves the choice of plant material, mutagen and optimizing dosage treatment. For seed propagated crops, seed are typically mutagenized, but examples also exist for pollen and in vitro mutagenesis (Serrat et al. 2014; Till et al. 2004, 2007). In most cases mutagenesis optimizations are performed to estimate the appropriate treatment conditions. When irradiation is used, optimization is typically performed by measuring either lethality or growth reduction in radio-sensitivity assays where different dosages are tested. A reduction of growth rate in the early seedling stage of many seed propagated crops serves as an easy indicator (Mba et al. 2010; Lee et al. 2014). A range of 30–50 % growth reduction is typically chosen. As with any mutagenesis treatment, a compromise needs to be made between achieving the highest possible density of induced mutations and an increased survival rate and suitable fecundity (Lee et al. 2014). Because methods for the careful measurement of mutation density and spectra have only recently become available and not yet broadly applied, there are little data to support that choosing similar growth reduction parameters in gamma or X-ray irradiated material correlates from species to species, genotype to genotype or even from irradiation treatment to irradiation treatment. This may be due to the fact that growth

reduction characteristics tend to be transient and not heritable. They are likely the cause of gross physiological responses to the stress of irradiation versus a direct response to DNA damage. For example, exposure to gamma irradiation has been shown to have direct effects on plant cell wall structure (Kovacs and Keresztes 2002). In Arabidopsis plants harboring a mutant repair endonuclease, high dosages of gamma irradiation caused developmental arrest and cell cycle arrest (Preuss and Britt 2003). DNA damage induced cell-cycle defects may therefore also play a role in measured growth reductions in early seedling stages of irradiated material. Nevertheless, growth measurements remain the simplest route to dosage optimization and have the added appeal that they can be applied in almost any facility as little infrastructure or expertise are required. For chemical mutagens such as ethyl methanesulphonate (EMS), both concentration and duration of treatment are adjusted during dose optimization. Different treatments are tested and survivability and fertility typically measured. Where possible, embryo lethality in the M_2 seed can be used as an indicator for the efficiency of mutagenesis (Till et al. 2003). As with physical mutagenesis, a balance is to be struck between survivability and the frequency of desired mutations. There are additional issues with chemical mutagenesis that can be considered. Cytotoxicity may limit the efficacy of the use of specific mutagens with certain species or genotypes, necessitating trials with different mutagens (Till et al. 2007).

Following the mutagenesis procedure, plants are evaluated for heritable traits that are different from the non-mutagenized parental line. When using either physical or chemical mutagenesis, the first generation (referred to as the M_1) is a genetic mosaic owing to the fact that each cell that is treated with mutagen is capable of accumulating different mutations. This chimeric generation is therefore not suitable for phenotypic or genotypic evaluation as different tissues may have different genotypes and thus be expressing different traits. The M_1 generation is also subject to physiological abnormalities caused by treatments. Furthermore, only one or a few cells will contribute to the production of the next generation, and heritable mutations are desired in most circumstances. Once potentially interesting traits are recovered, work is undertaken to fix the novel allele(s) in the population and ensure stability and heritability of the desired traits. Typically after 5–10 generations of self-crossing, stable plants are produced and moved forward to multi-location varietal testing before official release. Crossing is not a requisite for varietal release and many examples can be found in the Mutant Variety Database where plants are released directly after novel traits are fixed and stable.

8.4 Vegetatively-Propagated Crops and Tissue Culture

Tissue culture methodologies have many uses in plant breeding including allowing rapid generation cycling of plant materials and facilitating the production of doubled haploidy (Maluszynski et al. 2003; Zheng et al. 2013). Tissue culture materials are the primary source for mutagenesis of vegetatively-propagated plants. Such

plants can be categorized further as facultative and obligate vegetatively propagated. The former have the ability to undergo meiotic propagation but the efficiency of this is often reduced compared to seed propagated crops. Obligate vegetatively propagated species, such as the triploid banana, have lost the ability to undergo meiotic propagation and therefore must be mitotically (clonally) propagated.

Tissue-culture mutagenesis for breeding follows the same principles as with seed mutagenesis. An important difference is the choice of material to mutagenize. These include shoot tips, nodal cuttings, adventitious and lateral shoots, and cell and organ cultures. The exact methods available depend on the species and sometimes even the genotype used (Danso et al. 2010; Jankowicz-Cieslak et al. 2012). Mutagenesis optimization involves increasing dosage of treatments and measuring the survivability of materials and the ability to generate whole plants (Mba et al. 2010). However, when mutagenizing multi-cellular tissues the resulting material will be potentially chimeric and steps must be undertaken to make these genotypically homogeneous before phenotypic screening takes place. Guidelines for this typically involve several rounds of tissue cutting and propagation to reduce the number of genotypically heterogeneous cells. Another issue to consider is that most induced mutations will be in a heterozygous state. Additional steps may be needed to make mutations homozygous before phenotypes are expressed. This can be accomplished through meiotic reproduction or by producing doubled haploids. Interestingly, this may not always be needed. Crops that are propagated primarily vegetatively, such as banana and potato, cannot expunge deleterious alleles from their genome through the mechanisms of genetic recombination, meiosis and independent assortment. Thus, deleterious alleles accumulate in their genomes and can affect plant fitness and cause inbreeding depression. Referred to as Muller's ratchet, this phenomenon has been used to explain why sexual reproduction is predominant in multi-cellular life on earth (Felsenstein 1974; Muller 1932, 1964). Indeed, inspection of the genomes of vegetatively propagated crops has revealed a high accumulation of deleterious alleles in a heterozygous state (Jankowicz-Cieslak et al. 2012; Xu et al. 2011). Therefore, many alleles may be hemizygous and exist in a viable haploid form. Mutagenesis of the remaining functional copy would be sufficient to reveal otherwise recessive phenotypes. The converse to this is that haploid organisms tolerate a much lower level of induced mutations than their diploid or polyploid relatives. How the density of induced mutations and the rate of observable phenotypes affect the optimal population size for the breeding objectives remains to a large extent experimentally untested.

A further point of consideration when mutagenizing multi-cellular tissues is an observation termed diplontic drift or selection, whereby cells of differing fitness for division compete with each other, with the successful cell genotype ultimately being fixed in the plant meristem (Balkema 1972; Gaul 1958; Klekowski and Kazarinovafukshansky 1984). While previously observed only with natural mutations, recent studies with EMS mutagenized banana shoot apical meristems suggest that tissues can become homogenous rapidly, presumably through a similar mechanism (Jankowicz-Cieslak et al. 2012). There are several practical considerations arising from this work. A challenge of tissue culture mutagenesis lies in chimerism

that results from treatment of multi-cellular tissues. In clonal propagation schemes, once chimeras are dissolved, all subsequent material derived from a tissue will be genotypically identical. The consequence of this is that once material becomes genotypically homogeneous, any subsequent culturing is simply extra work and does not contribute toward increasing the size of the population of genotypically distinct individuals. Thus, if genotypic homogeneity happens quickly after mutagenesis, any culturing efforts to expand the mutant population may be wasted. Another point of consideration is that if diplontic selection is acting on mutagenized tissues, optimizations to increase the density of accumulated mutant alleles may become difficult. If cells accumulate mutations at different frequencies, it may be those that escape heavy mutation that remain fit and populate the meristematic central zone. Advances in mutation discovery for TILLING, and more recently next generation sequencing, are enabling a more careful study of the poorly understood processes of mutation accumulation in cultured tissues. One way to potentially avoid issues involving chimeric sectors and possible diplontic selection is to generate cell cultures for mutagenesis. In doing this, whole plants are generated from single mutagenized cells.

A final point to consider when opting for tissue-culture mutagenesis is that the act of culturing tissues can itself be mutagenic. The phenomenon of somaclonal variation has been described whereby novel phenotypes spontaneously arise. Most of these are detrimental but some can be used for variety development (Jain et al. 1998; Karp 1995). While the mechanisms of this variation have been a source of debate for many years, in several systems it is established that variation is caused by epigenetic changes induced by the stress of culture conditions (Kaeppler et al. 2000; Miguel and Marum 2011). Such tissue culture stress has been exploited to induce genetic transposition and utilized as an insertional mutagen for reverse genetic approaches. A classic example of this is the Tos17 system of rice (Hirochika et al. 1996).

The strength of traditional mutation breeding lies in the fact that it is easy to implement and does not require advanced techniques or large laboratory infrastructure for success. The ability to screen many thousands of mutant individuals rapidly for traits of interest greatly improves the chances of success. This is because while induced alleles may be produced at a frequency that is orders of magnitude higher than spontaneous mutations, the probability of recovering a mutation generating the desired phenotype is expected to be relatively low (Sim et al. 2012; Till et al. 2003). Mass screening techniques applied to cultured tissues can be especially powerful as trait development is independent of potentially lengthy field propagation cycles (Lebeda and Svabova 2010). An additional strength of traditional mutation breeding is that induced mutant alleles have been broadly utilized in crops for many decades. There is no intellectual property associated with inducing mutations, and most countries do not require regulation of mutagenized material. This, and the fact that continual and increasing breeding efforts are needed to meet the growing pressures on global food security, suggests that the use of forward genetics will continue to remain important for many years to come. Emerging technologies will help increase the efficiency of forward breeding.

8.5 Cloning Genes and Marker Development in Forward Genetic Strategies

The identification of the genetic mutation causative for an observed trait is the final step in the forward genetic process. While optional, this is an important step for several reasons. Characterizing the causative allelic variation establishes the in vivo function of genes. Studying multiple alleles of a gene can provide a nuanced understanding of gene function. While broadening the knowledge base of plant gene function, such research may not be of immediate importance to some breeders. Recovery of the causative mutation, however, provides a perfect genetic marker for introgression of alleles into different genetic backgrounds. Cloning causative mutations, however, can be challenging and take many years. A traditional way of locating alleles of interest is to develop and analyze segregation in mapping populations. This has been used for example to clone disease resistance genes in wheat (Huang et al. 2003). The requisite of mapping populations, however, is beginning to change with the availability of annotated reference genomes and the advent of high-throughput genomics technologies. For example, a microarray based technique was employed to clone a 532 base pair deletion in the *AtHKT1* gene of *Arabidopsis thaliana* that controls sodium accumulation (Gong et al. 2004). However, array based methods are often not suitable for the discovery of rare induced point mutations. Next generation sequencing is particularly flexible for the discovery of a broad spectrum of induced mutations and has been used in the rapid cloning of genes in the model species *Arabidopsis thaliana* (Cuperus et al. 2010; Schneeberger et al. 2009). Suitable strategies are also being developed for crops. For example, Ryohei Terauchi and colleagues have described a method for the rapid identification of causative recessive EMS induced mutations in rice, known as MutMap, that makes use of a bulked segregant analysis strategy (Abe et al. 2012). The method has been further adapted so that F1 crosses are not required (Fekih et al. 2013).

While such approaches allow rapid cloning of genes, whole genome sequencing is not without cost and these costs escalate with escalating genome size. It is noteworthy that most of the nucleotide sequence in plant genomes will not result in heritable phenotypes when mutated. Much of the expense of whole genome sequencing goes towards the recovery of unimportant nucleotides. A good compromise, therefore, is to focus on the coding regions of genomes. The number of base pairs comprising the entire collection of exonic sequences (the exome) does not vary dramatically from crop to crop with the exception of changes in ploidy. Therefore the cost of exome sequencing will not vary much from species to species. For crops such as rice and sorghum, sequencing 20 million base pairs out of a total of ~430 and 700 Mbs, respectively, is sufficient to recover induced mutations in coding regions (Henry et al. 2014, B.J. Till, I.M. Henry and L. Comai, unpublished). An added advantage is that production of a reference genome is not required as sequencing a transcriptome is sufficient for designing capture probes and for downstream data analysis. A disadvantage, however, in applying this to a MutMap type method for gene cloning in forward screens is that mutations outside of coding sequence

such as in promoters and regulatory RNAs will go undetected. However, many functional alleles are predicted to occur in coding regions and, when using deletion mutagens, there is an increased chance that lesions will span one or more coding sequence. Thus exome capture sequencing holds great promise for facilitating the cloning and marker development for mutant alleles important for breeding.

8.6 Reverse Genetics for Crop Functional Genomics and Breeding

The term reverse genetics describes the process of first identifying lesions in genes or regulatory regions and then observing the phenotypic consequences of the lesions (Fig. 8.1). Hence it is the *reverse* of traditional *forward* genetics that begins with a phenotype and ends with recovering causative alleles. It is a hypothesis driven approach whereby the functions of specific genes or gene regions are tested in vivo. Reverse genetic methods for creating nucleotide variation in plant genomes can be broadly divided into recombinant DNA based and non-recombinant based approaches (Table 8.2). Non-recombinant based techniques include biological, physical and chemical mutagenesis. Historically, reverse genetics in crops had its origins in the exploitation of biological agents, specifically transposable elements. First discovered in maize by Barbara McClintock, transposons have the ability to move naturally and insert into new regions of a genome (McClintock 1950). Found in all organisms, transposons have been exploited to disrupt and tag genes. The first example of this in forward genetics is the cloning of the *white* locus of *Drosophila melanogaster* (Bingham et al. 1981). For reverse-genetics, a large population of plants harboring transposon insertions across the genome are produced. The population is typically screened for insertions in candidate genes by PCR with gene-specific and transposon-specific primers. The maize Mutator, Ac/Ds and rice Tos17 are examples of endogenous transposons that have been used for reverse genetics (Conrad et al. 2008; Hirochika 2001; Hunter et al. 2014; Meeley and Briggs 1995; Till et al. 2001).

While facile recombinant DNA technologies in crops lagged behind developments in model organisms, such approaches are now common for many plants. Efforts have been made to transfer transposon systems into different species for both forward and reverse genetics. In rice, for example, maize Ds and Spm transposon insertional mutagenesis resources have been developed using transgenic techniques (Kolesnik et al. 2004; Kumar et al. 2005; Wei et al. 2013). Perhaps the best example of recombinant technologies being used to generate mutant libraries for reverse genetics in plants comes from Agrobacterium transferred DNA (T-DNA) insertions in *Arabidopsis thaliana*. By 2003, over 225,000 insertion events were produced with sequence location determined for more than 88,000 (Alonso et al. 2003). Large populations are now available allowing a fully in silico resource for recovering insertions in the Arabidopsis genome. Similar T-DNA strategies have

Table 8.2 Recombinant and non-recombinant methods of mutagenesis

Method	Recombinant (R) or non recombinant (NR)	Mutation specificity	Mutation heritability	Allelic variations	References
Chemical mutagenesis (e.g. EMS, NaN_3, DEB)	NR	Random/minor bias	+	Point mutations, small indels	Jankowicz-Cieslak et al. (2011)
Biological mutagenesis (e.g. transposons)	NR or R	Random or biased	+	Insertions and deletions	Hirochika (2001)
Physical mutagenesis (e.g. X-ray, gamma ray)	NR	Random/unknown	+	Wide range: e.g. deletions, insertions, translocations, SNPs	Jankowicz-Cieslak et al. (2011)
Insertional constructs (e.g. T-DNA)	R	Some bias in insertion site	+	Primarily insertions	Jeon et al. (2000)
Overexpression constructs (e.g. promoter traps)	R	Some bias	+	Insertions and transcriptional activation	Ichikawa et al. (2006)
Transient silencing (e.g. RNAi)	R	Targeted to specific sequence(s)	–	Transcriptional reduction	Bilichak et al. (2014)
Genome editing (e.g. TALENs, ZFN, CRISPR/Cas)	R	Targeted to specific sequence(s)	+ (not experimentally validated in all systems)	Point mutations, deletions	Gaj et al. (2013)

been developed for rice (Jeon et al. 2000). In addition, other reverse genetics methods utilizing recombinant techniques have been developed for plants, such as RNA interference which allows transient evaluation of gene function but does not provide heritable variation that is the requisite for breeding (Wesley et al. 2001). Newer approaches that utilize recombinant techniques for precision genome editing are discussed below. Considerations when choosing a recombinant based mutagenesis technique include the efficiency of the system, regulatory issues, intellectual property and stability of the lesions.

A major breakthrough in the area of non-recombinant reverse genetics came in the late 1990s with the development of a method known as Targeted Induced Local Lesions IN Genomes (TILLING). TILLING combines traditional mutation induction with high-throughput mutation discovery. DNA libraries are prepared from large mutant populations typically generated using chemical mutagens (Fig. 8.1). TILLING was feasible by the late 1990s because SNP discovery methods had advanced enough to consider a reverse genetic approach based on induced point mutations. The first technology to be employed was denaturing HPLC. This was used for the discovery of induced point mutations in EMS mutagenized populations of *Arabidopsis thaliana* and also *Drosophila melanogaster* (Bentley et al. 2000; McCallum et al. 2000b). Advantages of TILLING include the ability to induce and recover mutations in any gene, broad applicability to most species and the provision of a wide spectrum of induced alleles that can have varying effects on gene function. As a reverse genetic strategy, it has further advantages over forward induced mutation approaches in that lesions can be discovered in the absence of observable phenotypes. This can be of particular importance in polyploid species where mutations in multiple homeologous gene copies may need to be combined before phenotypes can be observed, and in dioecious species where obtaining homozygous mutations is challenging (Bielecka et al. 2014; Slade et al. 2005). The development of highly efficient mutation discovery methods utilizing single-strand-specific nucleases allowed for the development of community based TILLING services where researchers could order mutations in their favourite genes (Till et al. 2003). TILLING services now exist for a range of plants including arabidopsis, rice, tomato, wheat, tetraploid Arabidopsis, *Lotus japonicus*, *Medicago truncatula*, *Brassica rapa*, pea, pepper, melon, cucumber, (Ref URL: (UTILLdb; TILLINGCore; RevGenUK)). With the advent of next generation sequencing technologies, higher throughput approaches are being developed such as TILLING by Sequencing and exome capture strategies that are allowing the development of fully *in silico* resources (Kettleborough et al. 2013; Tsai et al. 2011).

TILLING has been widely adopted and examples now exist for over 20 plant and animal species (Jankowicz-Cieslak et al. 2011; Kurowska et al. 2011). This may be due in part to the fact that standardized protocols for chemical mutagenesis exist for many organisms and have been widely used for decades. Another important feature is the scalability of the mutation discovery process. While high-throughput next generation sequencing may be an efficient large-scale approach for TILLING, low-cost single-gene methods have been developed that utilize self-extracted nucleases and standard gel electrophoresis that are suitable for most laboratories, including

those in developing countries (Huynh et al. 2013). Studies to compare sensitivity and accuracy of high-throughput versus lower cost and lower throughput assays in seed and vegetatively propagated crop plants have shown that low cost methods are sufficiently accurate when sample pooling is reduced (Dong et al. 2009; Uauy et al. 2009). Issues such as polyploidy and the frequency of natural background polymorphisms can, however, increase false discovery errors when using electrophoresis platforms with lower base pair resolution. One advantage of next generation sequencing techniques is that the availability of gene sequences is not always necessary. De novo references can be created from wild-type plants to identify induced variation in mutated samples (Monson-Miller et al. 2012). However, the mutation discovery step will not be a bottleneck for most single laboratory projects where a small number of genes are to be studied. This is because plants harboring potentially interesting mutations will need to be grown, segregation of the mutation(s) examined, genetic crosses carried out and phenotypic analysis performed. These steps typically take months whereas the time-scale for mutation discovery using the various platforms ranges from days to weeks.

Table 8.3 provides some examples of TILLING projects. The majority of mutagenized populations used for TILLING to date have been developed exclusively for seed propagated crops using either seed or pollen mutagenesis (Jankowicz-Cieslak et al. 2011). Advantages of building such a population include facile multiplication of material, ease of storage of the mutant population and downstream accessibility to mutant germplasm. The feasibility of large-scale pollen mutagenesis for TILLING is limited, with outcrossing species producing large volumes of pollen, such as maize, being the best candidates for this approach (Till et al. 2004). Fewer examples exist for vegetatively propagated species, presumably owing to various bottlenecks associated with the production and maintenance of large mutant populations (Jankowicz-Cieslak et al. 2012; Wang et al. 2012). When working with vegetatively propagated crops, generating the appropriate size starting population (the M_0), plus the post-mutagenesis workload and the tissue maintenance can be costly and time consuming. One potential improvement comes from TILLING studies in vegetatively propagated triploid banana (Jankowicz-Cieslak et al. 2012). The authors showed that when mutagenizing isolated meristems, chimerism in tissues developing from the treated meristems was rapidly reduced, and no genotypic heterogeneity was observed in plants produced from cultured materials one month after EMS treatment. This could dramatically reduce the time and effort for developing mutagenized populations through tissue culture and is being investigated in other systems such as cassava (J. Jankowicz-Cieslak, S. Bado and B.J. Till, unpublished).

8.7 DNA Lesions Caused by Commonly Used Mutagens

One useful feature of chemical mutagens such as EMS is that their effect on DNA sequence is well-characterized and highly predictable. This allows for estimations of the frequency of recovering desired alleles and for developing optimal population

Table 8.3 Examples of reported mutation density and spectrum in different species

Species (common name), ploidy level	Mutagen	Mutation density (kb)	Transitions G/C>A/T	Transitions A/T>G/C	Transversions	References
Brassica rapa (field mustard), 2x	EMS	1/56 and 1/67	n.a	n.a	n.a	Stephenson et al. (2010)
Avena sativa (oat), 6x	EMS	1/24 kb	94.4	0	5.6	Chawade et al. (2010)
Cucumis melo (melon), 2x	EMS	1/573	97.8	0	2.2	Dahmani-Mardas et al. (2010)
Hordeum vulgare (barley), 2x	EMS	1/500	n.a	n.a	n.a	Gottwald et al. (2009)
Hordeum vulgare (barley), 2x	EMS	1/1000	70	10	20	Caldwell et al. (2004)
Triticum aestivum (hexaploid wheat), 6x	EMS	1/23.3 to 1/37.5 kb	99.2	0	0.8	Dong et al. (2009)
Triticum durum (tetraploid wheat), 4x	EMS	1/51 kb				Uauy et al. (2009)
Glycine max (soybean), 4x	EMS or MNU	1/140 to 1/550				Cooper et al. (2008)
Oryza sativa japonica (rice), 2x	EMS and Az-MNU	1/265 to 1/294	70.4 to 66.7	0	29.6 to 33.3	Till et al. (2007)
Oryza sativa japonica (rice), 2x	gamma ray	1/6190 kb	n.a	n.a	50	Sato et al. (2006)
Arabidopsis thaliana, 2x	EMS	~1/200	100	n.a	n.a	Greene et al. (2003)

sizes. For example, in many plants EMS produces nearly 100 % G:C to A:T transition mutations (Table 8.3). This information aids the development of computational tools to define regions of a gene that have the highest probability of accumulating deleterious mutations (McCallum et al. 2000a). Interestingly, the spectrum of EMS alleles can deviate appreciably from ~100 % transition changes in some species. This may reflect biological differences in mutation induction and repair. Large data sets produced by TILLING projects using EMS also provide information on the density of mutations that can accumulate in different genomes. While there are many lab-to-lab variations that can affect the measurement of mutation density, a

clear trend has emerged where higher densities are achievable with increasing ploidy (Table 8.3).

Contrary to some chemical mutagens which generate nearly exclusively point mutations, physical mutagens produce large and small DNA insertions/deletions, translocations, inversions and point mutations. Thus choice and dose of mutagen will affect the type of alleles produced and some mutagens may be more suitable for certain objectives. Fast neutron irradiation has been used in reverse-genetic projects to recover gene deletions. Li and colleagues described the development of mutant rice and Arabidopsis populations and a mutation screening method that utilized PCR to find deleted genomic regions (Li et al. 2001). Known as Deleteagene^{Tm}, the work showed the recovery of deletions from ~0.8 to 12 kilobases with deletions in 84 % of gene targets in 51, 840 Arabidopsis lines. This highlights a balance to be made with different mutagens. Deletions will on average be more deleterious than point mutations, and so a single plant can accumulate far fewer deletions than point mutations. This means that a much larger population is required to recover mutations in every gene. By contrast, a population of 3000 EMS mutagenized Arabidopsis lines is sufficient to recover putatively deleterious mutations in most genes (Henikoff et al. 2004). Other work has been done to develop fast neutron based reverse genetics.

A modified PCR screening strategy was developed for deletion screening in *Medicago truncatula*. Known as De-TILLING (for deletion TILLING), the authors designed assays to screen for specific deletions between 0.2 and 2.2 kb (Rogers et al. 2009). Such PCR based techniques have an inherent ascertainment bias in that the choice of primers dictates a range of observable deletions. A more genome wide approach was developed for soybean. Here, a combination of array Comparative Genome Hybridization and exome capture sequencing was used to identify deletions due to treatment with fast neutrons. This favors the discovery of large insertions and deletions, and in the published study a range of copy number variations between 986 to almost 3 million base pairs were recovered (Bolon et al. 2011). To date, over 20,000 independent mutant lines have been characterized (Ref URL: (SoyBase)). Whole genome sequencing may be the best technology to reduce ascertainment bias in mutation discovery assays. Belfield and colleagues took this approach to characterize fast neutron mutagenized populations of *Arabidopsis thaliana* (Belfield et al. 2012). In this work, 108 induced mutations were recovered with 38/39 deletions being less than 56 base pairs and the remainder 69 recovered mutations being single base pair insertions or substitutions. In the common bean, a paired-end sequencing method revealed a similar abundance of small mutations versus large deletions (O'Rourke et al. 2013). As more studies are performed using methods that accurately measure all mutation types, a clearer picture should emerge with regard to variations in mutation density and spectra due to genotype and dosage effects.

As with studies of fast neutron mutagenesis, the choice of mutation discovery method can lead to potentially misrepresentative data sets when evaluating other mutagens. For example, Sato and colleagues used a low-cost enzymatic mismatch cleavage strategy to search for mutations in rice lines produced from gamma irradia-

tion of seed (Sato et al. 2006). Of 6 mutations recovered, 4 were single nucleotide changes, one a 2 bp deletion and one a 4 bp deletion. However, gel-based systems utilizing enzymatic mismatch cleavage are suitable for the recovery of SNPs and deletions no larger than about 50 base pairs (Comai et al. 2004). Such single-strand-specific nuclease based methods are ideally suited for the recovery of induced or natural small polymorphisms and have been broadly used in TILLING approaches and for assays such as screening for loss of heterozygosity in doubled haploid plants (Hofinger et al. 2013; Jankowicz-Cieslak et al. 2011). A different picture emerges from studies aimed at characterizing a larger spectrum of gamma induced mutations. Using PCR and TAIL-PCR based screening, Morita and colleagues found approximately 82 % of isolated mutations in rice to be single base events or deletions up to 16 base pairs with the remaining mutations larger deletions between 9.4 and 129.7 kilobases (Morita et al. 2009). Studies employing deletion mapping in rice using oligonucleotide microarrays suggest gamma induced deletions up to approximately 500 kb (Bruce et al. 2009). Preliminary studies using exome capture sequencing in gamma irradiated sorghum suggest that the majority of lesions produced may be deletions larger than 100 kb (Bradley J. Till, Isabelle Henry, and Luca Comai, unpublished). This corresponds to work in maize where deletions of 232 kb and 1.2 Mb were recovered (Yuan et al. 2014). More studies are required before significant trends can be observed to link dosage and dose rate with mutation spectrum and density. This approach will also help elucidate the spectrum and density of mutations produced by X-ray, electron beam and ion beam irradiation.

Ion beam irradiation has in recent decades been emerging as a useful tool for plant mutation breeding (Tanaka 2009). Mutation spectra can vary based on treatment parameters. For example, in studies in *Arabidopsis thaliana* where *transparent testa* and *glaborous* loci were sequenced from treated lines expressing phenotypes, carbon ion treatment resulted in about 50 % rearrangements (including deletions, inversions, translocations and insertions) and 50 % point mutations (Shikazono et al. 2005). The same study showed that treatment with electrons provided about 75 % point mutations and 25 % rearrangements however carbon ion treatment produced a mutation density about 20 fold higher than that observed in electron treated material. Heavy-ions as well as fast neutrons produce radiation with much higher linear energy transfer (LET) when compared to gamma and X-ray irradiation. This has led to hypotheses regarding types of mutations that can be stably induced by different mutagens based on the type and repair of DNA lesions (Du et al. 2014). Yet, the reported spectrum of induced mutations for these mutagens is quite variable with many small mutations reported for all treatments. Clearly, as more data are produced, more accurate conclusions can be drawn. Ultimately the researcher and breeder must choose the type and density of mutations desired. If an allelic series of point mutations is desired, EMS is likely the best mutagen for many crops. If only seeking knockouts, deletion inducing mutagens may be best, however it is not yet clear which one will allow the accumulation of the highest density of induced lesions before lethality.

8.8 Combining the Power of Forward and Reverse Genetics

Both forward and reverse genetics have their relative strengths and weaknesses. A powerful tool therefore is to consider doing both simultaneously. In TILLING projects, the first example of this comes from *Lotus japonicus* where Perry and colleagues at the John Innes Centre created a phenotype indexed population for reverse genetics (Perry et al. 2003). When performing reverse genetic screens in genes hypothesized to be involved in root symbiosis, they were able to recover useful genic lesions from a sub-population expressing root phenotypes. Typically one would screen the entire mutant collection to recover such alleles. In this example, however, the alleles were recovered in a set of 275 plants rather than the entire population of 3697, providing over a 10 fold reduction in the effort of mutation discovery. Other examples can be found for barley, tomato, pea, brachypodium and flax (Dalmais et al. 2008; Menda et al. 2004; Talame et al. 2008). Similar resources exist for other reverse genetic projects such as the soybean fast neutron work described earlier in this chapter (Bolon et al. 2011). The availability of fully genotype and phenotype indexed mutant databases provides an excellent tool for rapid gene function studies and thereby in developing efficient strategies for the introgression of useful alleles in breeding.

8.9 Targeted Versus Random Approaches

To a large extent, many of the mutagens described thus far can be considered random in that at some probability all genes in a genome could be mutated. Local biases do exist in some cases (Greene et al. 2003; Krysan et al. 2002; Liu et al. 2009). Random mutagenesis means that the researcher has little control over the precise location where mutations will accumulate when the material is treated. In contrast, targeted mutagenesis or genome editing, allows the generation of a new allelic variants in a precise region of the genome. For many years, homologous recombination was held in high hope but remained elusive in higher plants. Breakthrough technologies came in the form of engineered nucleases such as zinc finger nucleases and TAL effector nucleases that can be manipulated to generate novel nucleotide variation at specific genomic regions (Gaj et al. 2013). The field is still emerging and more recent advances have been described. For example the CRISPR/Cas system utilizes the bacterial type II Clustered Regularly Interspaced Short Palindromic Repeats (CRISR) and CRISPR-associated (Cas) immune system (Belhaj et al. 2013). The Cas9 nuclease acts on genomic DNA in locations programed by a single guide RNA to create mutations via either homology directed repair pathways or non-homologous end-joining. Point mutations and deletions are possible with the system.

More work needs to be reported in order to determine the feasibility of producing heritable mutations in whole plants in a range of different species. However, tar-

geted genome editing is potentially very powerful as only the desired genomic change is introduced while the rest of the genome remains unchanged. This is of great interest to plant breeders as their breeding efforts aim to target specific traits with minimal disruption of the finely tuned genetic background of elite lines. This can be especially very efficient if only a narrow range of allele types provide the desired phenotype, and these allele types are known in advance. One disadvantage of this approach is that to date researchers have very little knowledge with regard to annotated genes with established in vivo functions. For example, by 2012 it was estimated that only 702 genes in rice had been functionally characterized (Yamamoto et al. 2012). Thus while targeted genome editing may open new doors to basic research and breeding, random mutagenesis will remain an important approach for years to come.

8.10 Conclusions and Prospects

There is no doubt that newly-emerging technologies hold great promise to enhance the efficiency of both forward and reverse genetics for functional genomics and plant breeding. The great progress in high-throughput and automated phenomics is a major step to enable more efficient forward genetic screens whether employing natural variation from diverse germplasm, wide crosses or induced mutations. These approaches are currently expensive and computationally intensive, but costs are expected to drop and this will result in greater uptake by breeders and researchers in both developed and developing countries. A topic not covered in this chapter, genomic selection, is being used to increase the efficiency of harnessing natural nucleotide variation and genetic linkage to speed up the process of forward breeding of quantitative traits (Heffner et al. 2009). Leveraging growing datasets of annotated plant genomes (and trait data) will also improve the efficiency of forward mutation breeding and functional genomics. Success in rapid cloning of causative gene lesions suggests that this will be a powerful approach for many species, especially if reduced representation genome sequencing can be adapted effectively and implemented at a reasonable cost in larger genome crops. As more reverse genetics projects utilize genome sequencing technologies, more precise data sets on density and spectrum of induced mutations will be produced. This will allow informed choice of mutagens and dosage treatments to generate desired alleles. One can envision the creation of in silico TILLING libraries in many crops where researchers can screen databases containing catalogues of gene mutations and observed phenotypes. New ideas and innovations are needed if the global scientific community is to address the current and expected growing pressures on global food security with great rigor.

Acknowledgements We thank Dr. Brian P. Forster for his critical reading of this work and for helpful suggestions. Funding for the work on exome capture sequencing of gamma irradiated sorghum was provided by the Food and Agriculture Organization of the United Nations and the International Atomic Energy Agency through their Joint FAO/IAEA Programme of Nuclear

Techniques in Food and Agriculture. This work is part of IAEA Coordinated Research Project D24012.

References

Abe A, Kosugi S, Yoshida K et al (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol 30(2):174–178. doi:10.1038/nbt.2095 nbt.2095 [pii]

Ahloowalia BS, Maluszynski M, Nichterlein K (2004) Global impact of mutation-derived varieties. Euphytica 135(2):187–204

Alonso JM, Stepanova AN, Leisse TJ et al (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301(5633):653–657. doi:10.1126/science.1086391

Balkema GH (1972) Diplontic drift in chimeric plants. Radiat Bot 12(1):51–55

Belfield EJ, Gan X, Mithani A et al (2012) Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of *Arabidopsis thaliana*. Genome Res 22(7):1306–1315. doi:10.1101/gr.131474.111

Belhaj K, Chaparro-Garcia A, Kamoun S, Nekrasov V (2013) Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods 9(1):39. doi:10.1186/1746-4811-9-39

Bentley A, MacLennan B, Calvo J, Dearolf CR (2000) Targeted recovery of mutations in Drosophila. Genetics 156(3):1169–1173

Bhattacharyya MK, Smith AM, Ellis TH et al (1990) The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. Cell 60(1):115–122

Bielecka M, Kaminski F, Adams I et al (2014) Targeted mutation of delta12 and delta15 desaturase genes in hemp produce major alterations in seed fatty acid composition including a high oleic hemp oil. Plant Biotechnol J. doi:10.1111/pbi.12167

Bilichak A, Yao Y, Kovalchuk I (2014) Transient down-regulation of the RNA silencing machinery increases efficiency of agrobacterium-mediated transformation of Arabidopsis. Plant Biotechnol J:n/a-n/a. doi:10.1111/pbi.12165

Bingham PM, Levis R, Rubin GM (1981) Cloning of DNA sequences from the white locus of D. melanogaster by a novel and general method. Cell 25(3):693–704

Bolon YT, Haun WJ, Xu WW et al (2011) Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. Plant Physiol 156(1):240–253. doi:10.1104/pp.110.170811

Broertjes C, Ballego JM (1967) Mutation breeding of *Dahlia variabilis*. Euphytica 16(2):171–176. doi:10.1007/BF00043451

Brown PD, McKenzie RIH, Mikaelsen K (1980) Agronomic, genetic, and cytologic evaluation of a vigorous new semidwarf oat1. Crop Sci 20(3):303–306. doi:10.2135/cropsci1980.0011183X002000030003x

Bruce M, Hess A, Bai J et al (2009) Detection of genomic deletions in rice using oligonucleotide microarrays. BMC Genome 10:129. doi:10.1186/1471-2164-10-129, 1471-2164-10-129 [pii]

Caldwell DG, McCallum N, Shaw P et al (2004) A structured mutant population for forward and reverse genetics in barley (*Hordeum vulgare* L.). Plant J 40(1):143–150

Chawade A, Sikora P, Brautigam M et al (2010) Development and characterization of an oat TILLING-population and identification of mutations in lignin and beta-glucan biosynthesis genes. BMC Plant Biol 10:86. doi:10.1186/1471-2229-10-86, 1471-2229-10-86 [pii]

Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. Nature 353(6339):31–37. doi:10.1038/353031a0

Comai L, Young K, Till BJ et al (2004) Efficient discovery of DNA polymorphisms in natural populations by ecotill. Plant J 37(5):778–786, 1999 [pii]

Conrad LJ, Kikuchi K, Brutnell TP (2008) Transposon tagging in cereal crops. In: The handbook of plant functional genomics. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, pp 267–290. doi:10.1002/9783527622542ch13

Coolhaas C (1952) Large-scale use of F1 hybrids in 'Vorstenlanden' tobacco. Euphytica 1(1):3–9. doi:10.1007/BF01902797

Cooper JL, Till BJ, Laport RG et al (2008) TILLING to detect induced mutations in soybean. BMC Plant Biol 8. doi:Artn 9. doi 10.1186/1471-2229-8-9

Cuperus JT, Montgomery TA, Fahlgren N et al (2010) Identification of MIR390a precursor processing-defective mutants in Arabidopsis by direct genome sequencing. Proc Natl Acad Sci U S A 107(1):466–471. doi:10.1073/pnas.0913203107

Dahmani-Mardas F, Troadec C, Boualem A et al (2010) Engineering melon plants with improved fruit shelf life using the TILLING approach. PLoS One 5(12), e15776. doi:10.1371/journal.pone.0015776

Dalmais M, Schmidt J, Le Signor C et al (2008) UTILLdb, a *Pisum sativum* in silico forward and reverse genetics tool. Genome Biol 9(2): doi:Artn R43. doi 10.1186/Gb-2008-9-2-R43

Danso KE, Elegba W, Oduro V, Kpentey P (2010) Comparative study of 2,4-D and picloram on friable embryogenic calli and somatic embryos development in cassava (*Manihot esculenta* Crantz). Int J Integr Biol 10(2):94–100

Do KT (2009) Socio-economic impacts of mutant rice varieties in Southern Vietnam. In: Shu QY (ed) Induced plant mutations in the genomics era. FAO, Rome, pp 65–68

Dong C, Dalton-Morgan J, Vincent K, Sharp P (2009) A modified TILLING method for wheat breeding. Plant Genome 2(1):39–47

Du Y, Li W, Yu L et al (2014) Mutagenic effects of carbon-ion irradiation on dry *Arabidopsis thaliana* seeds. Mutat Res 759:28–36. doi:10.1016/j.mrgentox.2013.07.018

Ellis TH, Hofer JM, Timmerman-Vaughan GM et al (2011) Mendel, 150 years on. Trends Plant Sci 16(11):590–596. doi:10.1016/j.tplants.2011.06.006

Fekih R, Takagi H, Tamiru M et al (2013) MutMap+: genetic mapping and mutant identification without crossing in rice. PLoS One 8(7), e68529. doi:10.1371/journal.pone.0068529

Felsenstein J (1974) The evolutionary advantage of recombination. Genetics 78(2):737–756

Gaj T, Gersbach CA, Barbas CF (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol 31(7):397–405. doi:10.1016/j.tibtech.2013.04.004

Gaul H (1958) Present aspects of induced mutations in plant breeding. Euphytica 7:275–289

Gong JM, Waner DA, Horie T et al (2004) Microarray-based rapid cloning of an ion accumulation deletion mutant in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A 101(43):15404–15409. doi:10.1073/pnas.0404780101

Gottwald S, Bauer P, Komatsuda T et al (2009) TILLING in the two-rowed barley cultivar 'Barke' reveals preferred sites of functional diversity in the gene HvHox1. BMC Res Notes 2:258. doi:10.1186/1756-0500-2-258, 1756-0500-2-258 [pii]

Greene EA, Codomo CA, Taylor NE et al (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. Genetics 164(2):731–740

Hamid MA, Azad MAK, Howlidar MAR (2006) Development of three varieties of groundnut with improved quantitative and qualitative traits through induced mutation. Plant Mut Rep 1(2):14–17

Hartwell LH, Culotti J, Reid B (1970) Genetic control of the cell-division cycle in yeast. I. detection of mutants. Proc Natl Acad Sci U S A 66(2):352–359

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. Crop Sci 49(1):1–12. doi:10.2135/cropsci2008.08.0512

Hekstra G, Broertjes C (1968) Mutation breeding in bulbous iris. Euphytica 17(3):345–351. doi:10.1007/BF00056234

Henikoff S, Till BJ, Comai L (2004) TILLING. Traditional mutagenesis meets functional genomics. Plant Physiol 135(2):630–636. doi:10.1104/pp.104.041061

Henry IM, Nagalakshmi U, Lieberman MC et al (2014) Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. Plant Cell. doi:10.1105/tpc.113.121590

Hirai S, Osada T, Hasegawa K, Hiraki M (1969) New varieties of malting barley produced by radiation breeding, vol 12. Rep. Res. Lab. Kirin Brewery Co., Takasaki, pp 69–77

Hirochika H (2001) Contribution of the Tos17 retrotransposon to rice functional genomics. Curr Opin Plant Biol 4(2):118–122

Hirochika H, Sugimoto K, Otsuki Y et al (1996) Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci U S A 93(15):7783–7788

Hofinger BJ, Huynh OA, Jankowicz-Cieslak J et al (2013) Validation of doubled haploid plants by enzymatic mismatch cleavage. Plant Methods 9(1):43. doi:10.1186/1746-4811-9-43

Huang L, Brooks SA, Li W et al (2003) Map-based cloning of leaf rust resistance gene Lr21 from the large and polyploid genome of bread wheat. Genetics 164(2):655–664

Hunter CT, Suzuki M, Saunders J et al (2014) Phenotype to genotype using forward-genetic Mu-seq for identification and functional classification of maize mutants. Front Plant Sci 4:545. doi:10.3389/fpls.2013.00545

Huynh OA, Hofinger BJ, Beshir MM et al (2013) Do-it-yourself molecular biology for plant breeding: low-cost tools for developing countries. In: Plant genetics and breeding technologies, Vienna. Medimond – Monduzzi Editore Internat Proceedings Division, Pianoro, pp 33–36

Ichikawa T, Nakazawa M, Kawashima M et al (2006) The FOX hunting system: an alternative gain-of-function gene hunting technique. Plant J 48(6):974–985. doi:10.1111/j.1365-313X.2006.02924.x

Jain SM, Brar DS, Ahloowalia BS (1998) Somaclonal variation and induced mutations in crop improvement. In: Current plant science and biotechnology in agriculture, vol 32, Springer, Netherlands

Jankowicz-Cieslak J, Huynh OA, Bado S et al (2011) Reverse-genetics by TILLING expands through the plant kingdom. Emir J Food Agric 23(4):290–300

Jankowicz-Cieslak J, Huynh OA, Brozynska M et al (2012) Induction, rapid fixation and retention of mutations in vegetatively propagated banana. Plant Biotechnol J 10(9):1056–1066. doi:10.1111/j.1467-7652.2012.00733.x

Jeon JS, Lee S, Jung KH et al (2000) T-DNA insertional mutagenesis for functional genomics in rice. Plant J 22(6):561–570

Kaeppler SM, Kaeppler HF, Rhee Y (2000) Epigenetic aspects of somaclonal variation in plants. Plant Mol Biol 43(2–3):179–188

Karp A (1995) Somaclonal variation as a tool for crop improvement. Euphytica 85(1–3):295–302. doi:10.1007/BF00023959

Kenaschuk EO (1977) Dufferin flax. Can J Plant Sci 57(3):977–978

KenyaGazette http://kenyalaw.org/kenya_gazette/gazette/volume/840#781. Accessed 5 May 2014

Kettleborough RN, Busch-Nentwich EM, Harvey SA et al (2013) A systematic genome-wide analysis of zebrafish protein-coding gene function. Nature 496(7446):494–497. doi:10.1038/nature11992

Klekowski EJ, Kazarinovafukshansky N (1984) Shoot apical meristems and mutation – selective loss of disadvantageous cell genotypes. Am J Bot 71(1):28–34

Kolesnik T, Szeverenyi I, Bachmann D et al (2004) Establishing an efficient Ac/Ds tagging system in rice: large-scale analysis of Ds flanking sequences. Plant J 37(2):301–314

Konzak CF (1957) Genetic effects of radiation on higher plants. Q Rev Biol 32(1):27–45

Koornneef M, van der Veen JH (1980) Induction and analysis of gibberellin sensitive mutants in Arabidopsis thaliana (L.) heynh. TAG Theor Appl Genet Theoretische und angewandte Genetik 58(6):257–263. doi:10.1007/BF00265176

Kovacs E, Keresztes A (2002) Effect of gamma and UV-B/C radiation on plant cells. Micron 33(2):199–210

Krysan PJ, Young JC, Jester PJ et al (2002) Characterization of T-DNA insertion sites in Arabidopsis thaliana and the implications for saturation mutagenesis. Omics J Integr Biol 6(2):163–174. doi:10.1089/153623102760092760

Kumar CS, Wing RA, Sundaresan V (2005) Efficient insertional mutagenesis in rice using the maize En/Spm elements. Plant J 44(5):879–892. doi:10.1111/j.1365-313X.2005.02570.x

Kume T, Amano E, Nakanishi TM, Chino M (2002) Economic scale of utilization of radiation (II): agriculture comparison between Japan and the USA. J Nucl Sci Technol 39:1106–1113

Kurowska M, Daszkowska-Golec A, Gruszka D et al (2011) TILLING – a shortcut in functional genomics. J Appl Genet 52(4):371–390. doi:10.1007/s13353-011-0061-1

Lawrence CW (1968) Radiation-induced polygenic mutation in *Arabidopsis thaliana.* II. Analysis of lines selected for flowering time. Heredity 23(4):573–589

Lebeda A, Svabova L (2010) In vitro screening methods for assessing plant disease resistance. In: Mass screening techniques for selecting crops resistant to disease. Int Atomic Energy Agency, Vienna, pp 5–46

Lee LS, Till B, Hill H et al (2014) Mutation and mutation screening. In: Henry RJ, Furtado A (eds). Cereal genomics, vol 1099. Methods in molecular biology. Humana Press, pp 77–95. doi:10.1007/978-1-62703-715-0_8

Li X, Song Y, Century K et al (2001) A fast neutron deletion mutagenesis-based reverse genetics system for plants. Plant J 27(3):235–242

Lindgren DT, Viehmeyer G, Uhlinger R (1996) 'Prairie Petite' Lilac. Hort Sci 31(1):166

Liu LX, Zhao LS, Guo HJ et al (2007) A salt tolerant mutant wheat cultivar 'H6756'. Plant Mut Rep 1(3):50–51

Liu S, Yeh CT, Ji T et al (2009) Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. PLoS Genet 5(11), e1000733. doi:10.1371/journal.pgen.1000733

Maluszynski M, Kasha KJ, Forster BP, Szarejko I (2003) Doubled haploid production in cop plants. A manual. Kluwer Academic Publishers, Dordrecht/Boston/London

Mba C, Afza R, Bado S, Jain SH (2010) Induced mutagenesis in plants using physical and chemical agents. In: Davey MR, Anthony P (eds) Plant cell culture: essential methods. John Wiley & Sons, Ltd, Chichester West Sussex, pp 111–130

McCallum CM, Comai L, Greene EA, Henikoff S (2000a) Choosing optimal regions for TILLING. Plant Physiol 123(2):439–442

McCallum CM, Comai L, Greene EA, Henikoff S (2000b) Targeted screening for induced mutations. Nat Biotechnol 18(4):455–457

McClintock B (1950) The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A 36(6):344–355

Meeley RB, Briggs SP (1995) Reverse genetics for maize. Maize Genet Coop Newsl 69:67–82

Menda N, Semel Y, Peled D et al (2004) In silico screening of a saturated mutation library of tomato. Plant J 38(5):861–872. doi:10.1111/j.1365-313X.2004.02088.x

Mendel G (1865) Versuche über Pflanzen-Hybriden. Naturforschender Verein in Brünn Verhandlungen 4:3–47

Miguel C, Marum L (2011) An epigenetic view of plant cells cultured in vitro: somaclonal variation and beyond. J Exp Bot 62(11):3713–3725. doi:10.1093/jxb/err155

Monson-Miller J, Sanchez-Mendez DC, Fass J et al (2012) Reference genome-independent assessment of mutation density using restriction enzyme-phased sequencing. BMC Genet 13:72. doi:10.1186/1471-2164-13-72

Morita R, Kusaba M, Iida S et al (2009) Molecular characterization of mutations induced by gamma irradiation in rice. Genes Genet Syst 84(5):361–370

Muller HJ (1927) Artificial transmutation of the gene. Science 66(1699):84–87. doi:10.1126/science.66.1699.84, 66/1699/84 [pii]

Muller HJ (1932) Some genetic aspects of sex. Am Nat 66(703):118–138

Muller HJ (1964) The relation of recombination to mutational advance. Mut Res 1(1):2–9

Murray MJ (1969) Successful use of irradiation breeding to obtain Verticillium-resistant strains of peppermint, *Mentha piperita* L. In: Induced mutations in plants. Int Atomic Energy Agency, Vienna, pp 345–371

Murray MJ (1971) Additional observations on mutation breeding to obtain Verticillium-resistant strains of peppermint, *Mentha piperita* L. In: Mutation breeding for disease resistance. Int Atomic Energy Agency, Vienna, pp 171–195

Murray MJ, Todd WA (1972) Registration of Todd's Mitcham peppermint. Crop Sci 12:128

MVGS http://mvgs.iaea.org/AboutMutantVarities.aspx. Accessed 5 May 2014

Nusslein-Volhard C, Wieschaus E (1980) Mutations affecting segment number and polarity in Drosophila. Nature 287(5785):795–801

O'Rourke JA, Iniguez LP, Bucciarelli B et al (2013) A re-sequencing based assessment of genomic heterogeneity and fast neutron-induced deletions in a common bean cultivar. Front Plant Sci 4:210. doi:10.3389/fpls.2013.00210

Perry JA, Wang TL, Welham TJ et al (2003) A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. Plant Physiol 131(3):866–871

Preuss SB, Britt AB (2003) A DNA-damage-induced cell cycle checkpoint in Arabidopsis. Genetics 164(1):323–334

*RevGen*UK http://revgenuk.jic.ac.uk/. Accessed 5 May 2014

Rogers C, Wen J, Chen R et al (2009) Deletion based reverse genetics in *Medicago truncatula*. Plant Phys. doi:10.1104/pp.109.142919, pp.109.142919 [pii]

Rutger JN, Bryant RJ (2004) Registration of aromatic se rice germplasm. Crop Sci 44:363–364

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74(12):5463–5467

Sato Y, Shirasawa K, Takahashi Y et al (2006) Mutant selection from progeny of gamma-ray-irradiated rice by DNA heteroduplex cleavage using Brassica petiole extract. Breed Sci 56:179–183

Schneeberger K, Ossowski S, Lanz C et al (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat Meth 6(8):550–551. doi:10.1038/nmeth0809-550

Serrat X, Esteban R, Guibourt N et al (2014) EMS mutagenesis in mature seed-derived rice calli as a new method for rapidly obtaining TILLING mutant populations. Plant Methods 10(1):5. doi:10.1186/1746-4811-10-5

Shikazono N, Suzuki C, Kitamura S et al (2005) Analysis of mutations induced by carbon ions in *Arabidopsis thaliana*. J Exp Bot 56(412):587–596. doi:10.1093/jxb/eri047

Sigurbjoernsson B, Micke A (1974) Philosophy and accomplishments of mutation breeding. In: Polyploidy and induced mutations in plant breeding. Proceedings of two meetings joint FAO/IAEA Division and EUCARPIA, Bari 1972. Int Atomic Energy Agency, Vienna, pp 303–343

Sim NL, Kumar P, Hu J et al (2012) SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 40 (Web Server issue):W452–457. doi:10.1093/nar/gks539

Skadhauge B, Doderer A (2006) Barley for the production of flavour-stable beer. Scand Brew Rev 63(1):8–11

Slade AJ, Fuerstenberg SI, Loeffler D et al (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat Biotechnol 23(1):75–81

Somerville C, Somerville S (1999) Plant functional genomics. Science 285(5426):380–383

SoyBase http://www.soybase.org/mutants/about.php. Accessed 5 May 2014

Stadler LJ (1928) Mutations in barley induced by x-rays and radium. Science 68(1756):186–187. doi:10.1126/science.68.1756.186, 68/1756/186 [pii]

Stadler LJ (1929) Chromosome number and the mutation rate in Avena and Triticum. Proc Natl Acad Sci U S A 15(12):876–881

Stephenson P, Baker D, Girin T et al (2010) A rich TILLING resource for studying gene function in *Brassica rapa*. BMC Plant Biol 10:62. doi:10.1186/1471-2229-10-62, 1471-2229-10-62 [pii]

Talame V, Bovina R, Sanguineti MC et al (2008) TILLMore, a resource for the discovery of chemically induced mutants in barley. Plant Biotechnol J 6(5):477–485

Tanaka A (2009) Establishment of ion beam technology for breeding. In: Shu QY (ed) Induced plant mutations in the genomics era. FAO, Rome, pp 216–219
Till B, Schmitz-Linneweber C, Williams-Carrier R, Barkan A (2001) CRS1 is a novel group II intron splicing factor that was derived from a domain of ancient origin. RNA (New York, NY) 7(9):1227–1238
Till BJ, Reynolds SH, Greene EA et al (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res 13(3):524–530. doi:10.1101/Gr.977903
Till BJ, Reynolds SH, Weil C et al (2004) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4:12. doi:10.1186/1471-2229-4-12, 1471-2229-4-12 [pii]
Till BJ, Cooper J, Tai TH et al (2007) Discovery of chemically induced mutations in rice by TILLING. BMC Plant Biol 7:19. doi:10.1186/1471-2229-7-19, 1471-2229-7-19 [pii]
TILLINGCore http://tilling.ucdavis.edu/index.php/Main_Page. Accessed 5 May 2014
Todd AJ (1990) Performance of disease resistant peppermint mutants in the USA. Mutat Breed Newsl 36:14–15
Todd WA, Green RJ, Horner CE (1977) Registration of Murray Mitcham peppermint. Crop Sci 17:188
Tollenaar D (1934) Untersuchungen ueber Mutation bei Tabak. I. Entstehungsweise und Wesen kuenstlich erzeugter Gen-Mutanten. Genetics 16(1–2):111–152
Tollenaar D (1938) Untersuchungen ueber Mutation bei Tabak. II. Einige kuenstlich erzeugte Chromosom-Mutanten. Genetics 20(3–4):285–294
Tsai H, Howell T, Nitcher R et al (2011) Discovery of rare mutations in populations: TILLING by sequencing. Plant Physiol 156(3):1257–1268. doi:10.1104/pp.110.169748
Uauy C, Paraiso F, Colasuonno P et al (2009) A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. BMC Plant Biol 9:115. doi:10.1186/1471-2229-9-115, 1471-2229-9-115 [pii]
UTILLdb http://urgv.evry.inra.fr/UTILLdb. Accessed 5 May 2014
Wang TL, Uauy C, Robson F, Till B (2012) TILLING in extremis. Plant Biotechnol J 10(7):761–772. doi:10.1111/j.1467-7652.2012.00708.x
Wei FJ, Droc G, Guiderdoni E, Hsing YI (2013) International consortium of rice mutagenesis: resources and beyond. Rice (N Y) 6(1):39. doi:10.1186/1939-8433-6-39
Wesley SV, Helliwell CA, Smith NA et al (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. Plant J 27(6):581–590
Xu X, Pan S, Cheng S et al (2011) Genome sequence and analysis of the tuber crop potato. Nature 475(7355):189–195
Yamamoto E, Yonemaru J, Yamamoto T et al (2012) OGRO: the overview of functionally characterized genes in rice online database. Rice 5. doi:Artn 26 doi: 10.1186/1939-8433-5-26
Yuan L, Dou Y, Kianian SF et al (2014) Deletion mutagenesis identifies a haploinsufficient role for gamma-zein in opaque2 endosperm modification. Plant Physiol 164(1):119–130. doi:10.1104/pp. 113.230961
Zheng Z, Wang HB, Chen GD et al (2013) A procedure allowing up to eight generations of wheat and nine generations of barley per annum. Euphytica 191(2):311–316. doi:10.1007/s10681-013-0909-z

Chapter 9
Doubled Haploid Breeding in Cereals

D. Gavin Humphreys and Ron E. Knox

Abstract Doubled haploid (DH) production has become an important tool in plant breeding largely due to its capacity to produce completely homozygous plants in one generation. Not only are traits fixed for selection but the multiple generations of inbreeding required using traditional breeding methods are circumvented. A major concern of implementing doubled haploid breeding is production costs, which can be divided into two major categories: operational limitations, such as methods of labeling plants and biological limitations, such as the proportion of germinating embryos. Operational efficiencies have been improved and biological impediments reduced to make DH breeding cost effective. However, prior to implementing a DH breeding program, the breeder should consider factors such as the potential for linkage drag, types of crosses to be used and whether production resources are sufficient to produce the DH populations necessary for success. Doubled haploid technology can be integrated with marker-assisted breeding for greater efficiency and to craft the DH population for particular traits. The technology can also be used to accelerate development of germplasm with new genes of interest and to generate cytogenetic stocks. To date, hundreds of DH-derived cultivars have been developed worldwide. In Canada, as much as 30 % of the spring wheat hectarage has been sown to cultivars developed using DH technology. The future for DH breeding is promising because robust DH protocols are available for an ever-growing number of crops and future applications will see a closer integration with molecular-marker and gene-splicing technologies.

Keywords Cytogenetics • DH breeding • Doubled haploid • Haploid • Maize pollen • Marker assisted breeding • True breeding • Wheat

D.G. Humphreys (✉)
Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, K.W. Neatby Building 960 Carling Avenue, Ottawa, ON K1A 0C6, Canada
e-mail: gavin.humphreys@canada.ca

R.E. Knox
Swift-Current Research and Development Centre, Agriculture and Agri-Food Canada, P.O. Box 1030, Swift Current, SK S9H 3X2, Canada
e-mail: ron.knox@canada.ca

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_9

9.1 Introduction

The term *haploid* is used to describe a plant which has the gametic number of chromosomes (n). Doubled haploid (DH) describes a haploid plant in which the chromosome number has been doubled (2n) (Kasha and Maluszynski 2003). The phenomenon of haploid plants was first described over 90 years ago in Jimson weed, *Datura stramonium* (Blakeslee et al. 1922). Chase (1952) was likely among the first to use haploids in a breeding program, generating haploids in maize via parthenogenesis and subsequently doubling their chromosome number to produce DH lines. The possibility of targeted induction and use of doubled haploidy in plant breeding arose in 1964 with the in vitro production of anther culture derived haploids from *Datura* (Guha and Maheshwari 1964). Thus, the production and use of double haploids has been a viable technology utilized in plant breeding for decades. It has been estimated that protocols for in vitro haploid production are available for over 250 plant species (see: Maluszynski et al. 2003). In plant species where doubled haploid production is feasible, it represents the quickest path to homozygosity (Collard et al. 2005). In crops where varieties are essentially highly purified inbred lines (e.g. wheat or barley), the use of DH technologies greatly improves breeding efficiency and represents an accelerated means to deliver new cultivars to producers. In cross-pollinated crops, the production of doubled haploid lines is an effective alternative to traditionally produced inbred lines. The DH lines can then be evaluated as parents for combining ability with other inbred lines, to generate hybrid lines that are evaluated for their potential as cultivars. In hybrid variety production, the in vitro process of haploid production can also provide a selection process to purge genotypes with an excessively strong expression of inbreeding depression (Murovec and Bohanec 2012).

Haploid production in vitro in major cereals is conducted either through induction of maternal or paternal haploids. Maternally-derived haploids are generated using pollination with the same species (e.g. maize) (Gieger 2009) or another species which may or may not be related to the target species. Wide hybridization has been deployed with particular success in barley and wheat breeding (Knox et al. 2000; Maluszynski et al. 2003; Niu et al. 2014). For barley, the wild relative *Hordeum bulbosum* L. is widely used to stimulate haploid embryo development (Kasha and Kao 1970). While in common wheat, it was demonstrated over 25 years ago that wheat fertilized with maize pollen could stimulate growth of haploid embryos (Laurie and Bennett 1988). In addition, unrelated pollen donors such as *Hordeum bulbosum* (Barclay 1975), pearl millet (Laurie 1989), teosinte (Suenaga et al. 1998; Ushiyama et al. 1991), *Tripsacum* grass (Mujeeb-Kazi et al. 1995) and Job's tears (Mochida and Tsujimoto 2001) have also been shown to be useful for wheat haploid production. In certain circumstances, gynogenesis, or the production of haploid embryos entirely from the female gametophyte can be used. In vitro culture of unfertilized female flower parts (e.g. ovules, ovaries or whole flower buds) is used to produce gynogenetic haploids. However, this method is normally employed only when other methods such as wide hybridization are unproductive and its use in

breeding has been mostly limited to onion and sugar beet (Murovec and Bohanec 2012). Paternally-derived haploids can be generated through androgenesis, which is the production and regeneration of haploids and doubled haploids from male gametic cells (Foroughi-Wehr et al. 1982). Androgenic-based approaches are highly desirable and have great potential for plant breeding; however, limitations, such as albinism and/or genotype dependency, can impede the use of this approach in some cereals (Al-Ashkar 2013; Jain et al. 1996).

The objective of this chapter is to provide an overview of the application of DH technology to cereal breeding with a strong emphasis on wheat. The chapter aims to provide salient details to those considering initiating DH breeding or research programs involving DH populations. The chapter provides a brief history of DH breeding and an overview of DH production methods including microspore culture, anther culture and the wheat maize-pollen techniques. Advantages and challenges associated with DH breeding are presented. Important considerations for breeding programs are outlined including: costs associated with DH breeding; impact of DH breeding on recombination; type of cross and selection of parental plants and DH population size. The chapter presents an overview of how doubled haploidy can facilitate the recovery of complex traits and be used in conjunction with marker-assisted breeding and genomic selection to further increase the effectiveness of DH breeding. Consideration is given to the use of DH methodology in the development of cytogenetic stocks and germplasm development. Finally, the ultimate value of DH breeding is demonstrated through a description of wheat and barley DH cultivars that have been developed and released. An appendix is included which provides a detailed wheat maize-pollen protocol for those who wish to use this method in their laboratory.

9.2 Improving Doubled Haploid Throughput with Methodology: A Wheat Example

9.2.1 Anther Culture

Androgenesis is the production of haploids from anther culture or isolated microspore culture. Anther culture is well established in barley and is reportedly less technically demanding than isolated microspore culture (Szarejko 2003a). Successful anther culture benefits from the fact that a fraction of the pollen grains in situ are embryogenic (Tadesse et al. 2012). When cultured on artificial media under appropriate conditions, the embryogenic pollen grains can develop into calli or embryoids. Calli can be induced to give rise to embryoids with suitable culture media. Embryoids will develop into plantlets although regenerant fertility and numbers of green plants can impact the number of doubled haploids lines produced (Cistué et al. 2003).

Careful attention to a number of procedural details will improve the response of anther culture. Anthers must be harvested from healthy, stress-free donor plants because plant stress significantly impacts androgenic response. Donor plants require optimal conditions of temperature, light intensity, photoperiod, nutrition and water. Pesticide application is discouraged two weeks before spike collection (Szarejko 2003a). Spikes of wheat or barley are selected that contain anthers with pollen at the mid- to late-uninucleate stage (Cistué et al. 2003; Szarejko 2003a; Tadesse et al. 2012). The stage of pollen development is critical to allow the reprogramming of gametophytic development into pollen embryogenesis and/or embryogenic calli production in receptive pollen grains. Spikes or anthers are subjected to an induction-stress treatment to increase the efficiency of androgenic response. In barley, cold shock and carbohydrate starvation have been used successfully (Cistué et al. 2003; Szarejko 2003a), and cold shock and auxin treatment have been reported to stimulate wheat androgenesis (El-Hennawy et al. 2011; Grauda et al. 2010; Kim and Baenziger 2005; Pauk et al. 2003). Following induction stress, anthers are transferred to a culture medium for approximately one month although induction treatment times and culture media vary for the various published protocols and among genotypes used (see: Kim and Baenziger 2005; Maluszynski et al. 2003). Anther culture is concluded when small calli or embryoids are observed as white beads in the culture medium. When calli or embryoids are 1.5–2 mm in size, they are transferred to a solid regeneration medium where they will be cultured for 3–4 weeks. Improved root formation in developing barley and wheat regenerants has been reported when they are transferred to jars or vials containing regeneration medium without growth hormones for approximately one week before transplanting into soil (Pauk et al 2003; Szarejko 2003a). In barley, spontaneous doubling of anther culture derived plantlets can be expected with as many as 70–90 % of regenerated plants spontaneously doubled (Szarejko et al. 1997); however, in wheat anther culture, regenerants are often treated with colchicine (Kim and Baenziger 2005; Pauk et al. 2003) in order to improve the proportion of doubled haploid plants (Navarro-Alvarez et al. 1994). After approximately 5–6 weeks, plantlets are carefully freed from the agar, rinsed with tap water and transferred to a soilless mix or potting compost. Plantlets should be covered for approximately 1–2 weeks to prevent desiccation. After this time, the plantlets can be transferred to larger pots to develop further. In the case of winter wheat, plants need to be vernalized to induce seed set which is completed before final transplanting.

Many researchers have adopted anther culture because of the high numbers of pollen grains within each anther that can theoretically produce DH plants compared to one haploid plant per floret with intergeneric crossing approaches. Anther culture is reportedly more cost-effective than the intergeneric crossing methods for the large scale production of doubled haploids (Snape et al. 1986), and in barley spontaneous doubling of haploids eliminates the need for the use of highly toxic chemical doubling agents such as colchicine. However, the anther culture method is highly genotype dependent (El-Hennawy et al. 2011; Grauda et al. 2010; Tadesse et al. 2012) and high frequencies of albino plants among regenerants is a significant impediment (Redha and Talaat 2008). Anther culture-derived plants can range from

pale green to variegated (green and white sectors) to completely unpigmented (Ankele et al. 2005). Barnabas (2003) reported that across 32 wheat genotypes, the proportion of green wheat plants per 100 anthers ranged from 1.9 % to 43.8 % with an average response of 11.8 %. The unpredictability in wheat androgenic response is a major limitation to the adoption of the anther culture method in commercial wheat breeding programs. While anther culture has been widely used in barley DH breeding (see Sect. 9.12; Thomas et al. 2003), the approach has not been widely adopted in wheat DH breeding programs.

9.2.2 Microspore Culture

Microspore culture involves harvesting immature microspores that are totipotent. Staging of the harvest is important to optimizing the efficiency of the microspore approach (Zheng 2003). Plants are grown in a chamber to provide optimal conditions for growth. Sporophyte induction from microspores requires a stress on the plants. Considering wheat, tillers are excised with spikes in the boot, and held in a refrigerator at 4 °C for 21 days (Asif et al. 2014). The cool temperature provides conditions to induce sporophytes and reduce albinism. Staging is performed by microscopic examination of a representative floret from the spikes to ensure microspores are present at the late uninucleate stage (Zheng 2003). Immature spikes are harvested and homogenized in a Waring blender (Asif et al. 2014). The homogenate is processed to isolate the microspores for culture on an induction medium. Embryos that form are transferred to a regeneration medium where the embryos germinate and form roots and shoots. Subsequently the plantlets are transferred to a solid medium.

The major disadvantage of microspore culture in certain species, including wheat, is the genotype dependency (Murovec and Bohanec 2012). Some genotypes are amenable to microspore culture, whereas other genotypes resist culture. Currently the genetic factors involved in controlling microspore culture amenability are unknown. Therefore, from a breeding perspective, the unpredictability of whether crosses will generate doubled haploids or not, makes the method inefficient and too costly to be sustainable, particularly when compared to the wide-cross method which is less genotype dependent. Another disadvantage of the microspore approach is difficulty in adequately staging the microspores for culturing. Although experience will reduce the difficulty of staging, plants must be monitored for the appropriate stage by sampling florets and looking microscopically for microspores at the late uninucleate stage (Asif et al. 2014; Eudes and Chugh 2009).

Advantages to the microspore approach are mainly the spontaneous doubling of chromosomes to reconstitute the diploid condition and the high volume of progeny. The rate of spontaneous doubling ranges from 25 % to 70 % in wheat (Castillo et al. 2009). Without spontaneous doubling of chromosomes, a great deal of labor is involved, along with the use of the hazardous chemical colchicine. Nevertheless, the approach has potential over the wide-cross method for greater productivity and

efficiency. Moreover, the microspore approach is more amenable to applications such as gene editing.

9.2.3 *Wheat-Maize-Pollen Method*

The wide-cross approach in wheat involves crossing maize pollen onto emasculated wheat spikes (Laurie and Bennett 1988; Fig. 9.1). A detailed description of the method that is used at several Agriculture and Agri-Food Canada research centers (e.g. Morden, Swift Current and Ottawa) is provided in the Appendix. The major appeal to the approach is its consistency for producing haploids across a range of wheat genotypes (Sadasivaiah et al. 1999). While some genotypes are more productive than others, haploid plants can be generated from all. This consistency is important as a breeding tool, because resources dedicated to generating doubled haploids are not being wasted on unresponsive crosses.

Maximizing productivity from the maize-pollen method depends on a plant production system that generates strong, healthy wheat plants for use as parents in DH production (Knox et al. 2000). It is also important to make sure maize plants are grown under optimized conditions for consistent, high quality pollen production.

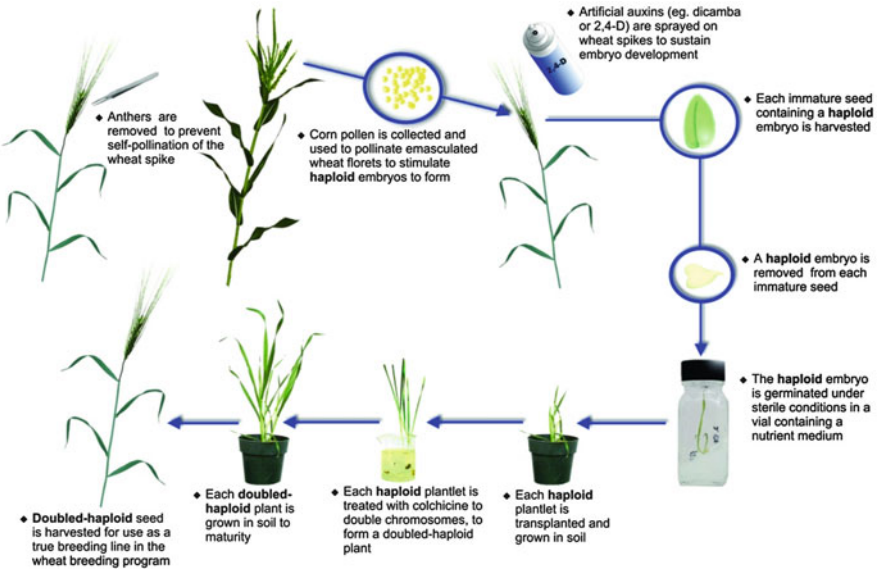

Fig. 9.1 The process of doubled haploid plant production using the maize-pollen method demonstrating pollination, application of growth regulator, haploid embryo culture, chromosome doubling and doubled haploid production. True breeding lines (lines that produce identical offspring) are produced in one generation (8–9 months) compared to 6 generations (3 years) with a traditional pedigree approach using shuttle breeding with 2 generations per year

Embryo rescue is critical to the success of this approach (Laurie and Bennett 1988). Therefore, excellent sterile techniques are necessary to achieve high rates of production, thereby minimizing contamination by fungi and bacteria that can impede embryo survival. Haploid plants and plants treated with colchicine for chromosome doubling are often weak, and require extra care and attention for their successful culture. As a starting point, we discuss a process used in wheat that meets the criteria for a productive doubled haploid system.

Growth of spring wheat is optimized under cool temperature conditions with high light intensity. A long light cycle satisfies any requirement for day length in day-length sensitive genotypes and optimizes rate of development. Genotypes with a winter growth habit require a vernalisation treatment (Zheng 2003). Growing plants in a growth chamber allows the type of control necessary to optimize plant performance and development. A growth chamber set at 15 °C during a dark cycle of 8 h and 18 °C during a light cycle of 16 h is a good compromise between duration of development and optimizing the robustness of plants (Knox et al. 2000). Light quantity and quality are important, both for producing robust plants and optimizing the rate of phenological development (Dash and Dash 2009). A traditional incandescent and fluorescent combination at a light intensity of 400 μmol m^{-2} s^{-1} is effective for wheat development, but other lighting types may be suitable after testing on the species of interest (Knox et al. 2000). Humidity controlled at 70 % will provide an adequate balance of plant growth and minimize development of diseases such as powdery mildew.

Throughput is important in breeding and can be optimized by growing plants in appropriately-sized pots and transplanting to larger pots as plants grow. Knox et al. (2000) suggested three pot changes for durum wheat. For wheat, two pot size changes, one at the three-leaf stage and another at tillering, after the initial planting in a small container has since been found to be sufficient. Sizing pots to plant development stage helps improve productivity in a high throughput doubled haploid system because it allows watering without a lot of time-consuming individual plant attention. A potting medium that provides good drainage but good nutrient holding capacity is important. Repotting plants has an additional advantage of providing fresh potting medium and diluting the buildup of fertilizer salts and compounds that hinder plant growth. Pre-germinating F_1 or other seed used as DH parents should be considered if germination rate is thought to be a problem. To achieve full growth potential of the plants, it is important to fertilize regularly and provide adequate water.

The development of fungal and insect pests, such as powdery mildew or aphids, reduces the potential productivity of haploid embryos on the plant; therefore, their control is crucial. However, chemical control measures can also reduce plant productivity and a balance must be achieved between the pest level and level of control.

It is not only important to optimize the growth conditions for the female plant, wheat, in our example, but also the male plant, maize (Knox et al. 2000). Good, consistent and continuous pollen production is essential for the timely pollination of female wheat plants. Maize plants perform well in a growth chamber with high light

intensity that is illuminated for 16 h at 22 °C and dark for 8 h at 20 °C. As with the female plants, sizing the pots to the size of the plant not only reduces over-watering problems but also reduces growth chamber space requirements (see Appendix). Pest control is also important, but chemicals to control pests can affect pollen viability, therefore use of chemicals should be kept to a minimum.

To optimize embryo production, timing of emasculation is important. A higher proportion of embryos per emasculated floret is achieved by removing anthers as late as possible before dehiscence. Generally, the anthers should be lime green to greenish yellow. Often the uppermost spikelets are more advanced and may show anthers starting to extrude. All crosses are different, with some requiring emasculation at a much earlier stage of spike emergence from the boot than others. Thus, each cross must be closely monitored.

In wheat there are two main methods of achieving anther removal. In the authors' experience, the method of clipping back the florets with scissors to expose the anthers reduces embryo production compared to the method of removing anthers through inserting forceps between an intact lemma and palea. The plant can tolerate some trimming such as the clipping of awns and the removal of tertiary florets from each spikelet. Placing a glassine bag on the spike is important to prevent stray pollen from pollenating florets.

Embryo production is further optimized with the use of fresh maize pollen. We have found pollen is best when collected about 2 h after supplemental lighting has started the day cycle. The pollen is a vibrant yellow color with a fine grain. Pollination on day 2 post-emasculation of a wheat spike, when stigmas are fluffy, is generally the most favorable pollen application time.

Critical to optimizing haploid production from the wide-cross method is the application of auxin. We have found the artificial auxin, dicamba, to be superior in producing germinated haploid embryos (Knox et al. 2000). Timing of application and concentration need to be optimized for maximum embryo production. We have found a 100 mgL^{-1} concentration of dicamba applied on both days 3 and 4 after emasculation using an aerosol spray bottle to be effective.

Until now we have discussed factors relating to the efficiency of haploid embryo production, i.e. conditions that maximize the number of embryos produced per emasculated spike. Under optimized conditions, one can expect 60–80 % of florets on a wheat spike to produce embryos (Knox et al. 2000). Another area of attrition is the embryo-culture-to-germination stage. The two major factors that interfere with conversion of embryos to haploid plantlets is failure of embryos to germinate and contamination. Factors that affect germination include, as alluded to previously, the type of hormone treatments while the embryos are developing on the spike, but also the timing of embryo rescue. We have found that rescuing embryos 16–19 days post-emasculation is the best compromise between maximizing embryo growth while minimizing the necrosis that begins at about 19 days. The type of medium affects germination of the embryos and development of plantlets. Use of Gamborg's B5 nutrient medium with minimal organics has worked successfully in the authors' experience. A chilling stress of freshly-cultured embryos helps to promote germination. We use 5 °C for 48–72 h. Following the cold treatment, we have found stratifi-

cation with rapid temperature cycling to further improve germination of embryos. We use a 6 h cycle of 11 °C dark and 18 °C with lights on. Germinated embryos, while still in vials, are placed in a chamber at 12 °C for 8 h darkness and 13 °C with 16 h light (the difference in temperature being from heat generated by the lights).

The other factor that reduces efficiency is contamination of embryos. Careful sterilization of the medium, vials containing the medium, work surface, instruments and caryopses themselves is critically important. Wheat caryopses are particularly difficult to surface disinfest because of the hairy brush and crease in which surface tension of water prevents penetration of the sterilizing solution. Therefore it is important to use solutions that will break water surface tension, such as alcohol, as part of the disinfestation process. Because high throughput is important to reduce overall breeding costs, it is necessary to utilize efficient but effective techniques. We use a syringe to improve the rate of surface disinfestation of caryopses, whereby the barrel of the syringe is loaded with the caryopses. The plunger is placed into the syringe barrel and used to draw up and expel the sterilizing solutions and rinses. Embryos are excised with the aid of a dissecting scope and placed in a vial on sterile nutrient medium and capped.

Germinated embryos that show both shoot and root development can be removed from the vials and placed into a potting mix that has good water absorption. It is important that these fine tender plants do not desiccate upon removal from the vials. To assist their transition, the haploid plantlets can be placed in trays, misted with water, and covered with a transparent cover within the growth chamber. The cover can be gradually opened to acclimatize the plants. Again using small potting containers appropriate to the size of the plantlets helps prevent overwatering and improves the survival of the plantlets.

Haploids developed using the wide-cross method must be subjected to a chromosome doubling treatment. We have found that colchicine continues to be the most effective treatment to achieve chromosome doubling and minimize attrition of plants post-treatment. We apply colchicine when the plants start to tiller, which coincides phenologically with a switch in development from the vegetative phase to the reproductive phase. The plants seem to be advanced enough to sustain the shock of the chromosome doubling treatment, while at the same time early enough to affect cells that give rise to eventual reproductive structures of florets, spikelets and sometimes whole spikes. The doubled plants are treated similar to the haploid plants to reintroduce them into the growth chamber. The plants regrow and meristematic cells that successfully experienced chromosome doubling give rise to fertile chimeras of florets, spikelets or spikes.

Often F_1 DH parent plants (or whichever generation of parent) deriving from multiple crosses are handled in concurrent streams, particularly in the wide-cross system. Often there are peaks and troughs in productivity of a doubled haploid program for various reasons, such as a pest infestation, growth chamber failure and so on. To mitigate the impact on any one population in the system, it is better to plant F_1 DH parental plants derived from a number crosses at the same time and spread the planting over several days to a few weeks. That way each of a few F_1 plants from multiple crosses experiences the low production issues, rather than all the F_1 plants

from one cross experiencing the issues which would be the case if only F_1 plants from one cross were planted at a time sequentially. The disadvantage to this approach is the extended production of doubled haploids as they come out the other end of the pipeline. Production duration of doubled plants of a population from a particular cross is further extended by the differential rate of embryo germination, recovery of plants doubled with colchicine and differences in plant maturity.

Critical to the high throughput required for doubled haploid production in a breeding program is the implementation of a tracking system. To be able to expend more energy on production of doubled haploid plants and less time documenting the plants at each transitional stage, a documentation system that can follow individual plants through the process is essential. As an example, we use a system whereby each cross is documented with a cross number in a computer file and each F_1 parental plant is assigned a unique number. As embryos are rescued from a parental plant, they are assigned a unique code. A combination of letters and numbers helps to make the plant identifier easy to read. For example a doubled haploid plant designated 1501AA10 would indicate it was derived from an F_1 parental plant crossed in 2015 (**15**01AA10), the cross being the first cross in a series of crosses (15**01**AA10), the F_1 plant being the first plant from which an embryo was rescued (1501**AA**10), and the plant being derived from the tenth rescued embryo (1501AA**10**) of the AA F_1 plant. By printing the number with a barcode on a pot marker and attaching it to the vial in which an embryo was rescued, the pot marker can follow the embryo through the entire doubled haploid process without any transcribing of information. Thus the pot marker is transferred to the pot in which the haploid plantlet is transplanted, tied to the plant as it is treated with colchicine, transferred to the pot in which the colchicine treated plant is transplanted, and applied to the envelope into which the doubled haploid spikes are harvested.

Another time-consuming process is the identification of spikes on F_1 parental plants being processed for the variety of manipulations including emasculation, maize pollination and multiple-hormone treatments. The process can be sped up by uniquely identifying each spike among a group of plants and multiple spikes on each plant. We identify the spikes with the cross and plant number on a colored tag that is uniquely assigned to a particular day. Then, when it is time to pollinate at the same time all spikes emasculated on the same day across a range of plants, the plants with the same color of tag can be assembled. Rather than writing treatments on tags, the tags are stamped or punched with a unique symbol.

9.3 Production Facilities and Resources

Doubled haploid production, whether through microspore culture or wide-cross approaches, is a sophisticated technology that requires appropriate segregating plant stocks, equipment, and appropriately-trained personnel for high throughput. Personnel must be trained in sterile techniques and plant husbandry. Knowledge of computer and basic software use are important for documenting and tracking large

numbers of unique lines. Proper facilities and equipment are essential for optimizing the environment and procedures necessary for high throughput. The ideal facility for DH production will have controlled environment rooms of some combination of greenhouse and growth chambers. There needs to be a header house equipped for handling potting media and processing plants at the front end of the growth facilities. A laboratory equipped to handle toxic chemicals and materials under sterile conditions is essential.

Haploid plants tend to be less vigorous than normal plants, therefore it is important to use materials that enhance their survival. For example, it may be necessary to pasteurize or sterilize potting media to prevent the media from being a source of pests and pathogens. Similarly it is important to periodically clean pots and equipment used in potting. The preferred facility will have an autoclave or pasteurizer and pot-washing facilities. As an alternative to pot washing, disposable containers can be used. The choice is a trade-off between investments in labor as opposed to supplies. The facility should be equipped with adequate benching for planting and transplanting and dust control for protection of workers. Air-borne pests are a persistent problem within growth rooms, therefore, the facility will also be used to handle pesticides. As a result, the header house facility should be equipped with a fume hood for mixing pesticides. The facility will need space to stock potting media, potting containers, fertilizers, pesticides and tools. There is no one system of handling plants, but some requirements are important to consider. The type of potting media should resist compaction and maintain porosity with good drainage and nutrient-supplying capability. As previously mentioned, sizing planting containers to plant size can save on expensive growth space and reduce labor by allowing general or automated watering and fertilizing. With pots optimized to plant size there are fewer issues with waterlogging or pots drying out too much, conditions that can set plants back or result in death.

Because plants exposed to optimal growing conditions experience greater survival rates and segregating source plants generate haploid embryos at higher rates, growth facilities are critical to efficient DH production. Examples of conditions for desirable wheat plants have already been noted, but the principles are that light, temperature, water and nutrients be optimized for the species under culture. Not only should light quality and intensity be considered, but also the cycling duration of light and dark may be critical to promote proper growth and development. To achieve these growth requirements, facilities with artificial light and light control and temperature cooling and heating are essential to maximize output. Water quality is often taken for granted, but under intensive production conditions, minerals or contaminants in water can affect production. For example, water provided by a utility may have high levels of chlorine that can be toxic to plants under continuous use. Water quality may need to be monitored, or a suitable filter or processing system implemented to ensure a consistent supply of pure water. In the case of doubled haploid production of lines with winter growth habit, growth facilities that can provide sufficiently low temperatures for vernalization are required.

Proper nutrition of the plants is necessary at all stages of DH production in a high-throughput system where rates of production are highly dependent on

well-cared-for plants. Too much or too little fertilizer or the wrong type impacts optimum growth. Optimization of the fertilization for the potting system and growing media can require some experimentation and may include the regulation of fertilizer application using metering devices in the watering system. Accumulation of fertilizer salts can become a problem if drainage is not adequate.

Pests are common in artificial growth environments and they include insects and fungal pathogens, among others. Good integrated pest control routines are important and should include sanitation, pesticides, avoidance and other appropriate means. By using integrated control methods, the use of pesticides can be minimized. By reducing the number of pesticide applications using the lowest possible rates through timely application, negative impact on plants is reduced. Facilities should be designed to exclude insects and pathogen inoculum and allow ventilation for pesticides. Growth chambers and greenhouse rooms cleaned regularly and other opportunities to break disease cycles will improve plant performance. Consideration should be given to applying pesticides at a stage that is less sensitive to phytotoxic effects of the pesticides. For example, in the wheat-maize system, anthesis and early caryopsis development appear to be sensitive to the application of fungicides and insecticides. Therefore, control applications preferably should be performed prior to anthesis.

In the case of production of F_1 or emasculated parental plants used for maize pollination, isolation of florets from stray pollen is important. Usually this is achieved by bagging spikes after they are emasculated. As a check on the purity, it is valuable to monitor DH families for uniformity of morphological traits. Segregating lines are obvious outcross or self-pollination derivatives. With the wide-cross method, application of growth regulators is important for sustaining caryopsis development. Synthetic auxins such as dicamba or 2,4-D are used. Although different schemes have been devised to deliver the growth regulators, we apply the compounds through misting. However, containing the mist requires a ventilation hood. Similarly, pollination presents problems as aerosols that contribute to concerns such as allergic reactions of workers. Ventilation, for example the use of a negative pressure snorkel over the work area, can minimize pollen escape into the work environment.

Paramount to the production of a segregating DH population is segregating source materials. These materials are usually F_1 plants, but could be derived from any generation where there is a reasonable expectation of segregation. The DH process can be used to fix all loci in cultivars intended as parents in a genetic study to eliminate the potential confounding heterozygosity in the resulting population. The maximum genetic variation will be achieved by producing DH lines from F_1 plants. Single cross F_1 plants derived from inbred or DH parents are homogeneously heterozygous. However, a round of selection may be considered on F_2 plants with DH progeny produced either on the F_2 plant or on a sample of F_3 derivatives of the most favorable F_2 plants. In the former case, it would be impractical to perform selection in the field, which would limit the number of traits that could be selected, particularly if selection is performed prior to gamete sampling. In the latter case, however, the F_2 plants could be grown in the field and selected for a variety of traits including

resistance to diseases and insect pests known to occur in the target environment. When creating doubled haploids from top cross F_1, backcross F_1, or F_1 progeny from a four-way cross, caution must be taken to carry out sufficient sampling of plants to obtain the variation for traits targeted within the cross. If more than a very few loci are under consideration in breeding objectives for a cross, doubled haploidy may not be the best approach with other than simple crosses. In wheat, with the maize-pollen method, for example, large, healthy, vigorous plants are generated because they generally produce high rates of embryos. Using large plants with multiple spikes, many DH lines can be produced from only a few F_1 plants. However, using only a few F_1 plants runs contrary to requirements of sampling non-simple cross F_1 plants. Similarly with microspore culture, only a few spikes need be harvested to generate large numbers of haploid progeny; however, it is suggested that additional F_1 plants should be grown with a single spike used from each plant. Populations of segregating F_1 plants could be pre-selected using markers prior to use in DH production. The costs and benefits of different strategies will be discussed in more detail later.

As one can see from the facility requirements and plant production techniques, substantial expertise is required in understanding plant growth and development of the species under doubled haploid production. These factors include a basic understanding of the biology of the crop, with expertise on the various factors needed to optimize the growing environment. Personnel also need the ability to recognize abiotic and biotic stresses on the plants and have the knowledge and experience of what to do to minimize and control abiotic and biotic stresses, such as handling and application of pesticides, and an understanding of fertilizer and fertilization rates. The personnel must also know how to operate equipment such as growth rooms and greenhouses.

In addition to adequate growth facilities to generate and grow segregating plants, a clean laboratory with the capability to carry out sterile technique is important. Equipment for sterile technique includes an autoclave to sterilize media, vials and other glassware, and utensils. A laminar flow bench with a HEPA filtered air stream across the working surface effectively provides the sterile environment for procedures such as embryo rescue or culture transfers. A balance should be available to weigh chemicals. Sufficient precision in the balance is needed to weigh small quantities such as growth regulator compounds. A microwave oven or hot plate is necessary to prepare solid media and a pH testing capability is required. A media pump greatly improves the rate at which media is dispensed. Adequate refrigeration capacity is required not only to store perishables such as media, but to condition or stress using cold, haploid embryos, in the case of the maize-pollen method or microspores. Microscopes are a requirement for staging microspores or rescuing embryos. The most common approach to chromosome doubling is the use of colchicine. For the preparation and use of this and other chemicals, a fume hood should be available. Colchicine is applied to plant roots in solution for an extended period of time. An air pump, to aerate the colchicine solution and rinse water, will improve the survival of treated plants. In the use of colchicine, care must be taken to work out the time-concentration relationship so as to provide adequate opportunity for

doubling meristematic cells, but to avoid phytotoxicity. The relationship needs to be determined for the particular conditions employed. For example, water temperature can change cell permeability, so the procedure needs to be customized for the temperature conditions to be used.

Routine glassware and plasticware are required, but also specialized containers may be needed. For example, specialized culture vessels can be used in microspore culture to improve productivity of the system. High volumes of glassware, particularly vials, require automated cleaning systems to reduce labor cost, or disposable products may be used if the cost offset can be justified in terms of reduced labor.

As already mentioned, a high-throughput DH system tailored for breeding will require a tracking system. A computer with proper software and a printer that can directly print pot markers are required. Pot markers may be barcoded to facilitate tracking and documentation. It is important to minimize transcribing of information at various steps of production. Use of grid systems for positional identification, color coding and multi-use labels can assist with tracking DH materials without the need to write out information. As previously mentioned, in the case of the maize-pollen method in wheat, we use computer printed pot markers with a unique identifier that are fixed to the vials in which embryos are rescued. When the embryo germinates, the pot marker is removed from the vial and placed in a potting tray cell with the plantlet. When the plantlet is ready for colchicine treatment, the pot marker is strung to the plantlet through the process of root washing, colchicine solution and rinse water root emersion. The marker is then placed in the greenhouse pot with the colchicine treated plant. The pot marker follows the plant through pot size changes and is placed in the harvest envelope.

9.4 Advantages of Doubled Haploid Breeding

The principal advantage to the use of DH breeding is the ability to produce completely homozygous plants in one generation (Guzy-Wrobelska and Szarekjo 2003), that in turn produce completely homogeneous progeny. In contrast, multiple generations of selection in segregating populations followed by inbreeding and further selection are required to achieve true breeding lines with traditional breeding approaches. Doubled haploid lines save the time and cost of multiple generations of selfing required using traditional-breeding methods and traditional inbreeding only provides a high degree of homozygosity not complete homozygosity as with DH methods. The ability to produce true breeding lines in one generation allows one to fix genes quickly and select for traits without the problems of ongoing segregation and allelic dominance. In programs that use off-season nurseries to advance filial generations, DH breeding results in fewer generations being sent to costly off-season nurseries.

In the case of self-pollinating crops such as wheat or barley, products of DH production can be evaluated and used as final cultivars, whereas in cross-pollinating crops, such as maize, DH lines can be used as parental lines for hybrid production.

In both cases, double-haploid breeding provides shorter cycling times as compared to traditional breeding approaches. The shorter cycling times are particularly important in cereals that require vernalization, such as winter wheat because each generation requires up to 9 weeks longer than a generation for cereals that do not require vernalization. It has been estimated that DH breeding can save up to 5 years in the development of new wheat cultivars (Barkley and Chumley 2012). The accelerated development of cultivars also represents the accelerated development of new parents for plant breeding programs because the superior newly developed DH lines and cultivars are often used as the parents in future crosses. Thus, new elite crosses are made earlier than would have been possible without the accelerated delivery of the improved parents. Over the course of a plant breeder's career, the benefits will compound from doubled haploid breeding similar to the effect of compounding interest on money re-invested in a savings plan. Not only is the doubled haploid method of benefit to breeding cultivars in general, it is useful in germplasm development. The DH method permits the fixing of a gene or genes from wide crosses with wild relatives in a single generation (Cao et al. 2014). The DH lines developed from the wide crosses can subsequently be used in crosses with elite breeding lines (Mujeeb-Kazi 2003).

The advantages of DH production can be enhanced further through the integration of doubled haploidy with other breeding technologies. For example, off-season increase of DH lines in a spring wheat breeding program accelerates the advancement of DH lines through the breeding pipeline. Normally fewer than 30 DH_1 seeds are produced and an increase generation is required before yield testing can commence. Through the use of off-season nurseries, it is possible to produce adequate DH_2 seed in the winter off-season to begin yield evaluation the next local field season. For example, in the Agriculture & Agri-Food Canada spring wheat breeding programs, approximately 25 DH_1 seeds sent to the winter nursery provider in New Zealand returns well over 200 g of clean DH_2 seed that is used for yield testing in the next field season in Canada.

The use of molecular markers can greatly improve the efficiency of DH breeding and the genotypic make-up of double-haploid breeding production. Haploid plants having only the 1n complement of chromosomes, have only one allele of each gene. Heterozygotes cannot occur. Thus, screening with DNA markers at the haploid stage of DH breeding allows the direct selection for the desired allele for traits of interest such as disease resistance, abiotic stress tolerance or end-use quality. Molecular markers in wheat are available for an ever-expanding number of traits and genes for a particular trait such as rust resistance (Liu et al. 2014; Somers and Humphreys 2009) so the potential to stack or pyramid genes with marker-assisted breeding (MAB) and DH production is also increasing (Wessels and Botes 2014). Marker-assisted breeding, for recessively inherited traits at the haploid stage of DH production, is much more efficient than phenotypic selection at the whole-plant level. Not only is the integration of MAB and doubled haploidy an advantage for selecting recessive alleles, but also for the use of markers that are recessive when applied to diploids. Marker-assisted breeding is most efficient at the haploid stage of DH production prior to chromosome doubling. With selection at the haploid

stage, the population developed is greatly enriched for the trait of interest. If a perfect marker is used, that is to say, a molecular marker that is completely linked to the trait of interest, all DH lines would possess the trait. However, care should be taken when using MAB and DH production because selection for a given trait (e.g. leaf rust resistance) could potentially reduce the genetic variability for other important traits such as grain yield or end-use quality. A means to offset potential reductions in genetic variability due to MAB, could include the generation of large DH populations, where possible, to ensure adequate genetic variability is present for other important traits. The use of *elite* crosses from high yielding, adapted parents would tend to favor the proportion of DH lines with desirable attributes.

In wheat, DH breeding in conjunction with MAB can greatly improve breeding efficiency and permit the development of improved cultivars (e.g. Humphreys et al. 2010); however, high throughput screening is necessary because the time available to complete marker-assisted screening is limited to the interval between transfer of DH haploids into soil and colchicine treatment. In wheat, this time interval is approximately three weeks. In cereals where spontaneous doubling is the method used to generate doubled haploids, marker assessment may best be applied to the doubled plants.

9.5 Challenges in Doubled Haploid Breeding

While DH breeding approaches have important advantages over traditional methods, there are noteworthy challenges. Doubled haploid lines are often generated from F_1 seed, in which case the DH_1 population is similar to a F_2 segregating population from a traditional breeding system; however, DH populations are usually considerably smaller than F_2 populations. While a F_2 population in wheat breeding can feasibly be thousands of plants, DH populations are usually limited to hundreds. Thus, F_2 populations sample a greater proportion of the genetic variability. Fortunately, it has been demonstrated that DH-derived populations are not biased in their segregation when compared to traditionally-developed populations (Friedt et al. 1986; Knox et al. 1998). Because DH lines are produced in one generation, the limitation to the amount of recombination may be a disadvantage. Recombination is critical to create new genetic combinations and breakup deleterious linkages, which are characteristics of the new genetic variability required to develop improved cultivars. Doubled haploid lines are completely homozygous so further selection within a DH line is not possible. In contrast, reselection of breeding lines developed through traditional methods is possible because heterozygosity declines each generation but complete homozygosity is never fully attained.

The old plant breeders' adage *garbage in, garbage out* appears to be especially true in DH breeding systems. Marker-assisted breeding combined with less recombination could result in the selection of less desirable attributes closely linked to the marker selected trait. For example, Humphreys et al. (2010) developed the wheat cultivar Burnside using both DH breeding combined with marker-assisted breeding

for the Gpc-B1 high protein gene derived from *Triticum turgidum* ssp. *dicoccoides* (Joppa et al. 1997). The Gpc-B1 marker was applied to haploid lines to develop the spring wheat cultivar with improved grain protein content. Burnside has approximately 1 % higher protein content compared to parental materials with earlier maturity and similar grain yield potential. However, the cultivar also had slightly lower test weight, an attribute that appears to be associated with the high protein gene (Gpc-B1) because lower test weight has been observed in other cultivars that possess this gene (Fox et al. 2006).

One of the greatest challenges to the adoption of DH breeding methodology is the resources that are required. Doubled haploid breeding requires both traditional field staff as well as highly-skilled laboratory staff who may be more costly than field staff. A dedicated laboratory is required that includes facilities, equipment, and supplies for preparation of sterile media as well as the sterile manipulation and culture of embryos. Special culture cabinets for developing embryos are also required as well as facilities for treating haploids with highly toxic chemicals (e.g. 2,4-D; DMSO and colchicine in the case of wheat) and storage/disposal of the toxic wastes. These requirements were outlined in detail earlier.

Thomas et al. (2003) indicated that conversion from a wholly traditional breeding system to DH breeding should expect a *ramp-up* time because not all DH protocols are readily transferred to a new laboratory. The genotype of the DH parents and the health and physiological status of the parent plants can have a critical impact on the successful production of haploid and consequently doubled haploid plants. Different breeding organizations have tried to initiate a DH breeding program only to find that it was not possible to produce enough DH lines to justify the investment in the program. Further, the time advantage of DH breeding over traditional breeding approaches for a spring crop that incorporates shuttle breeding to advance filial generations of populations can be minimal. This is particularly true if a high level of homozygousity is not required in finished cultivars (Thomas et al. 2003).

9.6 Cost Considerations

Factors affecting the production cost of doubled haploids can be divided into two major categories. The first being operational factors within the limitations of a particular approach. Examples of operational factors would include: automated watering versus hand-watering, labeling systems that minimize transcribing and similar procedures. The second area consists of factors related to the biological limitations to production. As examples in wheat, with the maize-pollen method there are the issues of effectiveness of stimulation of embryo development and rate of embryo germination that affect efficiency (Knox et al. 2000). With microspore culture the proportion of microspores converting to sporophytes and the proportion green plants (non-albino) are rate-limiting factors (Agache et al. 1989; Asif et al. 2014; Henry and de Buyser 1985).

We have indicated previously many operational factors that can improve throughput efficiency which reduces the cost of DH lines. No matter which approach is used, an improved understanding of biological limitations is needed to further reduce the cost of production. In the wide-cross approach, improvements can be made in the number of embryos per spike, the germination rate of the embryos and the recovery of plants from the colchicine treatment. With the microspore method, the genotype dependency must be addressed along with albinism.

However, even at the most optimum level of efficiency of the wide-cross method, the fixed upper limit of doubled haploid production is lower than that of microspore culture. The simple reason for the greater potential efficiency of microspore culture is based on the number of gametes. The wide-cross method is based on a modest number of egg cells per spike, whereas the microspore culture method is based on an abundance of microspores per spike.

With the maize-pollen method in wheat, the proportion of pollinated florets generating embryos can routinely be 70 % or more. Variables associated with this proportion are environmental, wheat genotype and maize genotype. Further research is required to understand which genetic factors in both maize and wheat contribute to embryo development response and to determine what other environmental variables could be improved. If genetic control of embryo formation is simply inherited, markers could be developed to genes controlling the response. Similarly if genetic factors can be identified in maize that stimulate wheat embryo development, again genetic markers could be developed. Prescreening of parents planned for crosses targeted for DH production could be tested with markers to determine those with the best DH potential. Similarly with microspore culture, determining the genetic factors behind recalcitrance to sporophyte initiation would allow prescreening of crosses so that only known performers are used for DH production. In the longer term such genetic information could lead to a better understanding of the biology of limiting factors with the development of procedures that overcome the limitations. Knowledge of genetic factors behind albinism in microspore culture and embryo germination recalcitrance in the maize pollen approach could also assist in developing improved procedures. In the meantime, empirical research to test environmental conditions and chemical triggers and their various combinations will be needed to improve on each approach to DH production. Another area where research could reduce the cost of production, is determining inexpensive substitutes for chemicals, particularly in the case of microspore culture.

The cost of breeding with doubled haploids compared to other breeding approaches is an important consideration. A straight cost comparison can be done on the number of cultivars generated from a DH breeding system compared to a classical-breeding system, but there are other details that need to be considered before adopting or dismissing the application of double haploidy in breeding. Currently with hexaploid wheat, the cost of breeding a commercial cultivar through an optimized maize-pollen based DH program using the current best technology is very similar to the cost of classical breeding using a modified pedigree program. This comparison is derived empirically on the basis of resources invested per commercial cultivar generated. In contrast, for example, the cost to produce a DH-derived

durum commercial cultivar is nearly twice as high compared to traditional breeding, largely because of embryo germination recalcitrance, which leads to durum wheat embryo germination being at one-half the rate of hexaploid wheat. In terms of cost, for many crops the DH option will only become more favorable in the future relative to classical approaches, but DH production is not a replacement for classical breeding. As previously mentioned, doubled haploid breeding is a tool with certain advantages and disadvantages that must be weighed out as with any other breeding approach in the overall cost of achieving breeding objectives.

With traditional inbreeding systems, there not only is the repetitive procedural costs to achieve true-breeding lines, but there is the opportunity cost in the delay of development of new or improved trait advances in cultivars. The cycle time of DH breeding is shorter than traditional methods, therefore, new traits such as improved grain protein content in Canadian bread wheat can be introduced and benefit farmers earlier than with a classically-developed cultivar (Humphreys et al. 2010). Conversely, in terms of pre-breeding and germplasm development, there is value in ongoing recombination and segregation of a classical breeding approach. Multiple recombinations allow for greater opportunity to reduce linkage drag between favorable and unfavorable linked traits. The ongoing segregation also allows greater opportunity to recover lines with overall favorable combinations of traits compared to a DH system i.e. population size can be smaller in a classical breeding scenario.

Other considerations aside from cost itself are the more precise characterization achieved with doubled haploids over segregating lines, the level of technological sophistication and skilled labor required to produce DH lines, and the use of technologies such as gene editing or transformation. Another consideration may be genotypic buffering and monoculture. Traditionally bred cultivars often have residual heterogeneity that may buffer the cultivar somewhat against abiotic and biotic stresses, whereas a doubled haploid cultivar has no such buffering capacity. However, it may be argued that on a large scale, the deployment of several different doubled haploid cultivars will provide buffering, and genetic erosion can be minimized through ample deployment of cultivars.

9.7 Breeding Considerations

9.7.1 Reduction in Recombination

Genetic recombination is critical for generating variation utilized in genetic studies and plant breeding. Doubled haploids are often produced after one generation of crossing from F_1 parental plants, resulting in limited recombination that is crucial for the development of improved germplasm and cultivars. Snape (1976) made theoretical considerations between DH and single seed decent (SSD) derived barley populations. He concluded that theoretically there were no differences between the progenies generated from these two breeding methods in the absence of linkage.

However, when linkage was present, the SSD method provided greater opportunity for recombination; thus, the occurrence of recombinants could be higher. Reduced recombination is a disadvantage of doubled haploid breeding compared to SSD in cases where the plant breeder is interested in breaking up deleterious linkage blocks. Conversely, reduced recombination of DH breeding may be an advantage with elite crosses in which desirable linkage blocks already exist and the DH process provides a means to merge desirable linkage blocks from different parents.

It has been observed that recombination frequencies are generally higher in pollen mother cells (Devaux et al. 1995; Guzy-Wróbelska et al. 2007) suggesting that the use of anther or microspore culture for DH production could improve the potential for higher genotypic variability among DH progeny. Guzy-Wrobleska and Szarejko (2003a) reported that anther culture derived wheat DH lines had greater variation for 1000 kernel weight and grain yield/plant compared to maize pollen derived wheat DH lines from the same cross. However, very small differences were observed when the best 10 % of lines from both methods were compared in single row field experiments. Consequently, it was demonstrated that DH lines with agronomic improvements can be developed using either the wide-cross or anther-culture methods.

9.7.2 Choice of Cross

Relatively little literature exists regarding choice of cross specifically for DH production. One can speculate that as with any breeding approach, a superior mid-parent value is highly desirable for a breeding initiative with the objective of generating an improved commercial cultivar. It is possible that elite × elite crosses would compensate for limitations in recombination associated with an F_1-based DH breeding approach because such crosses would also tend to have high mid-parent values.

In conservative breeding programs where complex traits such as grain quality or preharvest sprouting resistance are important, it is likely that crosses are often made between parents with a high coefficient of parentage in order to minimize variation while bringing together complementary traits. In crosses of related parents, recombination is less important, and may even be detrimental by breaking up desirable linkage groups. Such crosses fit well with a DH system where crossing over is more limited than traditional breeding protocols. Guzy-Wrobleska and Szarejko (2003a) found that DH lines from crosses of locally-adapted cultivars performed significantly better for certain traits compared to the checks. Conversely, the genetic variability in a DH_1 population will be higher if the parents are less related. Germplasm development, where the focus is on introducing a new trait, often from exotic germplasm, mid-parent value would not be as important as ensuring the new trait is retained while improving the general genetic background. Nevertheless, DH methodology can readily be used to fix desirable genes from non-elite or backcross F_1 parents. For example, DH lines were generated from crosses between Canadian

spring wheat and US winter wheat cultivars in an effort to quickly fix the resistance to the orange wheat blossom midge in the Canadian spring wheat background.

Choice of cross may be dictated by doubled haploid performance. With anther and microspore culture there is evidence that in vitro response is under genetic control. In anther and microspore culture, immature pollen gives rise to embryoids, from which plantlets are regenerated. Embryoid initiation and plantlet regeneration have been shown to be under independent genetic control (Agache et al. 1989; Henry and de Buyser 1985). In addition, segregation distortion in barley DH populations has been observed to be generally in favor of alleles of the parent which has the better in vitro culture performance (Foisset and Delourne 1996). Therefore, choice of crosses that include at least one parent with exceptional in vitro culturability may improve DH response which would be particularly important when large DH populations are required.

9.7.3 Parental Plants and Population Size

The success of a DH breeding program depends on multiple factors including the filial generation of the parental materials used for DH production and the size of the DH population. While in theory wheat DH lines could be produced from a plant of any generation or genetic background, genotype effects exist so DH response will depend in part on genetic background (Sayed et al. 2002; Tadesse et al 2012) and the DH method used. The more inbred the parental material, the less genotypic variation one would anticipate to be present in the resultant DH population. In an effort to maximize genotypic variability and shorten the breeding cycle, DH populations are often generated from F_1 plants. Backcross F_1 plants may be preferred where it is desirable to enrich specific parental backgrounds such as recovering a specific end use quality or seed coat color. As mentioned earlier, the use of F_1 as the DH parental source could reduce recombination events that create desirable genetic variability. Snape and Simpson (1981) reported on the theoretical and practical effects of linkage on the segregation of traits within DH lines derived from F_1, F_2, F_3 and intermated F_2 (S_3) generations of barley, and found significant improvements in genetic variation for spike emergence, plant height, number of kernels per spike and spikes per plant, when DH production was delayed to the F_2 generation. The improvements were ascribed to the breakup of repulsion linkages and creation of new allelic configurations at unlinked loci. Conversely, Iyamabo and Hayes (1995) compared F_1 and F_2 derived barley DH lines and reported that the additional round of recombination giving rise to the F_2 did not lead to large performance differences between the two DH populations. It appears that choice of filial generation is one of the numerous factors that impact on the characteristics of the DH breeding population generated. The use of selected F_2 or F_3 lines provides an opportunity for further recombination and selection for trait(s) of interest in the DH parental materials; however, there is an associated delay in DH production and breeding progress that must be considered.

A consideration of minimum population size will depend in part on the number of unlinked loci that are to be fixed. Jansen (1992) outlined theoretical population sizes required if, for example, one wanted a high probability of fixing two unlinked loci. In this case, at least 16 DH lines must be produced. If the trait of interest were controlled by five unlinked loci, a DH population of 203 would be required. However, in the presence of linkage, these population sizes would need to be increased. Alternatively, DH production could be delayed for a generation to allow additional recombination and selection for highly heritable traits in plants to be used as DH parents. Tadesse et al (2012) suggested that for the improvement of quantitatively inherited traits such as grain yield, large DH populations should be generated using a limited number of elite × elite crosses. Evidence from actual spring wheat DH breeding results is rather ambiguous. The first DH wheat cultivar registered in Canada, McKenzie (Graf et al. 2003) was selected from a population of only six anther culture-derived DH lines. In contrast, the spring wheat cultivar Carberry (DePauw et al. 2011b) was selected from a population of 649 DH lines developed using the wheat-maize pollen method on a single cross of elite parents. Both wheat cultivars have been grown on significant hectarage in western Canada.

9.8 Doubled Haploidy and Recovery of Complex Traits

Many traits considered within breeding programs are multigenic and quantitatively inherited. In wheat, examples of such traits are Fusarium head blight resistance, preharvest sprouting resistance, protein content and many others (McCartney et al. 2005). These traits often interact with the environment, and the genes involved provide marginal incremental improvement in the trait. The small differences in phenotype are often difficult to assess. With such traits, it is important to minimize variation. Traditional breeding methods typically involve segregating families and such segregation makes distinguishing small differences in a trait more difficult. Because doubled haploid lines are true breeding and the genotype is fixed, variation of a given genotype can be attributed to environment and random effects. Consequently, testing of doubled haploid lines for quantitative traits can be done across environments and years with greater confidence in the results than with segregating lines. Nevertheless, it is still important to minimize the random variation component and maximize sampling of environments over years and locations within the target region prior to deployment of the DH line as a cultivar.

9.9 Doubled Haploidy and Marker Assisted Breeding

Molecular markers are available for an ever-increasing number of traits in a wide range of crops. For example, in wheat there are DNA markers available for disease and insect resistance, as well as grain quality and agronomic traits (Lui et al. 2014;

Randhawa et al. 2013). A doubled haploid production system can be integrated with molecular breeding to fix qualitative traits and improve gene frequencies of quantitative traits. The greatest efficiency of molecular breeding with doubled haploid production is achieved by applying markers to haploid plants. While marker testing can be performed at any stage, provided adequate lab personnel and resources are available; however, with delayed testing, resources are used to promote lines that could have been discarded at an earlier stage.

The mechanics of applying marker-assisted breeding (MAB) to doubled haploid production requires good planning. Considering the maize pollen method in wheat, doubled haploid production and marker testing must be coordinated so that a small portion of leaf tissue can be sampled from haploid plantlets (Fig. 9.2). This is a narrow window between when plantlets are transferred out of sterile media culture into potting media but prior to chromosome doubling. By sampling this stage, fewer plants need be treated with colchicine and replanted after the colchicine step. A reliable and efficient system of plant identification is crucial for effective marker assessment so that correct plants will be culled based on marker results. This approach is used within the wheat breeding program at the Agriculture & Agri-Food Canada Research Centre in Swift Current, Saskatchewan, Canada. Parents of crosses are evaluated with a panel of markers listed by Randhawa et al. (2013). Crosses that are segregating for traits to which markers are available are scrutinized for marker polymorphism. For example, a cross between the wheat stem sawfly (*Cephus cinctus*) resistant parent BW925 carrying *Sst1* and the Fusarium head blight (*Fusarium* spp.) resistant parent Carberry (DePauw et al. 2011b) carrying *Fhb1* were considered for MAB. Markers for *Sst1* (Cook et al. 2004; Houshmand et al. 2007) and *Fhb1* (Cuthbert et al. 2006) applied to the parents showed polymorphism. Crosses that show polymorphism are used to generate haploid plants that are sampled for DNA and tested for the molecular markers. In the example, F_1 plants from the BW925/Carberry cross were used to generate haploid plantlets using the maize-pollen method. The haploid plantlets were grown in labelled multi-cell growth flats. Tissue was carefully sampled from each haploid plantlet to maintain the identity of each sample with each plant. The DNA was extracted and markers for stem solidness and FHB resistance were evaluated for each of the haploid plantlets. The haploid plantlets were culled based on marker data. A total of 394 haploid plantlets were evaluated and 48 % were retained while 52 % were discarded because they lacked the favorable molecular-marker variant for *Sst1*. Similarly 57 % of the progeny expressed the favorable molecular variant for *Fhb1*, while 43 % did not. The *Fhb1* markers were not used for selection but rather for marker validation. The haploid plantlets remaining after selection with the *Sst1* marker were treated with colchicine to double the chromosomes and generate a doubled haploid population with an enriched allele frequency for *Sst1*. A line, BW5008, from this cross is presently in varietal registration testing in western Canada.

Where possible, segregating F_1 plants generated from backcrosses or topcrosses (i.e. three-way crosses) should be pre-screened with markers to ensure the trait(s) of interest are present. A marker for high grain protein content was employed to select among the segregating backcross F_1 parents used for doubled haploid production in

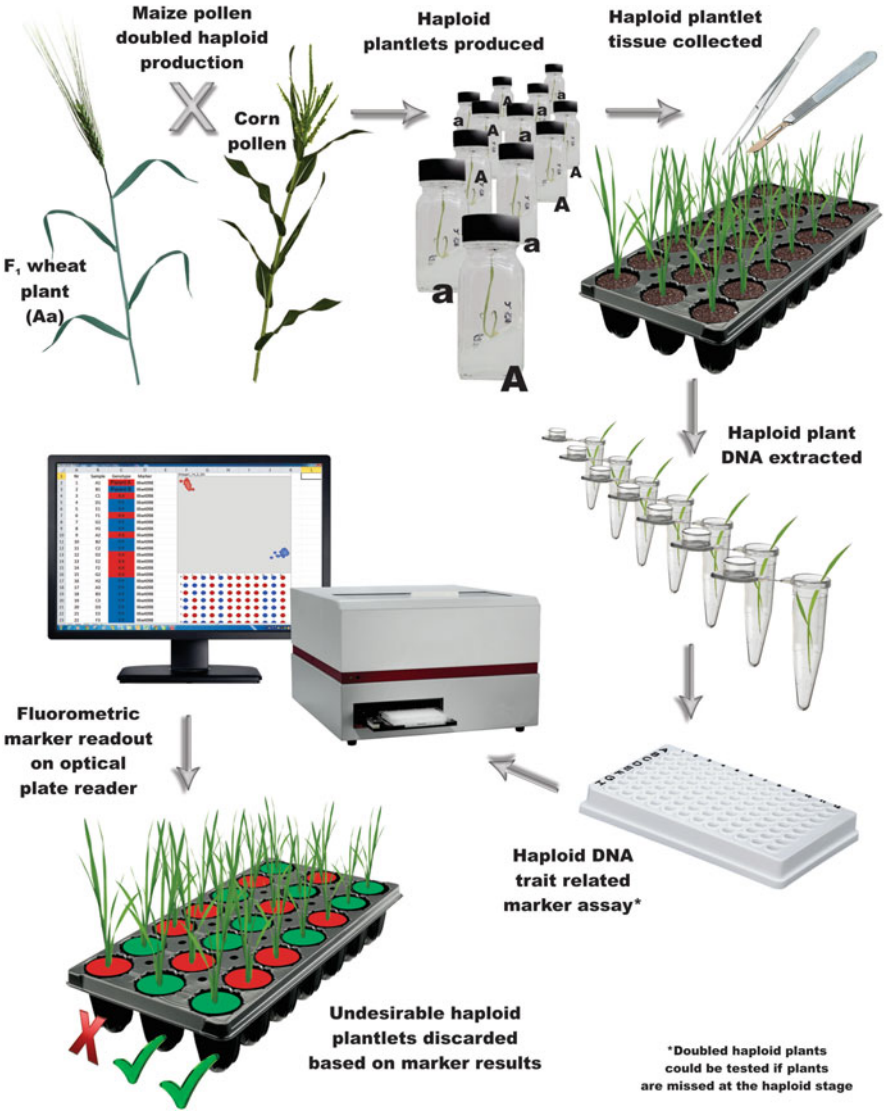

Fig. 9.2 The merging of doubled haploid production and marker-assisted breeding demonstrating the production of haploid plants using the maize-pollen method followed by tissue sampling, DNA marker analysis, and selection of favorable haploid plants that are further processed with chromosome doubling and doubled haploid generation

the development of the Canada Western Extra Strong hard red wheat Burnside (Humphreys et al. 2010). Testing segregating F_1 plants with markers is an extremely powerful method of gene enrichment and substantially reduces the attrition of lines coming out of the other end of the breeding pipeline. Testing the segregating F_1 plants will eliminate one-half for each marker tested. If the cross is configured such

that a double dose of the favorable allele is incorporated, undesirable heterozygotes are eliminated and none of the valuable DH lines will need to be discarded based on the locus. If the cross is configured such that the favorable allele is only in the heterozygous condition, half the F_1 plants will be eliminated with the marker and only one-half the DH lines will be eliminated instead of three quarters. Eliminating a few F_1 plants with a DNA marker is far more cost effective than eliminating many DH lines. The enriched population of DH lines resulting from testing F_1 lines will improve the success rate for identifying superior genotypes for other trait combinations when the DH lines are field evaluated. Nevertheless, marker-assisted selection requires considerably greater haploid production. If one marker is applied to haploids from a F_1 plant where the favorable allele is in the heterozygous condition, 50 % of haploids will not have the allele of interest and will be discarded; thus, in order to produce a population of, for example, 200 doubled haploids, approximately 400 haploids must be generated. To maximize stacking of genes for resistance to multiple pests Somers et al. (2007) used a complex crossing scheme in which they applied DNA markers to identify progeny carrying the desirable alleles. They then used the segregating progeny plants for the production of DH lines. The DH lines were then phenotyped in the field. Similarly, Wessels and Botes (2014) used markers on lines generated from crosses designed to recombine rust resistance genes to identify progeny possessing the accumulated favorable alleles. In addition to using markers on the segregating parental plants, DH plants were also subjected to marker evaluation.

9.10 Doubled Haploidy, Genetics and Opportunities for Genomic Selection

Doubled haploid lines do not just have a direct advantage to breeding, they advance breeding through their use in genetic analysis. Doubled haploid lines can be considered as immortal lines because they can theoretically be increased forever without segregating further due to their complete homozygosity. In this way, DH lines are a highly-valuable tool for developing genetic populations that are used to measure phenotypic value over environments (Huang et al. 2006; McCartney et al. 2005). Similar to the requirement within breeding programs to be able to assess quantitative traits without the concern of additional variation from segregating loci, so too is the advantage of doubled haploids in phenotyping of genetic populations. With the greater precision in phenotypic data, quantitative trait loci can be better defined, ultimately leading to improved markers for use in molecular breeding. With the unlimited seed supply of identical lines, doubled haploids can be tested genetically once and remain relevant for all subsequent analyses of mapping traits. For example, the RL4452 × AC Domain spring wheat DH population was developed and evaluated for agronomic and end-use quality traits. The population was subsequently found to be valuable for assessing other traits such as preharvest-sprouting resistance (McCartney et al. 2005, 2006; Rasul et al. 2009). The population of lines

and genetic information did not have to be regenerated as would be necessary if a segregating population of lines were used. The genetic database for a doubled haploid population can continue to increase as new markers and genetic information come on stream. Likewise the population can be tested in an unlimited number of nurseries and experiments to acquire phenotypic data. By removing heterozygotes from any genetic equation, the analysis is simplified, including complex analyses such as epistasis, genotype by environment interactions (Zhang et al. 2008; Zhao et al. 2013) and genome assembly (Zhang et al. 2014).

With the current feasibility to develop large DH populations, there is potential to use these populations to dissect quantitative trait loci (QTL) and assist in fine mapping. Recombinants can be identified within the QTL and extracted from the full population for further intensive phenotyping without the confounding effect of further segregation that would occur if, for example, a F_2-derived mapping population were used.

The immortal nature of doubled haploid populations makes them useful for generating data for modeling as is needed for the application of genomic selection. Genomic selection involves the development of training populations with which to model selection criteria for lines within a breeding program. The greater precision in phenotyping attributed to doubled haploid populations should improve the resolution of genomic selection models.

9.11 Double Haploidy and Germplasm Development

While DH technology has obvious value in breeding programs aimed at developing new cultivars, the technology can also be effectively deployed in the development of novel germplasm for use in genetic research or plant breeding. Landraces are valuable resources for broadening the genetic base of elite germplasm; however, the use of landraces can be impeded due to their genetic heterogeneity and genetic load. Strigens et al. (2013) showed that the generation of inbreds from European maize landraces using DH technology allowed for the characterization and use of the genetic diversity present in maize gene bank accessions. Mujeeb-Kazi et al (1995) have outlined strategies in which DH technology can be used to assist the transfer of genetics from wheat relatives (e.g. *Aegilops tauschii*) into hexaploid backgrounds for use in germplasm and cultivar development programs. Doubled haploids with improved *Cochliobolus sativus* spot blotch resistance were developed by crossing synthetic hexaploids developed from *Triticum turgidum*×*A. tauschii* combinations with susceptible hexaploid wheat cultivars, followed by doubled haploid production to quickly fix the resistance in new germplasm (Mujeeb-Kazi 2003). Traditionally backcross breeding methods have been used to move traits into adapted backgrounds. Forster and Thomas (2005) proposed the backcross conversion method that employs backcrossing, marker assisted selection and DH methodologies to greatly accelerate the introduction of new genes into adapted backgrounds.

Cytogenetic stocks can be developed and cytogenetic research facilitated using DH technology. Mujeeb-Kazi (2003) reported that DH technology can be used to conduct genetic analyses to identify physical gene locations through complete or partial monosomic analysis. The DH approach simplifies the conventional monosomic analysis and the monosomic germplasm generated using DH methodology is stable, which enables global distribution encouraging experimental repetition and collaboration. Cao et al. (2014) reported on the development of wheat aneuploids from a cross of *T. aestivum* × *Thinopyrum intermedium*. DH technology allowed the accelerated development of stable aneuploids for cytogenetic research and for use as parents enabling the transfer of genes from *Th. intermedium* to common wheat. Finally, the development of DH lines from triticale × common wheat crosses was proposed to enable the transfer of genetic material from rye to common wheat (Badiyal et al. 2014; Zhang et al. 2006).

Doubled haploid methodology has been used to effectively develop and isolate mutants for use in plant breeding. Haploid tissue can simplify the creation of genetic variability, its identification and isolation. Doubled haploid systems offer benefits for generation and selection of genetic mutants including: the opportunity to screen for both recessive and dominant mutations in the first generation after mutagen treatment, and direct fixation of mutants through DH methodology and the prospect to select for mutants in vitro at the haploid level. The topic has been thoroughly reviewed previously by Szarejko (2003b).

9.12 Doubled Haploid Cultivar Development and Adoption

The first cultivar developed using doubled haploid methodology was the Canadian barley cultivar Mingo (Thomas et al. 2003) and the cultivar McKenzie (Graf et al. 2003) was the first wheat variety developed using double haploidy in Canada. Since that time, over 25 additional wheat cultivars have been commercialized in western Canada (DePauw et al. 2011a; see Table 9.1). Thomas et al. (2003) listed over 200 cultivars that have been produced using various DH methods, which was believed to be an underestimate of the true number of cultivars because an accurate global audit of the total number of DH cultivars is not possible. At that time, the majority of cultivars had been developed in barley followed by rapeseed and wheat. For the major cereals, the wide-crossing methods (*Hordeum bulbosum* in barley; maize pollen in wheat) are the preferred methods, whereas microspore culture is preferred in *Brassica* (rapeseed). In wheat, anther culture is still not widely used because a robust, genotype-independent protocol has yet to be developed; thus, the maize-pollen method has been widely adopted.

The commercial uptake of DH varieties has varied between countries. The winter wheat cultivar Savanah occupied only 6 % of hectarage in the United Kingdom. In

Table 9.1 List of registered Canadian wheat and barley cultivars developed using doubled haploid breeding

Year	Name	Method	Marketing class[a]	Breeding institution[b]	References
A. List of Canadian wheat cultivars					
1997	McKenzie	Anther culture	CWRS	Saskatchewan Wheat Pool	Graf et al. (2003)
2000	Kanata	Maize pollination	CWHWS	AAFC-CRC	Humphreys et al. (2006)
2000	Snowbird	Maize pollination	CWHWS	AAFC-CRC	Humphreys et al. (2007)
2000	Superb	Maize pollination	CWRS	AAFC-CRC	Townley-Smith et al. (2010)
2000	AC Andrew	Anther culture	CWSWS	AAFC-LRC	Sadasivaiah et al. (2004)
2001	Warthog	Maize pollination	CEHRW	Hyland Seeds – DAS	
2002	Bhishaj	Maize pollination	CWSWS	AAFC-LRC	Randhawa et al. (2011)
2002	FT Wonder	Maize pollination	CESRW	AAFC-ECORC	
2003	Burnside	Maize pollination	CWES	AAFC-CRC	Humphreys et al. (2010)
2003	Lillian	Maize pollination	CWRS	AAFC-SPARC	DePauw et al. (2005)
2005	Ashley	Maize pollination	CEWW	AAFC-ECORC	
2005	Emmit srw	Maize pollination	CESRW	Hyland Seeds – DAS	
2005	FT Action	Maize pollination	CEWW	AAFC-ECORC	
2006	Alvena	Maize pollination	CWRS	AAFC-SPARC	Knox et al. (2008)
2006	CDC Abound	Maize pollination	CWRS	University of Saskatchewan	
2007	Ava sww	Maize pollination	CEWW	Hyland Seeds – DAS	
2007	Carnaval	Maize pollination	CEHRW	Hyland Seeds – DAS	
2007	Waskada	Maize pollination	CWRS	AAFC-CRC	Fox et al. (2009)
2007	Wentworth hrw	Maize pollination	CEHRW	Hyland Seeds – DAS	
2007	Glencross	Maize pollination	CWES	AAFC-CRC	
2008	Accipiter	Maize pollination	CWGP (winter)	University of Saskatchewan	Fowler (2011)

(continued)

Table 9.1 (continued)

Year	Name	Method	Marketing class[a]	Breeding institution[b]	References
2008	Broadview	Maize pollination	CWGP (winter)	AAFC-LRC	
2008	Peregrine	Maize pollination	CWGP (winter)	University of Saskatchewan	Fowler (2010)
2008	Stettler	Maize pollination	CWRS	AAFC-SPARC	DePauw et al. (2009)
2009	Carberry	Maize pollination	CWRS	AAFC-SPARC	DePauw et al. (2011b)
2009	HY116-SRW	Maize pollination	CESRW	Hyland Seeds – DAS	
2009	HY124-SRW	Maize pollination	CESRW	Hyland Seeds – DAS	
2009	Muchmore	Maize pollination	CWRS	AAFC-SPARC	DePauw et al. (2011c)
2009	Snowstar	Maize pollination	CWHWS	AAFC-CRC	Humphreys et al. (2013)
2009	Shaw	Maize pollination	CWRS	AAFC-CRC	Fox et al. (2013)
2009	Sunrise	Maize pollination	CWGP (winter)	University of Saskatchewan	Fowler (2012)
2010	Flourish	Maize pollination	CWRW	AAFC-LRC	Graf et al. (2012)
2010	Transcend	Maize pollination	CWAD	AAFC-SPARC	Singh et al. (2012)
2011	Emerson	Maize pollination	CWRW	AAFC-LRC	Graf et al. (2013)
2011	OAC Emmy	Maize pollination	CEWW	University of Guelph	Tamburic-Ilincic and Smid (2013a)
2011	Pintail	Maize pollination	CWGP (winter)	AARD-FCDC	
2011	Swainson	Maize pollination	CWGP (winter)	University of Saskatchewan	Fowler (2013)
2011	HY271-SRW	Maize pollination	CESRW	Hyland Seeds – DAS	
2012	AAC Raymore	Maize pollination	CWAD	AAFC-SPARC	Singh et al. (2014)
2012	HY300-HRW	Maize pollination	CEHRW	Hyland Seeds – DAS	
2012	OAC Flight	Maize pollination	CESRW	University of Guelph	Tamburic-Ilincic and Smid (2013b)

(continued)

Table 9.1 (continued)

Year	Name	Method	Marketing class[a]	Breeding institution[b]	References
2013	CDC Chase	Maize pollination	CWRW	University of Saskatchewan	Fowler (2014)
2013	HY301-HRW	Maize pollination	CEHRW	Hyland Seeds – DAS	
2013	AAC Prevail	Maize pollination	CWRS	AAFC-CRC	
2013	AAC Durafield	Maize pollination	CWAD	AAFC-SPARC	
2014	AAC Connery	Maize pollination	CWRS	AAFC-SPARC	
2014	AAC Elevate	Maize pollination	CWRS	AAFC-LRC	
B. List of Canadian barley cultivars					
1999	Westech	*Bulbosum* pollination	Six-row feed	AAFC-ECORC	Choo et al. (1999)
2003	Cyane	Anther culture	Six-row feed	La Coop Fédérée	
2003	Païdia	Anther culture	Six-row feed	La Coop Fédérée	
2003	Perseis	Anther culture	Six-row feed	La Coop Fédérée	
2007	Corcy	Anther culture	Six-row feed	La Coop Fédérée	
2009	Selena	Anther culture	Six-row feed	La Coop Fédérée	
2009	Norman	Anther culture	Two-row malting	AAFC-Brandon	Legge et al. (2011)
2010	Cerveza	Anther culture	Two-row malting	AAFC-Brandon	Legge et al. (2013a)
2011	Taylor	Anther culture	Two-row hulless malting	AAFC-Brandon	Legge et al. (2013b)

[a]Marketing classes:

Class acronym	Full name of marketing class
CEHRW	Canada Eastern Hard Red Winter
CESRW	Canada Eastern Soft Red Winter
CEWW	Canada Eastern White Winter
CWAD	Canada Western Amber Durum
CWES	Canada Western Extra Strong
CWGP	Canada Western General Purpose
CWHWS	Canada Western Hard White Spring
CWRS	Canada Western Red Spring

(continued)

Table 9.1 (continued)

Year	Name	Method	Marketing class[a]	Breeding institution[b]	References
CWRW	Canada Western Red Winter				
CWSWS	Canada Western Soft White Spring				

[b]Breeding Institutions: *AAFC* Agriculture & Agri-Food Canada, *CRC* Cereal Research Centre, *ECORC* Eastern Cereal and Oilseed Research Centre, *SPARC* Semi-arid Prairie Research Centre, *LRC* Lethbridge Research Centre, *DAS* Dow AgroSciences, *AARD-FCDC* Alberta Agriculture and Rural Development-Field Crop Development Centre

contrast, the spring wheat DH cultivar Lillian occupied up to 18 % (approximately 1.6 million ha) of the spring wheat hectarage in western Canada. The Canada Western Hard White Spring wheat class that reached approximately 450,000 ha of production in 2005, was composed entirely of DH cultivars from 2004 to 2012. DePauw et al. (2011a) reported that DH cultivars have occupied up to 30 % of the total wheat hectarage in western Canada. The blackleg resistant cultivar Quantum occupied as much as 30 % of the rapeseed hectarage in western Canada (Stringam et al. 1995).

9.13 Future Needs in Doubled Haploidy

The needs for DH methodology will depend in part on the crop type. For example, in wheat, a robust, genotype-independent anther or microspore culture DH protocol is still not available. Albinism can still be problematic in some protocols, such as barley anther culture. Methods that require regeneration or embryo generation from callus can suffer from unacceptable somaclonal variation. Thus, ongoing development and improvement of methodology is required. Improved production efficiencies through the use of barcode and computerized labeling and the invention of methods that require even less handling of materials are desirable. The expanded commercial availability of DH production services e.g. Haplotech (www.haplotech.com) and Heartland Plant Innovations (www.heartlandinnovations.com) should ensure the ongoing development and increased application of the technology in plant breeding. Future developments in DH technologies will require a closer integration of DH production with marker-assisted breeding to facilitate selection for major genes as well as quantitatively inherited traits. Due to the advantages that the DH method provides for the application of induced mutagenesis and transgene incorporation, these technologies will likely be deployed to a greater extent in the future; however, their use will be limited to crops where these technologies are accepted by producers, end users and the consumer.

9.14 Conclusions and Prospects

We have demonstrated how marker-assisted breeding can be applied to haploid plants. This will become more sophisticated with the application of array or genotype by sequencing technologies in which the whole genome can be characterized. However, with these genotyping technologies the rate of attrition can be high. The high attrition rates in doubled haploid populations occur because only 50 % of the population is retained for each locus considered. For example only 1 plant would remain from 1000, if 10 loci were evaluated. This outcome is without considering the population size required to account for a reasonable probability of ensuring that 1 remaining plant will have all 10 desired loci.

Microspore culture, with its greater potential to generate haploids and to bring the cost of doubled haploid production down, will be important to the success of genotype-based breeding of DH lines. As the cost of genotyping technologies drops, the feasibility is greatly increased of using such technology for full genome assessment of the genetic potential of each breeding line at an early generation for the purpose of discarding genotypes with poor genetic potential. Genotyping technologies will give the opportunity to measure the genetic potential of a cross between two parents, i.e. selecting at a genotypic level. A choice may be made to abandon the cross of particular parents for the DH stream based on assessment of the parental combining potential. Crosses can be designed based on genotypic information where the focus is on maintaining population size for selection for complex traits such as yield and protein and relegated to a classical inbreeding stream for germplasm development. Other crosses predicted to have a high combining ability would be entered in the DH stream for commercial cultivar development.

Technology for single-gene editing has advanced in recent years to become an option as a breeding tool. Being able to edit a gene sequence in a haploid to modify traits greatly enhances this technology over working with a diploid. With a diploid, gene editing would likely result in a heterozygote or hemizygote depending on the nature of the editing. Such conditions can mask expression of the gene and additional work is required to obtain homozygous true breeding lines. However, by editing the haploid with chromosome doubling, the plant will be in the homozygous condition and expression of the gene will be immediately apparent. Similar to gene editing, transformation technologies are enhanced by working with a haploid compared to a diploid because the transformed locus will undergo doubling along with all other genes. Microspore culture has an advantage over the maize-pollen method in genetic modification systems. As single cells, microspores can be manipulated for transformation or gene editing and all cells in the plant derived from the modified micropsore will be genetically identical. In contrast, it would be technically more challenging to genetically modify the egg cell for subsequent haploid development with the wide-cross approach. The modification would have to be done on a multicellular embryo which would create genetic chimeras that would give rise to variable seeds. A subsequent selection stage would have to be applied not to the

DH_1 (the first generation of doubled haploid) plant as with microspore culture, but to all DH_1 seed produced by each DH_1 plant.

Acknowledgements The authors acknowledge the generous assistance of Dr. Robert Graf, AAFC-Lethbridge for providing the list of wheat DH cultivars, and Dr. A. Choo, Dr. B. Legge and C. Azar for providing information regarding barley DH cultivars. The appendix is based on the wheat maize-pollen doubled haploid manual of Agriculture and Agri-Food Canada-Cereal Research Centre, Prepared by D. Jones and T. Malasiuk of the Agriculture and Agri-Food Canada, Cereal Research Centre, Morden, Manitoba, Canada and G. Humphreys of the Agriculture and Agri-Food Canada, Ottawa Research and Development Centre, Ontario, Canada, Figures 1 and 2 were prepared by Brad Meyer Agriculture and Agri-Food Canada, Swift Current, Saskatchewan.

Appendix: Wheat Maize-Pollen Doubled Haploid Manual of Agriculture and Agri-Food Canada

Maize Pollen Donor Plants

Maize is planted each week to ensure a constant supply of pollen. Approximately 25–30 seeds are placed in a 10 cm paper towel lined plastic pot containing a 6:1 mixture of soil and turface (a soil amendment similar to cat litter) (Fig. 9.3a). From newly planted seeds, plants will take approximately 6–8 weeks before tasseling, so the planting of maize must be started before the planting of spring wheat parent plants. Once the seedlings are approximately 7–10 cm tall, they are transplanted individually into 15 cm clay pots containing the same soil mixture (Fig. 9.3b). The number of seedlings that are transplanted is dependent upon (1) the number of

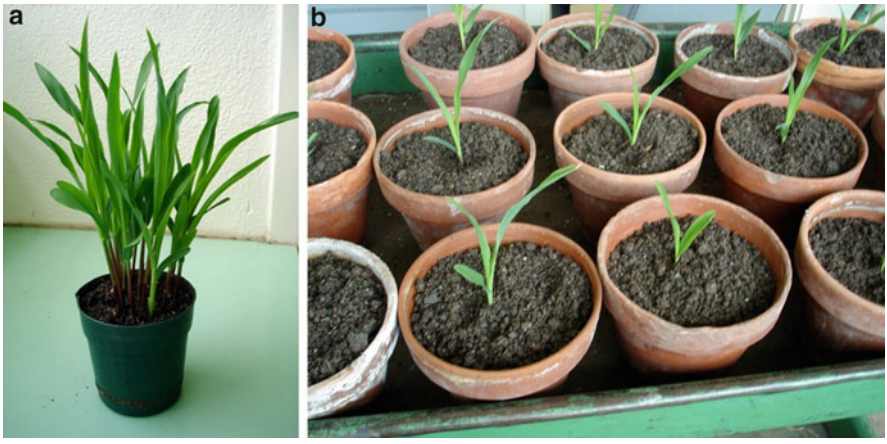

Fig. 9.3 (**a**). Maize 1 week after seeding 25–30 seeds in a 10 cm pot. (**b**) One-week-old maize seedlings transplanted into 15 cm clay pots

parental wheat plants, and (2) the number of spikes being emasculated. If 15 plants are transplanted per week, then at any given time there will be approximately 120 pots of maize at different stages of development. Once the maize is in clay pots the plants should be fertilized every week with a fertilizer solution of 20-20-20 (approx. 10 gL^{-1}) and 2 gL^{-1} of ammonium sulphate (21-0-0). The plants at all stages are grown at 22 °C light/20 °C dark with a photoperiod of 16 h. It is important to have the lights turn on around 7:00 a.m. as pollen shed is affected by light and temperature. The maize pollen starts to shed between 1 and 2 h after the lights are illuminated. Before or shortly after the lights turn on, strip all the old anthers off the tassels. By removing all the old anthers, any new anthers that start to protrude are easily noticed and use of fresh pollen is ensured. As soon as the new anthers start to emerge, either shake the pollen from the tassels onto a large Petri dish or move the maize plants to the area where the pollination will be performed. Once the pollen starts to shed, the maize plants should not be moved as the pollen is easily lost from the anthers. When the maize plants start to silk, the period of pollen shed is nearly complete and the plants should be discarded although, some plants may be kept for seed production. In which case, tap the tassels of one plant onto the silks of the others. Repeat with several plants so that there is a high level of inter-crossing. Isolate the maize plants for seed production in a separate chamber because they tend to harbor mites and aphids which are difficult to control on maize and undesirable on pollen donor plants.

Wheat Parental Plants

Seed of wheat plants to be used to produce doubled haploid lines are started in 8 cm plastic pots containing Sunshine 5 (a soilless mix containing peat moss, perlite and vermiculite). Seed can be planted directly or germinated in dampened filter paper lined Petri dishes. Transplanting from the Petri dish occurs when the shoot is approximately 3–5 cm long. Once the potted plants have reached the two leaf stage they are transplanted into a 25 cm plastic pot (8 L pots commonly used by nurseries) containing a 6:1 mixture of soil and turface (Fig. 9.4). At this stage, the plants are watered with 20-20-20 and ammonium sulphate (21-0-0). It is important to grow large, healthy parent plants with many tillers. The plants are grown at a 16 °C light/15 °C dark with 16 h of light. They are fertilized daily with a siphon mixer of a ratio of 16:1 water to fertilizer (fertilizer 20-20-20, 200 g/20 L). Depending on the genotype of the wheat parent plant, 6–8 weeks is required for spike emergence. If aphids appear they should be spot controlled with a pyrethroid such as Raid especially in the later stages to minimize pesticide exposure of technical staff as the plants will be handled often. Other pesticides (e.g. Pirliss, Cygon, Malathion) and fungicides (e.g. Tilt) do not appear to interfere with doubled haploid production, but there are handling and safety issues associated with their use.

Fig. 9.4 Wheat doubled haploid parental plants with segregating gametes (e.g. F_1) grown in 8 L plastic containers

Crossing

Emasculation

The timing of emasculation is critical when pollinating wheat with maize. A floret approximately 2/3 of the way up the spike should be checked to determine if the spike is ready. If the anthers are starting to change from a light green/yellow color to a bright yellow color and the stigma is fluffy, the spike should be emasculated (Fig. 9.5), which is often the day before natural pollen shed would occur. If the spike has awns, they should be clipped fairly close to the glumes before starting to emasculate. Starting at the bottom of the spike, using fine tipped forceps, remove the central florets of each spikelet leaving only the two lateral florets (Fig. 9.5a). Remove the anthers from the lateral florets being careful not to damage the female structures (Fig. 9.5b). Continue this procedure up to the top of the spike. Remove the very top spikelet and if any of the spikelets near the top or bottom of the spike are too small to work with, they can also be removed. Finish emasculating the other side of the spike. Once the emasculation of the spike is complete, it should be bagged with a glassine bag and closed with a paper clip (Fig. 9.4). Tag the spike with a crossing tag marked with the cross and the date of emasculation. Place the marked tag around the stem at the first internode. A second tag of a different color is placed on the tiller at the second internode. This is an indicator tag and nothing is to be marked on it. It merely alerts the person crossing that something must be done to that particular spike, i.e. pollination or spraying with dicamba.

Fig. 9.5 Spike preparation and emasculation. (**a**) Removal of center floret with forceps. (**b**) Removal of wheat anthers with forceps. (**c**) The *yellow* color of the anthers with a slight *green* tinge indicating the appropriate time for emasculation

Pollination

Being careful not to disturb other anthers, select anthers from the maize plants when they are starting to emerge (Fig. 9.6). The end of the anther should look like it is beginning to open. Remove several anthers from the maize plant. Remove the glassine bag from the emasculated wheat spike to be pollinated (the day after emasculation) and beginning with the bottom of the spike, carefully open the floret. Using tweezers, take a maize anther at its bottom end (opposite of the open end), tip the anther over the open floret and the pollen should flow easily from the anther to the stigma in the floret. One maize anther may be used for several florets. Alternatively, if maize pollen was collected in a Petri dish, use a soft brush to transfer the pollen from the dish to stigmas of the wheat floret. Once the entire spike is pollinated, mark the date of pollination on the crossing tag and replace the bag over the spike. The crossing bag may be left on the spike until the time of excision. The pollinated wheat plants are returned the growth cabinet. Ensure the lights in the growth cabinet are at a sufficient distance from the wheat spikes so the temperature is not elevated in the crossing bags.

Fig. 9.6 Maize anthers ready for pollination

Hormone Spray

The day following pollination, the spikes should be sprayed with a dicamba solution (100 mgL^{-1}). Remove the bag from the spike, and using a small hand-held sprayer, (a pump or a pressurized sprayer) apply the dicamba liberally to the spike, until a run off of liquid is seen (about 2 passes with the pressurized sprayer) (Fig. 9.7). Replace the crossing bag. Mark the tag to indicate the spike was treated. Remove the indicator tag as nothing else needs to be done to the spike until the time of excision.

It is important to keep the parent plants with the crossed spikes under the same conditions as each other (i.e. 15 °C dark/16 °C light, 16 h light) as differing temperatures will affect the timing of excision. If warmer conditions were used the spike would be ready to excise at an earlier date.

The crossing procedure and the hormone spray is a three day process. In order to maximize production, lab work could be scheduled for 7 days a week. If this is not possible there are some modifications to the procedure to allow for a 5-day work week schedule. These include:

1. Reduce the temperature of the cabinet to 9 °C day and night on Friday. This will slow down the development of the plants over the weekend. Spikes which were emasculated on Friday are pollinated on Monday. Spikes that were pollinated on Friday receive the dicamba treatment on Monday.

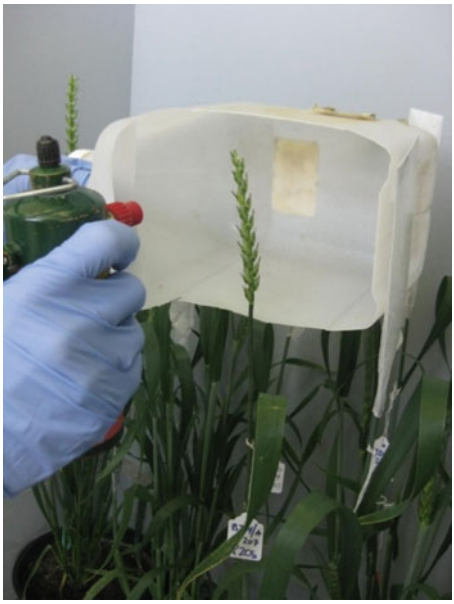

Fig. 9.7 Spray application of dicamba

2. Return the cabinet to normal operating temperature on Monday morning.
3. Slight adjustment of the timing of excision is required. Spikes that were pollinated on Monday and Tuesday are excised at 18 days post pollination.

The timing of the application of dicamba is critical to this procedure. Dicamba treatment approximately 24 h after pollination produces the best results in terms of the number of embryos recovered per spike.

Excision and Embryo Rescue

Culturing of the embryos should be done under sterile conditions in a laminar flow hood. Equipment needed for excision in the laminar flow hood include:

1. Dissecting microscope and light source
2. Sterilized distilled water (for rinsing after bleach/ethanol)
3. 70 % ethanol solution (for sterilization)
4. 50 % bleach solution (for sterilization)
5. 95 % ethanol solution (for wiping down surfaces in the hood)
6. 90 % alcohol for instrument sterilization
7. 3 vials for sterilization of caryopses (1 vial for sterilized water, 1 for 50 % bleach, 1 for 70 % ethanol solution) and 1 vial for 90 % ethanol solution to disinfect scalpel and forceps during excision.
8. Beaker for waste from sterilizing caryopses

9. Pair of forceps
10. Scalpel
11. Cheesecloth for wiping down surfaces with alcohol
12. Sterilized containers of media
13. 10 mL syringes

At 16 days after pollination, cut off the appropriate spikes retaining approximately 15–20 cm of the stem with the spike and place the stems in a container of water until they are ready to be excised. Carefully spread the lemma and palea open with forceps and remove the caryopsis and place into the barrel of a 10 mL syringe. Repeat for all caryopses on the spike. Be careful not to squeeze the caryopsis as they are fluid filled and lack endosperm. Identify the cross in the syringe by affixing the crossing tag to the syringe with tape or by writing the corresponding number of the syringe onto the applicable crossing tag. If more than one spike from a plant has been harvested, they can be combined, however, be sure to keep different genotypes as well as those that have been treated differently, separate. Surface sterilize the caryopses by filling the syringe with the 70 % ethanol solution for 40 s (Fig. 9.8a). Discard the ethanol and rinse the caryopses with sterilized distilled water. Next, draw in the 50 % bleach solution (Javex, approx. 6 % sodium hypochlorite), hold for 60 s, discard and rinse twice with sterilized distilled water.

Place caryopses onto viewing field of dissecting microscope. Dip the forceps and the scalpel in the 90 % ethanol to sterilize and allow to dry momentarily. Using forceps and a scalpel, carefully open each caryopsis. This can be achieved in one of two ways: (1) At the embryo end, hold the caryopsis with the forceps, using the scalpel cut the brush end of the caryopsis. Tilt the open end of the caryopsis towards the microscope with the forceps; (2) Steady the caryopsis with the forceps or scalpel and use another set of forceps to grab and tear the seed coat about two-thirds of the way from the embryo end and pull back exposing the embryo. The embryo is a

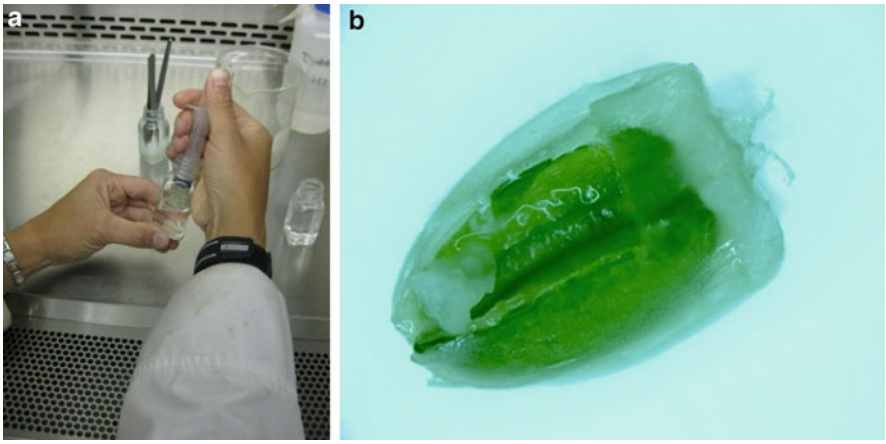

Fig. 9.8 (**a**) Sterilization of caryopses. (**b**) Wheat caryopsis with developing embryo

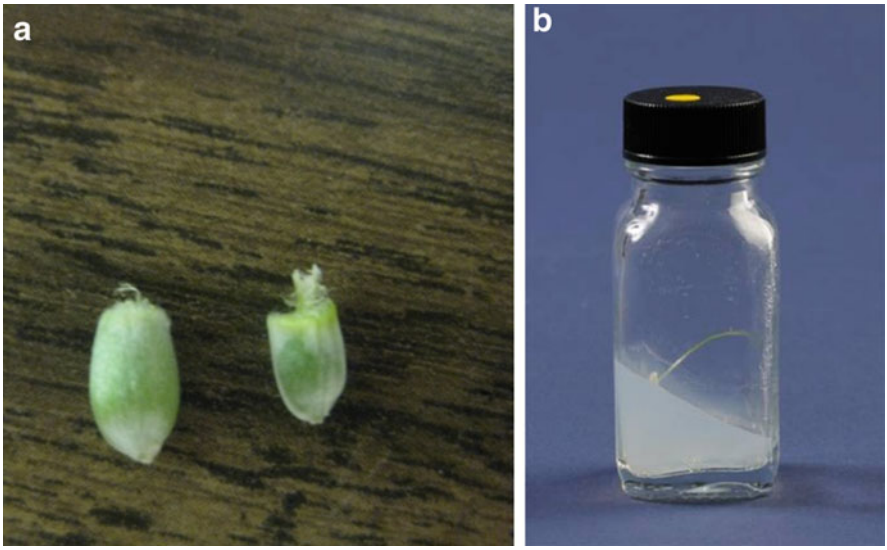

Fig. 9.9 (**a**) Selfed wheat caryopsis with endosperm development (*left*) compared to a caryopsis with a haploid embryo and no endosperm development (*right*). (**b**) Vial containing developing haploid plantlet on culture media

small white or somewhat translucent structure in the caryopsis. Because there is no solid endosperm present, the embryo is easy to see (Figs. 9.8b and 9.9a). Once the embryo is found, carefully touch the forceps to the embryo. The embryo will stick to the forceps and then it can be placed on the sterilized growth media. Open a vial of media in the flow hood, and touch the embryo to the media surface about half way up the slant. Keep the vials containing embryos from the different parental crosses separated. Only one embryo is placed in each vial (Fig. 9.9b).

Once excision is complete, the embryos are given a 3 day cold and darkness treatment in a fridge at 4 °C. After the cold treatment, the embryos are placed in the dark at room temperature for 2 days then placed under light banks for a 16-h light/8 h dark cycle at room temperature. A combination of fluorescent and incandescent light is used. The temperature under the lights is usually around 23–25 °C. Each week, all the vials under the light banks are checked to see if the embryos are growing and are ready for transplanting.

Second Cold Treatment

If the embryos fail to grow after 2 weeks under the light banks, they are given a second cold treatment. The dormant 2-week-old embryos are placed back in the fridge for 2 weeks. At this time the vials are placed back under the light banks (a separate set of light banks are used to regulate the flow of vials from the first light treatment to the fridge and back under the lights after a second cold treatment.)

These vials are checked for growth periodically and returned to the main light banks when significant shoots and roots are visible. After 4 months, those embryos which have received a second cold treatment and were returned to under the light banks and still show no signs of growth are discarded.

Medium

The medium constituents are agar, sucrose and Gamborg's B5 medium. From one liter of media, 120 one-oz Quorpak vials are filled with approximately 8 mL of medium. The medium is autoclaved in the vials and cooled on a slant.

Transplanting Haploid Plants

Once the embryos grow to a size of 2–4 cm, they are ready to be transplanted. Haploids are transplanted into 14 cm long plastic Conetainers (like those used in the forestry industry) containing Sunshine 5 soilless mix which has been loosely packed and watered.

To remove the haploid from the vial, use a blunt spoonula and dig around the root, freeing it completely, and gently tip the vial over and withdraw the plantlet on the instrument. Dip the root into water to completely remove the agar from the plantlet. A hole is punched in the Sunshine 5 medium and the plantlet is placed so that the crown is covered. Firm the mix around the plantlet.

The tray of Conetainers is subsequently set into a tray of water and covered with a plastic lid to provide a humid environment. The lid is left on for 3 days or until the haploids are established. These haploid plantlets are kept at 15 °C with a 16 h light/8 h dark cycle. It is important to keep the developing haploid plantlets cool so that tillering is encouraged.

Colchicine Treatment – Doubling

Plant Preparation

When the haploid plantlets are between the three- and five-leaf stage, they are ready for the colchicine treatment. In order to identify the plants when they are removed from the cones, the pot stake has a hole punched, and an elastic string (approximately 10 cm long) is fed through this hole and tied in a knot at the ends. The elastic is pulled around the base of the plant, the stake is then fed through the loop of the elastic pulled through to ensure a snug fit. This step can be done the day before the plants are to be colchicine treated.

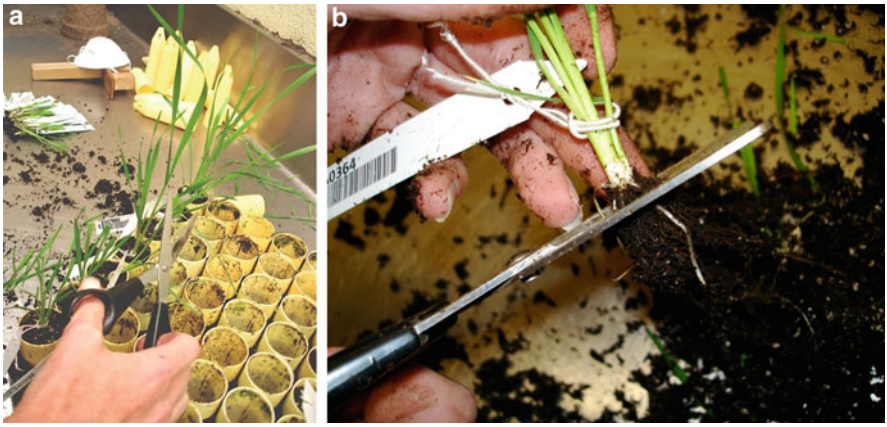

Fig. 9.10 (**a**) Trimming leaves of haploids. (**b**) Trimming roots of haploids

When preparing the plants, about 1/3 of the leaf tissue is trimmed off (Fig. 9.10a). Next, the plants are removed from the cones and the soil is stripped from the roots. The roots are cut back leaving about 5 mm and given a quick rinse with tap water (Fig. 9.10b). It is important to prepare the plants first thing in the morning because the colchicine treatment process requires steps throughout the day.

Colchicine Treatment and Rinse

Once the plant's leaves and roots have been trimmed they are placed in a beaker (a 250 ml beaker will hold about 100 plants) and are ready for the colchicine treatment (Fig. 9.11).

When working with colchicine, it must be handled very carefully as it is highly toxic. It is recommended to wear two pairs of nitrile gloves. Six drops (from an eye dropper) of DMSO (dimethyl sulfoxide) is added to 60 ml of colchicine (1 drop DMSO/10 ml of colchicine). This solution is transferred to the beaker containing the trimmed haploids (the crowns of the plants should be submerged). An aerator (a fish tank air pump will do) with a rubber tube leading from it, is placed in the beaker. The beaker is placed in a plastic pail and covered with a lid. A hole is made in the lid to allow the air tube to pass through. The covered pail is placed in a fume hood with the sash ajar to allow only for the air tube to run from the beaker to the air pump. This step ensures that the procedure is exposed to as little light as possible as well as containment of any chemical which may percolate from the beaker. The plants are treated for 3½h at room temperature. Pour the used colchicine from the beaker into a labelled container for use in a subsequent day's treatment (may be used twice, no additional DMSO is required). Once the colchicine solution has been used twice it is discarded into a container designated for toxic waste disposal.

Fig. 9.11 Plants in container for colchicine treatment

To rinse residual colchicine from the plants, place the aeration tube back into the beaker (in sink) along with a rubber hose leading from the faucet (a backflow valve should be installed on the faucet). Have the water running at a flow that is strong enough to allow the water to trickle out of the beaker. Adjust the cold and hot water taps so that the temperature of the rinse water is between 15–18 °C. Rinse for a minimum of 2 h.

Replanting Colchicine Treated Haploids

After rinsing, the colchicine treated haploid plants are replanted into 20 cm long Conetainers filled with Sunshine 5 medium and then moistened. This procedure is similar to the initial planting of the young haploids where the cones are watered, holes are punched into the surface of the soil medium to allow for replanting up to the crown and the soilless mix is gently firmed around the plant. The stakes are slid off the plant before transplanting and stuck into the edge of the cone. The trays are then covered with a plastic lid and set in tubs filled with water overnight to allow for sufficient water uptake by the Sunshine 5. The tops are taken off the trays 3 days after colchicine treatment. The colchicine treated plants are placed in a growth cabinet set at 15 °C with a cycle of 16 h of light and 8 h of darkness. They are fertilized daily with a 20-20-20 fertilizer solution (approx. 10 g L^{-1} stock solution) siphoned into the watering line (diluted 16:1).

Solution Preparation

Dicamba 100 mgL^{-1} (100 ppm)

Items Required

250 ml labelled brown glass bottle (with cap)	10 mL graduated cylinder
50 ml beaker	500 mL graduated cylinder
2 cm stir bar	Small weigh boat
25 mg dicamba	Stir plate
15 mL 95 % ethanol	Analytical balance
235 mL double distilled water	

Preparation Instructions Into a 50 mL beaker add the stir bar, 15 mL 95 % ethanol and 25 mg dicamba, stir to dissolve. Pour the dicamba solution into the 250 mL brown bottle, triple rinse the 50 mL beaker and add the rinseate and the remaining water to the 250 bottle.

Colchicine 0.2 % (w/v)

Items Required

500 mL labelled brown glass bottle (with lid)	500 mL distilled water
2 cm stir bar	500 mL graduated cylinder
1 g colchicine	Stir plate

Preparation Instructions Prepare the colchicine in a fume hood being sure to wear a lab coat, eye protection and two pairs of nitrile gloves. Pour approximately 200 mL of distilled water into the 500 mL bottle along with the stir bar. Carefully pour the pre-measured 1 g container of colchicine into the 500 mL bottle, using the water in the beaker to triple rinse out the container. Pour the remaining water into the 500 mL bottle and cap. The colchicine can be stirred with the magnetic stirrer outside of the fume hood.

Embryo Culture Medium

Items Required

2 L beaker	0.1 M HCl
40 g sucrose	Weigh boats
2 L distilled water	Analytical balance
20 g agar	240 1-oz Quorpak vials

6.4 g Gamborg's B5	Autoclavable boxes
5 cm stir bar	Pipette
Hot plate	500 mL graduated cylinder
pH meter	Thermometer
0.1 M NaOH	

Preparation Instructions Into a 2 L beaker place the stir bar and add 2 L of distilled water. Place the beaker on the hot plate and begin heating. Measure out the agar, sucrose and Gamborg's B5 solids. Turn on the stir bar. Add the agar to the 2 L beaker of water. Continue heating until the temperature reaches 85–90 °C. Turn off the hot plate and stirrer and remove the beaker to cool it down. Once the temperature drops below 80 °C place the beaker back onto the stirrer and continue stirring while adding the sucrose and Gamborg's B5 solids. Place the pH electrode in the medium and adjust the pH to 6.10 using the NaOH or HCl. Dispense the medium into vials. This can be done by pouring medium into a small beaker (400 mL size) and pouring into vials by hand (approximately 8 mL per vial). A faster and more accurate way is to use an automatic dispenser. Cap the vials, but not too tightly and place them in an autoclavable container. Autoclave the vials on the liquid cycle (15 psi, 121 °C) for 15 min. Cool the vials in the autoclavable containers on a slant overnight, then store under refrigeration.

References

Agache S, Bachelier J, de Buyser J et al (1989) Genetic analysis of anther-culture response in wheat using aneuploid, chromosome substitution and translocation lines. Theor Appl Genet 77:7–11

Al-Ashkar I (2013) Anther culture response and salt tolerance in some wheat genotypes. Ann Agric Sci 58(2):139–145

Ankele E, Heberle-Bors E, Pfosser MF et al (2005) Searching for mechanisms to albino plant formation in cereals. Acta Physiol Plant 27(4B):651–664

Asif M, Eudes F, Randhawa H et al (2014) Induction medium osmolality improves microspore embryogenesis in wheat and triticale. In Vitro Cell Dev Biol Plant 50:121–126. doi:10.1007/s11627-013-9545-5

Badiyal A, Chaudhary HK, Jamwal NS et al (2014) Interactive genotypic influence of triticale and wheat on their crossability and haploid induction under varied agroclimatic regimes. Cereal Res Commun 1(1):1–10. doi:10.1556/CRC.2014.0017

Barclay IR (1975) High frequencies of haploid production in wheat (*Triticum aestivum*) by chromosome elimination. Nature 256:410–411

Barkley A, Chumley FG (2012) A DH laboratory for Kansas wheat breeding: an economic analysis of biotechnology adoption. Int Food Agribus Rev 15(2):99–120

Barnabas B (2003) Protocol for producing doubled haploid plants from anther culture of wheat (*Triticum aestivum* L.). In: Maluszynski M, Kasha K, Forster BP, Szarejko I (eds) Doubled haploid production in crop plants. A manual. Kluwer Academic Publishers, Norwell, pp 65–70

Blakeslee AF, Belling J, Farnham ME et al (1922) A haploid mutant in the Jimson weed, *Datura stramonium*. Science 55:646–647

Cao W, Wolf D, Fedak G et al (2014) Development of aneuploids by means of doubled haploid technology-corn pollination method. Proceedings of the Canadian Society of Agronomy and Canadian Society of Horticultural Science joint annual conference, pp 15–16
Castillo AM, Cistué L, Vallés MP et al (2009) Chromosome doubling in monocots. In: Touraev A, Forster BP, Jain SM (eds) Advances in haploid production in higher plants. Springer, Berlin, pp 329–338
Chase SS (1952) Production of homozygous diploids of maize from monoploids. Agron J 44:263–267
Choo TM, Ye JM, Martin RA et al (1999) AC Westech barley. Can J Plant Sci 79:249–250
Cistué L, Vallés MP, Echavarri B et al (2003) Barley anther culture. In: Maluszynski M, Kasha K, Forster BP, Szarejko I (eds) Doubled haploid production in crop plants. A manual. Kluwer Academic Publishers, Norwell, pp 29–34
Collard BCY, Jahufer MZZ, Brouwer JB et al (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169–196
Cook JP, Wichman DM, Martin JM et al (2004) Identification of microsatellite markers associated with a stem solidness locus in wheat. Crop Sci 44:1397–1402
Cuthbert PA, Somers DJ, Thomas J et al (2006) Fine mapping *Fhb1*, a major gene controlling fusarium head blight resistance in bread wheat (*Triticum aestivum* L.). Theor Appl Genet 112:1465–1472
Dash MC, Dash SP (2009) Fundamentals of ecology, 3rd edn. Tata McGraw-Hill, New Delhi
DePauw RM, Townley-Smith TF, Humphreys G et al (2005) Lillian hard red spring wheat. Can J Plant Sci 85:397–401
DePauw RM, Knox RE, Clarke FR et al (2009) Stettler hard red spring wheat. Can J Plant Sci 89:945–951
DePauw RM, Knox RE, Humphreys DG et al (2011a) New breeding tools impact Canadian commercial farmer fields. Czech J Genet Plant Breed 47(1):S28–S34
DePauw RM, Knox RE, McCaig TN et al (2011b) Carberry hard red spring wheat. Can J Plant Sci 91:529–534
DePauw RM, Knox RE, McCaig TN et al (2011c) Muchmore hard red spring wheat. Can J Plant Sci 91:797–803
Devaux P, Kilian A, Klienhofs A (1995) Comparative mapping of the barley genome with male and female recombination-derived, doubled haploid populations. Mol Gen Genet 249:600–608
El-Hennawy MA, Abdalla AF, Shafey SA et al (2011) Production of doubled haploid wheat lines (*Triticum aestivum* L.) using anther culture technique. Ann Agric Sci 56:63–72
Eudes F, Chugh A (2009) An overview of triticale doubled haploids. In: Touraev A, Forster B, Jain S (eds) Advances in haploid production in higher plants. Springer, Berlin, pp 87–96
Foisset N, Delourme R (1996) Segregation distortion in androgenic plants. In: Jain SM, Sopory SK SK, Veilleux RE (eds) In vitro haploid production in higher plants. Kluwer Academic Publishers, Netherlands, pp 189–201
Foroughi-Wehr L, Friedt W, Wenzel G (1982) On the genetic improvement of androgenetic haploid formation in *Hordeum vulgare*. Theor Appl Genet 62:233–239
Forster BP, Thomas WTB (2005) Double haploids in genetics and plant breeding. Plant Breed Rev 25:57–88
Fowler DB (2010) Peregrine hard red winter wheat. Can J Plant Sci 90:853–856
Fowler DB (2011) Accipiter hard red winter wheat. Can J Plant Sci 91:363–365
Fowler DB (2012) Sunrise soft red winter wheat. Can J Plant Sci 92:195–198
Fowler DB (2013) Swainson hard red winter wheat. Can J Plant Sci 93:1257–1259
Fowler DB (2014) CDC chase hard red winter wheat. Can J Plant Sci 94:183–186
Fox SL, Townley-Smith TF, Humphreys DG et al (2006) Somerset hard red spring wheat. Can J Plant Sci 86(1):163–167. doi:10.4141/P05-117

Fox SL, Thomas JB, Wise IL et al (2009) Waskada hard red spring wheat. Can J Plant Sci 89:929–936
Fox SL, Wise IL, Smith MA et al (2013) Shaw hard red spring wheat. Can J Plant Sci 93:299–305
Friedt W, Breun S, Zucher S et al (1986) Comparative value of andro-genetic doubled haploid and conventionally selected barley lines. Plant Breed 97:56–63
Gieger HH (2009) DHs. In: Bennetzen J, Hake S (eds) Handbook of maize. Volume II: genetics and genomics. Springer Science and Business Media, New York, pp 641–657
Graf RJ, Hucl P, Orshinsky BR et al (2003) McKenzie hard red spring wheat. Can J Plant Sci 83:565–569
Graf RJ, Thomas JB, Beres BL et al (2012) Flourish hard red winter wheat. Can J Plant Sci 92:183–189
Graf RJ, Beres BL, Laroche A et al (2013) Emerson hard red winter wheat. Can J Plant Sci 93:741–748
Grauda D, Lepse N, Strazdiņa V et al (2010) Obtaining of doubled haploid lines by anther culture method for the Latvian wheat breeding. Agron Res 8:545–552
Guha S, Maheshwari SC (1964) *In vitro* production of embryos from anthers of *Datura*. Nature 204(4957):497
Guzy-Wrobelska J, Szarekjo I (2003) Molecular and agronomic evaluation of wheat DH lines obtained through maize pollination and anther culture methods. Plant Breed 122:305–313
Guzy-Wróbelska J, Labocha-Pawlowska A, Kwasniewski M et al (2007) Different recombination frequencies in wheat doubled haploid populations obtained through maize pollination and anther culture. Euphytica 156(1):173–183. doi:10.1007/s10681-007-9364-z
Henry J, de Buyser J (1985) Effect of the 1B/1R translocation on anther-culture ability in wheat (*Triticum aestivum* L.). Plant Cell Rep 4:307–310
Houshmand S, Knox RE, Clarke FR et al (2007) Microsatellite markers flanking a stem solidness gene on chromosome 3BL in durum wheat. Mol Breed 20:261–270
Huang XQ, Cloutier S, Lycar L et al (2006) Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.). Theor Appl Genet 113(4):753–766
Humphreys DG, Townley-Smith TF, Czarnecki E et al (2006) Kanata hard white spring wheat. Can J Plant Sci 87:879–882
Humphreys DG, Townley-Smith TF, Czarnecki E et al (2007) Snowbird hard white spring wheat. Can J Plant Sci 87:301–305
Humphreys DG, Townley-Smith TF, Lukow OM et al (2010) Burnside extra strong hard red spring wheat. Can J Plant Sci 90(1):79–84. doi:10.4141/CJPS09100
Humphreys DG, Townley-Smith TF, Czarnecki E et al (2013) Snowstar hard white spring wheat. Can J Plant Sci 93(1):143–148. doi:10.4141/cjps2012-14693
Iyamabo OE, Hayes PM (1995) Effects of selection and opportunities for recombination in doubled-haploid populations of barley (*Hordeum vulgare* L.). Plant Breed 114:131–136
Jain SM, Sopory SK, Veilleux RE (eds) (1996) In vitro haploid production in higher plants, Vol 4-Cereals. Kluwer Academic Publishers, Dordrecht
Jansen RC (1992) On the selection for specific genes in doubled haploids. Heredity 69:92–95
Joppa LR, Du C, Hart GE et al (1997) Mapping gene(s) for grain protein in tetraploid wheat (*Triticum turgidum* L.) using a population of recombinant inbred chromosome lines. Crop Sci 37:1586–1589
Kasha K, Maluszynski M (2003) Production of doubled haploids in crop plants. An introduction. In: Maluszynski M, Kasha K, Forster BP, Szarejko I (eds) Doubled haploid production in crop plants. A manual. Kluwer Academic Publishers, Norwell, pp 1–4
Kasha KJ, Kao KN (1970) High frequency haploid production in barley (*Hordeum vulgare* L.). Nature 225:874–876
Kim KM, Baenziger PS (2005) A simple wheat haploid and doubled haploid production system using anther culture. In Vitro Cell Dev Biol Plant 41:22–27

Knox RE, Fernandez MR, Brule-Babel AL et al (1998) Inheritance of common bunt resistance in androgenetically derived doubled haploid and random inbred populations of wheat. Crop Sci 38:1119–1124
Knox R, Clarke JM, DePauw RM (2000) Dicamba and growth condition effects on DH production in durum wheat crossed with maize. Plant Breed 119:289–298
Knox RE, DePauw RM, Clarke FR et al (2008) Alvena hard red spring wheat. Can J Plant Sci 88:513–518
Laurie DA (1989) The frequency of fertilization in wheat×pearl millet crosses. Genome 32:1063–1067
Laurie DA, Bennett MD (1988) The production of haploid wheat plants from wheat×maize crosses. Theor Appl Genet 76:393–397
Legge WG, Tucker JR, Bizimungu B et al (2011) Norman barley. Can J Plant Sci 91:1105–1113
Legge WG, Tucker JR, Bizimungu B et al (2013a) Cerveza barley. Can J Plant Sci 93:557–564
Legge WG, Tucker JR, Bizimungu B et al (2013b) Taylor barley. Can J Plant Sci 93:969–977
Liu S, Rudd JC, Bai G et al (2014) Molecular markers linked to genes important for hard winter wheat production and marketing in the U.S. Great Plains. Crop Sci 54:1304–1321
Maluszynski M, Kasha K, Forster BP et al (eds) (2003) Doubled haploid production in crop plants. A manual. Kluwer Academic Publishers, Norwell
McCartney CA, Somers DJ, Humphreys DG et al (2005) Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452x'AC Domain'. Genome 48(5):870–883
McCartney CA, Somers DJ, Lukow OM et al (2006) QTL analysis of quality traits in the spring wheat cross RL4452×'AC Domain'. Plant Breed 125(6):565–575
Mochida K, Tsujimoto H (2001) Production of wheat doubled haploids by pollination with Job's tears (*Coix lachryma-jobi* L.). J Hered 92(1):81–83
Mujeeb-Kazi A (2003) Wheat improvement facilitated by novel genetic diversity and *in vitro* technology. Plant Tissue Cult 13(2):179–210
Mujeeb-Kazi A, Riera-Lizarazu O, William M (1995) Production of polyhaploid wheat plants using maize and tripsacum. In: Mujeeb-Kazi A, Hettel GP (eds) Utilizing wild grass biodiversity in wheat improvement: 15 years of wide cross research at CIMMYT. CIMMYT Res Rep No 2. Mexico, DF, CIMMYT, pp 47–65
Murovec J, Bohanec B (2012) Haploids and doubled haploids in plant breeding. In: Abdurakhmonov I (ed) Plant breeding. INTECH. ISBN:978-953-307-932-5. doi:10.5772/29982
Navarro-Alvarez W, Baenziger PS, Eskridge KM et al (1994) Addition of colchicine to wheat anther culture media. Plant Breed 112:192–198
Niu Z, Jiang A, Hammad A et al (2014) Review of doubled haploid production in durum and common wheat through wheat×maize hybridization. Plant Breed 113:313–320
Pauk J, Mihaly R, Puolimatka M (2003) Protocol for wheat (*Triticum aestivum* L.) anther culture. In: Maluszynski M, Kasha K, Forster BP, Szarejko I (eds) Doubled haploid production in crop plants. A manual. Kluwer Academic Publishers, Norwell, pp 59–64
Randhawa HS, Sadasivaiah R, Graf RJ et al (2011) Bhishaj soft white spring wheat. Can J Plant Sci 91:805–810. doi:10.4141/cjps10144
Randhawa HS, Asif M, Pozniak C et al (2013) Application of molecular markers to wheat breeding in Canada. Plant Breed 132:458–471
Rasul G, Humphreys DG, Brûlé-Babel AL et al (2009) Mapping QTLs for pre-harvest sprouting traits in the spring wheat cross 'RL4452/AC Domain'. Euphytica 168(3):363–378
Redha A, Talaat A (2008) Improvement of green plant regeneration by manipulation of anther culture induction medium of hexaploid wheat. Plant Cell Tiss Org Cult 92:141–146
Sadasivaiah RS, Orshinsky BR, Kozub GC (1999) Production of wheat haploids using anther culture and wheat×maize hybridization techniques. Cereal Res Commun 27:33–40
Sadasivaiah RS, Perkovic SM, Pearson DC et al (2004) Registration of 'AC Andrew' wheat. Crop Sci 44:696–697

Sayed H, Kayyal H, Ramsey L et al (2002) Segregation distortion in doubled haploid lines of barley (*Hordeum vulgare* L.) detected by simple sequence repeat (SSR). Euphytica 225:265–272

Singh AK, Clarke JM, Knox RE et al (2012) Transcend durum wheat. Can J Plant Sci 92:809–813

Singh AK, Clarke JM, Knox RE et al (2014) AAC Raymore durum wheat. Can J Plant Sci 94(7):1289–1296. doi:10.4141/CJPS-2014-048

Snape JW (1976) A theoretical comparison of diploidised haploid and single seed decent populations. Heredity 36:275–277

Snape JW, Simpson E (1981) The genetical expectations of doubled haploid lines derived from different filial generations. Theor Appl Genet 60:123–128

Snape JW, Simpson E, Parker BB et al (1986) Criteria for the selection and used doubled haploid systems in cereal breeding programmes. In: Horn W, Jensen CJ, Odenbach W, Shieder O (eds) Genetic manipulation in plant breeding. Walter de Gruyter, Berlin, pp 217–229

Somers DJ, Humphreys G (2009) State of QTL detection and marker-assisted selection in wheat improvement. In: Carver B (ed) Wheat: science and trade. Blackwell Publishing Inc, USA, pp 309–326

Somers DJ, McCartney CA, DePauw RM et al (2007) Molecular breeding for multiple pest resistance in wheat. Dev Plant Breed 12:667–676. doi:10.1007/1-4020-5497-1_81

Strigens A, Schipprack W, Reif JC et al (2013) Unlocking the genetic diversity of maize landraces with doubled haploids opens new avenues for breeding. PLoS One 8(2):e57234. doi:10.1371/journal.pone.0057234

Stringam GR, Degenhardt DF, Thiagarajah MR et al (1995) Development of an agronomically superior blackleg resistant canola cultivar in *Brassica napus* L. using doubled haploidy. Can J Plant Sci 75(2):437–439

Suenaga K, Morshedi AR, Darvey NL (1998) Evaluation of teosinte lines as pollen parents for wheat haploid production. Cereal Res Comm 26:119–125

Szarejko I (2003a) Anther culture for doubled haploid production in barley (*Hordeum vulgare* L.). In: Maluszynski M, Kasha K, Forster BP, Szarejko I (eds) Doubled haploid production in crop plants. A manual. Kluwer Academic Publishers, Norwell, pp 35–42

Szarejko I (2003b) Doubled haploid mutant production. In: Maluszynski M, Kasha K, Forster BP, Szarejko I (eds) Doubled haploid production in crop plants. A manual. Kluwer Academic Publishers, Norwell, pp 351–361

Szarejko I, Falk DE, Janusz A (1997) Cytological and genetic evaluation of anther culture derived doubled haploids in barley. J Appl Genet 38(4):437–452

Tadesse W, Inagaki M, Tawkaz S et al (2012) Recent advances and application of doubled haploids in wheat breeding. Afr J Biotechnol 11(89):15484–15492. doi:10.5897/AJB12.2124

Tamburic-Ilincic L, Smid A (2013a) OAC Emmy soft white winter wheat. Can J Plant Sci 93:965–967

Tamburic-Ilincic L, Smid A (2013b) OAC Flight soft red winter wheat. Can J Plant Sci 93:1261–1263

Thomas WTB, Forster BP, Gertsson B (2003) Doubled haploids in breeding. In: Maluszynski M et al (eds) Doubled haploid production in crop plants. A manual. Kluwer Academic Publishers, Norwell, pp 337–350

Townley-Smith TF, Humphreys DG, Czarnecki E et al (2010) Superb hard red spring wheat. Can J Plant Sci 90:347–352

Ushiyama T, Shimizu T, Kuwabara T (1991) High frequency of haploid production of wheat through intergenetic cross with teosinte. Japan J Breed 41:353–357

Wessels E, Botes WC (2014) Accelerating resistance breeding in wheat by integrating marker-assisted selection and doubled haploid technology. South Afr J Plant Soil 31(1):35–43. doi:10.1080/02571862.2014.903434

Zhang X, Wang XP, Jing J (2006) Characterization of wheat-triticale doubled haploid lines by cytological and biochemical markers. Plant Breed 117(1):7–12. doi:10.1111/j.1439-0523.1998.tb01439.x

Zhang K, Tian J, Zhao L et al (2008) Mapping QTLs with epistatic effects and QTL×environment interactions for plant height using a doubled haploid population in cultivated wheat. J Genet Genomics 35(2):119–127. doi:10.1016/S1673-8527(08)60017-x

Zhang H, Tan E, Suzuki Y et al (2014) Dramatic improvement in genome assembly achieved using doubled-haploid genomes. Sci Rep 4:6780, doi.org/10.1038/srep06780

Zhao X, Qin Y, Jia B et al (2013) Comparison and analysis of QTLs, epistatic effects and QTL×environment interactions for yield traits using DH and RILs populations in rice. J Integr Agric 12(2):198–208

Zheng MY (2003) Microspore culture in wheat (*Triticum aestivum*) – doubled haploid production via induced embryogenesis. Plant Cell Tiss Org Cult 73:213–230

Part III
In Vitro Culture and Transgenic Approaches

Chapter 10
Applications of In Vitro Techniques in Plant Breeding

Zul Zulkarnain, Tanya Tapingkae, and Acram Taji

Abstract Although the Green Revolution of the 1960s tripled the world's food supply, food distribution remains uneven and the problem of chronic under-nutrition for millions of people in the world persists. In the face of predicted increases in the world population to around 10 billion by 2050 and the challenges faced by agriculture as a result of climate change, providing adequate food and fiber for humanity is a pressing issue requiring urgent attention. Since more and more of the arable land is being used to house the growing world population, provide feed for stock to supply animal protein and to grow crops for bio-energy, how could agriculture keep pace and remain productive without further degradation of the soil or damage to the environment? Could biotechnology be a key to solving world hunger given the challenges of climate change and immense population growth? This chapter examines recent advances in the application of a number of biotechnological techniques used in in vitro plant breeding including embryo rescue, somatic embryogenesis, in vitro pollination, flowering and fertilization as well as protoplast and somatic hybridization. A special focus has been given to exploitation of somaclonal variation in production of plants with better yield attributes as well as the ability to better cope with biotic and abiotic stresses. These techniques have the potential to increase food supply. The chapter overviews our collective experience working in this field over the past 30 years.

Keywords In vitro flowering • Protoplast • Somaclonal variation • Somatic embryogenesis • Somatic hybridization

Z. Zulkarnain (✉)
Faculty of Agriculture, University of Jambi, Jambi 36361, Indonesia
e-mail: zulkarnain@unja.ac.id; dr.zulkarnain@yahoo.com

T. Tapingkae
Faculty of Agricultural Technology, Chiang Mai Rajabhat University, Chiang Mai, Thailand
e-mail: ttapingkae@hotmail.com

A. Taji
School of Agricultural, Computational and Environmental Sciences, The University of Southern Queensland, Toowoomba, QLD 4350, Australia
e-mail: acram.taji@usq.edu.au

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_10

10.1 Introduction

Research in plant tissue culture over the past several decades has led to the development of techniques now used commercially across the globe to rapidly multiply a wide range of crops and improve their production systems. Plant tissue culture technology began with Gottlieb Haberlandt's theory of cell totipotency at the beginning of twentieth century (see Vasil 2008). Following on from this, the discovery of auxins by Frits Warmolt Went in 1926 (see Pennazio 2002), and cytokinins by Folke Skoog and colleagues in the 1950s (see Kieber 2002), led to the first success of in vitro techniques in plant tissues. Since then, the technology has developed considerably and now plays a key role in genetic engineering and crop improvement. Plant tissue culture offers an array of techniques that complement conventional plant propagation and plant breeding methods. The most common reasons for the use of in vitro techniques has been for plant propagation, but its most important application in recent years has been to crop improvement using gene technology (Khan 2009; Takeda and Matsuoka 2008; Thakur et al. 2012). Techniques such as in vitro fertilization and protoplast fusion enable the recombination of genotypes otherwise limited by incompatibility (Sri Rama Murthy et al. 2012; Tapingkae et al. 2012; Wang et al. 2006). Conventional breeding can be hastened by exploiting increased genetic diversity resulting from somatic variability (Bairu et al. 2011; Nwauzoma and Jaja 2013).

This chapter is aimed at introducing the reader to recent advances in the application of in vitro breeding techniques including embryo rescue, and utilizing somatic embryogenesis and somaclonal variation in plant breeding and mass propagation. In vitro flowering, pollination and fertilization as well as protoplast culture technologies and their relevance to plant breeding are also reviewed. These methods are now well-established technologies that have made significant contributions to plant improvement and mass propagation in horticulture, agriculture, and to some extent forestry.

10.2 Embryo Rescue

The term *embryo rescue* is used to describe the in vitro techniques aiming to encourage the development of immature embryos into complete plants. This technique has been widely used to avoid embryo abortion in regenerated plants from hybridization. The technique of embryo rescue has become an important tool in plant breeding, allowing the formation of many interspecific and intergeneric crop species.

Embryo rescue, also known as *embryo culture*, involves the excising of embryos and placing them onto sterile culture medium. The technique was first developed by Tukey in 1933 who successfully grew the embryo of cherry on an artificial medium. Since then, the procedure has been applied in embryo rescue of many other crops, such as *Lilium* (Chi 2002; Proseviçius and Strikulyte 2004), *Gossypium* (Mehetre

and Aher 2004), *Malus* (Dantas et al. 2006), *Prunus* (Kukharchyk and Kastrickaya 2006), *Elaeis* (Alves et al. 2011), various tree fruits (Fathi and Jahani 2012) and *Capsicum* (Debbarama et al. 2013).

Major application of embryo culture in plant breeding has been for interspecific or intergeneric hybridization, in which the endosperm develops poorly or does not develop at all due to hybridization barriers. Embryo culture can also be applied to shorten the breeding cycle by overcoming dormancy in seeds. Factors such as endogenous inhibitors, light, temperatures, humidity or embryo immaturity often cause seed dormancy to occur. By removing the embryos from the influences of these factors, they may germinate and grow quickly, and as a consequence the breeding cycle is shortened.

10.2.1 Interspecific and Intergeneric Hybridization

The endosperm of interspecific or intergeneric crosses, as well as crosses between diploids and tetraploids, often develop poorly or not at all, resulting in nonviable embryos. Such embryos, however, may have the potential for initiating growth and subsequent development. By aseptically isolating and culturing them in a nutrient medium, this hybridization barrier may be overcome. Stebbins (1950) proposed two types of hybridization barriers, pre-fertilization and post-fertilization barriers. The pre-fertilization barriers include those mechanisms which prevent fertilization to occur and can be the result of geographical distance, apomixes as well as pollen-pistil incompatibilities; whereas post-fertilization barriers can be caused by ploidy differences, chromosome elimination and seed dormancy.

The embryo rescue technique has been successfully applied to overcome the post-fertilization barrier which has been a great hindrance to hybridization in plant breeding. This technique has been applied in rescuing young embryos of successful interspecific crosses of *Lycopersicon esculentum*×*L. peruvianum* (Thomas and Pratt 1981) and *Medicago sativa*×*M. rupestris* (McCoy 1985). Fertile hybrid progenies were produced from the intergeneric cross between *Brassica napus* and *Sinapsis alba* by combination of ovary culture and embryo rescue techniques. The hybrids possess important traits such as resistance to major insect pests, tolerance to high temperatures and drought as well as resistance to shattering (Brown et al. 1997; Momotaz et al. 1998). Other recent examples of the application of embryo rescue technique in plant breeding are summarized in Table 10.1.

10.2.2 Overcoming Embryo Abortion

Embryo abortion is a common phenomenon and has been a major problem in conventional plant breeding. This is primarily due to the failure of the endosperm to properly develop as nutritive tissue surrounding the embryo (Bhojwani and

Table 10.1 Recent examples of the use of embryo rescue technology in plant breeding

Species	Use of embryo culture	References
Allium cepa × *A. roylei*	Introgression desirable traits of *Allium roylei* into the *A. cepa* genome	Chuda and Adamus (2012)
Arachis hypogaea × *A. paraguariensis, A. hypogaea* × *A. appressipila*	Regeneration of interspecific hybrids with resistance to early leaf spot	Rao et al. (2003)
Capsicum annuum, *C. chinense*, and *C. frutescens*	Interspecific hybridization for crop improvement	Debbarama et al. (2013)
Elaeis oleifera × *E. guineensis*	Optimizing media composition for immature embryo culture	Alves et al. (2011)
Gossypium spp.	Introgression of desirable characteristics into commercial cotton	Mehetre and Aher (2004)
Hylocereus polyrhizus × *H. undatus*	Developing efficient method to rescue embryos following interspecific crosses	Cisneros and Tel-Zur (2010)
Lilium spp.	Producing interspecific hybrids with useful traits for flower market	Proseviçius and Strikulyte (2004)
Lilium spp.	Optimizing efficiency of embryo rescue methods in interspecific hybridization	Chi (2002)
Malus pumila	Rescuing immature embryos for rootstocks	Dantas et al. (2006)
Phaseolus vulgaris × *P. coccineus*	Rescuing globular stage embryos following interspecific hybridization	Barikissou and Baudoin (2011)
Phoenix dactylifera × *P. pusilla*	Rescuing and multiplication of interspecific hybrid zygotic embryos	Sudhersan et al. (2009); Sudhersan and Al-Shayji (2011)
Prunus spp.	Rescuing genetically unbalanced hybrids	Kukharchyk and Kastrickaya (2006)
Salix viminalis × *Populus alba, S. viminalis* × *P. violascens, S. viminalis* × *P. tremula*	Overcoming post-zygotic barriers caused by the deficiency of endosperm as nutritive tissue	Bagniewska-Zadworna et al. (2010)
Solanum pinnatisectum × *S. tuberosum*	Introgression of resistance to late blight from *Solanum pinnatisectum* into *S. tuberosum* genome	Ramon and Hanneman Jr. (2002)
Triticale	Optimizing plant regeneration and the generation of genetic variability	Maddock (1985)
Vaccinium spp.	Overcoming post-zygotic barrier during the production of interspecific hybrids	Pathirana et al. (2013)

(continued)

Table 10.1 (continued)

Species	Use of embryo culture	References
Vitis vinifera	Application of CPPU and BA to enhance embryo recovery	Nookaraju et al. (2007)
V. vinifera × *V. amurensis*	Resistance to downy mildew and anthracnose	Tian and Wang (2008)

Razdan 1996; Hu and Wang 1986). For this reason, embryo rescue technology has been widely used to regenerate complete plants under in vitro conditions (Reed 2005).

This method can also be used to rescue young embryos from intraspecific hybrids that normally produce unviable seeds. For instance in seedless triploid embryos resulting from crosses between diploids and tetraploids of the same species. By in vitro culture on aseptic nutrient medium the embryos may develop and grow into complete plants, thus overcoming postzygotic barriers such as endosperm failure.

Recovery of triploid hybrids by embryo rescue has also been reported to be successful from intraspecific crosses between Fujiminori ($2n=4x=76$) × Jingxiu ($2n=2x=38$) grape varieties (Yang et al. 2007). Intraspecific crosses between diploid (2n = 22) and tetraploid (2n = 44) daylily (*Hemerocallis*) was also performed in vivo to produce triploid hybrids (3n = 33). Following from this, the embryo rescue procedure was applied to regenerate plants from immature triploid hybrid embryos, which then were propagated in vitro and successfully acclimatized to ex vitro conditions (Zhiwu et al. 2009). Further, Guo et al. (2011) reported successful embryo culture of triploid grapes obtained from crosses between diploid and tetraploid cultivars. More recent achievements in the use of embryo culture technology coupled with triploid plant regeneration have been reported by Aleza et al. (2010, 2012) in the breeding program of seedless mandarin oranges.

10.2.3 Shortening Breeding Cycle by Overcoming Seed Dormancy

Embryo rescue technique has also been used to shorten breeding cycle in a number of fruit crops by overcoming seed dormancy. In species which require enough time for embryo maturity, seedlings cannot be raised just after fruit ripening. In addition, some species need a longer period to break their seed dormancy. For example, seeds of Brussels sprouts, rose, apple, oil palm and *Iris* do not germinate after fruit ripening. Culturing immature embryos on proper growth medium will result in immediate germination, and therefore shorten the breeding cycle.

The occurrence of seed dormancy may be due to internal factors such as endogenous inhibitors and embryo immaturity, or external factors such as light, temperatures and humidity. Debbarama et al. (2013) claimed that seed dormancy may be localized

in the seed coat or in the endosperm, or both. Therefore, the embryos will germinate and grow successfully and more rapidly when removed from these factors.

Application of various embryo rescue technique to wide area of plant research have also been reviewed by Bridgen (1994), Sharma et al. (1996), Chuanen and Guangmin (2005), and more recently by Fathi and Jahani (2012).

10.3 Somatic Embryogenesis

The term *somatic embryogenesis* refers to the process of embryo development from cells other than gametes (somatic cells) without a normal fertilization process. Since the embryos developed by circumventing the normal fertilization process, they are genetically identical to their parent tissue, and as such they are clones.

The phenomenon of somatic embryogenesis was first reported by Steward et al. (1958) on suspension culture of *Daucus carota*, and by Reinert (1959) on callus culture of the same species. It is worth noting that Krikorian and Simola (1999), commented on the pioneer work of Harry Waris on somatic embryogenesis of *Oenanthe aquatica* (Umbelliferae). They emphasized that Waris was one of the first researchers to observe and recognize somatic embryo production in aseptic culture (see also Simola 2000).

Somatic embryogenesis can be used in a number of ways. For example, large scale-clonal propagation of elite cultivars (Ahmad et al. 2011), producing artificial seeds (synthetic seeds) (Pintos et al. 2008), gene transfer for genetic improvement (Li et al. 2002), in vitro selection approaches for various biotic and abiotic stresses (Ahmad et al. 2011), and providing potential models for studying molecular, regulatory and morphogenetic events during plant embryogenesis (Kamle et al. 2011; Ravi and Anand 2012).

Slater et al. (2003) claimed that somatic embryos may be produced indirectly by involving the dedifferentiation of organized tissue into the callus mass prior to embryo formation, or embryos may be produced directly from organized tissue without an intervening callus phase. The anatomical and physiological features of embryos derived from somatic tissues are highly comparable to zygotic embryos derived through normal fertilization (Dobrowolska et al. 2012; Mathew and Philip 2003; Palada-Nicolaua and Hausman 2001). They both proceed through a series of distinct stages which span a period of several days. There is no difference in embryogenesis of dicots or monocots up to the octant stage. Following from this, however, the embryogenesis takes different pathways (Raghavan 1986). In monocots Godbole et al. (2002) reported that beyond the octant, embryogenesis encompasses globular, elongated, scutellar and coleoptile stages; while in dicots Mandal and Gupta (2002) claimed that the stages are globular, heart, torpedo and cotyledon or plantlet stages.

10.3.1 Somatic Embryogenesis for Mass Propagation

Somatic embryogenesis is the most attractive and practical way for rapid mass multiplication of agricultural crops. Deo et al. (2010) suggested that this technology offers many advantages over conventional micropropagation such as the opportunity of producing somatic embryos; shoot and root of somatic embryos are formed simultaneously avoiding the requirement of a rooting phase as in conventional micropropagation. Embryo formation and germination can be synchronized to maximize plantlet regeneration, the dormancy of somatic embryos can be induced to make long-term storage become possible, and it is easy to scale-up somatic embryos with less labor inputs.

Research on somatic embryogenesis had been done on various plant species such as *Allium sativum* (Luciani et al. 2006), *Dioscorea alata* (Belarmino and Gonzales 2008), *Bactris gasipaes* (de-Alencar et al. 2010), *Agapanthus praecox* ssp. *Minimus* (Yaacob et al. 2012) and *Theobroma cacao* (Quainoo and Dwomo 2012). Somatic embryogenesis had also been successfully applied to *Alstroemeria* (Khaleghi et al. 2008), *Dianthus caryophyll*us (Ali et al. 2008; Karami et al. 2007) and *Bauhinia variegata* (Banerjee et al. 2012). Successful somatic embryogenesis has also been reported in *Coffea* (Neuenschwander and Baumann 1992; Priyono 1993; Sreenath et al. 1995), but the rate of success was relatively low and plantlets regeneration was found to be difficult. This technique has also been successfully applied in clonal propagation of various gymnosperms such as *Picea* (Ahmad et al. 2011), *Ephedra foliata* (Dhiman et al. 2010) and *Araucaria angustifolia* (Steiner et al. 2012).

In spite of its advantages, however, somatic embryogenesis also has limitations. There is the tendency of somatic embryo development to be nonsynchronous resulting in embryos of various developmental stages to be present in the culture system (Kong et al. 2012; Zegzouti et al. 2001). However, this limitation could be overcome by incorporating abscisic acid (ABA) and mannitol in culture medium (Torres et al. 2001). Another problem concerning somatic embryogenesis is the instability of cell lines in culture that may induce morphological abnormalities such as pluricotyledony, multiplex apex formation, fused cotyledons, fasciation (Singh and Chaturvedi 2013; Zegzouti et al. 2001), slender stems, stubby structures and non-functional leaves (Benelli et al. 2010). In gymnosperm, the regenerated somatic embryos lack a lipid- and protein-rich megagametophyte, from which amino acids and sugars are mobilized during germination (Bornman et al. 2001).

10.3.2 Somatic Embryogenesis in Plant Breeding

Somatic embryogenesis has become an indispensable modern plant breeding component since this system provides an alternative platform in the development of new crops with many valuable agronomic properties. One of the benefits is that the somaclonal variations (discussed later in this chapter) which arise from the process

of somatic embryogenesis could be utilized as a source of genetic variability for producing novel varieties. The use of somatic embryogenesis as a preferred method for genetic improvement of valuable germplasm of a number of important crops has been emphasized by many researchers (Ashakiran et al. 2011; Ji et al. 2011; Kamle et al. 2011).

Somatic cell hybridization through protoplast fusion (discussed later in this chapter) has proved useful in shortening the cycle of plant breeding. This technique was first developed by Melchers and Labib (1974) through their investigation on tobacco tissue culture. Since then, somatic cell hybridization has been applied to a wide range of crop species such as crosses between *Triticum aestivum* × *Haynaldia villosa* (Zhou et al. 1996, 2002), *Hordeum vulgare* × *Daucus carota* (Kisaka et al. 2001), *Triticum aestivum* × *Setaria italica* (Cheng et al. 2004) and *Carica papaya* L × *Vasconcellea cauliflora* (Dinesh et al. 2013). The rate of success, however, varied among crosses.

There are three types of hybrids formed through somatic cell hybridization: symmetric hybrids which contain somatic chromosomes of both parents, asymmetric hybrids which preserve the genetic material of one parent, and cybrids which consist of nucleus of one parent and cytoplasm from both parents. Among the three types, Zhou et al. (2001) claimed that asymmetric somatic hybridization is superior to symmetric because the resultants possess comparatively fewer chromosomes (genes) from the donor, avoiding too many wild traits being introduced into the acceptor, thus the traits of the hybrids were closer to the goal of breeding. Further, Miko (2008) suggested that a few characters are passed on from parent to offspring by genes which are not part of a nuclear chromosome, but located in cell organelles in the cytoplasm (cytoplasmic genes). This suggests that regeneration of cybrids with the mixture of cytoplasm from both parents but having only one nuclear genome will help in transfer of cytoplasmic genetic information from donor to offspring. Thus, somatic cybridization can be applicable in plant breeding program

10.3.3 Somatic Embryogenesis for Production of Synthetic Seeds

Ara et al. (2000) defined artificial seed as encapsulated plant tissue that can be used as a seed that will germinate and grow into a complete plant under favorable in vitro or ex vitro conditions, and this capability is preserved even after storage. Somatic embryos have the potential to be used in producing artificial seed as first demonstrated by Kitto and Janick (1982) in *Daucus carota*. In addition to somatic embryos, axillary shoots, apical shoot tips, buds or stem, root segments, embryogenic calli, as well as protocorms or protocorm-like bodies may be used to produce artificial seeds (Siong et al. 2012; Vdovitchenko and Kuzovkina 2011). However, of various micropropagules used, somatic embryos and axillary shoot buds have been mainly favored for synthetic seed production (Siong et al. 2012).

The development of synthetic seed production technology is considered an effective and efficient method of propagation in a number of important crops. Explants such as somatic embryos, shoot tips or axillary buds are encapsulated in coating material such as hydrogel, alginate gel, ethylene glycol, or dimethylsulfoxide (Asmah et al. 2011), which can be developed into plants. The coating protects the explants from mechanical damage during handling and allows germination to occur like true seeds and sprout into seedlings under suitable conditions. It is important to note that materials used to encapsulate embryos should also be nontoxic, biocompatible and water soluble (Ara et al. 2000). For this reason, Saiprasad (2011) suggested the use of alginate gel for encapsulating synthetic seed since it has moderate viscosity, low toxicity, rapid gelification, low cost and bio-compatibility properties. In addition, alginate may enhance capsule formation and its rigidity provides better protection for encased explants against mechanical injury.

The successful application of artificial seed technology can only be achieved when there is efficient upstream production of micropropagules (somatic embryos or axillary shoots) as well as downstream germination protocols for high percentage plant regeneration. In this regard, Ara et al. (2000) and Ravi and Anand (2012) proposed simplified schematic procedures for the production of artificial seeds, which is summarized in Fig. 10.1.

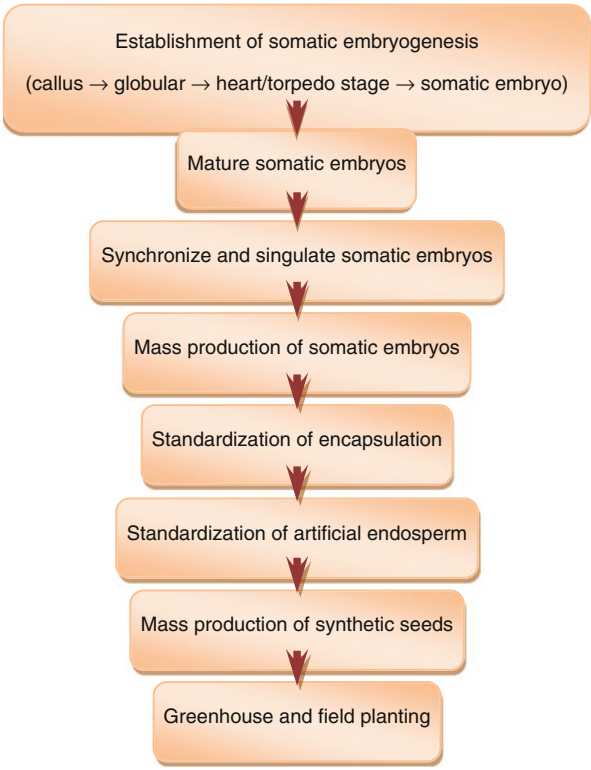

Fig. 10.1 Simplified procedure for the production of artificial seed

There are various advantages of artificial seeds such as: better clonal plants which could be propagated in large scale; preservation of elite and endangered or extinct or rare plant species; and consistent and synchronized harvesting of important agricultural crops (Khor and Loh 2005). In addition, ease of handling, potential long-term storage and low cost of production and subsequent propagation are other benefits (Bekheet 2006). Ravi and Anand (2012) suggested that artificial seeds can also be beneficial in understanding seed coat formation, fusion of endosperm in embryo development and seed germination, production of somatic hybrids in plants with unstable genotypes or those showing seed sterility; it can also be incorporated in embryo rescue technology.

Recent developments in tissue culture techniques have made synthetic seed technology considerably more diverse, and has been reported in a number of plant species, including *Asparagus officinalis* (Mamiya and Sakamoto 2001), *Geodorum densiflorum* (Datta et al. 2001), *Paulownia elongata* (Ipekci and Gozukirmizi 2003), *Dendrobium, Oncydium* and *Catlleya* orchids (Saiprasad and Polisetty 2003), *Gypsophila paniculata* (Rady and Hanafy 2004), *Cyclamen persicum* (Winkelmann et al. 2004), *Camellia sinensis* (Seran et al. 2005), *Allium sativum* (Bekheet 2006), *Pyrus communis* (Nower et al. 2007), *Oryza sativa* (Bidhan and Mandal 2008; Kumar et al. 2005) and *Brassica napus* (Zeynali et al. 2013). These results have shown that artificial seed production is potentially useful for the large scale propagation of superior hybrids of economically important species.

10.4 Somaclonal Variation

The existence of genetic variation is an important factor exploited in plant breeding. The desired variation, however, is often not available in the right combination or does not exist at all under in vitro conditions. Jain et al. (1998) suggested such variation can be induced in vitro from somatic cells or tissue resulting in somaclonal variation that could be used in combination with conventional breeding methods to create more genetic variability. Somaclonal variation has been a valuable tool in plant breeding; wherein variation in tissue culture regenerated plants from somatic cells can be used to develop crops with desirable traits. Characteristics for which somaclonal mutants can be improved during in vitro culture includes resistance to disease, herbicides and tolerance to environmental or chemical stress, as well as for increased production of secondary metabolites. Selection is done by employing a stress-causing agent in tissue culture containing dividing cells. An efficient method for obtaining plants with desired characteristics is to add selective agents that will alter other aspects of the phenotype is reviewed by Tapingkae et al. (2012).

Somaclonal variation has been associated with changes in chromosome numbers (polyploidy and aneuploidy), chromosome structure (translocations, deletions, insertions and duplications), point mutations, and DNA methylation (Nwauzoma and Jaja 2013; Rodriguez-Enriquez et al. 2011). The molecular basis of somaclonal variation is not precisely known; however, both genetic and epigenetic mechanisms

are suggested to play a role (Jiang et al. 2011). Changes in DNA methylation often give rise to epigenetic effects, which can affect expression of genes normally suppressed. Epigenetic variation is often unstable and can disappear either after plants are removed from culture or within a few clonal generations, whereas genetic variation is heritable (Biswas et al. 2009). Therefore, the success in applying somaclonal variation in plant breeding is dependent on the genetic stability of the selected somaclones.

10.4.1 Detection of Somaclonal Variation

The early detection of the presence of somaclonal variants could save valuable time and minimize the overall economic loss to the users of tissue-cultured planting materials. Various strategies were used to detect somaclonal variants, based on one or more determinants from among morphological traits, cytogenetic analysis (numerical and structural variation in the chromosomes) and molecular and biochemical genetic markers (Tan et al. 2013). Development and application of modern technologies based on molecular markers provide valuable tools for the detection of somaclonal variation. The use of molecular marker techniques to detect somaclonal variations has been applied successfully to several plant species, such as *Solanum tuberosum* (Ehsanpour et al. 2007), *Gossypium hirsutum* (Jin et al. 2008) and *Musa* spp. (Abdellatif et al. 2012). Research publications continue to increase rapidly and the field is gaining growing interest in a wide range of research areas. This has widely been covered in an extensive set of literature and the reader is referred to publications in broad reference works such as those of Davies (2010), Albrecht et al. (2012) and de Maagd and Hall (2013).

10.4.2 Genomics, Proteomics and Metabolomics

The technologies of genomics which assesses changes in the genome, proteomics which studies the total protein complement (the proteome) and the metabolomics which investigates the complement of small molecules (low molecular weight, <1500 Da) (Davies 2010), collectively are referred to as *omics*, have the potential to be utilized in the detection of somaclonal variation.

10.4.2.1 Genomics

Genetic fingerprinting is a powerful tool in the field of plant science, to be used, for example, for correct germplasm identification. When linked to metabolomics and proteomics (fingerprinting techniques on the plant's metabolites or protein composition) has the potential to elucidate data on phenotypic variation, caused by growth

conditions or environmental factors, and yield data on the genes involved in the pathways and enzymes involved in the synthesis of natural products (Terryn et al. 2006).

The development of high throughput genome technologies in the past decade has permitted a number of options to profile the epigenome of several organisms including plants such as Arabidopsis. These new methods are able to provide a detailed characterization of genomic DNA methylation and histone modifications at an unprecedented resolution, which can be integrated with transcriptomics data including the smRNA transcriptome (Wang et al. 2009).

Randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR) and inter-simple sequence repeat (ISSR) markers have been employed to assess the genetic fidelity or genetic variability in regenerated plants. Among the markers, ISSR and RAPD have been mostly preferred and well established in many plant species, including *Vitis* (Nookaraju and Agrawal 2012), *Musa* sp. (Aremu et al. 2013), *Cynara scolymus* (Rey et al. 2013), *Olea europaea* (Leva et al. 2012), *Pongamia pinnata* (Kesari et al. 2012) and *Punica granatum* L. (da Silva et al. 2013).

ISSR markers involve PCR amplification of the region between two closely placed simple repeat sequences that are inversely oriented. They are identified using primers designed from within the repeated region (Zietkiewicz et al. 1994). This technique is based on PCR amplification of inter-microsatellite sequences remaining popular due to their relative simplicity, reliability, cost effectiveness and highly discriminative nature (Agarwal et al. 2008). Hence it has been widely used in numerous fields such as genetic diversity, phylogenetic studies, ecology and evolutionary biology (Aremu et al. 2013; Aruna et al. 2012). However, reports about somaclonal variation are still controversial. For example, RAPD and ISSR analyses of leaf genomic DNA were able to distinguish between date palm (*Phoenix dactylifera*) cultivars (Zehdi et al. 2002), but according to Fki et al. (2011), they failed to reveal polymorphisms in genomic DNA that could be associated with somaclonal variations in tissue culture-derived date palms.

10.4.2.2 Proteomics

Modern technologies of gene sequencing, microarray experiments, gene and protein expression within the cell of an organism and information on molecular markers have been very useful in identifying regions on chromosomes that bring about variation in a trait, thereby providing tools that can lead to far more accurate selection processes for genetic improvement (Nimisha et al. 2013). Using proteomic analyses it is possible to identify proteins, their functions and interactions as well as their subcellular localization in a tissue. This can be useful for the determination of protein alteration during plant development, including somaclonal variation. Accordingly, use of proteomics has become increasingly common in cellular, genetic and physiological research (Miransari and Smith 2013). For example, the proteomic analysis of Arabidopsis cells using Polyacrylamide Gel Electrophoresis

(2D-PAGE) and matrix-assisted laser desorption/ionization (MALDI)-time-of-flight (TOF) (MALDI-TOF) peptide mass fingerprinting indicated the presence of 663 different proteins originated from 2943 2D-PAGE spots (Giavalisco et al. 2005).

Currently, three main techniques are used for quantitation in proteomics: two-dimensional 2D-PAGE linked to mass spectrometry for protein identification and stable-isotope labelling-free shotgun proteomics. With 2D-PAGE, proteome differences can be identified by gel-gel comparisons but care needs to be taken to ensure reproducibility, to avoid variable protein staining efficiency of individual gels, and to take account of bias against some protein classes such as membrane proteins. One can expect ca. 1000–2000 proteins to be separated and detected in a single gel. Another technique (shotgun proteomics) involves tryptic digestion of a complex protein sample followed by peptide analysis using liquid chromatography-mass spectrometry (LC–MS) techniques. In this case, the increased complexity of the sample caused by the high number of peptides produced is the primary challenge, but the capacity for throughput is high compared with that of 2D-PAGE with less bias against specific protein classes, as in the case of 2D-PAGE. Protein identification is difficult where whole genome sequences are not available. However, standard protein profiling technologies have two major limitations. First, only the most abundant proteins can be detected in a complex protein mixture and second, the high throughput analysis is mostly restricted to proteins and peptides in protein databases (Davies 2010).

More recently a comprehensive proteomic assay was conducted trying to find embryogenic-specific proteins. Proteomic approaches have been applied to investigate the protein profiles and their changes during callus differentiation of *Vanilla planifolia* (Palama et al. 2010). Morel et al. (2012) found 100 significantly- and differentially-expressed proteins in somatic embryogenesis of *Pinus pinaster*. They were mainly involved in carbohydrate or lipid metabolism and genetic information processing. Many storage proteins were identified (vicillin-like, legumin-like, LEA proteins); some of them were observed from the beginning of maturation. In *Cyphomandra betacea*, Correia and Canhoto (2012) reported that a better ability of the embryogenic cells to regulate the effects of stress conditions relied on an increase in metabolism, protein synthesis and stress related proteins.

10.4.2.3 Metabolomics

Metabolomics is the systematic characterization of small molecules and their concentrations in cells, tissues, or organisms, in ever more detail. The basis of metabolomics is making unbiased observations with highly reproducible analytical tools, followed by a biostatistical analysis to find correlations between all the available data. This can be illustrated by the identification of the active compounds in medicinal plants. Metabolomics intends to give a wide view of all metabolites in a specific sample at a specific time. Applications of metabolomics in the plant sciences are already extensive and broad. Information on the biochemical composition of plants can help identify biochemical changes as a potential basis of causality for

phenomena of interest such as visible (phenotype) or invisible (chemical/physiological) differences (reviewed by de Maagd and Hall 2013).

Most metabolomics experiments involve a combination of separation and detection technologies that can be used in serial or in parallel combinations, often referred to as *hyphenated approaches*. Therefore these analyses never give a complete picture, and one needs to know beforehand what class of compounds is considered relevant in order to choose the proper methodology. High performance liquid chromatography (HPLC) with UV, liquid chromatography-Mass spectrometry (LC-MS) or Gas chromatography-mass spectrometry (GC-MS) are most commonly used to monitor the presence of secondary metabolites (Moco and Vervoort 2012). Whereas, GC-MS is particularly appropriate for the analysis of volatile organic compounds and derivatives of primary and secondary metabolites, LC-MS is highly applicable to the analysis of a wide range of semi-polar compounds including many secondary metabolites with nutriceutical proprieties (Rea et al. 2011). Detailed coverage of the various technologies used for metabolomics can be found in Schauer and Fernie (2006) and Bowne et al. (2011).

Terryn et al. (2006) developed a comprehensive profiling approach that is based on functional genomics. This approach integrates cDNA-AFLP-based transcript profiling and targeted metabolic profiling. As this method requires no prior genetic knowledge or sequence databanks, it is appropriate to any plant species to unravel the biosynthesis of any metabolite of interest. This knowledge will then allow for metabolic engineering, as well as pave the way for so-called *combinatorial biochemistry*, with which novel metabolites could be produced in plants. However, this technology is quite expensive and needs specialized researchers for correct application as well as for interpretation of the data (de Maagd and Hall 2013). All these limitations are expected to decrease in the near future as the technology continues to develop fast.

New tools of functional genomics combined with metabolomics and proteomics will revolutionize the knowledge of pathways and enzymes involved in the synthesis of natural products, and thus allow a more focused approach for their production. With the increasing need for novel drugs, selection of somaclones with desired traits, using omics technologies will play an increasingly important role in the large-scale production of pharmaceuticals in plants.

10.5 In Vitro Flowering

An in vitro plant production system is considered a convenient tool to study the flowering mechanism, as the conditions can be manipulated and so provide a controlled system. It is possible to use in vitro flowering in genetic studies, biochemistry and studies of environmental factors affecting flower formation. It can also be applied as a tool to shorten breeding programs or can be adjusted to commercial production of specific compounds (Lin et al. 2003). In vitro flowering has been reported from both monocotyledonous (Joshi and Nadgauda 1997; Lin et al. 2003,

2004; Nadgauda et al. 1990; Ramanayake 2006) and dicotyledonous (Al-Khayri et al. 1991, 1992; Kanchanapoom et al. 2009; Yadav and Singh 2011) families. Several factors, such as genotype, culture medium, physical and environmental growth conditions, and factors inherent to the explants, affect in vitro flowering. More general reviews of in vitro flowering include Lakshmanan and Taji (2004), Ziv and Naor (2006) and Sri Rama Murthy et al. (2012). Successful in vitro flowering in *Dendrobium* orchid and *Dendrocalamus hamiltonii* has been reviewed by Zulkarnain et al. (2012).

Novel approaches involving in vitro flowering and molecular techniques offer unique opportunities to investigate the flowering process from new perspectives. For example, flowering time in Rosaceae species, such as apricot, sour and sweet cherry, apple, peach, almond and rose, were identified in breeding for climatic adaptation (reviewed by Campoy et al. 2011). These data may facilitate the development of breeding strategies aimed at releasing early or late flowering cultivars and controlling the phenological adaptation of new cultivars to climate change.

The endogenous signaling pathways responding to gibberellic acid (GA_3) (autonomous and age) and environmental signals (photoperiod and temperature) converge on genes that then activate floral homeotic genes (reviewed by Pineiro and Jarillo 2013; Pose et al. 2012; Song et al. 2013; Taoka et al. 2013). In vitro flowering can reduce the influence of environmental factors and allow for precise control of environmental factors and the application of plant growth regulators. It may also offer a means for studying the mechanism of flowering. Recently Chavan et al. (2013) in *Ceropegia panchganiensis* observed varied flower induction to BAP and sucrose concentrations and combinations.

10.5.1 In Vitro Pollination and Fertilization

In angiosperms, pollination is defined as the transfer of pollen grains from an anther to the stigma of the same or different flowers on the same plant (self-pollination) or flowers of other plants (cross-pollination). When a pollen grain makes contact with the stigma, it adheres, hydrates, and germinates, resulting in the growth of a pollen tube which convey the sperm cells to the embryo sac and egg in the ovule (Lord and Russell 2002). Following from this, the plant undergoes fertilization through the fusion of sperm and egg cells.

Pollination and fertilization are crucial events contributing to genetic diversity and providing a basis for the improvement of important agricultural crops. However, the possibilities for cross combination is sometimes limited by sexual incompatibility and incongruity, restricting successful interspecific hybridization (Vervaeke et al. 2002; Zenkteler 1970). Incompatibility may occur in intraspecific crosses as a result of the activity of S-alleles, whereas incongruity operates in interspecific crosses due to a lack of genetic information to complete pre- and post-pollination processes (Hogenboom 1973).

Stebbins (1958) separated such sexual barriers preventing interspecific hybridization into pre-fertilization (factors hindering effective fertilization) and post-fertilization (barriers occurring during or after syngamy) barriers. The pre-fertilization barrier is mostly confined within the style (Vervaeke et al. 2002), for example the failure of pollen to germinate on stigma, or failure of pollen tube to reach the ovule due to very long style, or slow growth of pollen tube which fails to reach the ovule before the ovary abscises (Taji et al. 2002). To overcome this type of sexual barrier, a number of techniques such as bud pollination, stump pollination, use of mentor pollen and grafting of the style have been developed (Van Tuyl and De Jeu 1997).

On the other hand, the post-fertilization barrier is mainly caused by the failure of hybrid embryo to gain its maturity because of embryo-endosperm incompatibility or poor embryo development (Taji et al. 2002). To overcome the post-fertilization barriers, Van Tuyl and De Jeu (1997) proposed ovary and ovary-slice culture, ovule culture and embryo culture.

Ovary and ovary-slice culture are applied when abortion takes place at a very young stage of seed development. In this technique, embryos are dissected or in some crops, where the ovary is large, slicing the ovary in small parts is a better option for rescuing the young seedlings in vitro. Ovule culture is applied when the mismatch between embryo and endosperm development occurs very early and ovary culture and/or ovary slice culture fails. In this method, ovules are dissected out of the ovaries and are cultured in vitro under controlled environment. Meanwhile, embryo rescue method may be applied when young fruits remain for a long time on the mother plant, and it is necessary to excise the entire embryo to prevent abortion. However, it is technically difficult to isolate the intact embryos due to their very small size, therefore ovaries with young embryos or entire fertilized ovules are often cultured (Van Tuyl and De Jeu 1997).

The technique of in vitro fertilization, in which isolated sperm and egg cells are induced to fuse under controlled conditions, removes much of the interfering barrier. In vitro pollination and fertilization have been used to overcome pre- and post-fertilization barriers in a number of genera. The use of in vitro fertilization in higher plants, therefore, has become an important contemporary research area in plant developmental and reproductive biology with potentially significant scientific applications (Wang et al. 2006).

Since it was first introduced by Kanta et al. (1962) on *Papaver somniforum*, the technology of in vitro pollination and fertilization have been successfully applied in various important agricultural crops including *Lycopersicon esculentum* (Sheeja and Mandal 2003), *Helianthus annuus* (Popielarska 2005) and *Melandrium album* (Zenkteler et al. 2005).

In plant breeding the technique of in vitro pollination and fertilization has potential applications at least in the following areas: (a) production of interspecific and intergeneric hybrids, (b) production of haploid plants through parthenogenesis, (c) overcoming self-incompatibility, and (d) overcoming cross-incompatibility. In addition, Taji et al. (2002) proposed that in vitro pollination and fertilization is also important in the study of pollen physiology and fertilization.

10.5.2 Production of Interspecific and Intergeneric Hybrids

In vitro pollination and fertilization have been successfully used in interspecific and intergeneric hybridization to produce novel cultivars with blended traits of both parents and to introgress useful traits of one species to another. This is made possible because under in vitro condition environmental factors can be controlled and optimized, while in nature such hybrids would not be formed readily. In vitro interspecific hybridization progenies have been created within the genus *Helianthus* (Weber et al. 2000), *Lilium* (Lim et al. 2008; Wang et al. 2012) and *Cucumis* (Skálová et al. 2010). Meanwhile, successful intergeneric hybridization has been achieved in *Brassica napus* using *Brassica elongata*, *B. fructiculosa, B. souliei, Diplotaxis tenuifolia, Hirschfeldia incana, Coincya monensis* and *Sinapis arvensis* (Siemens 2002). Further, intergeneric hybrids have also been successfully obtained between *Sandersonia aurantiaca* × *Gloriosa rothschildiana* (Nakamura et al. 2005), between rare and endangered orchids, *Renanthera imschootiana* × *Vanda coerulea* (Kishor and Sharma 2009) and between *Cyamopsis tetragonoloba* × *Cyamopsis serrata* and C. *tetragonoloba* × *C. senegalensis* (Ahlawat et al. 2013).

10.5.3 Production of Haploid Plants

The potential of haploid technology for plant breeding arose in 1964 with the first report by Guha and Maheshwari (1964) on haploid embryo production from in vitro culture of *Datura* anthers. This was followed by the successful effort of Nitsch and Nitsch (1969) on the regeneration of haploid plants from microspore culture of tobacco. Since then, many attempts have been made on over 250 plant species belonging to almost all families of the plant kingdom (Maluszynski et al. 2003).

As with anther or microspore culture, in vitro pollination and fertilization may also be used to produce haploid plants through parthenogenesis. This has been achieved in *Mimulus luteus* when the ovule was pollinated with pollen from *Torenia fournieri* (both from family Scrophulariaceae) (Taji et al. 2002). More recent achievement on the production of double haploid plants through ovule cultures resulted from the hybridization reported in cultivated species of *Cucurbita pepo* with three other Cucurbita species i.e. *C. moschata*, *C. ficifolia* and *C. martinezii* (Rakha et al. 2012). Though ovules offer possible alternative source for haploid or doubled-haploid production, the exploration of this tissue in producing haploid or doubled-haploids in breeding programs is still very limited. For more in-depth discussion on haploid plant production the reader is referred to Chap. 5 in this volume. Further reading on haploid higher plant production via in vitro technique can be found in Jain et al. (1996a, b, c, d, e).

10.5.4 Overcoming Self- and Cross-Incompatibility

Self-incompatibility is a physiological barrier preventing fusion of male and female gametes from the same individual (self-fertilization), thus preventing inbreeding and encouraging outcrossing. On the other hand, cross-incompatibility occurs when male and female gametes from different individual fail to fuse, and as such, prevent outcrossing and encourage inbreeding to take place. In vitro pollination and fertilization can be used to overcome these physiological barriers.

In this method, the entire ovule mass from an ovary, intact on the placenta, is cultured in vitro just after pollination and fertilization is completed. For example, *Cucumis sativus* which is sexually incompatible with almost all *Cucumis* species due to different chromosome number, n = 7 in *C. sativus* versus n = 12 in *C. melo* and most wild *Cucumis* species. This obstacle can be overcome through the use of embryo rescue and/or ovule culture (Skálová et al. 2008).

10.5.5 Pollen Physiology and Fertilization

The technique of in vitro pollination and fertilization is an effective tool in studying pollen physiology and fertilization. This has been reported by Chapman and Goring (2010) who found that successful fertilization in the Brassicaceae was regulated by the cellular systems in the pistil that guide the post-pollination events, from pollen capture on the stigmatic papillae to pollen tube guidance to the ovule, with the final release of the sperm cells to effect fertilization. Study on in vitro pollination and fertilization would also reveal information about the requirement of different constituents on the germination of pollen grains as demonstrated in *Solanum macranthum* (Mondal and Ghanta 2012), *Withania somnifera* (Ghanta and Mondal 2013) and *Cajanus cinereus*, *Rhyncosia rothii* and *R. aureus* (Jayaprakash and Sabesan 2013). Specific temperature requirement for in vitro pollen germination and pollen tube growth of cotton has also been reported by Kakani et al. (2005).

10.6 Protoplast Culture Technology

Protoplasts are described as naked plant cells obtained through the removal of the cellulosic cell wall. Removing the cell walls could be performed through enzymatic procedures as pioneered by Cocking (1960) using root tips of tomato (*Lycopersicon esculentum* Mill. var. Sutton's Best of All) seedlings and a fungal cellulase obtained from *Myrothecium verrucaria*. Since then, the technology has been applied in breeding programs aiming at producing progenies with useful agronomical traits.

The potential use of protoplast technology for the genetic improvement of many agricultural crops is immense. This technology has allowed not only intraspecific

hybridization to take place, but also the creation of interspecific and intergeneric hybrids as well as cybridization. Various desirable traits from donor plants have been successfully transferred to hybrids and cybrids using this technology (see Tables 10.2 and 10.3).

Table 10.2 Recent examples of the transfer of useful traits by protoplast fusion

Species	Useful traits transferred	References
Aspergillus niger (+) *A. ficuum*	Enhanced phytase production	Gunashree and Venkateswaran (2010)
Brassica napus (+) *B. rapa*	Increased biomass and yield	Qian et al. (2003)
Brassica napus (+) *Crambe abyssinica*	Increased erucic acid content in seeds	Wang et al. (2003)
Brassica napus (+) *Sinapsis arvensis*	Enhanced resistance to blackleg (*Leptosphaeria maculans*)	Hu et al. (2002a)
Brassica napus (+) *Orychophragmus violaceus*	Improved fatty acid composition in seeds	Hu et al. (2002b)
Curvularia lunata (+) *Helminthosporium gramineum*	Improved biocontrol efficiency against rice weeds	Zhang et al. (2007)
Citrus sinensis (+) *Fortunella crassifolia*	Increased plant vigor	Cheng et al. (2003)
Citrus sinensis (+) *Clausena lansium*	Production of triploid plants	Fu et al. (2003)
Raphanus sativus (+) *Brassica campestris*	Resistance to atrazine	Pelletier et al. (1983)
Saccharomyces cerevisiae (+) *Kluyveromyces marxianus*	Enhancing production of temperature-tolerance ethanol	Krishnamoorthy et al. (2010)
Solanum melongena (+) *S. aethiopicum*	Resistance to bacterial wilt (*Ralstonia solanacearum*)	Collonnier et al. (2001)
Solanum melongena (+) *S. sisymbrifolium*	Resistance to bacterial and fungal wilts	Collonnier et al. (2003)
Solanum nigrum (+) *S.tuberosum*	Resistance to atrazine	Binding et al. (1982)
Solanum nigrum (+) *S.lycopersicum*	Resistance to atrazine	Jain et al. (1988)
Solanum tuberosum (+) *S. tuberosum*	Resistance to potato virus Y	Gavrilenko et al. (2003)
Solanum tuberosum (+) *S. nigrum*	Resistance to late blight (*Phytophthora infestens*)	Zimnoch-Guzowska et al. (2003)
Solanum tuberosum (+) *S. stenotomum*	Resistance to bacterial wilt (*Ralstonia solanacearum*)	Fock et al. (2001)
Solanum tuberosum (+) *S. chacoense, S. tuberosum* (+) *S. cardiophyllum, S. tuberosum* (+) *S. pinnatisectum*	Resistance to late blight and Colorado potato beetle	Chen et al. (2008)
Trichoderma reesei (+) *T. harzianum*	Enhancing enzyme production and bio-control activity against soil-borne pathogens	Srinivasan et al. (2009)

Table 10.3 Recent examples of the applications of plant protoplasts technology

Species	Application	References
Arabidopsis thaliana	Mechanisms of gene recognition in plant pathogenicity	Leister and Katagiri (2000)
A. thaliana, Zea mays	Elucidation of plant signal transduction mechanisms	Sheen (2001)
Brassica carinata, B. rapa	Cell wall regeneration, cell division and production of microcolonies	Beránek et al. (2007)
B. carinata, B. napus	Studies on regenerative capacity and reliable protoplast culture protocol	Klíma et al. (2009)
B. chinensis	Electrophysiological studies of outward K+ channels	Fan et al. (2003)
B. napus	Resistance to *Leptosphaeria maculans*	Hu et al. (2002a)
B. oleracea ssp. *capitata*	Cold-tolerant cytoplasmic male-sterile	Sigareva and Earle (1997)
B. oleracea ssp. *italica*	Herbicide (glufosinate) resistance	Waterer et al. (2000)
B. rapa	Resistance to bacterial soft rot	Ren et al. (2000)
Cucurbita pepo	Viral pathogenicity	Choi et al. (2003)
Centella asiatica	Basis for the development of protoclones with high asiaticoside productivity	Aziz et al. (2006)
Citrus spp.	Production of tetraploids and triploids for scion and rootstock	Grosser and Gmitter Jr. (2010)
Helianthus annuus	Synthetic peptide import through the plasma membrane	Cormeau et al. (2002)
Hibiscus cannabinus	Viral replication processes	Liang et al. (2002)
Muscari neglectum	Somatic embryo development and plant regeneration	Karamian and Ranjbar (2010)
Nicotiana benthamiana	Viral recombination and replication	Shapka and Nagy (2004)
N. plumbaginifolia	Genetic basis of developmental regulation and specificity	Chesnokov et al. (2002)
N. tabacum	Regulation of osmotic water transport across cell membranes	Ding et al. (2004)
Phaseolus vulgaris	Electrophysiological studies of inward-rectifying K+ channels	Etherton et al. (2004)
Raphanus sativus	Immunocytochemical evaluation of aquaporin accumulation	Suga et al. (2003)
Vicia faba	Fluorometric analysis of photosynthetic electron transport	Goh et al. (2002)
Vigna radiata	Intracellular responses to drought and salinity stress	Kim et al. (2004)
V. unguiculata	Studies on plasma membrane organization	Vermeer et al. (2004)
Zea mays	Studies on transient gene expression and proteomics	Cheng et al. (2001)

10.6.1 Protoplast Isolation

Before dealing with protoplast culture, it is important to understand the method of protoplast isolation. The material should be handled carefully and gently in order to obtain uninjured and viable protoplasts as they are prone to breakage or injury during isolation. Protoplast isolation is achieved via mechanical and enzymatic procedures.

Mechanical isolation of protoplast was first demonstrated by von Klercker (1892) who successfully isolated protoplast from plasmolyzed cells of *Stratiates aloides*. Following from this, the method was then used to isolate protoplasts from onion bulb tissues. Onion scales were immersed in 1.0 M sucrose solution to induce plasmolysis and therefore protoplasts shrunk away from their enclosing cell walls. The plasmolyzed tissue was carefully cut at such a thickness, so that only cell walls are cut without disturbing the protoplasts. The shrunken protoplast were released by osmotic swelling when the tissue is placed in sucrose solution at a low concentration.

The enzymatic isolation of protoplast involves the use of enzymes to dissolve the plant cell wall. The wall consists of primary components such as cellulose, hemicellulose and pectic substances. Cellulose is a simple, linear polymer of D-glucose with β-1,4 linkage, and having molecular weight of 50,000–2,500,000 Dalton (depending on species). Hemicelluloses are less well defined and appear to be mixed polymers of glucose, galactose, mannose, arabinose and xylose, and most of them are linked with both β-1,4 and β-1,3 linkages. Meanwhile, pectin substances mostly found in middle lamella, which holds the cells together. Pectin is a polymer of methyl-D-galacturonate with α-1,4 linkages and having molecular weights of 25,000–360,000 Da. Therefore, cell wall degrading enzymes such as cellulases, hemicellulases and pectinases are used to treat source tissues in enzymatic protoplast isolation method.

The enzymatic method could be used as a two-step procedure (sequential method) or single-step procedure (direct method). In the two-step procedure, single cells are isolated from source tissues using pectinase, then cellulase is added to this cell suspension to digest the cell wall and release the protoplasts. However, with further refinement, a single-step procedure was developed for cell separation and cell wall degradation. In this method protoplasts are isolated directly by using two enzymes, cellulase and pectinase, simultaneously. Protoplast yield through direct method is high and, now this is the most frequently used method (Pati et al. 2008).

10.6.2 Protoplast Purification

Following isolation, protoplasts need to be purified by removing undigested materials (debris), burst protoplast and enzymes. This step is required to purify the population to obtain intact and viable protoplasts and to ensure the successful culture.

Debris can be removed by filtering the suspension through a steel or nylon mesh of 100 μ pore size, whereas enzymes are removed by centrifuging the solution at 600 rpm for 5 min. The protoplasts settle to the bottom of centrifuge tube while supernatant is removed using a pipette. The protoplast are then resuspended in a washing medium containing osmoticum with or without nutrient medium. The suspension is recentrifuged to allow the protoplast to settle down and the washing medium is decanted. Removing traces of enzyme by washing the protoplasts is carried out two or three times. Intact protoplasts are separated from broken ones by suspending the preparation in 20–40 % sucrose solution and centrifugation at 350 rpm for 3 min. Intact protoplasts are carefully removed from the surface of the solution using a pipette (Tomar and Dantu 2010).

The viability of the protoplasts is examined using a fluorescent microscope. For this reason, fluorescein diacetate (FDA) solution in acetone (5 mg L^{-1}) is added to the suspension to give final concentration of 0.01 %. After 5 min at room temperature, the protoplasts are examined. Only viable protoplasts can fluoresce under the fluorescent microscope (Tomar and Dantu 2010).

10.6.3 Protoplast Fusion

Protoplast fusion is known as a physical event, during which two or more protoplasts come in contact and adhere to one another, either spontaneously or with the aid of electrofusion procedure (De Filippis et al. 2000; Rakosy-Tican et al. 2001) or in the presence of fusiogenic agents such as polyethylene glycol (Klíma et al. 2009; Zhou et al. 2001) or by using positively charged ions (such as Ca^{++}) in a high pH solution (Tomar and Dantu 2010).

Protoplast fusion technology offers the possibility to transfer useful traits from one species to another, such as disease resistance, nitrogen fixation, rapid growth rate, more productivity, better protein quality, frost hardiness, drought resistance, herbicide resistance and heat and cold resistance. Specific genera have been targeted in somatic hybridization through protoplast fusion to transfer useful agronomic traits, as indicated in Table 10.2. Thus, this technology has the potential to improve commercially-important plant species. However, Taji et al. (2002) suggested that protoplast fusion should not be intended to replace conventional plant breeding methods, but rather to complement them. The main obstacles are the difficulty in regenerating fusion products and low fertility of recovered plants.

The fusion of cytoplasms from two protoplasts may or may not be followed by the fusion of their nuclei. When the fusion of nuclei does not take place, the cells become binucleate known as heterokaryon. However, when the nuclei fuse together the cells are known as hybrid or synkaryon or synkaryocyte. On the other hand, when cytoplasm fuse but genetic information from one of the two nuclei is lost, the cell becomes cybrid (cytoplasmic hybrid). Figure 10.2 represents schematic diagram of protoplast fusion resulting in hybrid and cybrid cells.

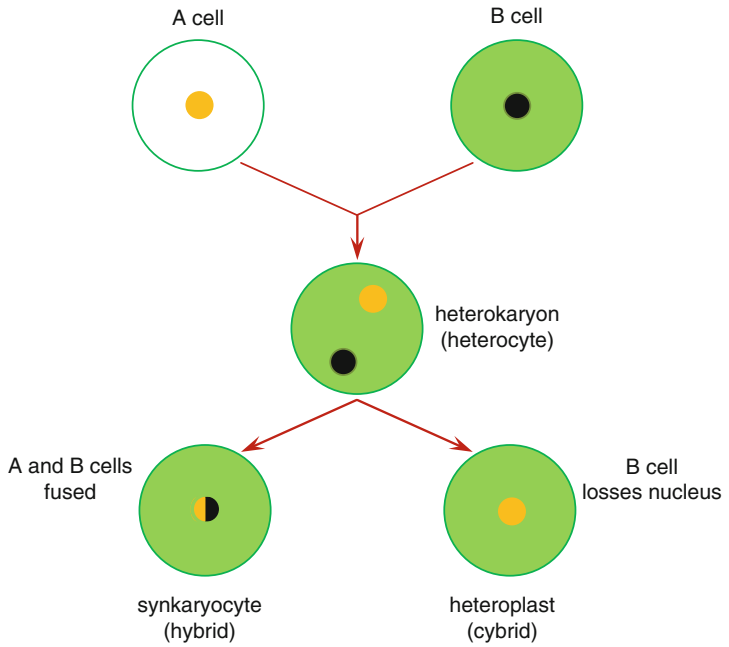

Fig. 10.2 Schematic diagram of hybrid and cybrid production through protoplast fusion

There are exceptions to the rule of chromosome theory of inheritance. One is that inherited characters may not be determined by genes located in chromosomes within the nucleus. A few characters are passed on from parents to offspring by genes which are not part of a nuclear chromosomes, but located in cell organelles in the cytoplasm (cytoplasmic genes) (Miko 2008). Tao et al. (2004) indicated that cytoplasmic and cytoplasmic-nuclear genomes interaction played important roles in yield, low temperature tolerance, and some important agronomic traits in *japonica* rice. Further Tao et al. (2011) detected that cytoplasmic genes had a significant effect on grain weight and filled-grain ratio on *indica* rice. In maize, Tang et al. (2013) reported that plant height and ear height are also controlled by cytoplasmic genes. These suggest that the selection of appropriate cytoplasmic germplasm is broadly important in plant breeding. Therefore, regeneration of cybrids with only one nuclear genome but having the mixture of cytoplasm from both parents will be very helpful in transfer of cytoplasmic genetic information from one plant to another. Thus, the production of cybrids can be applicable in plant breeding program.

Protoplast fusion may be used in somatic hybridization to create hybrid plants that may not be achieved through in vivo crossing techniques due to taxonomic or sexual barriers. However, it is important to note that even the creation of somatic hybrids from completely unrelated and incompatible species through protoplasts fusion is possible, the end products are often unbalanced (sterile, malformed and unstable), and therefore unviable (Chawla 2002).

10.6.4 Application of Protoplast Technology

Protoplasts provide a basic experimental system for ultrastructural, genetic and physiological studies. In addition, isolated protoplasts are also exploited in numerous studies involving synthesis of pharmaceutical products, and toxicological assessments. Several recent examples of the application of protoplasts technology are summarized in Table 10.3.

Cybrids produced as a result of protoplast fusion between *Nicotiana plumbaginifolia* TBR2 mutant and *N. tabacum* were resistant to high levels of herbicide atrazine (Menczel et al. 1986). They also found that these plants were male sterile due to the protruding stigma and shorter than normal filaments of the cybrid plants. Melchers et al. (1992) showed that when mitochondrial-inactivated tomato protoplasts were fused with nuclear-inactivated *Solanum* spp. protoplasts, cytoplasmic male sterility occurred in the resultant plants. Furthermore partial genome transfer can be obtained through asymmetric somatic hybridization when utilizing protoplast culture technology. Yamashita et al. (1989) produced asymmetric somatic hybrids of *Brassica* by fusing the inactivated *B. oleracea* protoplasts with X-irradiated *B. campestris* protoplasts.

10.7 Conclusion and Prospects

Advances in agricultural biotechnology are the cornerstone for feeding the growing world population, estimated to approach 10 billion by 2050. The impact of agriculture biotechnology has been immense since the first products of this technology were released, including crops that exhibited resistance to certain insects, tolerance to selected herbicides, and increased resistance to viral diseases, improving the yield of many crops dramatically and positively affecting the environment, the consumer and the farming companies. The next wave of agriculture biotechnology is further focused on producing crops that withstand biotic and abiotic stresses, as well as value adding to crops through increased levels of nutrients, vitamins and minerals. However, safety of products produced through genetic engineering has been a point of immense debate and hence such products, as yet, have not been widely embraced. The technologies described in this chapter have been utilized to complement classical plant breeding achieving the desired traits without the controversial method of genetic engineering. These technologies will further enhance crop productivity and produce novel plant-based biomaterials including potential energy resources in the face of limitations in natural resources and the inevitable impact of climate change on agriculture productivity.

In vitro techniques are now applied extensively in breeding programs because of their convenience and because of the potential of decreasing the breeding cycle. However, the lack of well-established in vitro techniques often prevents its practical use. A complete operation of in vitro screening and breeding involves variant or

mutant induction, selection, plant regeneration, acclimatization and assessment of in vivo plants. Extensive research into the establishment of each in vitro technique protocol for the success of plant screening and breeding programs is required. Recently, climate change has been affecting the growth and yield of plants all over the world. Growers may be forced to change to a new crop that is suitable for new environmental conditions such as extreme temperatures, drought, or chemical accumulated/contaminated soil. The endless possibilities for developing improvements in plants through technologies described in this chapter have the potential to help solve world hunger and problems faced in agriculture as a result of climate change.

References

Abdellatif KF, Hegazy AE, Aboshama HM et al (2012) Morphological and molecular characterization of somaclonal variations in tissue culture-derived banana plants. J Genet Eng Biotechnol 10(1):47–53

Agarwal M, Shrivastava N, Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep 27:617–631

Ahlawat A, Pahuja SK, Dhingra HR (2013) Studies on interspecific hybridization in *Cyamopsis* species. Afr J Agric Res 8(27):3590–3597

Ahmad I, Johansson S, Andrade G et al (2011) Somatic embryogenesis for mass propagation of elite Spruce families: effect of storage time on somatic embryogenesis initiation. BMC Proc 5(Suppl 7):P127

Albrecht CF, Stander MA, Grobbelaar MC et al (2012) LC-MS-based metabolomics assists with quality assessment and traceability of wild and cultivated plants of *Sutherlandia frutescens* (Fabaceae). S Afr J Bot 82:33–45

Aleza P, Juárez J, Cuenca J et al (2010) Recovery of citrus triploid hybrids by embryo rescue and flow cytometry from 2x × 2x sexual hybridisation and its application to extensive breeding programs. Plant Cell Rep 29(9):1023–1024

Aleza P, Juárez J, Cuenca J et al (2012) Extensive citrus triploid hybrid production by 2x × 4x sexual hybridizations and parent–effect on the length of the juvenile phase. Plant Cell Rep 31(9):1723–1735

Ali A, Afrasiab H, Naz S et al (2008) An efficient protocol for in vitro propagation of carnation (*Dianthus caryophyllus*). Pak J Bot 40:111–121

Al–Khayri JM, Huang FH, Morelock TE (1991) In vitro flowering in regenerated shoots of Spinach. Hortscience 26(11):1422

Al–Khayri JM, Huang FH, Morelock TE, Busharar TA (1992) In vitro seed production from sex–modified male spinach plants regenerated from callus cultures. Sci Hortic 52:277–282

Alves SAO, de Lemos OF, dos Santos Filho BG, da Silva ALL (2011) In vitro embryo rescue of interspecific hybrids of oil palm (*Elaeis oleifera* × *Elaeis guineensis*). J Biotechnol Biodivers 2(2):1–6

Ara H, Jaiswal U, Jaiswal VS (2000) Synthetic seed: prospects and limitations. Curr Sci 78(12):1438–1444

Aremu AO, Bairu MW, Szücová L et al (2013) Genetic fidelity in tissue–cultured 'Williams' bananas – the effect of high concentration of topolins and benzyladenine. Sci Hortic 161:324–327

Aruna C, Priya AR, Neeraja CN et al (2012) Diversity analysis using ISSR markers for resistance to shoot pests in sorghum. Crop Prot 35:110–117

Ashakiran K, Chidambareswaran M, Govindasamy V et al (2011) Somatic embryogenesis for *Agrobacterium* mediated transformation of tomato *Solanum lycopersicum*. Int J Biotechnol Appl 3(2):72–79

Asmah NH, Hasnida NH, Nashatul ZNA et al (2011) Synthetic seed technology for encapsulation and re growth of in vitro–derived *Acacia* hybrid shoot and axillary buds. Afr J Biotechnol 10(40):7820–7824

Aziz ZA, Davey MR, Lowe KC, Power JB (2006) Isolation and culture of protoplasts from the medicinal plant *Centella asiatica*. Rev Bras Plant Med 8:105–109

Bagniewska-Zadworna A, Zenkteler M et al (2010) A successful application of the embryo rescue technique as a model for studying crosses between *Salix viminalis* and *Populus* species. Aust J Bot 59(4):382–392

Bairu MW, Aremu AO, Van Staden J (2011) Somaclonal variation in plants: causes and detection methods. Plant Growth Regul 63(2):147–173

Banerjee P, Maity S, Banerjee N (2012) High frequency somatic embryogenesis and plantlet regeneration of *Bauhinia variegata*, a multipurpose tree legume. Indian J Fund Appl Life Sci 2:87–95

Barikissou E, Baudoin J–P (2011) Refinement of an in vitro culture technique for the rescue of globular embryos using microcutting for *P. vulgaris* L. and *P. coccineus* L. Tropicultura 29(4):218–224

Bekheet SA (2006) A synthetic seed method through encapsulation of in vitro proliferated bulblets of garlic (*Allium sativum* L.). Arab J Biotechnol 9(3):415–426

Belarmino MM, Gonzales JR (2008) Somatic embryogenesis and plant regeneration in purple food yam (*Dioscorea alata* L.). Ann Trop Res 30:22–33

Benelli C, Germanà MA, Ganino T et al (2010) Morphological and anatomical observations of abnormal somatic embryos from anther cultures of *Citrus reticulata*. Biol Plant 54(2):224–230

Beránek M, Bechyně M, Klíma M (2007) Protoplast isolation and fusion between *Brassica carinata* Braun and *Brassica rapa* L. Agric Trop Subtrop 40(1):1–7

Bhojwani SS, Razdan MK (1996) Plant tissue culture: theory and practice (rev ed). Studies in plant science, 5. Elsevier Science, Amsterdam

Bidhan R, Mandal AB (2008) Development of synthetic seeds involving androgenic and pro–embryos in elite *indica* rice. Indian J Biotechnol 7:515–519

Binding H, Jain S, Finger J et al (1982) Somatic hybridization of an atrazine resistant biotype of *Solanum nigrum* and *S. tuberosum*. I. Clonal variation in morphology and in atrazine sensitivity. Theor Appl Genet 63:273–277

Biswas MK, Dutt M, Roy UK et al (2009) Development and evaluation of in vitro somaclonal variation in strawberry for improved horticultural traits. Sci Hortic 122:409–416

Bornman CH, Dickens OSP, Botha AM (2001) The disadvantaged embryo. Acta Hortic 560:369–372

Bowne J, Bacic A, Tester M, Roessner U (2011) Abiotic stress and metabolomics. In: Hall RD (ed) Annual plant reviews (Vol. 43): Biology of plant metabolomics. Wiley-Blackwell, Oxford, pp 61–85

Bridgen MP (1994) A review of plant embryo culture. Hortscience 29(11):1243–1246

Brown J, Brown AP, Davis JB, Erickson D (1997) Intergeneric hybridization between *Sinapis alba* and *Brassica napus*. Euphytica 93:163–168

Campoy JA, Ruiz D, Egea J (2011) Dormancy in temperate fruit trees in a global warming context: a review. Sci Hortic 130:357–372

Chapman LA, Goring DR (2010) Pollen–pistil interactions regulating successful fertilization in the Brassicaceae. J Exp Bot 61(7):1987–1999

Chavan JJ, Gaikwad NB, Yadav SR (2013) High multiplication frequency and genetic stability analysis of *Ceropegia panchganiensis*, a threatened ornamental plant of Western Ghats: Conservation implications. Sci Hortic 161:134–142

Chawla HS (2002) Introduction to plant biotechnology, 2nd edn. Science Publishers, Inc, Enfield, New Hampshire
Chen Q, Li H, Shi YZ et al (2008) Development of an effective protoplast fusion system for production of new potatoes with disease and insect resistance using Mexican wild potato species as gene pools. Can J Plant Sci 88(4):611–619
Cheng SH, Sheen J, Gerrish C, Bolwell GP (2001) Molecular identification of phenylalanine ammonia–lyase as a substrate of a constitutively active *Arabidopsis* CDPK expressed in maize protoplasts. Fed Eur Biochem Soc Lett 503:185–188
Cheng YJ, Guo WW, Deng XX (2003) Molecular characterization of cytoplasmic and nuclear genomes in phenotypically abnormal Valencia orange (*Citrus sinensis*) plus Meiwa kumquat (*Fortunella crassifolia*) intergeneric somatic hybrids. Plant Cell Rep 21:445–451
Cheng A–X, Xia G–M, Chen H–M (2004) DNA transfer from wild millet to common wheat by asymmetric somatic hybridization. Acta Bot Sin 46(9):1114–1121
Chesnokov YV, Meister A, Manteuffel R (2002) A chimeric green fluorescent protein gene as an embryogenic marker in transgenic cell cultures of *Nicotiana plumbaginifolia*. Plant Sci 162:59–77
Chi H–S (2002) The efficiencies of various embryo rescue methods in interspecific crosses of *Lilium*. Bot Bull Acad Sin 43:139–146
Choi SK, Choi JK, Ryu KH (2003) Involvement of RNA2 for systemic infection of Cucumber Mosaic Virus isolated from lily on zucchini squash. Virus Res 97:1–6
Chuanen Z, Guangmin X (2005) Embryo rescue technology of distant hybrids in wheat. J Triticeae Crops 25(3):88–92
Chuda A, Adamus A (2012) Hybridization and molecular characterization of F1 *Allium cepa*×*Allium roylei* plants. Acta Biol Cracov Bot 54(2):25–31
Cisneros A, Tel–Zur N (2010) Embryo rescue and plant regeneration following interspecific crosses in the genus *Hylocereus* (Cactaceae). Euphytica 174:73–82
Cocking EC (1960) A method for the isolation of plant protoplasts and vacuoles. Nature 187:927–929
Collonnier C, Mulya K, Fock I et al (2001) Source of resistance against *Ralstonia solanacearum* in fertile somatic hybrids of eggplant (*Solanum melongena* L) with *Solanum aethiopicum* L. Plant Sci 160:301–303
Collonnier C, Fock I, Marie–Christine D et al (2003) Somatic hybrids between *Solanum melongena* and *S. sisymbrifolium*, as a useful source of resistance against bacterial and fungal wilts. Plant Sci 164:849–861
Cormeau J, Barthou H, Jauneau A et al (2002) Cellular import of synthetic peptide using a cell–permeable sequence in plant protoplasts. Plant Physiol Biochem 40:1081–1086
Correia SI, Canhoto JM (2012) Biotechnology of tamarillo (*Cyphomandra betacea*): From in vitro cloning to genetic transformation. Sci Hortic 148:161–168
da Silva JAT, Rana TS, Narzary D et al (2013) Pomegranate biology and biotechnology: a review. Sci Hortic 160:85–107
Dantas ACM, Boneti JI, Nodari RO, Guerra MP (2006) Embryo rescue from interspecific crosses in apple rootstocks. Pesq Agrop Bras 41(6):969–973
Datta KB, Kanjilal B, De Sarker D (2001) Artificial seed technology: development of a protocol in *Geodorum densiflorum* (Lam) Schltr. – an endangered orchid. Curr Sci 76:1142–1145
Davies H (2010) A role for "omics" technologies in food safety assessment. Food Control 21:1601–1610
De Filippis LF, Hampp R, Ziegler H (2000) Membrane permeability changes and ultrastructural abnormalities during protoplast fusion. J Plant Physiol 156:628–634
de Maagd RA, Hall RD (2013) The contribution of metabolomics research to the environmental risk assessment of genetically modified plants. Foundation Stichting Dienst Landbouwkundig Onderzoek (DLO) Research Institute Plant Research International, Wageningen

de-Alencar MS, Junior PCPF, da-Silva RA et al (2010) Morpho–anatomical characterization of embryogenic calluses from immature zygotic embryo of peach palm during somatic embryogenesis. Acta Sci-Agron 32:263–267

Debbarama C, Khanna VK, Tyagi W et al (2013) Wide hybridization and embryo-rescue for crop Improvement in *Capsicum*. Agrotechnology S11: 003. doi:10.4172/2168–9881S11–003

Deo PC, Tyagi AP, Taylor M et al (2010) Factors affecting somatic embryogenesis and transformation in modern plant breeding. South Pac J Nat Appl Sci 28:27–40

Dhiman M, Sharma V, Moitra S (2010) Somatic embryogenesis and plant regeneration in *Ephedra foliata* (Boiss.): a non coniferous Gymnosperm. Plant Tiss Cult Biotech 20(2):133–143

Dinesh MR, Veena GL, Vasugi C et al (2013) Intergeneric hybridization in papaya for 'PRSV' tolerance. Sci Hortic 161:357–360

Ding X, Iwasaki I, Kitagawa Y (2004) Overexpression of a lily PIP1 gene in tobacco increased the osmotic water permeability of leaf cells. Plant Cell Envir 27:177–186

Dobrowolska I, Majchrzak O, Baldwin TC, Kurczynska EU (2012) Differences in protodermal cell wall structure in zygotic and somatic embryos of *Daucus carota* (L.) cultured on solid and in liquid media. Protoplas 249(1):117–129

Ehsanpour AA, Madani S, Hoseini M (2007) Detection of somaclonal variation in Potato callus by UV-C radiation using RAPD-PCR. Gen Appl Plant Physiol 33(1–2):3–11

Etherton B, Heppner TJ, Cumming JR, Nelson MT (2004) Opposing effects of aluminium on inward–rectifier potassium currents in bean–tip protoplasts. J Membr Biol 198:15–22

Fan LM, Wang YF, Wu WH (2003) Outward K^+ channels in *Brassica chinensis* pollen protoplasts are regulated by external and internal pH. Protoplasma 220:143–152

Fathi H, Jahani (2012) Review of embryo culture in fruit trees. Ann Biol Res 3(9):4276–4281

Fki L, Bouaziz N, Kriaa W et al (2011) Multiple bud cultures of 'Barhee'™ date palm (*Phoenix dactylifera*) and physiological status of regenerated plants. J Plant Physiol 168:1694–1700

Fock I, Collonnier C, Luisetti J et al (2001) Use of *Solanum stenotomum* for introduction of resistance to bacterial wilt in somatic hybrids of potato. Plant Physiol Biochem 39:899–908

Fu CH, Guo WW, Liu JH, Deng XX (2003) Regeneration of *Citrus sinensis* (+) *Clausena lansium* intergeneric triploid and tetraploid somatic hybrids and their identification by molecular markers. In Vitro Cel Dev Biol Plant 39:360–364

Gavrilenko T, Thieme HU, Thieme T (2003) Fertile somatic hybrids of *Solanum etuberosum* (+) dihaploid *Solanum tuberosum* and their backcrossing progenies: relationships of genome dosage with tuber development and resistance to potato virus Y. Euphytica 131:323–332

Ghanta R, Mondal S (2013) Effect of some nutrients on in vitro pollen germination of *Withania somnifera* (L.) Dunal. Ann Plant Sci 2(6):182–187

Giavalisco P, Nordhoff E, Kreitler T et al (2005) Proteome analysis of *Arabidopsis thaliana* by two–dimensional gel electrophoresis and matrix assisted laser desorption/ionisation–time of flight mass spectrometry. Proteome 5:1902–1913

Godbole S, Sood A, Thakur R et al (2002) Somatic embryogenesis and it's conversion into plantlets in a multipurpose bamboo, *Dendrocalamus hamiltonii* Nees *et* Arn. Ex Munro. Curr Sci 83:885–889

Goh CH, Dietrich P, Steinmeyer R et al (2002) Parallel recordings of photosynthetic electron transport and K^+ channel activity in single guard cells. Plant J 32:623–630

Grosser JW, Gmitter FG Jr (2010) Protoplast fusion for production of tetraploids and triploids: applications for scion and rootstock breeding in citrus. Plant Cell Tiss Org Cult 104:343–357

Guha S, Maheshwari P (1964) In vitro production of embryos from anthers of *Datura*. Nature 204:497

Gunashree BS, Venkateswaran G (2010) Enhanced phytase production through interspecific protoplast fusion of *Aspergillus niger* CFR 335 and *Aspergillus ficuum* SGA 01 auxotrophic mutants. Enzym Microb Technol 46(7):562–567

Guo Y, Zhao Y, Li K et al (2011) Embryo rescue of crosses between diploid and tetraploid grape cultivars and production of triploid plants. Afr J Biotechnol 10(82):19005–19010

Hogenboom NG (1973) A model for incongruity in intimate partner relationships. Euphytica 22:219–233
Hu C, Wang P (1986) Embryo culture: technique and applications. In: Evans DA, Sharp WR, Ammirato PV (eds) Handbook of plant cell culture, vol 4. Macmillan, New York, pp 43–96
Hu Q, Andersen SB, Dixelius C, Hansen LN (2002a) Production of fertile intergeneric somatic hybrids between *Brassica napus* and *Sinapsis arvensis* for the enrichment of the rapeseed gene pool. Plant Cell Rep 21:147–152
Hu Q, Hansen LN, Laursen J et al (2002b) Intergeneric hybrids between *Brassica napus* and *Orychophragmus violaceus* containing traits of agronomic importance for oilseed rape breeding. Theor Appl Genet 105:834–840
Ipekci Z, Gozukirmizi N (2003) Direct somatic embryogenesis and synthetic seed production from *Paulownia elongata*. Plant Cell Rep 22:16–24
Jain SM, Shahin EA, Sun S (1988) Interspecific protoplast fusion for the transfer of atrazine resistance from Solanum nigrum to tomato (*Lycopersicon esculentum* L.). In: Puite KJ, Dons JJM, Huizing HJ et al (eds) Progress in plant protoplast research, vol 7. Springer, Dordrecht, pp 221–224
Jain SM, Sopory SK, Veilleux RE (1996a) In vitro haploid production in higher plants, vol 1 – Fundamental aspects and methods. Current plant science and biotechnology in agriculture. Springer, Dordrecht
Jain SM, Sopory SK, Veilleux RE (eds) (1996b) In vitro haploid production in higher plants, vol 2 – Applications current plants and biotechnology in agriculture. Springer, Dordrecht
Jain SM, Sopory SK, Veilleux RE (eds) (1996c) In vitro haploid production in higher plants, vol 3 – Important selected plants current plant science and biotechnology in agriculture. Springer, Dordrecht
Jain SM, Sopory SK, Veilleux RE (eds) (1996d) In vitro haploid production in higher plants, vol 4 – Cereals. current plant science and biotechnology in agriculture. Springer, Dordrecht
Jain SM, Sopory SK, Veilleux RE (eds) (1996e) In vitro haploid production in higher plants, vol 5 – Oil, ornamental and miscellaneous plants current plant science and biotechnology in agriculture. Springer, Dordrecht
Jain SM, Brar DS, Ahloowalia BS (eds) (1998) Somaclonal variation and induced mutations in crop improvement, vol 32. Current plant science and biotechnology in agriculture. Springer, Dordrecht
Jayaprakash P, Sabesan T (2013) In vitro pollen germination of some wild species of pigeonpea (*Cajanus cajan*) using PGM droplet technique. Indian J Genet Plant Breed 73(2):211–215
Ji A, Geng X, Zhang Y et al (2011) Advances in somatic embryogenesis research of horticultural plants. Am J Plant Sci 2:727–732
Jiang C, Mithani A, Gan X et al (2011) Regenerant *Arabidopsis* lineages display a distinct genome–wide spectrum of mutations conferring variant phenotypes. Curr Biol 21:1385–1390
Jin S, Mushke R, Zhu H et al (2008) Detection of somaclonal variation of cotton (*Gossypium hirsutum*) using cytogenetics, flow cytometry and molecular markers. Plant Cell Rep 27(8):1303–1316
Joshi M, Nadgauda RS (1997) Cytokinins and in vitro induction of flowering in bamboo: *Bambusa arundinaceae* (Retz.) Willd. Curr Sci 73(6):523–527
Kakani VG, Reddy KR, Koti S et al (2005) Differences in in vitro pollen germination and pollen tube growth of cotton cultivars in response to high temperature. Ann Bot 96(1):59–67
Kamle M, Bajpai A, Chandra R et al (2011) Somatic embryogenesis for crop improvement. GERF Bull Biosci 2(1):54–59
Kanchanapoom K, Posayapisit N, Kanchanapoom K (2009) In vitro flowering from cultured nodal explants of Rose (*Rosa hybrida* L.). Not Bot Hort Agrobot Cluj 37(2):261–263
Kanta K, Rangaswamy NS, Maheshwari P (1962) Test–tube fertilization in a flowering plant. Nature 194:1214–1217
Karami O, Esna-Ashari M, Piri K, Almasi P (2007) Efficient regeneration of carnation (*Dianthus caryophyllus* L.) via somatic embryogenesis. Propag Orn Plants 7:3–8

Karamian R, Ranjbar M (2010) Somatic embryogenesis and plantlet regeneration from protoplast culture of *Muscari neglectum* Guss. Afr J Biotechnol 10(22):4602–4607

Kesari V, Ramesh AM, Rangan L (2012) High frequency direct organogenesis and evaluation of genetic stability for in vitro regenerated *Pongamia pinnata*, a valuable biodiesel plant. Biomass Bioen 44:23–32

Khaleghi A, Khalighi A, Azadi P, Mii M (2008) Induction of embryogenic callus and plant regeneration from nodes of greenhouse grown plants of *Alstroemeria* cv. Fuego. J Food Agr Environ 6:374–377

Khan KH (2009) Gene transfer technologies in plants: Roles in improving crops. Recent Res Sci Technol 1(3):116–123

Khor E, Loh CS (2005) Artificial seeds. In: Nedović V, Willaert R (eds) Applications of cell immobilisation biotechnology. Springer, Dordrecht, pp 527–537

Kieber JJ (2002) Tribute to Folke Skoog: Recent advances in our understanding of cytokinin biology. J Plant Growth Regul 21(1):1–2

Kim YJ, Kim JE, Lee JH et al (2004) The Vr–PLC3 gene encodes a putative plasma membrane–localized phosphoinositide–specific phospholipase C whose expression is induced by abiotic stress in mung bean (*Vigna radiata* L.). Fed Eur Biochem Soc Lett 556(1-3):127–136

Kisaka H, Kisaka M, Kanno A, Kameya T (2001) Somatic hybridization between *Hordeum vulgare* L. (barley) and *Daucus carota* L. (carrot). In: Nagata T, Bajaj YPS (eds) Biotechnology in agriculture and forestry: somatic hybridization in crop improvement II, vol 49. Springer–Verlag, Berlin/Heidelberg, pp 3–7

Kishor R, Sharma GJ (2009) Intergeneric hybrid of two rare and endangered orchids, *Renanthera imschootiana* Rolfe and *Vanda coerulea* Griff. ex L. (Orchidaceae): synthesis and characterization. Euphytica 165(2):247–256

Kitto SK, Janick J (1982) Polyox as an artificial seed coat for asexual embryos. HortSci 17:488

Klíma M, Abraha E, Vyvadilová M, Bechyn M (2009) Protoplast culture and fusion between *Brassica carinata* and *Brassica napus*. Agric Trop Subtrop 42(1):34–45

Kong D-m, Shen H-l, Li N (2012) Influence of $AgNO_3$ on somatic embryo induction and development in Manchurian ash (*Fraxinus mandshurica* Rupr.). Afr J Biotechnol 11(1):120–125

Krikorian AD, Simola LK (1999) Totipotency, somatic embryogenesis, and Harry Waris (1893–1973). Physiol Plant 105(2):347–354

Krishnamoorthy RN, Narayanan R, Vijila K, Kumutha K (2010) Intergeneric protoplast fusion of yeast for high ethanol production from cheese industry waste – whey. J Yeast Fungal Res 1(15):81–87

Kukharchyk N, Kastrickaya M (2006) Embryo rescue techniques in *Prunus* L. breeding. J Fruit Orn Plant Res 14(Suppl. 1):129–135

Kumar MBA, Vakeswaran V, Krishnasamy V (2005) Enhancement of synthetic seed conversion to seedlings in hybrid rice. Plant Cell Tiss Org Cult 81:97–100

Lakshmanan P, Taji AM (2004) *In vitro* flowering. In: Goodman RM (ed) Encyclopedia of plant and crop science. Marcel Dekker, New York, pp 576–578

Leister RT, Katagiri F (2000) A resistance gene product of the nucleotide binding site–leucine rich repeats class can form a complex with bacterial avirulence proteins *in vivo*. Plant J 22:345–354

Leva AR, Petruccelli P, Rinaldi LMR (2012) Somaclonal variation in tissue culture: A case study with olive. In: Leva AR, Rinaldi LMR (eds) Recent advances in plant *in vitro* culture. InTech, Rijeka, pp 123–150

Li X, Krasnyanski SF, Korban SS (2002) Optimization of the *uidA* gene transfer into somatic embryos of rose via *Agrobacterium tumefaciens*. Plant Physiol Biochem 40:453–459

Liang XZ, Ding SW, Wong SM (2002) Development of kenaf (*Hibiscus cannabinus* L.) protoplast system for a replication study of hibiscus chlorotic ringspot virus. Plant Cell Rep 20:982–986

Lim K-B, Barba-Gonzalez R, Zhou S et al (2008) Interspecific hybridization in Lily (*Lilium*): Taxonomic and commercial aspects of using species hybrids in breeding. In: da Silva JAT (ed)

Floriculture, ornamental and plant biotechnology, vol V, Advances and topical issues. Global Science Books, Isleworth, pp 146–151
Lin CC, Lin CS, Chang WC (2003) In vitro flowering of *Bambusa edulis* and subsequent plantlet survival. Plant Cell Tiss Org Cult 72:71–78
Lin CS, Lin CC, Chang WC (2004) Effect of Thidiazuron on vegetative tissue derived somatic embryogenesis and in vitro flowering of bamboo *Bambusa edulis*. Plant Cell Tiss Org Cult 76:75–82
Lord EM, Russell SD (2002) The mechanisms of pollination and fertilization in plants. Annu Rev Cell Dev Biol 18:81–105
Luciani GF, Mary AK, Pellegrini C, Curvetto NR (2006) Effects of explants and growth regulators in garlic callus formation and plant regeneration. Plant Cell Tiss Org Cult 87:139–143
Maddock SE (1985) Cell culture, somatic embryogenesis and plant regeneration in wheat, barley, oat, rye and triticale. In: Bright SWJ, Jones MGK (eds) Advances in agriculture biotechnology, vol 15. Springer, Dordrecht, pp 131–174
Maluszynski M, Kasha KJ, Szarejko I (2003) Published doubled haploid protocols in plant species. In: Maluszynski M, Kasha KJ, Forster BP, Szarejko I (eds) Doubled haploid production in crop plants: a manual. Kluwer, Dordrecht, pp 309–335
Mamiya K, Sakamoto Y (2001) A method to produce encapsulatable units for synthetic seeds in *Asparagus officinalis*. Plant Cell Tiss Org Cult 64:27–32
Mandal AKA, Gupta SD (2002) Direct somatic embryogenesis of safflower: A scanning electron microscopic study. Curr Sci 83:1138–1140
Mathew MM, Philip VJ (2003) Somatic embryogenesis versus zygotic embryogenesis in *Ensete superbum*. Plant Cell Tiss Org Cult 72(3):267–275
McCoy TJ (1985) Interspecific hybridisation of *Medicago sativa* L. and *M. rupestris* M.B. using ovule–embryo culture. Can J Genet Cytol 27:238–245
Mehetre SS, Aher AR (2004) Embryo rescue: a tool to overcome incompatible interspecific hybridization in *Gossypium* Linn. – a review. Indian J Biotechnol 3:29–36
Melchers G, Labib G (1974) Somatic hybridisation of plants by fusion of protoplasts I. Selection of light resistant hybrids of "haploid" light sensitive varieties of tobacco. Mol Genet Genomics 135(4):277–294
Melchers G, Mohri Y, Watanabe K et al (1992) One-step generation of cytoplasmic male sterility by fusion of mitochondrial-inactivated tomato protoplasts with nuclear-inactivated *Solanum* protoplasts. Proc Natl Acad Sci 89:6832–6836
Menczel L, Polsby LS, Steinback KE, Maliage P (1986) Fusion-mediated transfer of triazine-resistant chloroplasts: characterization of *Nicotiana tabacum* cybrid plants. Mol Gen Genet 205:201–205
Miko I (2008) Non-nuclear genes and their inheritance. Nativ Educ 1(1):135
Miransari M, Smith DL (2013) Plant hormones and seed germination. Environ Exp Bot 99:110–121
Moco S, Vervoort J (2012) Chemical identification strategies by liquid chromatography (LC)–photo diode array (PDA)–Solid Phase Extraction (SPE)–Nuclear Magnetic Resonance (NMR)/Mass Spectrometry (MS). In: Hardy NW, Hall RD (eds) Plant metabolomics methods and protocols: Methods in molecular biology, vol 860. Humana Press, Ottawa, pp 287–316
Momotaz A, Kato M, Kakihara F (1998) Production of intergeneric hybrids between *Brassica* and *Sinapis soecies* by means of embryo rescue technique. Euphytica 103:123–130
Mondal S, Ghanta R (2012) Effect of sucrose and boric acid on in vitro pollen germination of *Solanum macranthum* Dunal. Indian J Fund Appl Life Sci 2(2):202–206
Morel A, Teyssier C, Beaufour M et al (2012) Somatic embryo maturation in Maritime Pine (*Pinus pinaster*): contribution of a 2–DE proteomic analysis for a better understanding. Paper presented at the IUFRO Working Party 2.09.02 Conference on integrating vegetative propagation, biotechnologies and genetic improvement for tree production and sustainable forest management, 25–28 Jun 2012, Brno, Czech Republic

Nadgauda RS, Parasharami VA, Mascarenhas AF (1990) Precocious flowering and seeding behaviour in tissue–cultured bamboos. Nature 344:335–336

Nakamura T, Kuwayama S, Tanaka S et al (2005) Production of intergeneric hybrid plants between *Sandersonia aurantiaca* and *Gloriosa rothschildiana* via ovule culture (Colchicaceae). Euphytica 142(3):283–289

Neuenschwander B, Baumann TW (1992) A novel type of somatic embryogenesis in *Coffea Arabica*. Plant Cell Rep 10:608–612

Nimisha S, Kherwar D, Ajay KM et al (2013) Molecular breeding to improve guava (*Psidium guajava* L.): current status and future prospective. Sci Hortic 164:578–588

Nitsch JP, Nitsch C (1969) Haploid plants from pollen grains. Science 163:85–87

Nookaraju A, Agrawal DC (2012) Genetic homogeneity of in vitro raised plants of grapevine cv. Crimson Seedless revealed by ISSR and microsatellite markers. S Afr J Bot 78:302–306

Nookaraju A, Barreto MS, Karibasappa GS, Agrawal DC (2007) Synergistic effect of CPPU and benzyladenine on embryo rescue in six stenospermocarpic cultivars of grapevine. Vitis 46(4):188–191

Nower AA, Ali EA, Rizkalla AA (2007) Synthetic seeds of pear (*Pyrus communis* L.) rootstock storage in vitro. Aust J Basic Appl Sci 1(3):262–270

Nwauzoma AB, Jaja ET (2013) A review of somaclonal variation in plantain (*Musa* spp.): mechanisms and applications. J Appl Biosci 67:5252–5260

Palada-Nicolaua M, Hausman J-F (2001) Comparison between somatic and zygotic embryo development in *Quercus robur* L. Plant Biosyst 135(1):47–55

Palama TL, Menard P, Fock YH et al (2010) Shoot differentiation from protocorm callus cultures of *Vanilla planifolia* (Orchidaceae): proteomic and metabolic responses at early stage. BMC Plant Biol 10:82

Pathirana R, Scalzo J, Frew T et al (2013) Ovule culture and embryo rescue facilitates interspecific hybridisation in blueberry (*Vaccinium* spp.). Paper presented at the 20th biennial meeting of The New Zealand Branch of The International Association for Plant Biotechnology, Waiheke Island, Auckland, 25–28 Feb 2013

Pati PK, Sharma M, Ahuja PS (2008) Protoplast isolation and culture. In: Keshavachanadran R, Peter KV (eds) Plant biotechnology: methods in tissue culture and gene transfer, vol 14. Universities Press Pvt. Ltd, Hyderabad, pp 115–132

Pelletier G, Primard C, Vedal F et al (1983) Intergeneric cytoplasmic hybridization in cruciferae by protoplast fusion. Mol Gen Genet 191:244–250

Pennazio S (2002) The discovery of the chemical nature of the plant hormone auxin. Riv Biol 95(2):289–308

Pineiro M, Jarillo JA (2013) Ubiquitination in the control of photoperiodic flowering. Plant Sci 198:98–109

Pintos B, Bueno MA, Cuenca B, Manzanera JA (2008) Synthetic seed production from encapsulated somatic embryos of cork oak (*Quercus suber* L.) and automated growth monitoring. Plant Cell Tiss Org Cult 95(2):217–225

Popielarska M (2005) In vitro pollination of isolated ovules of sunflower (*Helianthus annuus* L.). Acta Biol Cracov Bot 47(1):85–92

Pose D, Yant L, Schmid M (2012) The end of innocence: Flowering networks explode in complexity. Curr Opin Plant Biol 15:45–50

Priyono (1993) Embriogenesis somatik langsung pada kultur in vitro eksplan daun kopi arabika (*Coffea arabica*). Jurnal Pertanian Indonesia 3(1):16–20

Proseviçius J, Strikulyte L (2004) Interspecific hybridization and embryo rescue in breeding of lilies. Acta Univ Latv Biol 676:213–217

Qian W, Liu R, Meng J (2003) Genetic effects on biomass yield in interspecific hybrids between *Brassica napus* and *B. rapa*. Euphytica 134:9–15

Quainoo AK, Dwomo BI (2012) The effect of TDZ and 2, 4–D concentrations on the Induction of somatic embryo and embryogenesis in different cocoa genotypes. J Plant Stud 1(1):72–78

Rady MR, Hanafy MS (2004) Synthetic seed technology for encapsulation and regrowth of in vitro–derived *Gypsophila paniculata* L. shoot tips. Arab J Biotechnol 7(1):251–264
Raghavan V (1986) Embryogenesis in angiosperms: A developmental and experimental study. Cambridge University Press, Cambridge
Rakha MT, Metwally EI, Moustafa SA et al (2012) Evaluation of regenerated strains from six Cucurbita interspecific hybrids obtained through anther and ovule in vitro cultures. Aust J Crop Sci 6(1):23–30
Rakosy-Tican L, Hornok MLM (2001) An improved procedure for protoplast microelectrofusion and culture of *Nicotiana tabacum* intraspecific somatic hybrids: plant regeneration and initial proofs on organelle segregation. Plant Cell Tiss Org Cult 67:153–158
Ramanayake SMSD (2006) Flowering in bamboo: An enigma! Ceylon J Sci 35(2):95–105
Ramon M, Hanneman RE Jr (2002) Introgression of resistance to late blight (*Phytophthora infestense*) from *Solanum pinnatisectum* into *S. tuberosum* using embryo rescue and double pollination. Euphytica 127:421–435
Rao NK, Reddy LJ, Bramel P (2003) Potential of wild species for genetic enhancement of some semi-arid food crops. Genet Resour Crop Evol 50:707–721
Ravi D, Anand P (2012) Production and applications of artificial seeds: a review. Int Res J Biol Sci 1(5):74–78
Rea G, Antonacci A, Lambreva M et al (2011) Integrated plant biotechnologies applied to safer and healthier food production: The Nutra–Snack manufacturing chain. Trends Food Sci Technol 22:353–356
Reed SM (2005) Embryo rescue. In: Trigiano RN, Gray DJ (eds) Plant development and biotechnology. CRC Press, Boca Raton, pp 235–239
Reinert J (1959) Ueber die kontrolle der morphogenese und die induktion von adventiveembryonen an gewebekulturen aus karotten. Planta 53:318–333
Ren JP, Dickson MH, Earle ED (2000) Improved resistance to bacterial soft rot by protoplast fusion between *Brassica rapa* and *B. oleracea*. Theor Appl Genet 100:810–819
Rey NA, Papacchioli V, Tavazza R, Pagnotta MA (2013) Gauging the genetic changes occurring across globe artichoke micropropagation towards an appropriate variety registration and nursery production. Sci Hortic 156:121–126
Rodriguez-Enriquez J, Dickinson HG, Grant-Downton RT (2011) MicroRNA misregulation: an overlooked factor generating somaclonal variation? Trends Plant Sci 16:242–248
Saiprasad GVS (2011) Artificial seeds and their applications. Reson May 6(5):39–47
Saiprasad GVS, Polisetty R (2003) Propagation of three orchid genera using encapsulated protocorm-like bodies. In Vitro Cel Dev Biol Plant 39:42–48
Schauer S, Fernie AR (2006) Plant metabolomics: towards biological function and mechanism. Trends Plant Sci 11:508–516
Seran TH, Hirimburegama K, Gunasekare MTK (2005) Encapsulation of embryonic axes of *Camellia sinensis* (L.) O. Kuntze (tea) and subsequent in vitro germination. J Hortic Sci Biotechnol 80:154–158
Shapka N, Nagy PD (2004) The AU–rich RNA recombination hot spot sequence of Brome Mosaic Virus is functional in tombus viruses: implications for the mechanism of RNA recombination. J Virol 78:2288–2230
Sharma DR, Kaur R, Kumar K (1996) Embryo rescue in plants – a review. Euphytica 89:325–337
Sheeja TE, Mandal AKA (2003) In vitro flowering and fruiting in tomato (*Lycopersicon esculentum* Mill.). Asia Pac J Mol Biol Biotechnol 11(1):37–42
Sheen J (2001) Signal transduction in maize and *Arabidopsis* mesophyll protoplasts. Plant Physiol 127:1466–1475
Siemens J (2002) Interspecific hybridisation between wild relatives and *Brassica napus* to introduce new resistance traits into the oilseed rape gene pool. Czech J Genet Plant Breed 38(3–4):155–157

Sigareva MA, Earle ED (1997) Direct transfer of a cold-tolerant Ogura male-sterile cytoplasm into cabbage (*Brassica oleracea* ssp. *capitata*) via protoplast fusion. Theor Appl Genet 94:213–220

Simola L (2000) Harry Waris, A Pioneer in Somatic Embryogenesis. In: Jain SM, Gupta PK, Newton RJ (eds) Somatic Embryogenesis in Woody Plants, vol 6. Kluwer Academic Publishers, Dordrecht, pp 1–16

Singh M, Chaturvedi R (2013) Somatic embryogenesis in neem (*Azadirachta indica* A. Juss.): Current status and biotechnological perspectives. In: Aslam J, Srivastava PS, Sharma MP (eds) Somatic embryogenesis and gene expression. Narosa Publishing House, New Delhi, pp 35–55

Siong PK, Mohajer S, Taha RM (2012) Production of artificial seeds derived from encapsulated in vitro micro shoots of cauliflower. Brassica oleracea var botrytis Rom Biotechnol Lett 17(4):7549–7556

Skálová D, Navrátilová B, Lebeda A (2008) Embryo rescue of cucumber (*Cucumis sativus*), muskmelon (*C. melo*) and some wild *Cucumis* species (*C. anguria*, *C. zeyheri*, and *C. metuliferus*). J Appl Bot Food Qual 82:83–89

Skálová D, Navrátilová D, Ondřej V, Lebeda A (2010) Optimizing culture for in vitro pollination and fertilization in *Cucumis sativus* and *C. melo*. Acta Biol Cracov Bot 52(1):111–115

Slater A, Scott NW, Fowler MR (2003) Plant biotechnology. Oxford University Press, Oxford

Song YH, Ito S, Imaizumi T (2013) Flowering time regulation: Photoperiod– and temperature–sensing in leaves. Trends Plant Sci 18:575–583

Sreenath HL, Shanta HM, Babu KH, Naidu MM (1995) Somatic embryogenesis from integument (perisperm) cultures of coffee. Plant Cell Rep 14:670–673

Sri Rama Murthy K, Kondamudi R, Chalapathi Rao PV, Pullaiah T (2012) In vitro flowering – a review. J Agr Technol 8(5):1517–1536

Srinivasan TR, Sagadevan E, Subhankar C et al (2009) Inter–specific protoplast fusion in *Trichoderma* spp. for enhancing enzyme production and biocontrol activity. J Phytol 1(5):285–298

Stebbins GL (1950) Variation and evolution in plants. Columbia University Press, New York

Stebbins GL (1958) The inviability, weakness, and sterility of interspecific hybrids. In: Demerec M (ed) Advances in genetics, vol 9. Academic, New York, pp 147–215

Steiner N, Santa–Catarina C, Guerra MP et al (2012) A gymnosperm homolog of somatic embryogenesis receptor–like kinase–1 (SERK1) is expressed during somatic embryogenesis. Plant Cell Tiss Org Cult 109(1):41–50

Steward FC, Mapes MO, Mears K (1958) Growth and organized development of cultured cells. Am J Bot 45:693–713

Sudhersan C, Al-Shayji Y (2011) Interspecific hybridization and embryo rescue in date palm. In: Jain SM, Al-Khayri JM, Johnson DV (eds) Date palm biotechnology. Springer, Dordrecht, pp 567–584

Sudhersan C, Al-Shayji Y, Jibi MS (2009) Date crop improvement via TxD hybridization integrated with in vitro culture technique. Acta Hortic 829:219–224

Suga S, Murai M, Kuwagata T, Maeshima M (2003) Differences in aquaporin levels among cell types of radish and measurement of osmotic water permeability of individual protoplasts. Plant Cell Phys 44:277–286

Taji AM, Kumar K, Lakshmanan P (2002) *In vitro* plant breeding. Haworth Press, Inc., New York

Takeda S, Matsuoka M (2008) Genetic approaches to crop improvement: responding to environmental and population changes. Nat Rev Genet 9:444–457

Tan M, Yan M, Wang L et al (2013) Replication of pistillate plants of *Ricinus communis* L. and investigation of the sex stability and genetic variation of the somaclones. Ind Crops Prod 50:50–57

Tang Z, Yang Z, Hu Z et al (2013) Cytonuclear epistatic quantitative trait locus mapping for plant height and ear height in maize. Mol Breed 31:1–14

Tao D, Hu F, Yang J et al (2004) Cytoplasm and cytoplasm–nucleus interactions affect agronomic traits in *japonica* rice. Euphytica 135:129–134

Tao D, Xu P, Zhou J et al (2011) Cytoplasm affects grain weight and filled–grain ratio in *indica* rice. BMC Genet 12:53

Taoka K, Ohki I, Tsuji H et al (2013) Structure and function of florigen and the receptor complex. Trends Plant Sci 18(5):287–294

Tapingkae T, Zulkarnain Z, Kawaguchi M et al (2012) Somatic (asexual) procedures (haploids, protoplasts, cell selection) and their applications. In: Altman A, Hasegawa PM (eds) Plant biotechnology and agriculture: Prospects for the 21st century. Academic, Oxford, pp 139–162

Terryn N, Van Montagu M, Inze D, Goossens A (2006) Functional genomic approaches to study and engineer secondary metabolism in plant cell cultures. In: Bogers RJ, Craker LE, Lange D (eds) Medicinal and aromatic plants. Springer, Dordrecht, pp 291–300

Thakur AK, Chauhan DK, Parmar N, Verma V (2012) Role of genetic engineering in horticultural crop improvement – a review. Agr Rev 33(3):248–255

Thomas BR, Pratt D (1981) Efficient hybridization between *Lycopersicon esculentum* tomatoes and *Lycopersicon peruvianum* via embryo callus. Theor Appl Genet 59:215–219

Tian L, Wang Y (2008) Seedless grape breeding for disease resistance by using embryo rescue. Vitis 47(1):15–19

Tomar UK, Dantu PK (2010) Protoplast culture and somatic hybridization. In: Tripathi G (ed) Cellular and biochemical science. I.K. International House Pvt. Ltd., New Delhi, pp 876–891

Torres AC, Mfe'e Ze N, Cantliffe DJ (2001) Abscisic acid and osmotic induction of synchronous somatic embryo development of sweet potato. In Vitro Cel Dev Biol Plant 37(2):262–267

Van Tuyl JM, De Jeu MJ (1997) Methods for overcoming interspecific crossing barriers. pollen biotechnology for crop production and improvement. Cambridge University Press, London

Vasil IK (2008) A history of plant biotechnology: From the cell theory of Schleiden and Schwann to biotech crops. Plant Cell Rep 27:1423–1440

Vdovitchenko YM, Kuzovkina IN (2011) Artificial seeds as a way to produce ecologically clean herbal remedies and to preserve endangered plant species. Mosc Univ Biol Sci Bull 66(2):48–50

Vermeer JEM, van Munster EB, Vischer NO, Gadella TWJ (2004) Probing plasma membrane microdomains in cowpea protoplasts using lapidated GFP–fusion proteins and multimode FRET. J Microsc 214:190–200

Vervaeke I, Parton E, Deroose R, De Proft MP (2002) Controlling prefertilization barriers by in vitro pollination and fertilization of bromeliaceae. Paper presented at the proceedings of the 12th international eucarpia symposium, Section Ornamentals: Strategies for New Ornamentals II Melle, Belgium

von Klercker J (1892) Eine methode zur isolier lebender protoplasten. Ofvers Vetensk Akad Forhandl 49:363–474

Wang J, Huang L, Bao M-z, G-f L (2012) Production of interspecific hybrids between *Lilium longiflorum* and *L. lophophorum* var. *linearifolium* via ovule culture at early stage. Euphytica 167(1):45–55

Wang X, Elling AA, Li X et al (2009) Genome–wide and organ–specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. Plant Cell 21:1053–1069

Wang YP, Sonntag K, Rudloff E (2003) Development of rapeseed with high erucic acid content by asymmetric somatic hybridisation between *Brassica napus* and *Crambe abyssinica*. Theor Appl Genet 106:1147–1155

Wang YY, Kuang A, Russell SD, Tian HQ (2006) In vitro fertilization as a tool for investigating sexual reproduction of angiosperms. Sex Plant Reprod 19(3):103–115

Waterer D, Lee S, Scoles G, Keller W (2000) Field evaluation of herbicide-resistant transgenic broccoli. Hortscience 35:930–932

Weber S, Horn R, Friedt W (2000) High regeneration potential in vitro of sunflower (*Helianthus annuus* L.) lines derived from interspecific hybridization. Euphytica 116(3):271–280

Winkelmann T, Meyer L, Serek M (2004) Germination of encapsulated somatic embryos of *Cyclamen persicum*. Hortscience 39:1093–1097

Yaacob JS, Yussof AIM, Taha RM, Mohajer S (2012) Somatic embryogenesis and plant regeneration from bulb, leaf and root explants of African blue lily (*Agapanthus praecox* ssp. *minimus*). Aust J Crop Sci 6(10):1462–1470

Yadav K, Singh N (2011) In vitro flowering of shoots regenerated from cultured nodal explants of *Spilanthes acmella* Murr. – an ornamental cum medicinal herb. Analele Universităţii din Oradea – Fascicula Biologie 18(1):23–28

Yamashita Y, Terada R, Nishibayashi S, Shimamoto K (1989) Asymmetric somatic hybrids of *Brassica*: Partial transfer of *B. campestris* genome into *B. oleracea* by cell fusion. Theor Appl Genet 77:189–194

Yang D, Li W, Li S, Yang X et al (2007) In vitro embryo rescue culture of F_1 progenies from crosses between diploid and tetraploid grape varieties. Plant Grow Regul 51(1):63–71

Zegzouti R, Arnould M-F, Favre J-M (2001) Histological investigation of the multiplication steps in secondary somatic embryogenesis of *Quercus robur* L. Ann For Sci 58:681–690

Zehdi S, Trifi M, Ould Mohamed Salem A et al (2002) Survey of inter simple sequence repeat polymorphisms in Tunisian date palms (*Phoenix dactylifera* L.). J Genet Breed 56:77–83

Zenkteler M (1970) Test-tube fertilization of ovules in *Malandrium album* Mill. with pollen grains of several species of Caryophyllaceae family. Experient 23:775–776

Zenkteler M, Bagniewska-Zadworna A, Zenkteler E (2005) Embryological studies on ovules of *Melandrium album* pollinated in vitro with *Lychnis coronaria* pollen grains. Acta Biol Cracov Bot 47(1):135–138

Zeynali M, Zanjani BM, Saba J et al (2013) In vitro plant regeneration from alginate–encapsulated somatic embryos of rapeseed (*Brassica napus* cv. Tallayeh). Int J Trad Herb Med 1(1):13–18

Zhang ZB, Burgos NR, Zhang JP, Yu LQ (2007) Biological control agent for rice weeds from protoplast fusion between *Curvularia lunata* and *Helminthosporium gramineum*. Weed Sci 55(6):599–605

Zhiwu L, Pinkham L, Campbell NF et al (2009) Development of triploid daylily (*Hemerocallis*) germplasm by embryo rescue. Euphytica 169(3):313–318

Zhou A, Xia G, Chen H (1996) Asymetric somatic hybridization between *Triticum aestivum* L. and *Haynaldia villosa* Schur. Sci China Ser C Life Sci 39(6):617–627

Zhou A, Xia G, Chen H, Hu H (2001) Comparative study of symmetric and asymmetric somatic hybridization between common wheat and *Haynaldia villosa*. Sci China Ser C Life Sci 44(3):294–304

Zhou A-F, Xia G-M, Chen X-L, Chen H-M (2002) Production of somatic hybrid plants between two types of wheat protoplasts and the protoplasts of *Haynaldia villosa*. Acta Bot Sin 44(8):1004–1008

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20:176–183

Zimnoch-Guzowska E, Lebecka R, Kryszczuk A et al (2003) Resistance to *Phytophthora infestens* in somatic hybrids of *Solanum nigrum* L. and diploid potato. Theor Appl Genet 107:43–48

Ziv M, Naor V (2006) Flowering of geophytes in vitro. Prop Orn Plants 6:3–16

Zulkarnain Z, Tapingkae T, Taji AM (2012) In vitro breeding strategies in the development of Australian native plants. In: Peter KV (ed) Nutri-horticulture. Daya Publishing House, New Delhi, pp 75–97

Chapter 11
Induced Mutations in Plant Breeding

H. Dayton Wilde

Abstract Many important traits for plant domestication and improvement have resulted from human selection for novel alleles of structural or regulatory genes. In addition to naturally-occurring genetic mutations, novel alleles have been induced in plants by chemical and physical mutagenesis. The goal of mutagenesis is to induce genetic variation in cells that give rise to plants, while minimizing chimeras, sterility and lethality. For several crop species, chemically-mutagenized populations of a few thousand lines were sufficient for finding the desired phenotype. The efficiency of mutation breeding can be improved by screening plants at the genetic level, prior to phenotypic analysis. High-throughput physical methods and, increasingly, next-generation sequencing are being used to identify lines with induced mutations in candidate genes. An alternative approach to increasing the precision of mutation breeding is through gene-specific mutation using engineered nucleases. Allelic diversity in candidate genes, whether induced naturally or experimentally, can be a resource in breeding programs for developing new agricultural traits.

Keywords Alleles • Mutagenesis • Polymorphism • Reverse-genetics

11.1 Introduction

The breeding of plants for new traits requires genetic variation. Many phenotypes important in plant domestication and improvement have resulted from human selection for novel alleles of structural or regulatory genes (Olsen and Wendel 2013). Examples of qualitative traits based on novel alleles range from maize architecture and soybean growth habit to grape berry color and repetitive flowering in roses. Allelic variation in key genes can arise naturally from *spontaneous* mutation through replication errors, DNA damage or transposon activity.

H.D. Wilde (✉)
Department of Horticulture and the Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA 30602, USA
e-mail: dwilde@uga.edu

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_11

Table 11.1 Examples of natural and induced mutation of the same gene resulting in similar phenotypes. The natural and induced mutations are described in the publications listed under the mutation type

Plant	Trait	Gene	Natural mutation	Induced mutation
Rice	Semi-dwarf	*ga20-ox*	Spielmeyer et al. (2002)	Ashikari et al. (2002)
Tomato	Potyvirus resistance	*eIF4E*	Ruffel et al. (2005)	Piron et al. (2010)
Wheat	Starch quality	*waxy*	Marcoz-Ragot et al. (2000)	Slade et al. (2005)
Barley	Powdery mildew resistance	*mlo*	Piffanelli et al. (2004)	Jørgensen (1992)
Rapeseed	Oil quality	*fae1*	Wu et al. (2008)	Wang et al. (2008b)
Tobacco	Reduced nornicotine	*cyp82E*	Pakdeechanuan et al. (2012)	Julio et al. (2008)

Since the 1920s, plant breeders have taken advantage of physical and chemical mutagenesis to introduce genetic variation (Stadler 1928). Mechanistically, mutations in plant DNA have the same effect on plant phenotype whether they result from natural or human-directed processes. In both cases, gene activity can be altered by nucleotide substitution, the deletion or insertion of DNA sequence, or modification of *cis*-regulation. Table 11.1 shows traits in crop species that have been obtained from naturally or deliberately induced mutations in the same genes.

For example, semi-dwarf varieties of rice that enabled the Green Revolution were derived independently from natural and induced mutations in the gene for gibberellin 20-oxidase (Ashikari et al. 2002). In hexaploid wheat, natural and induced mutations in *waxy* homeologs have been combined by breeding to modify starch quality (Dong et al. 2009; Slade et al. 2005).

Mutagenesis is a useful tool for plant breeding that can be made more precise when combined with molecular technology and bioinformatics. The field of plant mutagenesis has been reviewed recently (Lee et al. 2014; Mba 2013; Pathirana 2012; Roychowdhury and Tah 2013; Sikora et al. 2011), and comprehensive discussion of certain aspects of this technology will be left to those reviews in order to focus on other topics. A reverse genetics platform for mutation discovery is discussed in detail in the chapter of this book entitled *Targeting induced local lesions in genomes (TILLING)*. In the present chapter, the application of mutagenesis to crop species, its successes and limitations, and future prospects for this approach will be discussed.

11.2 Application of Plant Mutagenesis

11.2.1 Mutation Induction

11.2.1.1 Mutagens

Mutations were first induced in plants with physical mutagens and this methodology has produced the majority of the varieties (77 %) listed in the FAO/IAEA Mutant Varieties Database (Maluszynski 2001). Physical mutagenesis is most commonly conducted with ionizing radiation produced directly by gamma and X-rays or indirectly by fast neutron bombardment. A detailed discussion of physical mutagenesis techniques is presented in Roychowdhury and Tah (2013). It is generally assumed that gamma radiation causes less chromosomal damage and more point mutations and short deletions compared to X-rays and fast neutrons (Sikora et al. 2011). However, a genome-wide analysis of mutations induced by fast neutron bombardment in *Arabidopsis thaliana* found a higher frequency of novel single nucleotide polymorphisms (SNPs) than deletions (Belfield et al. 2012).

Chemical mutagenesis has become a widely adopted approach because it does not require special facilities and the resulting mutations are primarily SNPs. Ethyl methane sulfonate (EMS) is currently the most commonly used chemical mutagen, but methylnitrosourea, sodium azide, diethyl sulfate and diepoxybutane have also proven effective (Sikora et al. 2011). The mode of action of a chemical can influence the mutations induced. For example, EMS selectively alkylates guanine bases, leading primarily to GC to AT transitions (Greene et al. 2003). The number of possible mutations induced in a gene by EMS can be predicted by its GC content (e.g. Harloff et al. 2012). Codon usage, therefore, could affect mutation frequency. Disadvantages of chemical mutagens compared to physical mutagens are reduced penetration through plant tissue and increased safety requirements. In addition to physical and chemical methods, mutations can be induced in plants through the introduction of active transposable elements, such as *mPing* (Hancock et al. 2011).

11.2.1.2 Target Tissues

Seeds are the typical target for mutagenesis, leading to 76 % of the mutant plant varieties (IAEA Mutant Varieties Database). For vegetatively-propagated crops, mutagenesis targets include buds, stem cuttings and tubers. The larger size of many vegetative propagules may require physical mutagenesis for effective penetration of the mutagen. The goal of mutagenesis is to induce high levels of genetic variation in cells that give rise to plants, while minimizing chimeras, sterility and lethality. The optimal mutagen dose is determined empirically in order to minimize the impact on growth and fertility.

Mutagenesis of seeds and vegetative propagules usually produces genetic chimeras, which can be dissociated by sexual or vegetative propagation (Mba 2013). One

approach to producing nonchimeric plants is through fertilization with mutagenized pollen, which has been successful in maize (Till et al. 2004), eucalyptus (McManus et al. 2006), cherry (Sigurbjörnsson and Micke 1974), and peach (Plovdiv 6, IAEA Mutant Variety Database). Alternatively, mutagenesis can be conducted with single cells or small cell clusters that are competent for plant regeneration. Examples of this strategy are the mutagenesis of rice zygotes and seed-derived calli (Serrat et al. 2014; Suzuki et al. 2008) and an embryogenic cell suspension of grapevine (Acanda et al. 2013).

11.2.2 Population Size

The size of a mutagenized population needed to detect a targeted phenotype depends on the mutation frequency and the ploidy level of the plant species. Polyploid plants have been found to tolerate higher mutation frequencies due to the redundant function of homeologous genes. Genetic screening studies conducted over the past decade provide empirical data on the rate of recovery of targeted phenotypes from chemically-mutagenized populations (Table 11.2). EMS-induced mutations that conferred novel phenotypes could be identified in populations of less than 2000 lines of several polyploid species, which were found to have high mutation densities. For example, EMS-induced mutations were found in wheat (Slade et al. 2005), oilseed rape (Wang et al. 2008b) and tobacco (Julio et al. 2008), on average, once every 24, 42, and 56 Kb, respectively.

In diploid species, mutations causing desired traits could be found in populations of 2000–8000 lines (Table 11.2). A commercial source of chemically-mutagenized plants (BenchBio) routinely produces 1000 to 5000 M1 lines, and recommends at least 3000 lines (http://www.benchbio.com). Mutant populations of 3515 lines and 5000 lines generated by BenchBio for flax and sunflower, respectively, were sufficient for the identification of several induced mutations in targeted genes. (Chantreau et al. 2013; Kumar et al. 2013). On the other hand, theoretical calculations of the population size needed to find induced mutations in diploid plants suggest that tens of thousands of M1 lines are necessary (Roychowdhury and Tah 2013; Sikora et al. 2011). Producing and maintaining progeny of tens of thousands of lines can sometimes be impractical for species that are large or long-lived. Genetic screening could help to focus efforts on lines with functional mutations.

11.2.3 Effect and Frequency of Mutation Types

Large deletions caused by physical mutagens can eliminate entire genes, whereas small deletions and insertions in genes alter protein translation by shifting reading frames. SNPs induced in genes by mutagens such as EMS or gamma rays have different consequences depending on the context of the coding sequence. A nucleotide

Table 11.2 Population size of EMS-mutagenized plants giving rise to novel phenotypes. Induced mutations were identified by sequencing altered alleles of candidate genes detected by TILLING. Where multiple mutants with targeted phenotypes were developed, one example of an induced mutation is listed

Plant	Mutagenized population	Induced mutation (s)	Novel phenotype
Sorghum[a]	768 M2 lines	Missense	Modified lignin
Potato[b]	864 M1 lines	Splice site	Modified starch
Wheat[c]	1152 M2 lines	Nonsense/missense	Modified starch
Tobacco[d]	1311 M2 lines	Nonsense	Low nornictine
Oilseed rape[e]	1344 M2 lines	Missense	Modified oil
Soybean[f]	1536 M2 lines	Missense	Low raffinose
Rice[g]	1860 M2 lines	Nonsense	Semi-dwarfism
Melon[h]	2368 M2 lines	Missense	Fruit color
Sorghum[i]	2709 M2 lines	Missense	Acyanogenic
Melon[j]	3000 M2 lines	Missense	Andromonoecy
Tomato[k]	3052 M2 lines	Missense	Delayed ripening
Pea[l]	3072 M2 lines	Missense	Reduced height
Flax[m]	3515 M2 lines	Nonsense	Modified lignin
Melon[n]	4704 M2 lines	Missense	Delayed ripening
Tomato[o]	4759 M3 lines	Splice site	Virus resistance
Tomato[p]	8025 M2 lines	Nonsense	Fruit color

[a]Xin et al. (2008)
[b]Muth et al. (2008)
[c]Slade et al. (2004): two mutated alleles combined for trait
[d]Julio et al. (2008)
[e]Wang et al. (2008b)
[f]Dierking and Bilyue (2009)
[g]Casella et al. (2013)
[h]González et al. (2011)
[i]Blomstedt (2011)
[j]Boualem et al. (2008)
[k]Okabe et al. (2011)
[l]Triques et al. (2007)
[m]Chantreau et al. (2013)
[n]Dahmani-Mardas et al. (2010)
[o]Piron et al. (2010)
[p]Gady et al. (2012)

triplet could be changed to encode a different amino acid (missense mutation), a translational stop signal (nonsense mutation), or the same amino acid (silent mutation). In addition, nucleotide substitution at the exon/intron border can cause the missplicing of mRNA (splice site mutation). Nonsense and splice site mutations can have severe effects on protein function. Depending on characteristics of the substituted amino acid, missense mutations have variable effects on protein activity. Between 25 and 30 % of missense mutations are predicted to have deleterious effects (Ng and Henikoff 2006).

The frequencies of different mutation types affect the probability of finding a mutagenized line with a trait of interest. Several studies have examined the

Table 11.3 Frequency of mutation types induced in plants with EMS. Mutations induced in exons and exon/intron borders were examined

Plant	Mutations examined	Mutation types (%)		
		Silent	Missense	Nonsense/splice site
Arabidopsis[a]	1890	45.0	50.1	4.9
Oilseed rape[b]	466	31.1	63.7	5.1
Oilseed rape[c]	432	53.9	49.1	4.2
Pea[d]	371	28.5	66.8	4.7
Tomato[e]	145	36.6	58.6	4.8
Flax[f]	141	27.7	66.7	5.7
Melon[g]	134	31.3	65.1	3.6
Pea[h]	60	31.7	65.0	3.3
Sunflower[i]	47	57.4	38.3	4.3
Tomato[j]	35	16.0	80.0	4.0
Potato[k]	15	26.7	66.7	6.7

[a]Greene et al. (2003)
[b]Harloff et al. (2012)
[c]Gilchrist et al. (2013)
[d]Dalmais et al. (2008)
[e]Piron et al. (2010)
[f]Chantreau et al. (2013)
[g]Dahmani-Mardas et al. (2010)
[h]Triques et al. (2007)
[i]Kumar et al. (2013)
[j]Okabe et al. (2011)
[k]Muth et al. (2008)

distribution of different types of point mutations caused by EMS (Table 11.3). Analysis of nearly 1900 mutations of *Arabidopsis thaliana* found that missense, nonsense, and splice site mutations comprised 50.1, 3.4, and 1.5 % of the point mutations, respectively (Greene et al. 2003). Investigations with other plant species generally found a similar distribution of mutation types, taking the smaller samples size into account. Missense mutations are responsible for the majority of novel phenotypes induced by EMS. Nonsense and splice site mutations, which are likely to result in a gene knockout, are found less frequently (Tables 11.2 and 11.3).

Silent or synonymous mutations can account for over 40 % of the mutations caused by EMS (Table 11.3). Although they do not affect the amino acid sequence, synonymous mutations have the potential to affect the level and structure of the encoded proteins (Delker and Quint 2011; Saunders and Deane 2010). Different codons specifying the same amino acid are not used at the same frequency. The deliberate introduction of unpreferred synonymous codons was found to alter levels of the encoded protein due to changes in translation (Carlini and Stephan 2003) or miRNA regulation (Mallory et al. 2004). Synonymous substitutions can also alter

protein function through folding changes that result from ribosomal stalling (Tsai et al. 2008). Harloff et al. (2012) examined synonymous mutations induced in oilseed rape that led to unpreferred codons. Two nucleotide changes were identified that resulted in synonymous codons that are infrequently used, GTG to GTA and TTG to TTA. Ten lines were found with induced mutations causing these changes in sinapine synthesis genes; phenotypic analysis was not reported. Missense mutations may also lead to codons used infrequently by a particular plant species.

11.2.4 Genetic Screening for Mutations

The efficiency of mutation breeding can be improved by screening mutagenized plants at the genetic level, prior to phenotypic analysis. Genetic screening allows lines with candidate gene mutations to be identified prior to expression of the trait. The development of a crop species with a novel trait can be conducted with selected lines instead of maintaining thousands of lines for phenotypic analysis. This approach would particularly useful for traits that are delayed in expression (e.g. fruit quality of woody species) and for traits that are not easily observed (e.g. modified lignin). In addition, genetic analysis permits the identification of lines that are heterozygous for mutations, which should be half of an M2 family after selfing. Heterozygous recessive lines could be missed by phenotypic analysis alone. Genetic screening can also detect induced mutations in polyploid species, which may not exhibit a phenotype due to the activity of homeologous genes with similar function.

Genetic screening approaches for identifying induced mutations have been termed precision breeding (e.g. Muth et al. 2008), targeted mutation breeding (e.g. Julio et al. 2008) or TILLING (e.g. McCallum et al. 2000). Deletions induced in genes by fast neutrons or gamma rays have been detected by PCR-based methods, using primers that flank the gene (Li et al. 2002; Wu et al. 2005). This method is applicable to deletions that are large enough to be detected in gel assays, but not SNPs or small insertion/deletions (indels).

To identify SNPs and indels, genetic screening approaches were developed that have become referred to as TILLING (Wang et al. 2012). Target genes (or portions of genes) are amplified by PCR and induced mutations are identified through DNA mismatch detection (Colbert et al. 2001; Gady et al. 2009), differences in secondary structure (Julio et al. 2008) or by next-generation sequencing (Tsai et al. 2011). Sikora et al. (2011) provide a detailed comparison of different mutation detection methods. The throughput of genetic screening can be increased with capillary electrophoresis, and this has been used for research (Julio et al. 2008; Suzuki et al. 2008) and commercial applications of TILLING (Loeffler et al. 2011).

11.3 Success and Limitations of Mutation Breeding

Mutation breeding has had measurable impacts on food production, quality enhancement and economics on a worldwide basis (Ahloowalia et al. 2004). The FAO/IAEA Mutant Variety Database currently lists over 3200 plant varieties developed from induced mutations. The mutation breeding of cereals, legumes, oil crops, ornamentals and fruit trees has produced significant benefits. For example, a mutated duram wheat variety Creso and varieties derived from it make up over half the wheat used for pasta in Italy (Ahloowalia et al. 2004). Rice varieties derived from mutation breeding are grown extensively in Asia and Australia, generating USD billions annually (Mba 2013). The mutant barley varieties Golden Promise and Diamant and varieties derived from them have been important to the brewing industry in Europe (Ahloowalia et al. 2004). A mutation-derived cotton variety, Lumian No. 1, has become one of the most widely grown cotton varieties in China. In the US, mutant varieties of grapefruit (Rio Star), sunflower (NuSun), and peppermint (Todd's Mitcham) are a major component of each crop. Mutation breeding has been an important tool for creating novelty in ornamental plants such as chrysanthemums, roses and dahlias (Fig. 11.1). The adoption of mutation-derived varieties in the marketplace is facilitated by the absence of government regulations that hinder other technologies like genetic engineering.

There are also limitations of mutation breeding to be considered. One issue is that large mutant populations must be generated due to the random nature of mutation induction. Greenhouse space, field space or labor can limit the number of M1 lines that can be produced and the number of their progeny that can be maintained for phenotypic analysis. This issue is more significant for large perennial plants, like fruit trees, than smaller annual species, like rice. Early line selection, based on genetic screening for candidate gene mutations, can be used to increase the efficiency of mutation breeding. Mutant populations could be obtained from commercial service providers or, for several species, from mutant stock collections that are publicly available (Mba 2013).

The production of genetic chimeras presents another concern for mutation breeding. For seed-propagated crops, chimeras can be eliminated by one to two generations of selfing (Mba 2013). However, for vegetatively-propagated crops, passage through several rounds of tissue culture may be required before nonchimeric shoots are obtained (Predieri 2001). Maintaining large numbers of lines in vitro can sometimes be cost-prohibitive.

Differences between seed-propagated plants and vegetatively-propagated plants also exist for dealing with background mutations, i.e. the induced mutations not responsible for the trait being targeted. In vegetatively-propagated plants, background mutations cannot be eliminated without genetic recombination. In seed-propagated species, backcrossing with the parental genotype has been used to reduce background mutations. For example, EMS-mutated tomato lines with delayed ripening were backcrossed or outcrossed twice to minimize background mutations (Colbert et al. 2011; McCallum et al. 2008).

Fig. 11.1 Mutagenesis of *Vitex* sp. for novel ornamental traits. Gamma irradiation of seeds of *V. agnus-castus* (**a**) resulted in a cultivar (**b**) with shorter, denser panicles and darker pink flowers. Gamma irradiation of seeds of *V. negundo* (**c**) produced a denser, more compact cultivar (**d**) with novel leaf morphology (DA Knauft and CD Robacker, unpublished results)

Finally, induced mutations are generally recessive because they impair one allele. For diploid plants capable of self-fertilization, obtaining homozygous recessive plants is relatively straightforward. Obtaining homozygosity in a recessive target gene is more difficult for outcrossing or polyploid plants, but approaches typically used for breeding recessive traits can be applied (e.g. Muth et al. 2008; Slade et al. 2004). Doubled haploid systems may also provide a means of developing homozygous recessive mutations (Szarejko and Forster 2007).

11.4 Future Prospects of Mutation Breeding

11.4.1 Next-Generation Sequencing to Detect Induced Mutations

The precision of mutation breeding can be increased by integrating molecular technology and bioinformatics. A variety of techniques have been applied over the past 15 years for detecting induced mutations in plant populations, including denaturing HPLC, LI-COR DNA analysis and high-resolution melting (HRM) analysis (Sikora et al. 2011). As the cost of DNA sequencing dropped and its accuracy improved, the potential of this approach for mutation discovery has been explored.

One of the first demonstrations of the use of next-generation sequencing (NGS) technology to discover induced and natural mutations was conducted by the plant biotechnology company Keygene (Rigola et al. 2009; Van Eijk and Van Tunen 2009). Using a Roche 454 platform, mutations in the *eIF4E* gene of tomato were detected by sequencing PCR-amplified DNA from 3000 EMS-mutagenized lines. Tsai et al. (2011) developed TILLING by Sequencing to identify induced mutations in crop species using an Illumina sequencing platform. In this application, a pool of PCR products is barcoded with a unique DNA adapter sequence that facilitates mutant line identification in multidimensional pools. Bioinformatics is used to evaluate the frequency, sequencing quality, intersection pattern in pools and statistical relevance of nucleotide changes.

A comparison of mutation discovery by Illumina sequencing and traditional TILLING (DNA mismatch detection) was conducted by Gilchrist et al. (2013). The level of sensitivity of the two approaches was found to be similar finding mutations in oilseed rape, but sequencing was more expensive. The authors suggested that the costs for mutation detection by sequencing could be reduced by: (a) barcoding samples during PCR amplification, instead of post-amplification, (b) using automated platforms for high complexity PCR amplification of many target regions in multiple lines, and (c) using platforms that provide more sequencing data per lane (e.g. Illumina HiSeq 2000).

Whole-genome sequencing is an alternative strategy being developed for the identification of induced mutations (Allen et al. 2013; Hwang et al. 2014; Nordström et al. 2013). While it currently cost-prohibitive to carry out with large populations, whole-genome sequencing will likely be an option for mutation detection in the near future (Wang et al. 2012). The detection of mutations by NGS is also an area of active research in the medical field (e.g. Rehm 2013), which is helping to push technology development.

11.4.2 Induced Mutation of Specific Genes

Technology will continue to improve for the identification of plants with novel alleles induced randomly by chemical or physical mutagenesis. A complementary approach being developed is the mutation of specific genes. The most promising methods for gene-specific mutagenesis involve the targeting of an endonuclease to a specific gene, where it creates a double-strand break (Voytas 2013). The repair of the double-strand break by non-homologous end-joining results in sequence deletion (or insertion) that can knock out gene activity.

Engineered nucleases have been developed that differ in the ease with which the nuclease can be targeted to a specific site. Meganucleases such as I-SceI have been used to introduce mutations in specific genes of crop species (e.g. Yang et al. 2009); however, manipulating the enzyme to target a specific DNA sequence can be difficult. Zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) combine a nuclease (FokI) with a gene targeting sequence that is designed separately. ZFNs and TALENs have been used to induce mutations in specific plant genes (Voytas 2013). The design of the DNA-binding peptide sequence is more straightforward for TALENs than ZFNs. For FokI activity, TALENs and ZFNs must be delivered as dimers.

To date, the most tractable engineered nuclease is the CRISPR-Cas9 system (e.g. Shan et al. 2013). Advantages of CRISPR-Cas9 are that it works as a monomer and that an easily manipulated mRNA sequence is used for targeting the nuclease. CRISPR-Cas9 has used to introduce gene-specific mutations in crop species such as rice, tobacco and sorghum (Jiang et al. 2013).

The regulatory status of plants with targeted mutations has not been settled (Hartung and Schiemann 2014; Lusser and Davies 2013). Designer nucleases are delivered into plant cells by gene transfer methods, prompting speculation that the resulting plants could become regulated in the EU and elsewhere as GMOs. However, nucleases can be expressed transiently or can be removed from the genome by recombination. In these cases, plants produced by targeted mutagenesis would be indistinguishable from plants derived through random mutagenesis. It has been argued the plants with targeted mutations should be evaluated according to the new trait and the resulting end product, not the process used to for development (Hartung and Schiemann 2014).

11.4.3 Screening Germplasm Collections for Natural Mutations

Screening techniques for finding induced mutations can be used to discover natural mutations in candidate genes in germplasm collections. The application of this reverse-genetics strategy to natural mutations has been termed ecotype TILLING (ecoTilling; Till 2014) or breeding with rare defective alleles (BRDA; Vanholme

Table 11.4 Discovery of allelic diversity by Ecotilling. Germplasm collections were screened for variation in structural genes and transcription factors (TFs)

Plant	Accessions	Target gene(s)	Trait	References
Rice	95	25 TFs	Drought tolerance	Yu et al. (2012)
Melon	113	*eIF4E*	Virus resistance	Nieto et al. (2007)
Brassica sp.	117	*FAE1*	Modified oil	Wang et al. (2010)
Maize	175	*IPT2*	Kernel weight	Weng et al. (2013)
Barley	210	*HSP116.8*	Drought tolerance	Xia et al. (2013)
Wheat	225	*Pina-D1, Pinb-D1*	Modified starch	Wang et al. (2008a)
Pepper	233	*eIF4E*	Virus resistance	Ibiza et al. (2010)
Sugar beet	268	*FL1, FT1, BTC1*	Early flowering	Frerichmann et al. (2013)
Rice	392	*CPK17,RMC,NHX1, Sal*T, *HKT1;5*	Salt tolerance	Negrão et al. (2012)
Poplar	768	*CAD, HCT, C3H, CCR, 4CL*	Modified lignin	Marroni et al. (2011)

et al. 2013). These approaches have been taken to search germplasm collections of a number of plant species for novel alleles in candidate genes (Table 11.4). NGS, HRM, and LI-COR analysis systems have been used to identify natural mutations, but procedures are modified for screening germplasm collections to account for the increased genetic heterogeneity.

11.5 Conclusions and Prospects

Mutagenesis has played a significant role in plant improvement and technological innovations will increase its precision in the future. Mutation breeding can be streamlined by screening for lines with candidate gene mutations that were induced either naturally or experimentally. Bioinformatics and sequence data from the growing number of sequenced plant genomes will simplify the identification of candidate genes for qualitative traits. Allelic diversity in these candidate genes can be a resource in breeding programs for developing new agricultural traits.

References

Acanda Y, Martínez Ó, Prado MJ et al (2013) EMS mutagenesis and qPCR-HRM prescreening for point mutations in an embryogenic cell suspension of grapevine. Plant Cell Rep 33(471):481

Ahloowalia BS, Maluszynski M, Nichterlein K (2004) Global impact of mutation-derived varieties. Euphytica 135:187–204

Allen RS, Nakasugi K, Doran RL et al (2013) Facile mutant identification via a single parental backcross method and application of whole genome sequencing based mapping pipelines. Front Plant Sci 4:362

Ashikari M, Sasaki A, Ueguchi-Tanaka M et al (2002) Loss-of-function of a rice gibberellin biosynthetic gene, GA20 oxidase (GA20ox-2), led to the rice 'green revolution'. Breed Sci 52:143–150

Belfield EJ, Gan X, Mithani A et al (2012) Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of *Arabidopsis thaliana*. Genome Res 22:1306–1315

Blomstedt CK, Gleadow RM, O'Donnell N et al (2011) A combined biochemical screen and TILLING approach identifies mutations in *Sorghum bicolor* L. Moench resulting in acyanogenic forage production. Plant Biotechnol J 10:54–66

Boualem A, Fergany M, Fernandez R et al (2008) A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. Science 321:836–838

Carlini DB, Stephan W (2003) In vivo introduction of unpreferred synonymous codons into the Drosophila Adh gene results in reduced levels of ADH protein. Genetics 163:239–243

Casella L, Greco R, Bruschi G et al (2013) TILLING in European rice: hunting mutations for crop improvement. Crop Sci 53:2550–2562

Chantreau M, Grec S, Gutierrez L et al (2013) PT-Flax (phenotyping and TILLinG of flax): development of a flax (*Linum usitatissimum* L.) mutant population and TILLinG platform for forward and reverse genetics. BMC Plant Biol 13:159

Colbert T, Till BJ, Tompa R et al (2001) High-throughput screening for induced point mutations. Plant Physiol 126:480–484

Colbert TG, Hurst SR, Slade AJ (2011) Tomatoes that soften more slowly postharvest due to non-transgenic alterations in an expansin gene. US Patent Application 20110113507

Dahmani-Mardas F, Troadec C, Boualem A et al (2010) Engineering melon plants with improved fruit shelf life using the TILLING approach. PLoS One 5, e15776

Dalmais M, Schmidt J, Le Signor C et al (2008) UTILLdb, a *Pisum sativum* in silico forward and reverse genetics tool. Genome Biol 9:R43

Delker C, Quint M (2011) Expression level polymorphisms: heritable traits shaping natural variation. Trends Plant Sci 16:481–488

Dierking E, Bilyeu K (2009) New sources of soybean seed meal and oil composition traits identified through TILLING. BMC Plant Biol 9:89

Dong C, Dalton-Morgan J, Vincent K, Sharp P (2009) A modified TILLING method for wheat breeding. Plant Genome 2:39–47

Frerichmann SL, Kirchhoff M, Müller AE et al (2013) EcoTILLING in Beta vulgaris reveals polymorphisms in the FLC-like gene BvFL1 that are associated with annuality and winter hardiness. BMC Plant Biol 13:52

Gady ALF, Hermans FWK, Van de Wal MHBJ et al (2009) Implementation of two high throughput techniques in a novel application: detecting point mutations in large EMS mutated plant populations. Plant Methods 5:13

Gady AL, Vriezen WH, Van de Wal MH et al (2012) Induced point mutations in the phytoene synthase 1 gene cause differences in carotenoid content during tomato fruit ripening. Mol Breed 29:801–812

Gilchrist EJ, Sidebottom CHD, Koh CS, MacInnes T et al (2013) A mutant *Brassica napus* (Canola) population for the identification of new genetic diversity via TILLING and next generation sequencing. PLoS One 8, e84303. doi:10.1371/journal.pone.0084303

González M, Xu M, Esteras C et al (2011) Towards a TILLING platform for functional genomics in Piel de Sapo melons. BMC Res Notes 4:289
Greene EA, Codomo CA, Taylor NE et al (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. Genetics 164:731–740
Hancock CN, Zhang F, Floyd K et al (2011) The rice miniature inverted repeat transposable element mPing is an effective insertional mutagen in soybean. Plant Physiol 157:552–562
Harloff HJ, Lemcke S, Mittasch J et al (2012) A mutation screening platform for rapeseed (*Brassica napus* L.) and the detection of sinapine biosynthesis mutants. Theor Appl Genet 124:957–969
Hartung F, Schiemann J (2014) Precise plant breeding using new genome editing techniques: opportunities, safety and regulation in the EU. Plant J. doi:10.1111/tpj.12413
Hwang SG, Hwang JG, Kim DS, Jang CS (2014) Genome-wide DNA polymorphism and transcriptome analysis of an early-maturing rice mutant. Genetica 142:73–85
Ibiza VP, Canizares J, Nuez F (2010) EcoTILLING in *Capsicum* species: searching for new virus resistances. BMC Genomics 11:631
Jiang W, Zhou H, Bi H et al (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res 41, e188. doi:10.1093/nar/gkt780
Jørgensen JH (1992) Discovery, characterization and exploitation of *Mlo* powdery mildew resistance in barley. Euphytica 63:141–152
Julio E, Laporte F, Reis S et al (2008) Reducing the content of nornicotine in tobacco via targeted mutation breeding. Mol Breed 21:369–381
Kumar AP, Boualem A, Bhattacharya A et al (2013) SMART-sunflower mutant population and reverse genetic tool for crop improvement. BMC Plant Biol 13:38
Lee LS, Till BJ, Hill H et al (2014) Mutation and mutation screening. In: Henry RJ, Furtado H (eds) Cereal genomics. Humana Press, New York, pp 77–95
Li X, Lassner M, Zhang Y (2002) Deleteagene: a fast neutron deletion mutagenesis-based gene knockout system for plants. Comp Funct Genomics 3:158–160
Loeffler DL, Dibya D, Kenseth J et al (2011) TILLING for mutation discovery in rice, wheat, lettuce, tomato and soybean using the advance™ Fs96 high throughput capillary electrophoresis system. In: Abstract of the plant & animal genomes XIX Conference. http://www.intl-pag.org/19/abstracts/P02c_PAGXIX_101.html
Lusser M, Davies HV (2013) Comparative regulatory approaches for groups of new plant breeding techniques. New Biotechnol 30:437–446
Mallory AC, Reinhart BJ, Jones-Rhoades MW et al (2004) MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5'region. EMBO J 23:3356–3364
Maluszynski M (2001) Officially released mutant varieties–the FAO/IAEA database. Plant Cell Tiss Org Cult 65:175–177
Marcoz-Ragot C, Gateau I, Koenig J et al (2000) Allelic variation of granule-bound starch synthase proteins in European bread wheat varieties. Plant Breed 119:305–309
Marroni F, Pinosio S, Di Centa E et al (2011) Large-scale detection of rare variants via pooled multiplexed next-generation sequencing: towards next-generation ecotilling. Plant J 67:736–745
Mba C (2013) Induced mutations unleash the potentials of plant genetic resources for food and agriculture. Agronomy 3:200–231
McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lessions in genomes (TILLING) for plant functional genomics. Plant Physiol 123:439–442
McCallum CM, Slade AJ, Colbert TG et al (2008) Tomatoes having reduced polygalacturonase activity caused by non-transgenic mutations in the polygalacturonase gene. US Patent 7393996
McManus LJ, Sasse J, Blomstedt CK, Bossinger G (2006) Pollen treatment for mutation induction in *Eucalyptus globulus* ssp. *globulus* (Myrtaceae). Aust J Bot 54:65–71
Muth J, Hartje S, Twyman RM et al (2008) Precision breeding for novel starch variants in potato. Plant Biotechnol J 6:576–584

Negrão S, Almadanim MC, Pires IS et al (2012) New allelic variants found in key rice salt-tolerance genes: an association study. Plant Biotechnol J 11:87–100
Ng PC, Henikoff S (2006) Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Hum Genet 7:61–80
Nieto C, Piron F, Dalmais M et al (2007) EcoTILLING for the identification of allelic variants of melon eIF4E, a factor that controls virus susceptibility. BMC Plant Biol 7:34
Nordström KJ, Albani MC, James GV et al (2013) Mutation identification by direct comparison of whole-genome sequencing data from mutant and wild-type individuals using k-mers. Nat Biotechnol 31:325–330
Okabe Y, Asamizu E, Saito T et al (2011) Tomato TILLING technology: development of a reverse genetics tool for the efficient isolation of mutants from Micro-Tom mutant libraries. Plant Cell Phys 52:1994–2005
Olsen KM, Wendel JF (2013) A bountiful harvest: genomic insights into crop domestication phenotypes. Annu Rev Plant Biol 64:47–70
Pakdeechanuan P, Teoh S, Shoji T, Hashimoto T (2012) Non-functionalization of two CYP82E nicotine N-demethylase genes abolishes nornicotine formation in *Nicotiana langsdorffii*. Plant Cell Phys 53:2038–2046
Pathirana R (2012) Plant mutation breeding in agriculture. In: Hemming D(ed) Plant sciences reviews. CAB International, Wellingford, England, pp 107–110
Piffanelli P, Ramsay L, Waugh R et al (2004) A barley cultivation-associated polymorphism conveys resistance to powdery mildew. Nature 430:887–891
Piron F, Nicola M, Mino S et al (2010) An induced mutation in tomato eIF4E leads to immunity to two potyviruses. PLoS One 5, e11313. doi:10.1371/journal.pone.0011313
Predieri S (2001) Mutation induction and tissue culture in improving fruits. Plant Cell Tiss Org Cult 64:185–210
Rehm HL (2013) Disease-targeted sequencing: a cornerstone in the clinic. Nat Rev Genet 14:295–300
Rigola D, van Oeveren J, Janssen A et al (2009) High throughput detection of induced mutations and natural variation using KeyPoint technology. PLoS One 4, e4761
Roychowdhury R, Tah J (2013) Mutagenesis- a potential approach for crop improvement. In: Hakeem KR, Ahmad P, Ozturk M (eds) Crop improvement. Springer, New York, pp 149–187
Ruffel S, Gallois JL, Lesage ML, Caranta C (2005) The recessive potyvirus resistance gene pot-1 is the tomato orthologue of the pepper pvr2-eIF4E gene. Mol Gen Genomics 274:346–353
Saunders R, Deane CM (2010) Synonymous codon usage influences the local protein structure observed. Nucleic Acid Res 38:6719–6728
Serrat X, Esteban R, Guibourt N et al (2014) EMS mutagenesis in mature seed-derived rice calli as a new method for rapidly obtaining TILLING mutant populations. Plant Meth 10:5
Shan Q, Wang Y, Li J et al (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol 31:686–688
Sigurbjörnsson B, Micke A (1974) Philosophy and accomplishments of mutation breeding. In: Polyploidy and induced mutations in plant breeding. IAEA, Vienna, pp 303–343
Sikora P, Chawade A, Larsson M et al (2011) Mutagenesis as a tool in plant genetics, functional genomics, and breeding. Int J Plant Genomics. doi:10.1155/2011/314829, ID 31482913
Slade AJ, Fuerstenberg SI, Loeffler D et al (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat Biotechnol 23:75–81
Spielmeyer W, Ellis MH, Chandler PM (2002) Semidwarf (sd-1), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. Proc Nat Acad Sci 99:9043–9048
Stadler LJ (1928) Genetic effects of X-rays in maize. Proc Nat Acad Sci 14:69–72
Suzuki T, Eiguchi M, Kumamaru T et al (2008) MNU-induced mutant pools and high performance TILLING enable finding of any gene mutation in rice. Mol Genet Genomics 279:213–223
Szarejko I, Forster BP (2007) Doubled haploidy and induced mutation. Euphytica 158:359–370
Till BJ (2014) Mining genetic resources via Ecotilling. In: Tuberosa R, Graner A, Frison E (eds) Genomics of plant genetic resources. Springer, Dordrecht, pp 349–365

Till BJ, Reynolds SH, Weil C et al (2004) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4:12

Triques K, Sturbois B, Gallais S et al (2007) Characterization of *Arabidopsis thaliana* mismatch specific endonucleases: application to mutation discovery by TILLING in pea. Plant J 51:1116–1125

Tsai CJ, Sauna ZE, Kimchi-Sarfaty C et al (2008) Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. J Mol Biol 383:281–291

Tsai H, Howell T, Nitcher R et al (2011) Discovery of rare mutations in populations: TILLING by sequencing. Plant Physiol 156:1257–1268

Van Eijk MJT, Van Tunen AJ (2009) High throughput screening of mutagenized populations. US Patent Application 20090170713

Vanholme B, Cesarino I, Goeminne G et al (2013) Breeding with rare defective alleles (BRDA): a natural *Populus nigra* HCT mutant with modified lignin as a case study. New Phytol 198:765–776

Voytas DF (2013) Plant genome engineering with sequence-specific nucleases. Annu Rev Plant Biol 64:327–350

Wang J, Sun JZ, Liu DC et al (2008a) Analysis of Pina and Pinb alleles in the micro-core collections of Chinese wheat germplasm by ecotilling and identification of a novel Pinb allele. J Cereal Sci 48:836–842

Wang N, Wang Y, Tian F et al (2008b) A functional genomics resource for *Brassica napus*: development of an EMS mutagenized population and discovery of FAE1 point mutations by TILLING. New Phytol 180:751–765

Wang NA, Shi L, Tian F et al (2010) Assessment of *FAE1* polymorphisms in three Brassica species using EcoTILLING and their association with differences in seed erucic acid contents. BMC Plant Biol 10:137

Wang TL, Uauy C, Robson F, Till B (2012) TILLING in extremis. Plant Biotechnol J 10:761–772

Weng J, Li B, Liu C et al (2013) A non-synonymous SNP within the isopentenyl transferase 2 locus is associated with kernel weight in Chinese maize inbreds (*Zea mays* L.). BMC Plant Biol 13:98

Wu G, Wu Y, Xiao L et al (2008) Zero erucic acid trait of rapeseed (*Brassica napus* L.) results from a deletion of four base pairs in the fatty acid elongase 1 gene. Theor Appl Genet 116:491–499

Wu JL, Wu C, Lei C et al (2005) Chemical-and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. Plant Mol Biol 59:85–97

Xia Y, Li R, Ning Z et al (2013) Single nucleotide polymorphisms in HSP116. 8 and their association with agronomic traits in barley. PLoS One 8, e56816

Xin Z, Wang ML, Barkley NA et al (2008) Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. BMC Plant Biol 8:103

Yang M, Djukanovic V, Stagg J et al (2009) Targeted mutagenesis in the progeny of maize transgenic plants. Plant Mol Biol 706:669–679

Yu S, Liao F, Wang F et al (2012) Identification of rice transcription factors associated with drought tolerance using the ecotilling method. PLoS One 7, e30765

Chapter 12
Alternative to Transgenesis: Cisgenesis and Intragenesis

Ankita Singh, Meenakshi Joshi, and E. Lamalakshmi Devi

Abstract As more transgenic crops are commercially released, environmental hazards and potential risk to human health have become a serious concern worldwide. This has led to the development of the two new concepts of cisgenesis and intragenesis as alternatives to transgenesis. Both these concepts, although different at the molecular level, make use of genes present in a sexually-compatible gene pool for crop improvement. The risks associated with alien gene introgression are mitigated to some extent. To date these new concepts have been employed in few important food, fruit, forage and woody crops and many more developments are in the pipeline. Studies have revealed that cisgenesis is similar to traditional breeding with additional advantages of transfer of desirable gene in a single step and less risk of linkage drag, while both intragenesis and transgenesis may lead to novel hazards like foreign gene introgression unassociated with conventional breeding. These novel techniques are safer than transgenesis; however, unintended and unexpected effects of random gene insertion and interaction between foreign genes and host genes in the transgenic organisms are reported. Research focus is now shifting to *gene targeting* to achieve *site-specific* mutagenesis to avoid potential risks associated with insertions at random locations (ectopic) in the genome. Acceptance of genetically modified organisms (GMOs) requires consumer awareness of safety issues with respect to the environment and human health.

Keywords Crop improvement • Cisgenesis • Gene modification • GMO • Intragenesis • Regulations • Safety issues • Transgenesis

A. Singh • M. Joshi
Department of Genetics and Plant Breeding, G.B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, India
e-mail: ankita.gbp@gmail.com; mjoshi0089@gmail.com

E.L. Devi (✉)
Department of Genetics and Plant Breeding, College of Agriculture, Tripura, Lembucherra, West Tripura, India
e-mail: elangbamlama@gmail.com

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_12

12.1 Introduction

Before agriculture was widely adopted, early humans fulfilled their energy and nutrition requirements by foraging, The shift from foraging to activities that relied on domesticated plants and animal resources occurred independently in several parts of the world 5000–11,000 years ago (Smith 1995). Since the practice of agriculture began, farmers have been altering the genetic makeup of the crops they grow. Early farmers selected the best looking plants and seeds and saved them to plant for the next season. Once the science of genetics became better understood, plant breeders used what they knew about plant genes to select for specific desirable traits to develop improved varieties. The advances in the methods and practices of plant breeding led over time to the development of new and more efficient methods of gene transfer and crop improvement.

Conventional plant breeding focuses on introgression of genes of value found in wild germplasm into the already domesticated ones. This approach *introgression breeding*, aims at incorporating a target gene from a donor plant into a recipient plant by interspecific hybridization followed by repeated backcrossing with the donor plant and selection for the traits of interest simultaneously. The end product of this process is a plant that has target gene from the donor plant along with the undesirable genes contributing to linkage drag (Jacobsen and Schouten 2007).

Induced translocation breeding is similar to introgression breeding except there is an absence of crossing over. Backcrossing and selection are followed by obtaining addition lines that contain the desirable gene along with the chromosome from the donor plant. These additional lines are then irradiated with ϒ-radiations and selection is carried out in subsequent generations. This method has been used in the development of rust and mildew resistant wheat; for example, Friebe et al. (1996) reported production of brown rust resistant wheat by induced translocation of a resistant gene from *Aegilops speltoides*. However, desirable plants are not frequent and the place of insertion is generally random. Thus, there is risk of the transfer of undesired traits along with the desired ones. This risk can be reduced by repeated backcrossing for many generations. The traditional introgression breeding is, thus, time consuming as it requires several generations of breeding. Also, the random insertion of a target gene may pose threats to the integrity and stability of whole genome accompanied with unpredictable consequences to the phenotype (Jacobsen and Schouten 2007).

Crop improvement has limited the genetic variability present in cultivated varieties. The efficiency of breeding for biotic and abiotic stresses and for nutritional and quality enrichment relies on the amount of variability for target traits. Mutation breeding is crucial to the generation of exploitable variation. However, most of the mutations are recessive. This feature does not pose any problem in seed propagated crops where homozygous recessive genotypes or that trait can be obtained. Recessive mutations are of no use in vegetatively propagated crops. For example, in potato breeding very few mutations viz. (tuber color, tuber shape and amylose-free starch)

have been found beneficial (van Harten 1998). Nevertheless, mutations have played a pivotal role in development of several successful cultivated varieties worldwide.

These challenges would be mitigated if we can specifically transfer genes of interest without the risk of any linkage drag. Transgenesis, cisgenesis and intragenesis are the new breeding techniques that have been developed to meet these challenges.

Transgenesis is the genetic modification of a recipient plant with one or more genes from any non-plant organism, or from a donor plant that is sexually incompatible with the recipient plant (Schouten et al. 2006a). There is growing concern among the general public about the introduction of foreign DNA into the food crops, leading to worldwide objections to transgenic plants (Dale 1992; Krishna et al. 2014; Lu and Yang 2009). The public outrage over the perceived potential hazards caused by foreign gene introgression have been answered by the two related techniques of cisgenic and intragenic modification that involve the import of genes from closely related, sexually compatible plant species, or that modify the plant genome without involving any foreign DNA (Hunter 2014). Additionally, the final intragenic and cisgenic plants should not contain foreign sequences such as selection marker genes and vector-backbone sequences.

This chapter aims at providing a comprehensive background of achievements, ongoing research and the concerns associated with the use of cisgenesis and intragenesis and the possible solutions that may lead to widescale acceptability of organisms modified through these techniques. It covers the need of cisgenesis and intragenesis as methods of gene transfer, basic definitions and principles involved, application of cis-/intragenesis in crop improvement, safety issues and concerns in commercial exploitation of organisms and future prospects.

12.2 Transgenesis

Transgenesis includes stable integration of genetic material (their own, foreign or artificial) stably integrated into the recipient genome that has been introduced by methods other than classical breeding. This includes gene sequences of any origin in the anti-sense orientation, any artificial combination of a coding sequence and a regulatory sequence, such as a promoter from another gene, or a synthetic gene. This transfer usually takes place between the species or genera or kingdom that does not intercross naturally. Thus, natural barriers are not respected in transgenesis. A transgenic plant is a plant with an insertion of DNA from another organism. This technology can be used as a tool for researchers to gather evidence about hypotheses and insert useful genes as a means of improving crops.

12.2.1 Development of Transgenics: General Outline

The main steps involved in the development of GM crops are as follows:

(a) Isolation of the gene(s) of interest: Based on knowledge about the structure, function or location on chromosomes, gene(s) for the desired trait is identified.
(b) Insertion of the gene to transfer vector: Most commonly used vector is plasmid from the naturally occurring soil bacterium, *Agrobacterium tumefaciens*. The gene(s) of interest is inserted into the plasmid using recombinant DNA (rDNA) techniques.
(c) Transformation of plants: The modified *A. tumefaciens* cells containing the plasmid with the new gene are mixed with plant cells or cut pieces of plants such as leaves or stems (explants). Some of the cells take up a piece of the plasmid known as the T-DNA (transferred-DNA). The *A. tumefaciens* inserts the desired genes into one of the plant's chromosomes to form GM (or transgenic) cells. The other most commonly used method to transfer DNA is particle bombardment (gene gun) where small particles coated with DNA molecules are bombarded into the cell.
(d) Selection of the transformed plant cells: After transformation, various methods are used to identify transformed cells. Selectable marker genes that confer antibiotic or herbicide resistance are most commonly used. Genes responsible for resistance are inserted into the vector and transferred along with the gene(s) conferring desired traits to the plant cells. When the cells are exposed to an antibiotic or herbicide, only the transformed cells (containing and expressing the selectable marker gene) will survive. The transformed cells are then regenerated to form whole plants using tissue-culture methods.
(e) Regeneration of transformed cells via tissue culture: Development of the cells into various plant parts to form whole plantlets by growing cells into nutrient medium. After rooting, shooting and hardening the plant is grown in the field (Table 12.1).

12.3 Cisgenesis

The concept of cisgenesis was introduced by Jochemsen and Schouten (2000). Schouten et al. (2006a) defined a cisgenic plant as *a crop plant that has been genetically modified with one or more genes isolated from a crossable donor plant*. Such a gene (cisgene) is an identical copy of the endogenous gene of the organism in which the promoter, introns and the terminator are present in the normal-sense orientation (Lusser et al. 2012). Thus, cisgenesis restricts any alteration whatsoever in the original gene.

Cisgenic plants can harbor one or more cisgenes, but they do not contain any transgenes (Shiva Prasad et al. 2013). The production process of cisgenic plants

Table 12.1 Traits improved in transgenics of some crops

Traits	Crops	References
Insect tolerance	Cotton (*Gossypium* spp.)	Finer and McMullen (1990); Firozabady et al. (1987); Umbeck et al. (1987)
	Tobacco (*Nicotiana* sp.)	Barton et al. (1987)
	Tomato (*Solanum lycopersicum*)	Vaeck et al. (1987)
Herbicide tolerance	Maize (*Zea mays*)	Anderson and Georgeson (1989; Newhouse et al. (1991)
	Tobacco (*Nicotiana* sp.)	Lyon et al. (1989); Streber and Willmitzer (1989)
	Sugar beet (*Beta vulgaris*)	D' Halluin et al. (1992)
	Wheat (*Triticum aestivum*)	Newhouse et al. (1992); Vasil et al. (1992)
	Alfalfa (*Medicago sativa)*	D' Halluin et al. (1990)
Virus resistance	Papaya (*Carica papaya*)	Fitch et al.(1992); Gonsalves et al. (1998)
	Squash (*Cucurbita pepo*)	Tricoli et al. (1995)
	Potato (*Solanum tuberosum*)	Hemenway et al. (1988)
Altered oil composition	Canola (*Brassica napus)*	Dehesh et al. (1996); Del Vecchio (1996); Froman and Ursin (2002); James et al. (2003); Roesler et al. (1997)
	Soybean (*Glycine max*)	Kinney and Knowlton (1998); Reddy and Thomas (1996)
Amino acid and protein quality and level	Maize (*Zea mays*)	Cromwell et al. (1967); O'Quinn et al. (2000); Yang et al. (2002); Young et al. (2004)
	Rice (*Oryza sativa*)	Katsube et al. (1999); Ye et al. (2000)
	Potato (*Solanum tuberosum*)	Atanassov et al. (2004); Chakraborty et al. (2000); Li et al. (2001); Yu and Ao (1997)

employs the same techniques as transgenesis, the difference being the use of a cisgene and not a transgene. Furthermore, in case of *Agrobacterium*-mediated transformation, the T-DNA borders originating from *Agrobacterium* may sometimes remain in the end-product (Fig. 12.1). This is referred to as *cisgenesis with T-DNA borders* in some texts. Unlike transgenesis, foreign genes such as selection marker genes and vector-backbone genes should be absent or eliminated from the primary transformants or their progeny to qualify as cisgenics.

This novel breeding approach integrates traditional breeding with modern biotechnology and dramatically speeds up the breeding process reducing the unwanted affects of linkage drag (Rommens 2007). The plants transformed through cisgenesis

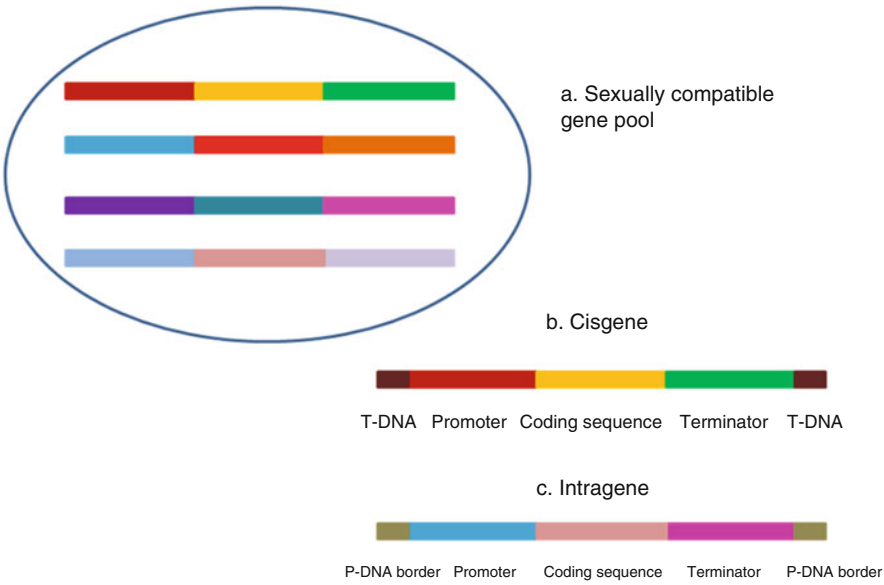

Fig. 12.1 Schematic representation of different strategies used for genetic modification according to definitions of cisgenesis and intragenesis. In cisgenesis, expression constructs must contain complete genetic elements from a sexually compatible gene pool i.e. promoter, coding sequence and terminator. In intragenesis, expression constructs can contain genetic elements from different genes within a sexually compatible gene pool. When an *Agrobacterium*-mediated transformation method is used, P-DNA or T-DNA borders can be inserted into the plant genome to generate cisgenic or intragenic approaches

should not be assessed as transgenics for environmental impacts because the plant genomes are modified without changing their gene pool (Hou et al. 2014).

12.4 Intragenesis

Intragenic classification was introduced by Rommens et al. (2004) as the isolation of specific genetic elements from a plant, recombination of these elements in vitro and insertion of the resulting expression cassettes into a plant belonging to the same sexual compatibility group. Unlike transgenesis, in the case of intragenesis, the inserted DNA can be a new combination of DNA fragments from the species itself or from a cross compatible species. This concept is less restrictive than cisgenesis as rearrangement of genetic elements is permitted. When using *Agrobacterium-mediated* transformation, the T-DNA border sequences should originate from the sexually compatible DNA pool (P-DNA borders) (Fig. 12.1). Plant transfer DNAs were discovered through intensive database searches and PCR analyses of sequences resembling T-DNA borders. Rommens et al. 2003 showed that P-DNA from some

species including tomato and potato supported DNA transfer from *Agrobacterium* to plant cells despite some sequence divergence.

Intragenesis is a new concept in plant breeding that is widely considered as a fine combination of conventional plant breeding and transgenesis for the improvement of crops. Unlike traditional breeding methods, it does not involve random gene insertion causing undesirable and unpredictable genome modifications. Intragenesis includes recombining native genetic elements in vitro and inserting the linked DNA back into the plant using marker-free transformation (Rommens 2004; Rommens et al. 2004). One of the elements, the plant-derived transfer (P-) DNA, replaces the *Agrobacterium* T-DNA by functioning as vehicle for DNA delivery into the plant cell (Rommens et al. 2005). Thus, intragenic modification incorporates neither uncharacterized DNA (as is the case with traditional breeding) nor foreign DNA (as is typical for transgenic modification) into a plant's genome.

Intragenesis also include RNA interference (RNAi) constructs (Schouten and Jacobsen 2008). Artificial miRNA technology has been used in the improvement of fruit crops (Molesini et al. 2012). This technology causes silencing of a gene of interest by generating an artificial micro RNA (amiRNA) directed using the gene sequence of natural miRNA. For example, small scale cloning and high-throughput sequencing analysis of small RNAs has allowed the identification of many conserved and non-conserved miRNAs present in tomato fruits. Researchers have demonstrated that some genes known to be involved in tomato fruit development and ripening are targets of miRNA/siRNA. For instance, colorless never ripe (CNR), a key gene in fruit ripening, is the target of miR156/157, and Lefsm1, a gene expressed specifically in the first stages of fruit formation, is the target of a siRNA identified in the fruit (Itaya 2008). The coupling of intragenesis and RNA interference technology is favorable when the alteration in endogenous processes can lead to crop improvement.

12.5 Special Techniques for Production of Cisgenic and Intrangenic Plants

In cisgesis and intragenesis, integration of genes from a cross compatible species into a desirable background takes place is much the same way as in transgenesis with a few additional steps. The major steps are isolating the DNA that codes for the protein we wish to express, inserting that DNA into a plasmid vector, transformation by an appropriate method, selectable marker elimination and, finally, selection of recombinants with the insertion.

Transgenic and cisgenic plants are produced by the same transformation techniques (Schouten et al. 2006a). Most of the cisgenic plants have been produced by the *Agrobacterium*-mediated transformation. The use of biolistic transformation is documented less frequently (Akhond and Machray 2009; Lusser et al. 2012).

Cisgenic and intragenic plants should also be free from other non-plant sequences, such as vector backbone and selection markers. Several strategies to either avoid or

remove marker genes have been described, mainly based on the target plant and the efficiency of the transformation method used. Marker excision can be further confirmed by molecular biology techniques.

Cisgenic and intragenic apples (Joshi et al. 2011; Vanblaere et al. 2014) as well as intragenic strawberries (Schaart et al. 2004) have been obtained using this marker-free technology. Co-transformation is another alternative to produce marker-free plants. This method is based on the integration of the marker gene and the transgene of interest in different positions of the plant's genome, allowing the segregation of both genes in different progeny. Co-transformation can be done using two transformation vectors, one including the gene of interest and one with the marker gene. Such an approach has been successfully used to generate cisgenic barley plants (Holme et al. 2012) and durum wheat (Gadaleta et al. 2008).

12.5.1 Selectable Marker Elimination

A crucial requirement for the cisgenic products to be environment friendly is the removal of selectable marker genes as the end-product should contain the desirable gene from cross compatible species. The strategies that may be applied for the removal of marker genes are discussed below.

12.5.1.1 Co-transformation

Modified protocols for plant cell transformation with *Agrobacterium* result in the integration of the selectable marker and the transgene into two different genomic loci, preferentially into two different chromosomes. This constellation allows segregation of the two loci by breeding, resulting in a cell line which has lost the resistance marker, but still contains the desired transgene (Woo et al. 2011). The technology is considered mature and efficient (up to 25 % of all co-transformed cell lines show marker segregation), but screening becomes more tedious and costly because four times more cell lines have to be checked. Functionally irrelevant marker gene sequences are clearly separated from the new transgenic trait, and do not appear anymore in the adult plant (Konig 2003). However, this technology is now patented and accessibility is restricted.

12.5.1.2 Excision by Homologous Recombination

Due to homologous recombination between identical DNA sequences flanking a central region, this central DNA fragment can be excised (Woo et al. 2009). Despite some initial success, the naturally-occurring recombination frequencies in plants are rather disappointing and the whole process is difficult to control. The technique,

at this stage, is not yet considered established but may be promising for future applications (Konig 2003).

12.5.1.3 Recombinase Induced Excision

The DNA sequence coding for selectable marker is removed by the recombinase enzyme that cuts two DNA recognition sequences and ligates the free ends after the eliminating the DNA sequence located in between. The Cre/lox system from bacteriophage P1 is predominantly used in plants relying on three different strategies (Zuo et al. 2001): Autoexcision, transient transfer and outbreeding with Cre carrying plant cell line.

Autoexcision The marker gene flanked by two lox-sequences is co-transformed with the Cre gene. The promoter that regulates the expression of the Cre gene is either tissue-specific, growth phase dependent or inducible. This is a new technology and complicated during execution. The transformation efficiency is too low for commercial application. Patents limit accessibility.

Transient Transfer Transiently the Cre-protein is directly transferred into the targeted plant cell via microinjection. Strains of *Agrobacterium* carrying two different T-DNA containing vectors harbor one vector carrying a positive and a negative selection marker gene whereas the other vector contains the gene of interest. After transformation a transient positive selection step is followed by a negative selection for marker gene integration. This approach leads to approximately 29 % of plants containing the desired DNA sequences (Rommens et al. 2004).

Outbreeding with Cre Carrying Plant Cell Line Generation F_1 plants carry both the Cre-sequences and the transgene. In generation F_2, outcrossing will eliminate the Cre-sequences. A single lox insertion site and the transgene will remain in the plant genome (Rommens et al. 2004). This technology is established and can be applied in many plants of commercial interest. However, it is certainly more resource consuming and less cost efficient than the nptII marker gene system (Konig 2003).

12.6 Cisgenesis: A Case Study

Recombinant DNA technology has been employed in the generation of fire blight and apple scab resistant cultivars of apple. However, in these cases a large number of exogenous genes coding for lytic enzymes from different sources (e.g. encoding lysozymes from bacteria, phages and animals) have been integrated in the apple genome. Also, genes and promoters derived from pathogens have been used (Gessler and Patocchi 2007). The presence of exogenous genes and selectable markers in these genetically modified apples has raised public concerns about their fitness for human consumption.

The first truly cisgenic apple cv. Gala, resistant to fungal pathogen *Venturia inequalis* causing apple scab, was reported by Vanblaere et al. (2011). In this work plants of the popular apple Gala were genetically modified by inserting the apple scab resistance gene Rvi6 (formerly HcrVf2) under control of its own regulatory sequences. This gene is derived from the scab-resistant apple Florina (originally from the wild apple accession *Malus floribunda* 821). The Rvi6 gene isolated from the wild relative *Malus floribunda* 821 has been shown to induce scab resistance in several apple cultivars (Szankowski et al. 2009). The gene Rvi6 was isolated along with 5' UTR and 3' UTR regions of the ORF and amplified from the BAC clone M18-5. The gene was then inserted in the binary vector PMF1. The T-DNA of this vector carried the cisgene Rvi6 along with its native up- and downstream regulatory sequences and a cassette flanked by recombination sites containing the R recombinase gene and a fusion of marker genes nptII and codA, allowing positive and negative selection on kanamycin- and 5-fluoro-cytosine selective medium respectively. The plasmids were then transferred for transformation into *Agrobacterium tumefaciens* through electroporation. The method is, thus, based on an *Agrobacterium-mediated* transformation of apple lines of cv. Gala. This was followed by regeneration on a kanamycin selective medium. The excision of the cassette carrying the transgenes was carried out by recombinase which is chemically activated by addition of dexamethasone to the medium and recombinants selected on a negative selection medium containing 5-fluorocytosine. These recombinants were putative cisgenic lines. Thus, a marker free system was developed for cisgenesis containing only the gene of interest.

The requirements for safety assessment of cisgenic and transgenic products are: (1) the assessment of the newly introduced gene sequences and their (intended) expression products and derived effects and (2) potential unintended effects caused by the genetic modification (Prins and Kok 2010). To address this concern molecular characterization of this product was requested by the (EC) Regulation No. 1829/2003. In the study on molecular characterization three cisgenic lines were analyzed to assess copy number, integration site, expression level and resistance to apple scab. For two of these lines, a single insertion was observed and, despite a very low expression of 0.07- and 0.002-fold compared with the natural expression of Florina, this was sufficient to induce plant reaction and reduce fungal growth by 80 % compared with the scab-susceptible Gala. Similar results for resistance and expression analysis were obtained also for the third line, although it was impossible to determine the copy number and TDNA integration site (Vanblaere et al. 2014). This is the first scientific report of the generation of a cisgenic crop and the further molecular characterization of the end-product as an attempt to address the safety concerns.

12.7 Intragenesis: A Case Study

An example of intragenic crop that uses the marker free transformation method is enhanced quality alfalfa. A silencing construct targeting the native caffeic acid o-methyltransferase (Comt) gene was positioned within an alfalfa-derived transfer DNA. Alfalfa plants were transformed and allowed to mature in the greenhouse. Polymerase chain reaction (PCR)-based genotyping of 1000 5-week old plants identified 2.4 % that contained the modified P-DNA. Stem sections were isolated from intragenic progeny plants derived from eight randomly chosen original transformants and assayed for lignin accumulation. Reduced lignin levels were confirmed in three cases out of eight. These reduced lignin levels have been found to enhance the value of alfalfa as feed for dairy cattle. Resulting plants may represent low-risk genetically engineered crops that should be cleared through the regulatory process in a timely and cost-effective manner.

12.8 Use of Cisgenesis/Intragenesis in Crop Improvement

The growing public concern about transgenic food products is based on the fact that these are usually produced by artificial means that defy nature. This concern is further fuelled by the fear of potential health hazards and a possibility of gene flow between genetically modified crop and its wild relatives. The regulatory bodies worldwide are, thus, apprehensive about allowing the use of different genetically modified crops. With the aim of meeting these reservations and at the same time ensuring an environmentally sound and efficient plant production, the two transformation concepts intragenesis and cisgenesis were developed as alternatives to transgenic crop development (Holme et al. 2013). The two concepts are based on the exclusive use of genetic material from the same species or genetic material from closely related species capable of sexual hybridization. Thus, these concepts are very similar to traditional plant breeding as they do not enhance the gene pool of recipient species (Schouten et al. 2006a). The intragenic/cisgenic concepts can also overcome limitations of classical breeding when it comes to improving traits with limited natural allelic variation within the sexually compatible gene pool. Higher expression level of a trait can be obtained by re-introducing the gene of the trait with its own promoter and terminator (cisgenesis) or with a promoter and terminator isolated from the sexually compatible gene pool (intragenesis). Lower expression levels can be obtained through different silencing constructs as is the case in intragenesis (Holme et al. 2013). In addition, the risk of linkage drag and the time required to deploy the gene of interest is greatly reduced as is the case in transgenesis. In recent times cisgenesis/intragenesis has emerged as the alternative that combines the best of both conventional plant breeding and transgenesis. This is evidenced by several cisgenic and transgenic products that have been developed and many more that are in the pipeline.

12.8.1 Cisgenesis/Intragenesis in Heterozygous Cultivars

In cultivars where two alleles at a locus are not the same (heterozygous cultivars) sexual reproduction produces new combination of genes in the progeny of desirable parents and the chances of recovery of original genotypes are low.

12.8.1.1 Commercially Widespread Clones

Heterozygous crops need to be propagated vegetatively commercially because outcrossing with desired variety creates new genotypes and original heterozygous genotype is never recovered. Thus, even though offspring can be produced through sexual mating, it is desirable to maintain an improved genotype asexually for commercial purposes. The direct insertion of the desired genotype through transformation is beneficial for the genotype to remain intact. Cisgenesis and intragenesis were first implemented in commercially widespread clones of crops like apple, strawberry, potato and grapevine for this reason. Much of the research in these areas is still being carried out in fruit crops (Rai and Shekhawat 2014). Transformations through these techniques have been brought about by the silencing of undesired gene activities or through enhancement of disease resistance (Table 12.2).

12.8.1.2 Woody Species

Woody plants are highly heterozygous, intolerant to inbreeding and have very long generation times, all making traditional breeding very slow and difficult. Genomic poplar clones encoding for gibberellic acid biosynthesis enzymes, catabolic enzymes and negative signal regulators were isolated with their native promoters and terminators and independently introduced into the poplar tree (Han et al. 2011). This helped in increasing the genetic variance in *Populus* as the overexpression of gibberellic acid biosynthesis enzymes produced genotypes with enhanced growth rate whereas overexpression of catabolic enzymes and negative signal regulators decreased the growth rate substantially. The *Populus* genotypes produced by this technique were not truly cisgenic as they contained Pat selection marker gene encoding phosphinothricin acetyl transferase.

12.8.1.3 Forage Crops

To date, cisgenic/intragenic approach to improve traits has been applied in only two forage crops viz. alfalfa and perennial rye grass (Table 12.3). This limited application is due to the potential risk of easy transfer of transgenes to wild species and relatives as these crops are perennial and outcrossing.

Table 12.2 Heterozygous crops with commercially widespread clones developed or under development through cisgenesis or intragenesis

Crop	Technique	Trait	Gene	References
Intragenic				
Potato (*Solanum tuberosum*)	Silencing	High amylopectin	GBSS (granule-bound starch synthase)	de Vetten et al. (2003)
Potato (*Solanum tuberosum*)	Silencing	Preventing enzymating browning (enhancing storage quality)	Ppo (polyphenol oxidase)	Rommens et al. (2004)
Potato (*Solanum tuberosum*)	Silencing	Limiting cold-induced degradation of starch, limiting acrylamide formation	Ppo, *R1* (water dikinase), PhL (amyloplast-targeted phosphorylase-L)	Rommens et al. (2006)
Potato (*Solanum tuberosum*)	Silencing	Limiting acrylamide formation	StAs1,StAS2 (Asparagine synthetase genes)	Rommens et al. (2008)
Potato (*Solanum tuberosum*)	Silencing	Limiting acrylamide formation	StAs1 (Asparagine synthetase gene)	Chawla et al. (2012)
Strawberry (*Fragaria* spp.)	Over-expression	Grey mould resistance	PGIP (Polygalacturonase Inhibiting Protein)	Schaart et al. (2004)
Cisgenic				
Potato (*Solanum tuberosum*)	Genes from related species	Late blight resistance	R-genes	Haverkort et al. (2009)
Apple (*Malus domestica*)	Genes from related species	Scab resistance	HcrVf2 (gene present in the scab resistance locus Vf)	Vanblaere et al. (2011)
Grapevine (*Vitis vinifera*)	Genes from related species	Fungal disease resistance	VVTL-1 (*Vitis vinifera* thaumatin-like protein)	Dhekney et al. (2011); Espinoza et al. (2013)

Table 12.3 Application of intragenesis in perennial forage crops

Intragenic crop	Technique	Trait	Gene	References
Perennial rye grass (*Lolium perenne*) (UD)	Overexpression	Drought tolerance	Lpvp1 (H^+-pyrophosphatase)	Bajaj et al. (2008)
Alfalfa (*Medicago sativa*) (D)	Silencing	Reduced lignin content	Comt (caffeic acid o-methyltransferase)	Weeks et al. (2008)

D developed, *UD* underdeveloped

Table 12.4 Application of cisgenesis in seed propagated, self-pollinating crops

Cisgenic crop	Technique	Trait	Gene	References
Durum wheat (*Triticum turgidum* var. *durum*)	Gene from related species	Improved baking quality	1Dy10(HMW glutenin subunit gene from bread wheat)	Gadaleta et al. (2008)
Barley (*Hordeum vulgare*)	Overexpression	Improved grain phytase activity	HvPAPhy_a (phytase gene)	Holme et al. (2012)

12.8.2 Cisgenesis/Intragenesis in Seed Propagating Self-pollinating Crops

The genetic transformation concepts of cisgenesis and intragenesis have been employed in self-pollinating crops (Table 12.4) that are propagated through seeds for achieving more accurate introgression of gene of interest in less time. It is particularly useful in avoiding linkage drag and in traits with limited natural allelic variation.

These examples clearly illustrate the use of cisgenesis and intragenesis as an alternative to transgenesis in precise transfer of gene of interest in an expedite manner.

12.9 Safety Issues of Cisgenesis, Intragenesis and Transgenesis

Genetic modification means that it is possible to modify and identify the existence of a change in the genetic material of an organism at the level of DNA through the presence of a novel DNA sequence. With the advent of new technologies (cisgenesis, intragenesis and transgenesis) genetic modification became possible but some consequences related to these should be focused on. Some major safety issues related to cisgenesis, transgenesis and intragenesis are described below:

12.9.1 Cisgenesis

In cisgenesis one or few genes from crossable species are introduced into a variety with a safe crop history. Expression of the Cis gene differs from the expression of the endogenous gene due to position effect and may lead to phenotypic differences. Cisgene expression should be checked and compared with the baseline. Conventionally bred plants and related species from which cisgenes have been isolated can be used as the baseline (Schaart and Visser 2009). Cisgenes can be

integrated into genes of the recipient genome due to a random integration event that may lead to insertional mutagenesis and deletion of the gene in which the cisgene has integrated (Schouten et al. 2006a). When no genes of the recipient have been mutated as a result of cisgene integration, and when the right border and left border have not become part of an open reading frame (ORF), the cisgenic plants are similar to conventionally bred plants used as the baseline. It can be argued that cisgenesis is safer than conventional breeding because it prevents the introduction of genes via linkage drag which could lead to unwanted traits (e.g. increase glycoalkaloid content to a higher level than allowed in the regulations for plant breeder's rights). Conventional breeding can also result in disruptions to ORFs and other molecular changes including deletions and recombinations. Therefore, the focus should be on the shifting to gene targeting to achieve site-specific mutagenesis. Schouten et al. (2006b) claim that cisgenic plants should be exempt from current GMO regulatory requirements as these are same as conventionally bred plants. It can be argued that some conventional breeding methods like induced mutagenesis (Ahloowalia and Maluszynski 2001; Ahloowalia et al. 2004; Brunner 1995) have an even higher probability of unexpected genetic changes (Mei et al. 1994; Shirley et al. 1992) than cisgenesis. On a biological level, the alterations can range from small changes (adding a single resistance gene) to extensive changes in which complex metabolic routes have been introduced. Depending on their specific modifications and characteristics, cisgenic crop varieties will be somewhere in the between transgenic and conventional bred plants (Kok et al. 2008).

12.9.2 Intragenesis

The cisgenic/intragenic approach is based on the assumption of cross-compatibility of the host plant and the plant used to provide the genes. The safety aspect depends on the germplasm used as a gene source and history of its use in the food chain. Development of new genetic elements may lead to chimeric genes that do not exist in nature. These new genes may have expression levels and patterns that do not correspond to that of the native gene. It can lead to position effect on the intragene expression. The consequences for food and feed safety of this deviation in gene expression should be evaluated on a case-by-case study and compared to the baseline. As baseline products, those from conventionally bred plants can be used. If intragenesis was used for the silencing of a single native gene, the consequences for food and feed safety are in general similar to that of crops obtained by mutation breeding in which the same gene is knocked-out. Such plants from mutation breeding programs can serve as a baseline. The random integration of intragenes may have an effect on the expression of genes of the recipient genome. This phenomenon is however also expected when translocation breeding, a traditional breeding method, is applied. To investigate whether or not the intragene has been inserted into a recipient gene, and has become part of an ORF, the part of the recipient genome that is flanking the inserted P-DNA site can be sequenced. The insertion of

an intragene may result in a mutation in the recipient genome at the site of insertion (Forsbach et al. 2003). Such a mutation usually leads to a disruption of gene function in the recipient genome and can thereby induce phenotypic effects. Next to this, the newly introduced gene may also influence the expression of genes of the recipient genome if they are located around the integration site. However, both effects of intragene integration are natural phenomena occurring during transposon transition (Greco et al. 2001) and translocation breeding, respectively (Papazova and Gecheff 2003).

12.9.3 Transgenesis

Transgenics are most affected by biosafety issues and have already become a crucial factor in constraining the further development of transgenic biotechnology and wider application of GM products in agriculture. There are quite a number of biosafety related concerns in general; the most important ones can be summarized as follows:

(a) Direct and indirect effects of toxic transgenes (e.g. the Bt insect-resistance gene) to non-target organisms (O'Callaghan et al. 2005; Oliveira et al. 2007)
(b) Influences of transgenes and GM plants on biodiversity, ecosystem functions and soil microbes (Giovannetti et al. 2005; Oliveira et al. 2007)
(c) Transgene escape to crop landraces and wild relatives through gene flow and its potential ecological consequences (Lu and Snow 2005; Mercer et al. 2007)
(d) Transgene escape can easily occur via gene flow that may result in potential ecological consequences if significant amounts of transgenes constantly move to non-GM crops and wild relative species.
(e) There are potential risks associated with the development of resistance to biotic-resistance transgenes in the target organisms (Dalecky et al. 2007; Li et al. 2007; Wu 2007).
(f) Unintended, unexpected effects of random gene insertion and interaction between foreign genes and host genes in the transgenic organisms (Kok and Kuiper 2003).

In addition, there are still some unknown involvements in potentially significant interactions between transgenic traits and their environments.

12.10 Advantages and Drawbacks

Cisgenesis and intragenesis have advantages over transgenesis as described in Table 12.5. A main advantage of cisgenesis/intragenes compared to conventional breeding is the saving of time necessary for breeding. This is especially important for crops which are vegetatively propagated, such as potato, strawberry or banana,

Table 12.5 Comparison of cisgenesis and intragenesis versus transgenesis

Cisgenesis and intragenesis	Transgenesis
Great potential to overcome the limitations of classical breeding.	Great potential to overcome the limitations of classical breeding.
Plants similar to conventional bred plants but linkage Drag is not associated with cisgenic plant. It has potential to overcome some of the limitations of classical breeding (Holme et al. 2013)	Plants different to conventional bred plants. It has potential to overcome many of the limitations of classical breeding.
Traits outside the sexually compatible gene pool cannot be introduced.	Traits outside the sexually compatible gene pool can be introduced.
Vector-backbone sequences are not present in transformants.	Some 20–80 % (depending on the species) of the transformants contain vector-backbone sequences, many transgenic lines have to be discarded.
Faster and more precise tool for transfer of genes between related species than classical backcross breeding.	Faster and more precise tool for transfer of genes between related species than classical backcross breeding.
Cisgenesis leads to hazards similar to conventional bred plants. Must be treated differently than transgenic GMO regulations.	Transgenesis creates novel hazards. So must be treated under strict GMO regulations.

and for crops with long generation times, such as fruit trees. For example, half a century was necessary for breeding of apples with scab resistance. By using cisgenesis or intragenesis, this time can be reduced to five years when isolated resistance genes are available. In addition, these have great potential to overcome the limitations of classical breeding and the transfer of genes between sexually compatible plants can be speeded up. The advantages are the following:

(a) Linkage drag associated with conventional breeding is avoided. Tightly linked inferior traits can be completely eliminated. The exchange of genetic material is precise.
(b) In intragenesis, higher expression is achieved by using a more efficient promoter and lower expression is achieved through gene silencing.
(c) Transgenic provide a wider gene pool for integration as genes from any species can be introgressed. In transgenic breeding there is no restriction on the source of variability so it provides better improvement opportunities. However, risks attached to cisgenics and intragenics are relatively less than transgenics.

12.11 Conclusions and Prospects

The gene pool exploited by intragenesis and cisgenesis is identical to the gene pool available for traditional breeding (Holme et al. 2013). Therefore, cisgenesis and intragenesis could be treated on a par with conventional breeding and exempted from regulations that govern transgenics (Bhargava and Carmona 2012).

The present and future potential use of these concepts to produce intragenic and cisgenic crops is encouraged by several surveys and focus group interviews in the US and Europe that clearly show that both intragenic and cisgenic crops are acceptable to a greater number of people than transgenic crops. Further encouragement for commercial application would be derived from less rigid regulatory measures compared to the current regulation for transgenic crops (Holme et al. 2013) Cisgenesis and intragenesis are still generating new risks because insertion at random locations (ectopic) in the genome could have unpredictable pleiotropic effects. The endogenous gene in the recipient plant is not altered. Therefore the focus must be on shifting to gene targeting to achieve site-specific mutagenesis. Cisgenesis and intragenesis have created an opportunity to initiate a new dialogue among scientists, breeders and consumers to discuss a new group of genetically modified crops which are consumer friendly. The future developments of intragenic/cisgenic crops will, however, depend on a lessening of rigid regulatory measures compared to the current regulation for transgenic crops (Holme et al. 2013).

Increased confidence in these novel crops may facilitate a qualified discussion of the potential environmental, economic, and health benefits of genetically engineered plants, and of their possible contribution to address the global need for more efficient and sustainable crop production (Holme et al. 2013).

However, in transgenics proper testing of transformants is necessary so that undesirable effects can be avoided. There is the need for the development of techniques for the efficient production of clean marker-free transgenic plants. Thus, the development of efficient techniques for the removal of selectable markers, as well as the directed integration of transgenes at safe locations in the genome, is of great interest to biotech companies. Therefore attention should be on the development of marker- and vector backbone-free plants (Tuteja et al. 2012). Genome sequences responsible for the trait including the endogenous promoters and terminators should be cloned.

All these techniques provide a means for gene integration and transformation and results in general changes in structure and/or re-organization of the plant's genome as a result of the modification. These strategies lead to major breakthroughs in the fields of biotechnology, genetics and plant breeding, which were not otherwise possible. GMOs can be released for marketing after approval, both for their introduction in the environment and for their application as food and feed so that known and unknown hazards can be avoided.

References

Ahloowalia BS, Maluszynski M (2001) Induced mutations – a new paradigm in plant breeding. Euphytica 118(2):167–173

Ahloowalia BS, Maluszynski M, Nichterlein K (2004) Global impact of mutation-derived varieties. Euphytica 135:187–204

Akhond M, Machray G (2009) Biotech crops: technologies, achievements and prospects. Euphytica 166:47–59

Anderson PC, Georgeson M (1989) Herbicide tolerant mutants of corn. Genome 31:994

Atanassov A, Bahieldin A, Brink J et al (2004) To reach the poor: results from the ISNAR-IFPRI next harvest study on genetically modified crops, public research, and policy implications (no.116). In: EPTD discussion paper. Environment and Production Technology Division, International Food Policy Research Institute, Washington, DC, pp 1–57

Bajaj S, Puthigae S, Templeton K et al (2008) Towards engineering drought tolerance in perennial ryegrass using its own genome. Paper presented at the 6th Canadian plant genomics workshop, Toronto, 23–26 June 2008

Barton K, Whitely H, Yang NS (1987) *Bacillus thuringiensis* d-endotoxin in transgenic *Nicotiana tabacum* provides resistance to lepidopteran insects. Plant Physiol 85:1103–1109

Bhargava A, Carmona FF (2012) Cisgenesis and intragenesis. In: Bhargava A, Srivastava S (eds) Biotechnology: new ideas, new developments : a textbook of modern technology. Nova Science Publishers, New York, pp 151–159

Brunner H (1995) Radiation induced mutations for plant selection. Appl Radiat Isot 56(6/7):589–594

Chakraborty S, Chakraborty N, Datta A (2000) Increased nutritive value of transgenic potato by expressing a nonallergenic seed albumin gene from *Amaranthus hypochondriacus*. Proc Natl Acad Sci U S A 97:3724–3729

Chawla R, Shakya R, Rommens CM (2012) Tuber-specific silencing of asparagine synthetase-1 reduces the acrylamide-forming potential of potatoes grown in the field without affecting tuber shape and yield. Plant Biotechnol J 10:913–924

Cromwell GL, Pickett RA, Beeson WM (1967) Nutritional value of opaque-2 corn for swine. J Anim Sci 26:1325–1331

D'Halluin K, Botterman J, De Greef W (1990) Transformation engineering of herbicide resistant alfalfa and evaluation under field conditions. Crop Sci 30:866

D'Halluin K, Bossut M, Bonne E et al (1992) Transformation of sugarbeet (*Beta vulgaris* L.) and evaluation of herbicide resistance in transgenic plants. Biotechnol 10:309

Dale PJ (1992) Spread of engineered genes to wild relatives. Plant Physiol 100(1):13–15

Dalecky A, Bourguet D, Ponsard S (2007) Does the European corn borer disperse enough for a sustainable control of resistance to Bt maize via the high dose/refuge strategy? Cah Agric 16:171–176

de Vetten N, Wolters A, Raemakers K et al (2003) A transformation method for obtaining marker-free plants of a cross-pollinating and vegetatively propagated crop. Nat Biotechnol 21:439–442

Dehesh K, Jones A, Knutzon DS, Voelker TA (1996) Production of high levels of 8:0 and 10:0 fatty acids in transgenic canola by overexpression of Ch FatB2, a thioesterase cDNA from *Cuphea hookeriana*. Plant J 9:167–172

Del Vecchio AJ (1996) High-laurate canola: how Calgene's program began, where it's headed. INFORM 7:230–243

Dhekney SA, Li ZT, Gray DJ (2011) Grapevines engineered to express cisgenic *Vitis vinifera* thaumatin-like protein exhibit fungal disease resistance. In Vitro Cell Dev Biol Plant 47:458–466

Espinoza C, Schlechter R, Herrera D et al (2013) Cisgenesis and intragenesis: new tools for improving crops. Biol Res 46:323–331

Finer JJ, Mc Mullen MD (1990) Transformation of cotton (*Gossypium hirsutum* L.) via particle bombardment. Plant Cell Rep 8:586–589
Firoozabady E, Debore DL, Merlo DJ et al (1987) Transformation of cotton, *Gossypium hirsutum* L. by *Agrobacterium tumfaciens* and regeneration of transgenic plants. Plant Mol Biol 10:566–578
Fitch MMM, Manshard RM, Gonsalves D et al (1992) Virus resistant papaya derived from tissues bombarded with the coat protein gene of papaya ringspot virus. Biotechnol 10:1466–1472
Forsbach A, Schubert D, Lechtenberg B (2003) A comprehensive characterization of single-copy T-DNA insertions in the *Arabidopsis thaliana* genome. Plant Mol Biol 52:161–176
Friebe B, Jiang J, Raupp WJ, McIntosh RA, Gill BS (1996) Characterization of wheat-alien translocations conferring resistance to diseases and pests: Current status. Euphytica 91:59–87
Froman B, Ursin V (2002) Genetic modification of oils for improved health benefits: production of long chain omega-3 fatty acids in plants. In: Abstracts of papers of the American chemical society 223:U35
Gadaleta A, Giancaspro A, Blechl AE, Blanco A (2008) A transgenic durum wheat line that is free of marker genes and expresses 1DY10. J Cereal Sci 48:439–445
Gessler C, Patocchi A (2007) Recombinant DNA technology in apple. Adv Biochem Eng Biotechnol 107:113–132
Giovannetti M, Sbrana C, Turrini A (2005) The impact of genetically modified crops on soil microbial communities. Riv Biol 98:393–417
Gonsalves D, Ferreira S, Manshardt R et al (1998) Transgenic virus resistant papaya: new hope for control of papaya ringspot virus in Hawaii. APSnet feature story for September 1998 on worldwide web. http://www.apsnet.org/publications/apsnetfeatures/Pages/Papaya.aspx
Greco R, Ouwerkerk PFB, Sallaud C et al (2001) Transposon insertional mutagenesis in rice. Plant Physiol 125:1175–1177
Han KM, Dharmawardhana P, Arias RS et al (2011) Gibberellin-associated cisgenes modify growth, stature and wood properties in *Populus*. Plant Biotechnol J 9:162–178
Haverkort AJ, Struik PC, Visser RGF, Jacobsen E (2009) Applied biotechnology to combat late blight in potato caused by *Phytophthora infestans*. Potato Res 52:249–264
Hemenway C, Fang RX, Kaniewske WK (1988) Analysis of the mechanism of protection in transgenic plants expressing the potato virus X coat protein or its antisense RNA. EMBO J 7:1273–1280
Holme IB, Dionisio G, Brinch-Pedersen H et al (2012) Cisgenic barley with improved phytase activity. Plant Biotechnol J 10:237–247
Holme IB, Wendt T, Holm PB (2013) Intragenesis and cisgenesis as alternatives to transgenic crop development. Plant Biotechnol J 11:395–407
Hou H, Atlihan N, Lu ZX (2014) New biotechnology enhances the application of cisgenesis in plant breeding. Front Plant Sci 389(5):1–5
http://www.apsnet.org/education/feature/papaya/Yop.htm. Accessed 29 Nov 2014
http://www.efsa.europa.eu/en/science/gmo/gmo_guidance/660.html. Accessed 15 Jul 2014
Hunter P (2014) "Genetically Modified Lite" placates public but not activists. EMBO Rep. doi:10.1002/embr.201338365
Itaya A (2008) Small RNAs in tomato fruit and leaf development. Biochim Biophys Acta 1779:99–107
Jacobsen E, Schouten HJ (2007) Cisgenesis strongly improves introgression breeding and induced translocation breeding of plants. Trends Biotechnol 25(5):219–223
James MJ, Ursin VM, Cleland LG (2003) Metabolism of stearidonic acid in human subjects: comparison with the metabolism of other n-3 fatty acids. Am J Clin Nutr 77:1140–1145
Jochemsen H, Schouten HJ (2000) Ethische beoordeling van genetische modificatie. In: Jochemsen H (ed.) Toetsen en Begrenzen. Een Ethische en Politieke Beoordeling van de Moderne Biotechnologie. Buijten and Schipperheijn. Amsterdam, pp 88–95

Joshi SG, Schaart JG, Groenwold R et al (2011) Functional analysis and expression profiling of HcrVf1 and HcrVf2 for development of scab resistant cisgenic and intragenic apples. Plant Mol Biol 75:579–591
Katsube T, Kurisaka N, Ogawa M et al (1999) Accumulation of soybean glycinin and its assembly with the glutelins in rice. Plant Physiol 120:1063–1074
Kinney AJ, Knowlton S (1998) Designer oils: the high oleic acid soybean. In: Roller S, Harlander S (eds) Genetic modification in the food industry. Blackie Academic and Professional, London, pp 193–213
Kok EJ, Kuiper HA (2003) Comparative safety assessment for biotech crops. Trends Biotechnol 21:439–444
Kok EJ, Keijer J, Meter GA, Kuiper HA (2008) Comparative safety assessment of plant-derived foods. Regul Toxicol Pharmacol 50(1):98–113
Konig A (2003) A framework for designing transgenic crops – science, safety and citizen's concerns. Nat Biotechnol 21:1274–1279
Krishna VV, Qaim M, Zilberman D (February 2014) Transgenic crops, production risk, and agrobiodiversity. ZEF- discussion papers on development policy no. 186. http://dx.doi.org10.2139/ssrn.2405466
Li L, Liu SM, Hu YL et al (2001) Increase of sulphurcontaining amino acids in transgenic potato with 10 ku zein gene from maize. Chin Sci Bull 46:482–484
Li GP, Wu KM, Gould F et al (2007) Increasing tolerance to Cry1Ac cotton from cotton bollworm *Helicoverpa armigera* was confirmed in Bt cotton farming area of China. Ecol Entomol 32:366–375
Lu BR, Snow AA (2005) Gene flow from genetically modified rice and its environmental consequences. Bioscience 55:669–678
Lu BR, Yang C (2009) Gene flow from genetically modified rice to its wild relatives: assessing potential ecological consequences. Biotechnol Adv 27(6):1083–1091
Lusser M, Parisi C, Plan D, Rodríguez-Cerezo E (2012) Deployment of new biotechnologies in plant breeding. Nat Biotechnol 30:3
Lyon BR, Llewellyn DJ, Huppatz JL et al (1989) Expression of a bacterial gene in transgenic tobacco plant confers resistance to herbicide 2,4-dichlorophenoxyacetic acid. Plant Mol Biol 13:533
Maluszynski M (2001) Officially released mutant varieties – the FAO/IAEA database. Plant Cell Tiss Org Cult 65:175–177
Mei M, Deng H, Lu Y et al (1994) Mutagenic effects of heavy-ion radiation in plants. Life Sci Space Res 14:363–372
Mercer KL, Andow DA, Wyse DL, Shaw RG (2007) Stress and domestication traits increase the relative fitness of crop-wild hybrids in sunflower. Ecol Lett 10:383–393
Molesini B, Youry P, Pandolfini T (2012) Fruit improvement using intragenesis and artificial microRNA. Trends Biotechnol 30(2):143–147
Newhouse K, Singh B, Shaner D, Stidham M (1991) Mutations in corn (*Zea mays* L.) conferring resistance to imidazonilone herbicides. Theor Appl Genet 83:65
Newhouse KE, Smith WA, Starrett MA et al (1992) Tolerance to imidazonilone herbicides in wheat. Plant Physiol 100:882
O'Callaghan M, Glare TR, Burgess EPJ, Malone LA (2005) Effects of plants genetically modified for insect resistance on non-target organisms. Annu Rev Entomol 50:271–292
Oliveira AR, Castro TR, Capalbo DMF, Delalibera I (2007) Toxicological evaluation of genetically modified cotton BollgardR and Dipel RWP on the non-target soil mite *Scheloribates praeincisus Acari*: Oribatida. Exp Appl Acarol 41:191–201
O'Quinn PR, Nelssen JL, Goodband RD et al (2000) Nutritional value of a genetically improved high-lysine, high-oil corn for young pigs. J Anim Sci 78:2144–2149
Papazova N, Gecheff K (2003) Position-dependent gene activity in cytologically reconstructed barley karyotypes. Cell Biol Int 27:247–248

Prins T, Kok E (2010) COGEM report: food and feed safety aspects of cisgenic crop plant varieties. http://edepot.wur.nl/157733
Rai MK, Shekhawat NS (2014) Recent advances in genetic engineering for improvement of fruit crops. Plant Cell Tiss Org Cult 116(1):1–15
Reddy AS, Thomas TL (1996) Expression of a cyanobacterial DELTA 6-desaturase gene results in gamma-linolenic acid production in transgenic plants. Nat Biotechnol 14:639–642
Roesler K, Shintani D, Savage L et al (1997) Targeting of the *Arabidopsis* homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. Plant Physiol 113:75–81
Rommens CM (2004) All-native DNA transformation: a new approach to plant genetic engineering. Trends Plant Sci 9:457–464
Rommens CM (2007) Intragenic crop improvement: combining the benefits of traditional breeding and genetic engineering. J Agr Food Chem 55:4281–4288
Rommens CM, Ye J, Humara JM et al (2003) Precise breeding. U.S. Patent Application No. 221,213
Rommens CM, Humara JM, Ye J et al (2004) Crop improvement through modification of the plant's own genome. Plant Physiol 135:421–431
Rommens CM, Bougri O, Yan H, Humara JM, Owen J, Swords K, Ye J (2005) Plant-derived transfer DNAs. Plant Physiol 139:1338–1349
Rommens CM, Ye J, Richael C, Swords K (2006) Improving potato storage and processing characteristics through all-native DNA transformation. J Agr Food Chem 54:9882–9887
Rommens CM, Yan H, Swords K, Richael C, Ye J (2008) Low-acrylamide French fries and potato chips. Plant Biotechnol J 6:843–853
Schaart JG, Krens FA, Pelgrom KTB et al (2004) Effective production of marker-free transgenic strawberry plants using inducible site-specific recombination and a bifunctional selectable marker gene. Plant Biotechnol J 2:233–240
Schaart JG, Visser RGF (2009) Novel plant breeding techniques – consequences of new genetic modification-based plant breeding techniques in comparison to conventional plant breeding. COGEM Research Report Number 2009-02, The Netherlands Commission on Genetic Modification
Schouten HJ, Jacobsen E (2008) Cisgenesis and intragenesis, sisters in innovative plant breeding. Trends Plant Sci 13:260–261
Schouten HJ, Krens FA, Jacobsen E (2006a) Cisgenic plants are similar to traditionally bred plants. EMBO Rep 7:750–753
Schouten HJ, Krens FA, Jacobsen E (2006b) Do cisgenic plants warrant less stringent oversight? Nat Biotechnol 24(7):753
Shirley BW, Hanley S, Goodman HM (1992) Effects of ionizing-radiation on a plant genome – analysis of two *Arabidopsis* transparent-testa mutations. Plant Cell 4(3):333–347
Shiva Prasad G, Raju B, Rajesh Kumar K (2013) Cisgenesis – an alternate approach for development of genetically modified crops. Annu Biol Res 4(9):109–115
Smith BD (1995) The emergence of agriculture. Scientific American Library, New York
Streber WR, Willmitzer L (1989) Transgenic tobacco plants expressing a bacterial detoxifying enzyme are resistant to 2,4-D. Biotechnol J 7:811
Szankowski I, Waidmann S, Degenhardt J et al (2009) Highly scab-resistant transgenic apple lines achieved by introgression of HcrVf2 controlled by different native promoter lengths. Tree Genet Genome 5(2):349–358
Tricoli D, Carney KJ, Russell PF et al (1995) Field evaluation of transgenic squash containing single or multiple virus coat protein gene constructs for resistance to cucumber mosaic virus, watermelon mosaic virus 2 and zucchini yellow mosaic virus. Biotechnol 13:1458–1465
Tuteja N, Verma S, Sahoo RK (2012) Recent advances in development of marker-free transgenic plants: regulation and biosafety concern. J Biosci 37:167–197
Umbeck P, Johnson P, Barton K, Swain W (1987) Genetically transformed cotton (*Gossypium hirsutum* L.) plants. Biotechnol 5:263–266

Vaeck M, Reynaerts A, Hoftey H et al (1987) Transgenic plants protected from insect attack. Nature 327:33–37

van Harten AM (1998) Mutation breeding, theory and practical application. Cambridge University Press, Cambridge UK

Vanblaere T, Szankowski I, Schaart J et al (2011) The development of a cisgenic apple plant. J Biotechnol 154:304–311

Vanblaere T, Flachowsky H, Gessler C, Giovanni AL (2014) Molecular characterization of cisgenic lines of apple 'Gala' carrying the Rvi6 scab resistance gene. Plant Biotechnol J 12:2–9

Vasil V, Castillo AM, Fromm ME, Vasil IK (1992) Herbicide resistant fertile wheat transgenic plants obtained by microprojectile bombardment of regenerable embryogenic callus. Biotechnol 10:67

Weeks JT, Ye J, Rommens CM (2008) Development of an in planta method for transformation of alfalfa (*Medicago sativa*). Transgenic Res 17:587–597

Woo HJ, Cho HS, Lim SH, Shin KS, Lee SM, Lee KJ, Kim DH, Cho YG (2009) Autoexcision of selectable marker genes from transgenic tobacco via a stress inducible FLP/FRT site-specific recombination system. Transgenic Res 18:455–465

Woo HJ, Suh SC, Cho YG (2011) Strategies for developing marker-free transgenic plants. Biotech Bioproc Eng 16(6):1053–1064

Wu KM (2007) Monitoring and management strategy for *Helicoverpa armigera* resistance to Bt cotton in China. J Invertebr Pathol 95:220–223

Yang SH, Moran DL, Jia HW et al (2002) Expression of a synthetic porcine alpha-lactalbumin gene in the kernels of transgenic maize. Transgenic Res 11:11–20

Ye X, Al-Babili S, Klöti A et al (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287(5451):303–305

Young TE, Giesler-Lee J, Gallie DR (2004) Senescence-induced expression of cytokinin reverses pistil abortion during maize flower development. Plant J 38:910–922

Yu J, Ao G (1997) Expression of 10 kDa sulfur-rich prolamin gene of rice in potato. Acta Bot Sin 39:329–334

Zuo J, Niu QW, Moller SG, Chua NH (2001) Chemical-regulated, site-specific DNA excision in transgenic plants. Nat Biotechnol 19:157–161

Chapter 13
Transgenic Pyramiding for Crop Improvement

Bingliang Wan

Abstract In recent years, transgenic *pyramiding* or *stacking* technology has gradually developed with the rapid development of genetically modified (GM) crops. This technology has unique advantages compared with the transgenic technology of a single gene. Transgenic pyramiding technology can simultaneously modify several traits of a crop, particularly the metabolic pathways and yield traits that are usually controlled by multiple genes. A batch of second generation GM crops with stacked traits has been commercialized, which mainly include resistances to pest and herbicide, and exhibit great application potential of transgenic pyramiding breeding. The present stacked GM crops are developed mainly through the integrated use of vector-based pyramiding and molecular marker assisted selection (MAS) based crossing/backcrossing pyramiding. Advances of vector based pyramiding technologies have been achieved on large capacity vectors, multiple-gene assembling, plastid transformation, polyprotein expression system and combinatorial genetic transformation. The combination of transgenic pyramiding and MAS-based pyramiding will help to effectively pyramid more targeted genes together. The advance of single nucleotide polymorphism (SNP) discovery and detecting technology has greatly promoted the MAS-based pyramiding in crop breeding. The rapid development of commercialized stacked GM crops, plant metabolic engineering and crop improvement for disease resistance have proved the successful applications of transgenic pyramiding. The present review discusses the advances of transgenic pyramiding technology, the application of transgenic pyramiding in crop improvement and the prospects and challenges of transgenic pyramiding breeding.

Keywords Crop improvement • Gene pyramiding • GM crops • MAS • Molecular marker • Single nucleotide polymorphism • Transgene

B. Wan (✉)
Hubei Key Laboratory of Food Crop Germplasm and Genetic Improvement, Food Crop Research Institute, Hubei Academy of Agricultural Science, Hubei Collaborative Innovation Center for Grain Industry, Wuhan 430064, China
e-mail: ricewanbl@126.com

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_13

13.1 Introduction

Transgenic technology has great potential for genetic improvement of crops, in which foreign genes that crops do not possess can be transferred into a crop's genome. Foreign genes can come from animals, microorganisms, and other species. Gene expression in crops can also be regulated by transgenic technology. Thus, genetically modified (GM) crops can gain new agronomic traits such as resistances to abiotic and biotic stress. In addition, metabolic pathways can be modified by applying transgenic technology. Considering the unique characteristics of transgenic technology contrary to conventional breeding, the research and production of transgenic crops develop rapidly. The global hectarage of GM crops has been increasing every year since the first GM crop was introduced in 1996. According to the International Service for the Acquisition of Agri-biotech Applications (ISAAA) report (James 2013), GM crops reached 175.2 million ha in 2013 (100 times greater than in 1996), distributed across 27 countries and included more than 10 kinds of crops. Although GM crops are developing rapidly, transgenic technology has not yet achieved its full potential. Given the constraints of transgenic technology, many GM crops currently have simple transgenic traits controlled by one or two genes, and thus belong to the first-generation transgenic crops. Research on GM crops with stacked traits is still at an early stage, but has exhibited broad application prospects in crop improvement.

Stacked trait GM crops are superior to the single trait GM crops in various aspects. First, stacked trait GM crops can pyramid multiple genes controlling different traits into a single transgenic crop variety and simultaneously improve several traits in the crop. For example, GM maize Agrisure® Viptera™ 3111 (produced by Syngenta) contains stacked insect-resistance genes (*cry1Ab*, *vip3Aa20*, and *mcry3A*) and herbicide-resistance genes (*mepsps* and *pat*) (GM Approval Database 2004). This GM crop is endowed with resistance to *Lepidoptera* insects, *Coleoptera* insects, glufosinate and glyphosate. Second, the pyramiding of several genes encoding Bt (*Bacillus thuringiensis*) toxin proteins with two or more modes of action in a single variety is an effective strategy for insect-resistance management and can delay the occurrence of insects with Bt-toxin resistance (Bates et al. 2005; Ghimire et al. 2011; Zhao et al. 2003). Based on the requests of the Environmental Protection Agency in the USA, 20 % (in the Corn Belt) or 50 % (in the Cotton Belt) of a farmer's corn acreage should be planted with non-Bt corns as refuge acreage. If corn farmers plant the GM corns with several Bt genes, the refuge acreage can be reduced to 10 %, allowing farmers to grow either 10 or 40 % more Bt corns in the same area. Third, the transgenic pyramiding is suitable for the modification of biological traits controlled by multiple genes, such as metabolic pathways, protein complexes and quantitative traits. The modification of a single gene has no significant effect on these traits.

The focus of this chapter is on the technology for transgenic pyramiding and its applications in crop improvements. It aims to provide an easily-understandable overview of the latest developments of techniques involved in large-capacity vectors, multiple-gene assembling, plastid transformation, polyprotein expression systems, combinatorial genetic transformation and MAS-based gene pyramiding. Also, applications in commercialized GM crops, plant metabolic engineering, and crop improvements for resistances to biotic and abiotic stresses.

13.2 Technology for Transgenic Pyramiding Breeding

Three methods have been established to pyramid multiple transgenes; namely, multi-gene transformation, retransformation and sexual hybridization of two or more transgenic events (Fig. 13.1). The multi-gene transformation method can be further classified into two types: single-vector transformation and multi-vector co-transformation. For the single-vector transformation method, multiple genes are inserted within a single T-DNA, or multiple T-DNAs are included on a binary vector, the plant transformation of a single-vector is simple, but the assembly of multiple genes into a conventional vector is difficult owing to vector instability, the lack of unique restriction sites and the limited vector capacity. By this method, transgenes in the transgenic plants are tightly linked and co-integrated in the same loci in the plant genome, so they will not segregate in offspring. In multi-vector co-transformation methods, multiple genes are cloned on a different vector, respectively. Multiple strains carrying a different vector are used for simultaneous infection of plant tissues, or multiple genes in different vectors are simultaneously transferred to plant cells. The transgenes from the co-transformation are usually co-integrated at the same position of the transgenic plant chromosomes, which may consequently be inherited together in the progeny (Halpin 2005).

Retransformation is a valid gene pyramiding method, with which multiple transgenes can be sequentially introduced into a plant. For retransformation, every transformation requires a new and different selectable marker from the original event, which is crucial in this process. The drawback of few available

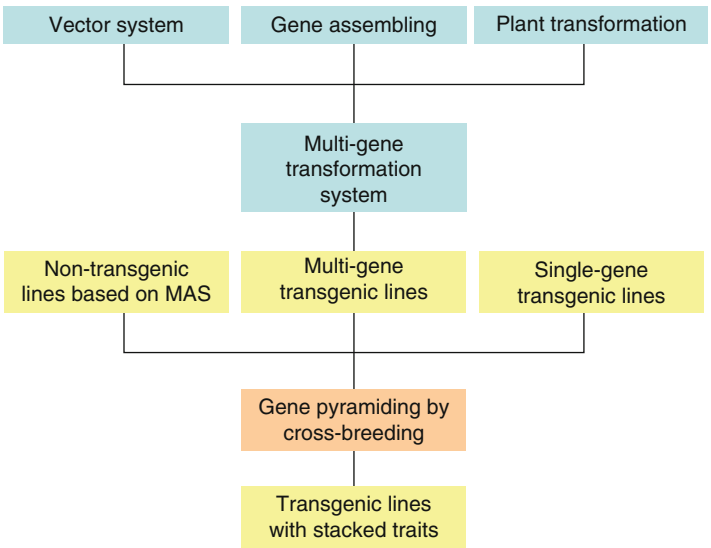

Fig. 13.1 Transgenic pyramiding methods in crop improvement

selective markers restricts the development of retransformation. The retransformant transgenes are randomly integrated and may segregate in the progeny. Sexual hybridization, which is flexible for the combination of transgenes but time-consuming, is a conventional breeding technology that pyramids transgenes from different events. Considering that the transgenes from different events are not linked and will segregate in the progeny, a large breeding population is required to obtain the plant with stacked genes.

MAS is an efficient technology for gene pyramiding in conventional breeding (Dokku et al. 2013; Hu et al. 2012; Tan et al. 2010). Generally, MAS-based pyramiding can be performed using two strategies: backcrossing and crossing (Fig. 13.2). The backcrossing is used for pyramiding genes into the genetic background of the recurrent parent to improve a few traits of the recurrent parent while the main traits of the recurrent parent remain unchanged. The crossing method is used for pyramiding genes into new genetic backgrounds through the genetic recombination of various parents to develop a new variety.

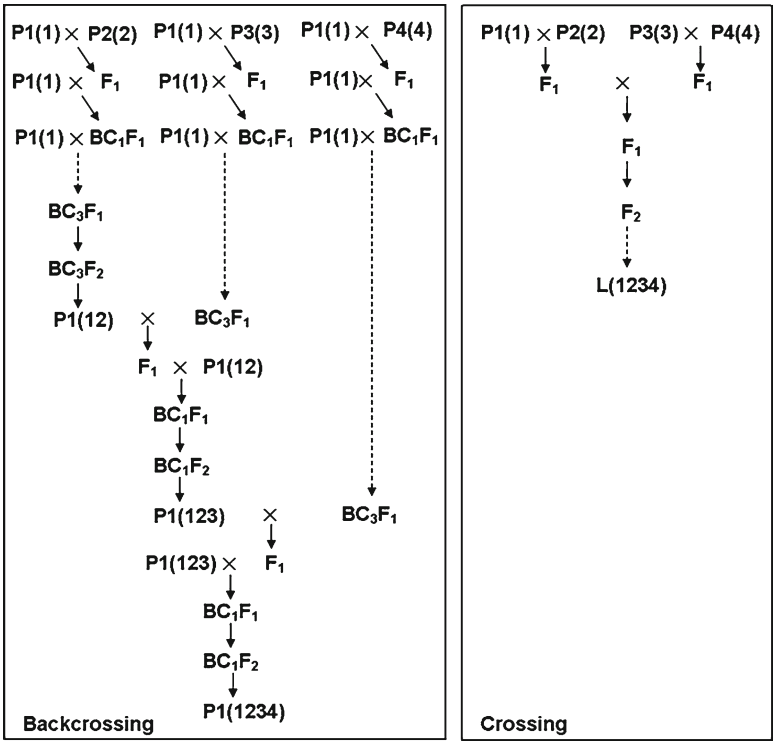

Fig. 13.2 Schemes of backcrossing/crossing pyramiding based on MAS. *P1(1)*, *P2(2)*, *P3(3)* and *P4(4)* represent four different parents carrying different genes (genes are indicated by the numbers in brackets). *P1(12)*, *P1(123)* and *P1(1234)* indicate the improved parent *P1(1)* carrying different pyramided genes. *L(1234)* indicates the new developed line carrying four pyramided genes

13.2.1 Transgenic Pyramiding Methods in Commercialized GM Crops

The pyramiding of transgenes in commercialized GM crops is mainly realized through gene stacking based on vector and sexual hybridization. The Genuity® SmartStax™ maize developed by Monsanto was derived from the stacking of four events (MON89034 × TC1507 × MON88017 × 59122) by conventional crossbreeding. Each event originated from the transformation of a vector with multiple genes. The event MON89034 was obtained from the transformation of vector PV-ZMIR254 containing *Lepidoptera* insect-resistance genes (*cry2Ab2* and *cry1A.105* from *Bacillus thuringiensis*). The event MON88017 was achieved from the transformation of vector PV-ZMIR39, which contains glyphosate-resistance gene (*cp4 epsps*) and *Coleoptera* insect-resistance gene (*cry3Bb1*). The event TC1507 was derived from the particle bombardment transformation of a linear fragment PHI8999A that comprises two expression cassettes of *Lepidoptera* insect-resistance gene (*cry1Fa*) and glufosinate-resistance gene (*pat*). The event 59122 contains *Coleoptera* insect-resistance, particularly corn rootworm-resistance genes (*cry34Ab1* and *cry35Ab1*) and glufosinate-resistance gene (*pat*). The Agrisure® Viptera™ 3111 maize developed by Syngenta, was also derived from four events (Bt11, GA21, MIR162, and MIR604) by conventional crossbreeding. The event Bt11 was obtained from the transformation of a vector with stacked genes (*cry1Ab* and *pat*), and has resistances to *Lepidoptera* insects and glufosinate herbicide. GA21 was acquired from the transformation of vector pDPG434 carrying glyphosate-resistance gene (*mepses*). The events, MIR162 and MIR604 were obtained from the transformation of vector pNOV1300 containing *Lepidoptera* insect-resistance gene (*vip3Aa20*) and pZM26 containing *Coleoptera* insect-resistance gene (*mcry3A*).

According to the examples provided, the transgenic pyramiding in commercialized GM crops with stacked traits is realized mainly through conventional crossbreeding of single events deriving from the vector transformation. Each event must be tested and approved before becoming commercially available, which is a complex, time-consuming, and arduous process. Considering that the single events used for crossbreeding have obtained approvals, the stacked products from the conventional crossbreeding do not involve new transformations, and no interaction effect among the stacked traits happens, the registration as a new GM product and new approvals are unnecessary for the stacked products from crossbreeding. Thus, this crossbreeding-based pyramiding can accelerate the commercialization of GM crops and reduce the research cost for GM crops. This accelerated test and approval procedure has become the standard in some countries (Taverniers et al. 2008).

In crop production, more stacked transgenes are not necessarily advantageous. More genes and traits in breeding mean a more complicated and extended breeding process. In addition, each stacked transgene is protected by intellectual property rights, thus more stacked transgenes mean greater cost for farmers who use them. Moreover, the required crop traits are not identical in different crop planting areas. With the crossbreeding method, the desired transgenic traits can be combined to

develop the transgenic products suitable to various needs of markets and crop productions. Each of four transgenic events (Bt11, GA21, MIR162 and MIR604) contain different transgenes, and the different combinations of these events developed into various GM varieties. Agrisure™ GT/CB/LL, obtained from the hybridization of events Bt11 and GA21 and contains stacked genes (*cry1Ab*, *pat*, and *mepsps*), providing resistances to *Lepidoptera* insect, glyphosate and glufosinate. Agrisure® Viptera™ 2100, from the combination of events Bt11 and MIR162 and with stacked genes (*cry1Ab*, *pat*, and *vip3Aa20*), endowing resistances to *Lepidoptera* and *Coleoptera* insects and glufosinate. Agrisure® Viptera™ 3110 was developed by stacking GA21 on the basis of Agrisure® Viptera™ 2100, adding a glyphosate-resistance gene (*mepsps*). Agrisure® Viptera™ 3100 was bred by stacking MIR604 on the basis of Agrisure® Viptera™ 2100, adding a *Lepidoptera* insect-resistance gene (*mcry3A*). Agrisure® Viptera™ 3111 was the result of four events (Bt11, GA21, MIR162 and MIR604) with five stacked genes (*cry1Ab*, *pat*, *vip3Aa20*, *mcry3A* and *mepsps*).

Pyramiding by conventional crossbreeding also has disadvantages. Transgenes in different events are not linked and have a tendency to segregate in subsequent breeding progenies. Moreover, transgene loci increase with increasing the crossing parents, resulting in the complex gene segregation. Therefore, a large and sufficient breeding population and more breeding generations are needed for the identification of plants with the stacked genes in offspring. In breeding processes, except for the target transgenic traits, other agronomic traits should also be selected. Thus, if the number of pyramided loci exceeds three, it is more difficult to obtain the plant that is homozygous for all transgene loci.

13.2.2 Development of Transgenic Pyramiding Technology

13.2.2.1 Large-Capacity Vectors

The length capacity of vectors for the inserted DNA fragments limits the assembling of multiple genes in a single vector. The maximum length of the inserted fragment of a conventional binary vector is about 50 kb. However, in actual practice, the assembly and transformation of a vector with an inserted DNA fragment of more than 30 kb is difficult. The binary bacterial artificial chromosome (BIBAC) and transformation-competent artificial chromosome (TAC) are two new kinds of vectors that can be used for genomic library construction and plant transformation. In contrast to conventional binary vector, BIBAC and TAC can contain an inserted DNA fragment of up to 200 kb, and are suitable for the transformation of large DNA fragments (Chang et al. 2003; Hamilton et al. 1999; Kong et al. 2006; Lin et al. 2003; Qu et al. 2003; Zhai et al. 2013). At present, these vectors are mainly used in the cloning and functional identification of large genome fragments.

Minichromosome is a newly developed gene expression platform for large DNA fragments. Minichromosome is a micro-artificial chromosome containing the struc-

ture and function unit of chromosomes, and exists as an episome in the cell period and cannot match host chromosomes, but can carry genes and transfer genetic information between generations (Goyal et al. 2009). The mammalian minichromosomes have been used as vector systems for gene therapy. However, the minichromosome construction was not successful in plants until the reports of telomere-mediated chromosomal truncation in maize in 2006 (Yu et al. 2006). Maize minichromosomes were obtained from A chromosomes and B chromosomes through telomere-mediated chromosome truncation technology (Yu et al. 2006, 2007). A 2.6 kb *Arabidopsis* telomere sequence was transformed into maize to cause chromosomal breakage at the site of integration, efficiently resulting in the truncated chromosomes. However, the truncated chromosomes are deficient, and the truncation of A chromosomes causes an obstacle for the growth and development of the plants that carry the deletion, because they lack essential genes. So the recovery of large truncated chromosomes is always difficult through genetic transformation, and moreover the truncated chromosomes could not be transmitted to the next generation for the gametophytic lethality at the haploid level. However, these problems can be overcome through polyploidization of transformants or by using polyploid plants as the target material for genetic transformation. The gametes of tetraploid plants will be diploid and the deficiency of truncated chromosomes would be compensated for by additional copies of the homolog. These problems do not exist in the truncated chromosomes from B chromosomes, because B chromosomes are basically inert, without any known active genes. Yu et al. (2007) demonstrated the transgene manipulation using minichromosomes. A transgenic plant carrying a 35S-l*ox66*-Cre expression transgenic cassette was crossed with a plant carrying the minichromosome, which contains the promoterless *lox71*-DsRed gene, and the recombination of *lox66* site with *lox71* site was successfully catalyzed by the Cre recombinase, resulting in the transfer of *DsRed* and *Cre* genes and the expression of *DsRed* gene. As the next-generation genetic vector, plant minichromosomes are in an initial development stage, and are only reported in several plants, including maize, rice and *Arabidopsis* (Murata et al. 2013; Xu et al. 2012a; Yu et al. 2006, 2007). Larger capacity for the insert DNA fragments of minichromosome can be expected. Although the exact maximum length is unknown, the length may reach the Mb level.

13.2.2.2 Multiple-Gene Assembling

Gene assembly is a limiting factor for multiple gene transformation. Conventional gene assembly is dependent on the specific restriction endonuclease sites. However, these sites usually cannot satisfy the gene assembly with the increase in the number of genes. So, a number of restriction endonuclease independent multiple-gene assembling technologies have been developed (Fig. 13.3). In theory, the most convenient way for multiple-gene assembly is seamless assembly or de novo synthesis of long chain DNA. At present, DNA fragments of 1–3 kb can be easily synthesized by conventional DNA synthesis technology, and then assembled into the long chain DNA by seamless assembly. The existing long chain DNA assembling technology includes

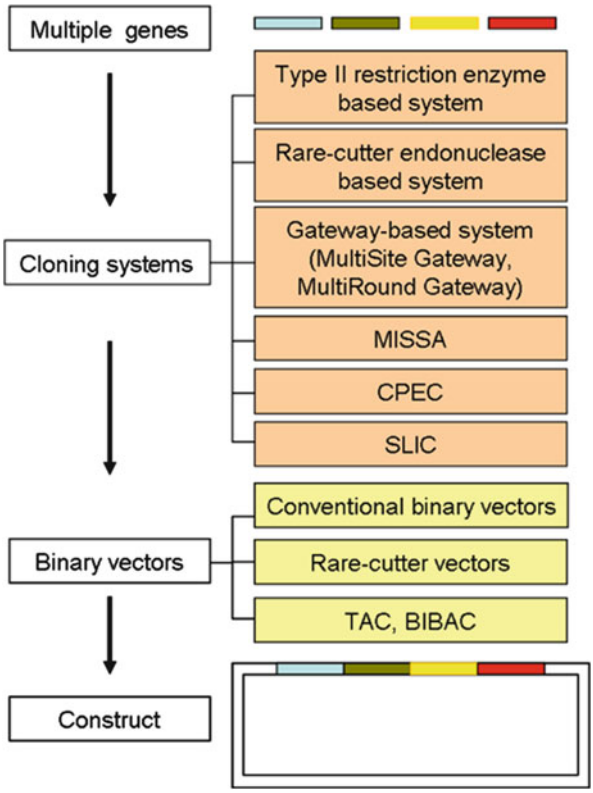

Fig. 13.3 Schemes of multiple-gene constructs for plant transformation

sequence-independent and ligation-independent cloning (SLIC), circular polymerase extension (CPEC), Gibson isothermal assembly, homologous recombination method (Gibson et al. 2008, 2009; Li and Elledge 2007, 2012; Quan and Tian 2009) and iterative assembly systems, such as GoldenBraid (Sarrion-Perdigones et al. 2011).

The CPEC method extends overlapping regions between the insert and vector fragments to form a complete circle plasmid using a polymerase extension mechanism, and can not only be used for the cloning of a single gene, but also be used for the cloning of libraries and metabolic pathways. Quan and Tian (2009) successfully assembled four genes of the metabolic pathway for synthesizing a biodegradable plastic material, poly (3HB-co-4HB) in *Escherichia coli* into the vector using CPEC. The CPEC reaction is completed in a single tube, and only needs 2–5 cycles, depending on the complexity of inserted fragments. The advantage of CPEC is convenience, efficiency and cost-effectiveness. The SLIC method assembles DNA fragments into vector through in vitro homologous recombination and single-strand annealing. The ssDNA overhangs in the insert and vector fragments are generated using exonuclease, and then these fragments can be assembled together by recombination in vitro. The SLIC method also allows an efficient and reproducible assembly of recombinant DNA with as many as 5–10 fragments simultaneously (Li and Elledge 2007, 2012).

Gibson isothermal assembly method assembles multiple overlapping DNA fragments by using exonuclease III and antibody-bound *Taq* DNA polymerase, which allows for one-step thermocycled in vitro recombination. All these technologies can assemble DNA fragments of up to 20 kb, and especially, Gibson isothermal assembly in *Saccharomyces cerevisiae* can even assemble DNA fragments of 0.5–1 Mb (Gibson et al. 2008, 2009; Li and Elledge 2012; Quan and Tian 2009).

At present, the popular recombinase-based vector systems for multi-gene assembly are Gateway-based systems, such as MultiSite Gateway (Sasaki et al. 2004) and MultiRound Gateway (Buntru et al. 2013). The Gateway system is based on the function of *Escherichia coli* bacteriophage lambda integrase and two sets of *att* sites. Two sets of recombination reactions are catalyzed by the *att*B × *att*P (BP) and *att*L × *att*R (LR) Clonase. BP Clonase is responsible for the recombination of *att*B sites with *att*P sites, and LR Clonase catalyzes the recombination of *att*L sites with *att*R sites. The *att*B sites are added to the two terminals of any interest DNA segment through PCR reaction with primers containing *att*B sites, and then the DNA segment is cloned into the donor vector with *att*P sites through BP reaction, resulting in an entry clone containing the insert DNA segment flanked with two *att*L sites. The insert in entry clone can be mobilized into any destination vector with *att*R site through the recombination of *att*L site with *att*R site. To facilitate multi-gene assembly, the Gateway variants, MultiSite Gateway and MultiRound Gateway were developed. Sasaki et al. (2004) assembled four DNA fragments into a single vector through MultiSite Gateway. By using six *att*B sites, four DNA segments were cloned into four donor vectors with corresponding specific *att*P sites through the BP reaction, creating four different entry clones, respectively; the four entry clones simultaneously reacted with the destination vector in the presence of LR clonase, resulting in the construct with the three inserts in a defined and oriented order (Sasaki et al. 2004). Compared with the MultiSite Gateway system, that only needs a single LR recombination reaction for multiple entry clones, the MultiRound Gateway system sequentially delivered multiple entry clones into a destination vector by multiple rounds of LR recombination reactions (Chen et al. 2006). These systems are becoming popular in multi-gene assembly because of the advantage of sequence independence, but they suffer from several drawbacks in that the assembly process is intricate, time-consuming and expensive. Chen et al. (2010) developed a multi-gene assembly system named MISSA (for multiple-round in vivo site-specific assembly), combining the Cre/*Lox*P recombinase system, phage λ site-specific recombinase system, and conjugal transfer of genes among bacteria together. MISSA consists of donor vectors and acceptor vectors, and the corresponding donor bacteria and acceptor bacteria. The gene transfer from a donor vector to an acceptor vector can be accomplished through in vivo conjugational transfer between the corresponding acceptor bacteria and donor bacteria. Therefore, target genes only need to be assembled into the donor vectors, and then multiple genes in the donor vectors can be transferred to the acceptor vectors individually through several rounds of conjugational transfer between the donor bacteria and the acceptor bacteria. Furthermore, the basic vectors of several acceptor vectors of this system are developed from BIBAC and TAC; thus, MISSA is not suitable only for the assembly of multiple-gene, but also for the assembly of large DNA fragment.

GoldenBraid is an iterative assembly system for multi-gene based on the use of a second type, IIS restriction enzymes. The second type IIS restriction enzyme recognizes asymmetric DNA sequences (the recognition site) and cleaves DNA at a defined position (the cleavage site) several base pairs away from the recognition site, no specific sequence requiring in the cleavage site, leaving a short overhang in the digested segments. The segments with complementing overhangs can be assembled together in the defined order, so the cleavage sites can be designed as the boundaries of DNA segments. In the GoldenBraid system, the parts (e.g. promoters, coding sequences, terminators, etc.) are created by PCR amplification, adding appropriate extensions to the primers to ensuring appropriate overhangs, and cloned into the entry clones flanked by the *Bsa* I sites. The digested part segments by *Bsa* I are assembled together with the digested destination plasmid containing a LacZ cassette flanked by the *Bsa* I sites in divergent orientation by incubating in a tube, creating a device (e.g. transcriptional units), where the *Bsa* I sites have disappeared. In order to assemble multiple devices together, another second type IIS restriction enzyme (e.g. Bsm BI) is added to the destination plasmids, so that *Bsa* I-assembled devices (the first order assembly) could similarly be assembled in second order destination plasmids. The GoldenBraid system consists of a set of four destination plasmids (pDGBs), All pDGBs contain a LacZ selection cassette flanked by four second Type IIS restriction sites (*Bsa* I, *Bsm* BI), but positioned in the inverted positions and orientations. The relative position of second type IIS restriction sites inside pDGBs inserts a double loop (*braid*) topology in the cloning design, so that the assemblies from first level become entry plasmids for second level assemblies and vice versa. Using this system, Sarrion-Perdigones et al. (2011) successfully assembled five devices in a T-DNA of binary vectors.

13.2.2.3 Plastid Transformation

By contrast with nuclear genes, plastid genes are organized as operons co-expressed as a single transcriptional unit. This gene organization can coordinate the expression of several genes in an operon. Thus, plastid transformation is particularly suitable for the study of metabolic pathways that are synergistically regulated by multiple genes. Bohmert-Tatarev et al. (2011) designed a plastid transformation system of tobacco (*Nicotiana tabacum*) for the production of renewable and biodegradable plastic polyhydroxybutyrate (PHB). Three genes in the PHB pathway were assembled in the *psbA* operon expression cassette to be transformed into the tobacco plastid genome. The BPH contents in T_0 and T_1 plants reached up to 18.8 and 17.3 % (dry weight) in leaf tissue samples, respectively. Homogentisate phytyltransferase, tocopherol cyclase, and γ-tocopherol methyltransferase are three key enzymes of tocochromanol pathway that provides tocopherols and tocotrienols, which are collectively referred as *vitamin E*. The genes that encode these three proteins were assembled in an operon construct to be transformed into the tomato chloroplast genome. The content of tocochromanol in transgenic plants is 10 times higher than that of the wild control (Lu et al. 2013).

Plastid transformation technology has been proved efficient in model plants, such as tobacco (Svab and Maliga 1993), tomato (Ruf et al. 2001) and potato (Sidorov et al. 1999); Kumar et al. (2004) also reported the high-efficient plastid transformation of cotton. However, plastid transformation is still not well developed in major crops like rice, maize, wheat and rape. Major challenges for the application of plastid transformation in crops include lower transformation efficiency, the difficulties to generate homoplastomic plants and to express transgenes in non-green plastids. In rice, two transformants were obtained on 100–120 bombarded plates, and no transplastomic plant of homoplasmy was obtained (Lee et al. 2006); The chloroplast transformants of oilseed rape were yielded at a frequency of 4 in 1000 bombarded cotyledon petioles (Hou et al. 2003). The transformation efficiency of wheat was estimated as 2 transplastomic plants per 42 bombarded plates of scutella and 1 transgenic line per 15 bombarded plates of immature inflorescences (Cui et al. 2011). However, up to 14 transplastomic lines of tobacco were obtained per bombarded leaf (Daniell et al. 2001). The vector inefficiency is one of the major obstacles of plastid transformation technology. The DNA fragments in vectors are transferred into the plastid genome through homologous recombination of the targeting sequence in vectors with the homologous sequence in plastids. So, the specificity of targeting sequence in vectors may be the important element influencing the transformation efficiency (Skarjinskaia et al. 2003). In addition, the explant, bombardment parameter and marker gene are also important for transformation efficiency and homoplasmy. The efforts for improving the plastid transformation of major crops should be concentrated on these respects.

13.2.2.4 Polyprotein Expression System

The polyprotein expression system based on linker peptide is another method of coordinating multi-gene expression. In this system, multiple genes linked by the linker peptide sequences are regulated by a promoter as a single open reading frame. After translation, the polyprotein is cleaved into its constituent protein units through the self-splicing function of linker peptide. The linker peptide 2A is a short peptide of 20 amino acids from foot-and-mouth disease virus, and mediates co-translational cleavage at its own carboxy terminus by an apparently enzyme-independent type of reaction. The efficiency of peptide 2A polyprotein expression system has been proven in human, mammalian, fungus and yeast (de Felipe et al. 2003; Fang et al. 2005; Kim et al. 2011; Ryan and Drew 1994; Suzuki et al. 2000), and also has been tested in plants. Ma and Mitra (2002) incorporated *CAT* and *GUS* gene into a single open reading frame with a copy of the FMDV 2A protein gene, and transformed tobacco; the freed CAT and GUS proteins were detected in transgenic plants with the cleavage efficiency ranging from 80–100 %. el Amrani et al. (2004) demonstrated that each protein from a 2A-polyprotein in plant cells only has its own independent targeting signals, and can correctly target various subcellular locations via either co-translational or posttranslational mechanisms. LP4 is another linker peptide, which comes from *Raphanus sativus* seeds, and has a recognition site and is

cleaved by a protease. The advantage of LP4 as a polyprotein linker exists in that the polyprotein is completely cleaved with few redundant amino acids left at two terminals of proteins. LP4/2A is a hybrid peptide that contains the first 9 amino acids of LP4 and 20 amino acids from 2A, and can efficiently produce the individual proteins that can accurately target to the respective cellular compartments in *Arabidopsis* (François et al. 2002, 2004). Sun et al. (2012) demonstrated that the expression level of the two genes linked by LP4/2A was higher than those linked by 2A in tobacco. Several polyprotein transformations in major crops were reported in recent years. Two antimicrobial protein encoding genes, *Dm-AMP1* and *Rs-AFP2*, linked by LP4 peptide sequence, were introduced into rice, and the proteins Dm-AMP1 and Rs-AFP2 were detected in the leaf extracts of transgenic plants; the disease resistance against *Magnaporthe oryzae* and *Rhizoctonia solani* of transgenic plants were improved by 90 and 79 %, respectively, as compared to the untransformed plants (Jha and Chattoo 2009). Ha et al. (2010) successfully obtained the transgenic rice accumulating 1.3 μg/g total carotenoids in seed endosperms through the transformation of two carotenoid biosynthetic genes, *Psy* and *Crtl*, linked by 2A peptide.

13.2.2.5 Combinatorial Genetic Transformation

Metabolic pathways usually involve multiple genes. Therefore, the function of every gene and gene combination in the entire metabolic pathway should be thoroughly understood. This knowledge is beneficial in searching for the limiting gene and gene combination when modifying metabolic pathways by metabolic engineering technology. White maize variety M37W lacks carotenoids in the endosperm because of the absence of phytoene synthase (PSY1). To accumulate carotenoids in maize, the carotenoid-synthesis pathways should be reconstructed in the maize endosperm. Five genes for carotenoid synthesis, namely, *Zmpsy1* (*Zea mays* phytoene synthase 1), *PacrtI* (*Pantoea ananatis* phytoene desaturase), *Gllycb* (*Gentiana lutea* lycopene γ-cyclase), *Glbch* (*Gentiana lutea* γ-carotene hydroxylase), and *ParacrtW* (*Paracoccus* γ-carotene ketolase), were assembled to five vectors, respectively, and co-transformed into maize variety M37W using bombardment technology. A series of transgenic lines containing one to five genes were obtained. The analysis of carotenoid content in the seeds of these transgenic lines indicated that all transgenic lines containing *Zmpsy1* gene (except the transgenic line Ph-2 without *Zmpsy1* gene) could synthesize carotenoid, but varying in the carotenoid content. The line Ph-3, which contains *Zmpsy1* and *PacrtI*, had the highest carotenoid and β-carotene content, suggesting that the combination of *Zmpsy1* and *PacrtI* are the most effective for the modification of the carotenoid synthesis pathway in maize (Zhu et al. 2008). This transformation method is known as the combinatorial genetic transformation designed to generate the population of transgenic plants containing random transgene combinations through one transformation experiment; and is appropriate for systematic analysis of metabolic pathway and other genetic networks that require the coordinated expression of multiple genes (Farre et al. 2012).

13.2.2.6 New Development of MAS Technology

The successful use of MAS technology depends on three factors: functionally clarified genes, tightly linked molecular markers and cost-effective identification technology of molecular markers. Single nucleotide polymorphism (SNP) markers are rapidly becoming the choice for genetic and breeding applications because of abundant polymorphism and automated detecting. However, the high cost and low efficiency of SNP discovery and detecting had been the main hurdles, especially for species with no reference sequence, until the next-generation sequencing (NGS) technology emerged. These NGS methods include reduced-representation sequencing using reduced-representation libraries (RRLs) or complexity reduction of polymorphic sequences (CRoPS), restriction-site-associated DNA sequencing (RAD-seq) and low coverage genotyping (Davey et al. 2011). In recent years, a large number of SNP markers have been developed in many crops using NGS technology, moreover some crops were resequenced (Davey et al. 2011; Seeb et al. 2011; Xu et al. 2012b). SNP markers have been widely used in genetic mapping, association analysis, genotyping, MAS-based breeding, and so on (Chen et al. 2011; Ferguson et al. 2012; Tiwari et al. 2014; Yang et al. 2012). Many intragenic SNPs are the source of genotypic mutation, and reflect the allelic variations. These SNP markers are designated as genic molecular markers (GMMs) or functional markers, and greatly facilitate the pyramiding of alleles as compared with SSR markers. In the base sequence of rice grain size gene *GS3*, the C to A mutation in the second exon is associated with enhanced grain length. A cleaved amplified polymorphic sequence (CAPS) based on C-A polymorphism was developed to identify the *GS3* alleles controlling various grain sizes in rice (Fan et al. 2009).

13.3 A Case of Pyramiding of Five Genes for Multiple Resistances in Rice

Rice production is affected by various environmental stresses, including diseases (rice blast), insects (rice leaf folder, stem borer and brown plant hopper) and weeds. So, the objective is to pyramid resistance genes against disease, insect and herbicide into a single variety, developing a variety carrying multiple resistances.

The rice restorer line R022 is an elite rice restorer line that contains brown planthopper resistance genes *Bph14* and *Bph15*. A backcrossing breeding scheme was designed to improve the resistance of R022 against rice blast, leaf folder, stem borer and herbicide. In this scheme, R022 was used as the recurrent parent. The transgenic rice T1C-19 was used as the donor of *cry1C* (encoding *Bacillus thuringiensis* toxin protein) and glufosinate resistance gene *bar* (encoding phosphinthricin acetyltransferase). T1C-19 was developed through the transformation of a construct containing *cry1C* and *bar* in the binary vector plasmid pC-1C* (Tang et al. 2006). The rice restorer line R2047 was used as the donor of rice blast resistance gene *Pi9*. Five genes, *cry1C*, *bar*, *Pi9*, *Bph14* and *Bph15* were stepwise pyramided into R022 through MAS-based crossing.

The F_1 generation was obtained by crossing between female parent R022 and male parent T1C-19, followed by three consecutive backcrosses using R022 as recurrent parent and three selfings to obtain BC_3F_4 lines. In each generation of backcrossing, the herbicide glufosinate was used for the identification of *bar* gene, the co-dominant molecular markers were used for the identification of genes, *cry1C*, *Pi9*, *Bph14* and *Bph15*. Finally, the line W425 in BC_3F_4 was identified according to their agronomic traits and genetic backgrounds, and was named KR022. KR022 was confirmed to contain four homozygous genes, *cry1C*, *bar*, *Bph14* and *Bph15* through glufosinate identification and PCR assay, and showed the strong resistance against rice leaf folder and stem borer, brown planthopper and glufosinate. No damage caused by leaf folder, stem borer and glufosinate was observed in KR022 and its hybrid rice in two years of field study, while the control R022 suffered serious damage by leaf folder and stem borer, and thoroughly withered in the glufosinate treatment. KR022 inherits the brown planthopper resistance from R022 (Wan et al. 2014).

Similarly, rice blast resistance genes *Pi9* from R2047 was pyramided into the R022 background through backcrossing. In the BC_3F_1, the plants containing homozygous *Bph14* and *Bph15*, heterozygous *Pi9* were identified to pollinate KR022 plants to produce F_1, and then the F_1 plants were backcrossed by KR022 and followed by three successive selfings. Finally, the line carrying five genes, *Pi9*, *cry1C*, *bar*, *Bph14* and *Bph15*, was selected in BC_3F_3. The improved line showed the enhanced blast resistance (resistance score of 3) as compared with KR022 (resistance score of 9).

13.4 Application of Transgenic Pyramiding Breeding

13.4.1 Commercialized Transgenic Pyramiding Breeding

In recent years, stacked traits have rapidly increased in global GM crops. A record of 43.7 million ha of stacked traits was grown globally in 2012, which is equivalent to 26 % of the global 170 million ha of GM crops in 2012 and 7.5 times higher than the 5.8 million ha of stacked traits in 2003. Stacked traits mainly exist in GM maize and cotton. In 2012, 40 million ha of GM maize were stacked, contributing 91.5 % to 43.7 million ha of stacked traits, and stacked GM cottons were 3.7 million ha (Table 13.1). Table 13.2 lists some of the worlds commercialized GM maize and cotton with stacked traits.

At present, the stacked traits of GM crops are mainly insect resistance and herbicide resistance. Insect resistance primarily comes from the *Bacillus thuringiensis* genes encoding toxic proteins, such as *cry1Ab*, *cry1Ac*, *cry2Ae*, *vip3A (a)*, *cry1F*, *cry3A*, *mcry3A*, *cry34Ab1* and *cry35Ab1*. Herbicide resistance comes from the glyphosate-resistance gene (*epsps*) encoding 5-enolpyruvylshikimate-3-phosphate synthase, the glufosinate-resistance genes (*pat* and *bar*) encoding phosphinothricin N-acetyltransferase.

Several stacked GM varieties have been introduced. Agrisure® Viptera™ 3111 maize, bred by Syngenta, possessed genes, *cry1Ab*, *vip3Aa20*, *mcry3A, pat* and *epsps. cry1Ab* and *vip3Aa20* offer the resistance to *Lepidoptera* insect aboveground; *mcry3A* offers the resistance to *Coleoptera* insect, particularly corn rootworm pest underground. The pyramiding of *cry1Ab*, *vip3Aa20* and *mcry3A* broadens the resistance spectrum of Agrisure® Viptera™ 3111. According to the data published by Syngenta (http://www.syngenta.com/), Agrisure® Viptera™ 3111 can resist 14 kinds of aboveground and underground insects.

Two modes of resistance to herbicides are offered by *pat* and *epsps* gene, and the pyramiding of these two genes provides a more flexible herbicide use of glufosinate and glyphosate. Agrisure® Viptera™ 3111 has gained the approvals for food in eight countries or districts, for feed in five countries, and for cultivation in two countries. Genuity® VT Triple Pro™ maize containing *epsps*, *cry1A.105*, *cry2Bb2* and *cry3Bb1*, developed by Monsanto, is endowed with resistances to glyphosate herbicide and *Lepidoptera* and *Coleoptera* insect. Genuity® VT Triple Pro™ has been approved for food use in ten countries or districts, for feed use in eight countries or districts, and for cultivation in five countries. Genuity® SmartStax™ maize was developed by Monsanto and Tow AgroSciences and possesses eight effect genes, including *cry1A105*, *cry2Ab1*, *cry3Bb1*, *cry1Fa2*, *cry34Ab1*, *cry35Ab1*, *epsps* and *pat*. These genes confer Genuity® SmartStax™ the resistances against *Lepidoptera* and *Coleoptera* insects and the herbicide glyphosate and glufosinate. Genuity® SmartStax™ has been approved for food or feed cultivation in eight countries or districts. Widestrike™ Roundup Ready Flex™ Cotton, jointly developed by Monsanto and Dow AgroSciences, possesses four effect genes, including *epsps, cry1F*, *cry1Ac* and *bar*, which endow Widestrike™ Roundup Ready Flex™ with the resistances against *Lepidoptera* insect and the herbicide glyphosate and glufosinate. Widestrike™ Roundup Ready Flex™ has obtained approvals for food or feed cultivation in four countries.

Table 13.1 Global area of stacked GM crops 2003–2012 (million ha)

Year	Area (million ha)	GM crops (%)
2003	5.8	8.0
2004	6.8	9.0
2005	10.1	11.0
2006	13.1	13.0
2007	21.8	17.4
2008	26.9	21.5
2009	28.7	21.4
2010	32.2	22.0
2011	42.2	26.0
2012	43.7	26.0

Data source: ISAAA briefs
http://www.isaaa.org/resources/publications/briefs/

Table 13.2 Summary of stacked GM crops with approvals

Trade name	Events	Effect genes	Countries/regions of approvals
Cotton			
WideStrike™ Cotton	281-24-236×3006-210-23 (MXB-13)	*pat*, *cry1F*, *cry1Ac*	Australia, Brazil, Costa Rica, European Union, Japan, Mexico, New Zealand, South Korea
WideStrike™ Roundup Ready™ Cotton	3006-210-23×281-24-236×MON1445	*bar*, *epsps*, *cry1F*, *cry1Ac*	Japan, Mexico, South Korea
VIPCOT™ Roundup Ready Flex™ Cotton	COT102×COT67B× MON88913	*vip3A(a)*, *epsps*, *cry1Ab*	Costa Rica
GlyTol™ Liberty Link™	GHB614×LLCotton25	*epsps*, *bar*	Brazil, Japan, Mexico, South Korea
VIPCOT™ Cotton	COT102×COT67B	*vip3A(a)*,*cry1Ab*	Costa Rica
Glytol™ x Twinlink™	GHB614×T304-40×GHB119	*bar*, *epsps*, *cry2Ae*, *cry1Ab*	Brazil, Mexico, South Korea
Fibermax™ Liberty Link™ Bollgard II™	LLCotton25×MON15985	*bar*, *cry2Ab2*, *cry1Ac*	Australia, Japan, Mexico, New Zealand, South Korea
Roundup Ready™ Bollgard II™ Cotton	MON15985×MON1445	*epsps*, *cry2Ab2*, *cry1Ac*	Australia, Costa Rica, European Union, Japan, Mexico, New Zealand, Philippines, South Korea
Roundup Ready™ Flex™ Bollgard II™ Cotton	MON88913×MON15985	*epsps*, *cry2Ab2*, *cry1Ac*	Australia, Brazil, Colombia, Costa Rica, Japan, Mexico, New Zealand, Philippines, South Africa, South Korea
TwinLink™ Cotton	T304-40×GHB119	*bar*, *cry2Ae*, *cry1Ab*	Brazil, Canada, South Korea
Maize			
Trade Name	Events	Effect genes	Countries/Regions of approvals
Agrisure® Duracade™ 5122	5307×MIR604×Bt11×TC1507×GA21	*ecry3.1Ab*, *mcry3A*, *cry1Ab*, *pat*, *cry1Fa2*, *epsps*	Japan, Taiwan

Agrisure® Duracade™ 5222	5307 × MIR604 × Bt11 × TC1507 × GA21 × MIR162	*ecry3.1Ab*, *mcry3A*, *cry1Ab*, *pat*, *cry1Fa2*, *epsps*, *vip3Aa20*	Japan
Herculex™ RW Roundup Ready™ 2	59122 × NK603	*epsps*, *cry34Ab1*, *cry35Ab1*, *pat*	Canada, European Union, Japan, Mexico, Philippines, South Africa, South Korea, Taiwan, Turkey
Agrisure® 3122	BT11 × 59122 × MIR604 × TC1507 × GA21	*cry1Ab*, *cry1Fa2*, *pat*, *epsps*, *mcr3A*, *cry34Ab1*, *cry35Ab1*	Canada, Japan, Mexico, Philippines, Taiwan, South Korea
Agrisure® Viptera™ 3110	Bt11 × MIR162 × GA21	*cry1Ab*, *vip3Aa20*, *pat*, *epsps*	Argentina, Brazil, Canada, Colombia, Japan, Mexico, Philippines, South Africa, South Korea, Taiwan, Uruguay
Agrisure® Viptera™ 3100	BT11 × MIR162 × MIR604	*cry1Ab*, *pat*, *mcry3A*, *vip3Aa20*	Japan
Agrisure® Viptera™ 3111, Agrisure® Viptera™ 4	Bt11 × MIR162 × MIR604 × GA21	*cry1Ab*, *pat*, *mcry3A*, *vip3Aa20*, *epsps*	Argentina, Canada, Colombia, Japan, Mexico, Philippines, South Africa, South Korea, Taiwan
Roundup Ready™ YieldGard™ maize	GA21 × MON810	*cry1Ab*, *epsps*	European Union, Japan, Philippines, South Africa, South Korea
YieldGard™ Plus with RR	MON863 × MON810 × NK603	*cry1Ab*, *cry3Bb1*, *epsps*	Canada, Colombia, European Union, Japan, Mexico, Philippines, South Africa, South Korea, Taiwan
Genuity® VT Triple Pro™	MON89034 × MON88017	*epsps*, *cry1A105*, *cry2Ab1*, *cry3Bb1*	Argentina, Brazil, Canada, European Union, Japan, Mexico, Paraguay, Philippines, South Korea, Taiwan
Genuity® SmartStax™	MON89034 × TC1507 × MON88017 × 59122	*epsps*, *cry1A105*, *cry2Ab1*, *cry3Bb1*, *cry1Fa2*, *cry34Ab1*, *cry35Ab1*, *pat*	Canada, Colombia, European Union, Japan, Mexico, South Africa, South Korea, Taiwan
Optimum™ Intrasect Xtreme	TC1507 × 59122 × MON810 × MIR604 × NK603	*cry1Fa2*, *epsps*, *pat*, *cry34Ab1*, *cry35Ab1*, *cry1Ab*, *mcry3A*	Canada, Japan, Mexico, South Korea, Taiwan

Source: ISAAA's GM approval database

13.4.2 Application of Transgenic Pyramiding Breeding in Metabolic Engineering

Although plant metabolic engineering is still at an early stage and has not been widely applied in commercialized GM crops, more successful cases of plant metabolic engineering research are reported. Table 13.3 lists a few engineered plants from the transgenic pyramiding breeding in recent years, in which some have potential for commercial application.

13.4.2.1 Polyunsaturated Fatty Acid Synthesis

Very long chain polyunsaturated fatty acids (VLCPUFAs) are nutritional substances with important function in human health (SanGiovanni and Chew 2005). Eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) are two of the most important VLCPUFAs for the human body. These two fatty acids occur primarily in ocean algae, and accumulate in the fat of deep-sea fishes. The commercial VLCPUFA are mainly refined from deep-sea fishes, but the deep-sea fish supply has been limited for overfishing, and thus cannot satisfy the market demand. Oilseeds are easily obtained with high yields and generally contain unsaturated fatty acids, such as oleic acid (OA), linoleic acid (LA), and α-linolenic acid (ALA). Oilseeds can be used as substrates for VLCPUFAs synthesis, but lack the capacity of elongation/unsaturation. Thus, to synthesize VLCPUFAs, biologists have tried to reconstruct its synthesis pathway in oilseeds through metabolic engineering.

In 2004, VLCPUFA synthesis pathway was reconstructed in *Arabidopsis thaliana* through the transformation of genes encoding a Δ^9-elongase from *Isochrysis galbana*, a Δ^8- desaturase from *Euglena gracilis,* and a Δ^5- desaturase from *Mortierella alpinato*. EPA successfully accumulated for the first time in the engineering of *Arabidopsis thaliana* (Qi et al. 2004). Subsequently, Abbadi et al. (2004) transformed tobacco and linseed with genes encoding a Δ^6-fatty acid desaturase from *Physcomitrella patens*, a Δ5-fatty acid desaturase from *Phaeodactylum tricornutum*, and a Δ6- prolonged desaturase from *Physcomitrella patens*, and detected approximately 0.5 % EPA in the transgenic seed oils. DHA was synthesized for the first time in seed oils in 2005. *Arabidopsis thaliana* was transformed with a Δ6/Δ5 desaturase from zebrafish. A transgenic plant accumulating 1.6 % AA and 3.2 % EPA was selected to be retransformed with genes encoding a Δ4-desaturase and a Δ5- desaturase from *Pavlova salina*, and 0.2–0.5 % DHA was detected in transgenic seed oils (Robert et al. 2005). These studies demonstrated the feasibility of DHA and EPA synthesis in engineering plants, but their yields should be improved. Wu et al. (2005) reported the transgenic production of significant amounts of EPA and AA in *Brassica juncea* by stepwise metabolic engineering strategy. Through a series of transformations with increasing the number of transgenes encoding desaturase and prolonged desaturase from different organisms, the transgenic plants were demonstrated to produce VLCPUFAs with up to 25 % AA and 15 % EPA of total seed fatty acids.

Table 13.3 Some cases of plant metabolic engineering through multi-gene pyramiding in recent years

Metabolic pathway	Plant	Stacked genes	Products	References
2C-methyl-D-erythritol 4-phosphate (MEP) pathway	Peppermint (*Mentha*×*piperita* L.)	2	up to 78 % oil yield increase	Lange et al. (2011)
Carotenoid synthesis	Maize (*Zea mays* L.)	5	142-fold increases in total carotenoid (156.14 μg/g DW)	Zhu et al. (2008)
	Canola (*Brassica napus* L.)	7	60–190 μg/g (FW) ketocarotenoids	Fujisawa et al. (2009)
	Lilium (*Lilium*×*formolongi*)	7	102.9 and 135.2 μg/g (FW) total carotenoid	Azadi et al. (2010)
	Tomato (*Solanum lycopersicum*)	2	3.12 mg/g astaxanthin in leaves, 16.1 mg/g esterified astaxanthin in fruits	Huang et al. (2013)
Fatty acid synthesis	Maize (*Zea mays* L.)	2	25.23 % increase in seed oil content	Zhang et al. (2010)
	Canola (*Brassica napus* L.)	2	20 % increase in seed oil content	Tan et al. (2011)
Polyunsaturated fatty acid synthesis	Arabidopsis (*Arabidopsis thaliana*)	7	13.2 % EPA of total seed fatty acids	Ruiz-Lopez et al. (2013)
	Arabidopsis (*Arabidopsis thaliana*)	7	11.0 % DHA of total seed fatty acids	Ruiz-Lopez et al. (2013)
	Canola (*Brassica napus* L.)	7	12 % DHA in seed oil	Petrie et al. (2012)
	Ethiopian mustard (*Brassica carinata*)	2	25 % EPA of total fatty acids	Cheng et al. (2010)
	Camelina *(Camelina sativa* L.)	5	24 % EPA of total fatty acids	Ruiz-Lopez et al. (2014)
	Camelina *(Camelina sativa* L.)	7	11 % EPA and 8 % DHA of total fatty acids	Ruiz-Lopez et al. (2014)
	Soybean (*Glycine max* (L.) Merr)	3	low linolenic acid soybean oil	Flores et al. (2008)
Starch synthesis	Barley (*Hordeum vulgare* L.)	3	amylose-only starch granules	Carciofi et al. (2012)
	Maize (*Zea mays* L.)	6	2.8–7.7 % increase in the endosperm starch content and a 37.8–43.7 % increase in the proportion of amylose	Jiang et al. (2013)
Tocochromanol synthesis	Maize (*Zea mays* L.)	2	up to 3 times increase of γ-tocopherol	Naqvi et al. (2011)
	Tomato (*Solanum lycopersicum*)	3	10 times increase of tocochromanol	Lu et al. (2013)
Vitamin synthesis	Maize (*Zea mays* L.)	4	59.32 μg/g (DW) β-carotene, 106.94 μg/g (DW) ascorbate, 1.94 μg/g (DW) folate	Naqvi et al. (2009)

Moreover, DHA synthesis pathway was reconstituted in plant seeds through adding genes in Δ4 pathway and successfully synthesized DHA despite of 0.2 % of total fatty acids only. To overcome the poor DHA synthesis, Petrie et al. (2012) designed a transgenic synthesis pathway from OA to DHA, which involved seven genes encoding desaturases and prolonged enzymes, including the Δ^{12}-desaturase from *Lachancea kluyveri*, the Δ^{15}-/ω^3-desaturase from *Pichia pastoris*, the Δ^6-desaturase from *Micromonas pusilla*, the Δ^5-elongases from *Pyramimonas cordata*, and the Δ^5- and Δ^4-desaturases from *Pavlova salina*. These seven genes were transformed into *Arabidopsis thaliana*, and DHA averaged 13.3 % in seed oils of transgenic plants. This content level, even exceeds the DHA content in common commercial fish oil, and thus satisfies the demand for the commercialized VLCPUFA product.

13.4.2.2 Carotene Synthesis

Vitamin A is an indispensable nutrient, and has an important physiological function in maintaining the normal growth and development of the human body. Vitamin A deficiency can induce many diseases, such as nyctalopia, and still remains a problem in the poor developing countries (WHO 2009). Vitamin A can be obtained from animal foods as retinol, and obtained from plant and animal foods as carotene that can be transformed into pro-vitamin A.

Rice is the primary world grain crop. However, rice cannot supply vitamin A to the human body because of the lack of synthesis of pro-vitamin A in the edible parts of rice. Thus, biologists attempted to reconstruct the carotene synthesis pathway in rice by overexpressing key genes in the carotene synthesis pathway in rice endosperms. Ye et al. (2000) used three genes encoding the key enzymes in β-carotene synthesis, namely phytoene desaturase (*crtI*) from *Erwinia uredovora*, phytoene synthase (*psy*) from daffodil, and lycopene β-cyclase (*lcy*) from *Narcissus pseudonarcissus*. These genes were introduced into rice under the control of an endosperm-specific glutenin promoter Gtl by double-vector co-transformation. The transgenic rice has yellow endosperms for the accumulation of β-carotene, and is called Golden Rice. However, the β-carotene content is only 1.6 μg/g in this Golden Rice. To increase the β-carotenoid content in transgenic rice endosperms, genes encoding phytoene synthases, the limiting step for β-carotene accumulation, were tested from different species. A maize phytoene synthase encoding gene (*psy*) that substantially increases carotenoid accumulation in a model plant system was identified to transform rice combined with the *Erwinia uredovora* carotene desaturase encoding gene (*crtI*) to develop Golden Rice 2. As a result, the total carotenoids in the endosperm of Golden Rice 2 increased up to 23-fold (maximum 37 μg/g) compared with the original Golden Rice, and β-carotene has a preferential accumulation (Paine et al. 2005). According to the American RDA, 100 g of Golden rice 2 consumed per day can satisfy 55–70 % of the vitamin requirements of an adult male (Tang et al. 2009).

The metabolic engineering of carotenoids has also been successful in potato and maize. A mini-pathway of β-carotene synthesis was reconstructed in potato through expressing three genes encoding phytoene synthase (CrtB), phytoene desaturase

(CrtI), and lycopene beta-cyclase (CrtY) from *Erwinia* under the control of tuber-specific promoter. Carotenoid largely accumulated in the transgenic potato tuber. The carotenoid content in the tuber increased 20-fold, and reached 114 μg/g dry weight, and the β-carotene content increased 3600-fold and reached 47 μg/g dry weight (Diretto et al. 2007). The overexpression of *crtB* and *crtI* from *Erwinia* under the control of the super γ-zein promoter in maize also resulted in an increase of total carotenoids of up to 34-fold with a preferential accumulation of β-carotene in the transgenic plants (Aluru et al. 2008). Naqvi et al. (2009) developed an elite transgenic corn containing four transgenes for the metabolism of β-carotene, ascorbate and folate, including two genes in the pathways of the β-carotene synthesis (*psy1* from maize controlled by the wheat LMW glutenin promoter and *crtI* from *Pantoea ananatis* under the control of barley D-hordein promoter), one gene involved in the synthesis of ascorbate (*dhar* encoding rice dehydroascorbate reductase under the control of barley D-hordein promoter), and one gene involved in the metabolism of folate (*folE* encoding *Escherichia coli*. GTP cyclohydrolase under the control of barley D-hordein promoter). The contents of β-carotene, ascorbate, and folate in the seeds of transgenic maize increased up to 169-fold (59.32 μg/g DW), 6-fold (106.94 μg/g DW) and 2-fold (1.94 μg/g DW) compared with the wild control, respectively.

13.4.2.3 Starch Synthesis

Starch composes the main edible part of food crops, and is the main human energy source. Starch in plants exists in two forms: amylose and amylopectin, those contents and proportion involve both crop yield and palatability. Starch synthesis is a complex pathway controlled by many genes encoding ADP-glucose pyrophosphorylase (AGP), granule-bound starch synthase (GBSS), soluble starch synthase (SS), starch branching enzyme (SBE), starch debranching enzyme (DBE), etc. The expression of two starch branching enzymes, SBE IIa and SBE IIb were suppressed in barley through RNAi-mediated silencing technology. The transgenic lines, where both SBE IIa and SBE IIb expression were reduced by >80 %, was observed with a high amylose phenotype (>70 %) (Regina et al. 2010). All genes encoding starch branching enzymes (SBE I, SBE IIa, SBE IIb) were further simultaneously suppressed in barley, resulting in the production of amylose-only starch granules in the endosperm of transgenic lines. In these lines, a very high content of resistance starch (RS) (65 %) was observed, which is 2.2-fold higher than control (29 %) (Carciofi et al. 2012). In maize, the expression of six genes involving starch synthesis was modified, these genes include the overexpressed *Bt2*, *Sh2*, *Sh1* and *GbssIIa* (to enhance the activity of sucrose synthase, AGPase and granule-bound starch synthase) and the suppressed *SbeI* and *SbeIIb*. The transgenic plants expressing all six genes showed a 2.8–7.7 % increase in the endosperm starch content and a 37.8–43.7 % increase in the proportion of amylose. Moreover, the 100-grain weight and ear weight of transgenic plants had up to 20.1–34.7 % and 13.9–19.0 % increases, respectively (Jiang et al. 2013).

13.4.3 Gene Pyramiding for Resistances against Disease, Drought and Salt

Disease resistance improvement of crops may be the most successfully used field of gene pyramiding. A number of works have been done on this aspect. Since the 1990s, the International Rice Research Institute (IRRI) has developed a series of rice near-isogenic lines (NILs) (named IRBB) carrying stacked various rice bacterial blight (BB) resistance genes (Huang et al. 1997). These NILs have been widely used in rice breeding programs. Recently, IRBB60 (*Xa4*/*xa5*/*xa13*/*Xa21*) was used as the donor of BB resistance genes to improve the BB resistance of an elite rice cultivar in India (Dokku et al. 2013). Jin 23B is a cytoplasmic male sterile (CMS) rice maintainer line. Its sterile line, Jin 23A, has been widely used in hybrid rice production in China. However, Jin 23B is highly susceptible to rice blast. In a backcross breeding scheme, three blast resistance genes, *Pi1*, *Pi2*, and *D12*, were introgressed into Jin 23B through MAS, and the improved line Jin 23B (*Pi1*/*Pi2*/*D12*) was obtained from BC_4F_3 families. The blast resistance of Jin 23B (*Pi1*/*Pi2*/*D12*) and its hybrid rice, Jinyou 402 (*Pi1*/*Pi2*/*D12*), significantly increased compared with their counterparts, Jin 23B and its hybrid rice, Jinyou 402. The resistance scores of leaf blast at tillering stage for Jin 23B (*Pi1*/*Pi2*/*D12*) and Jinyou 402 (*Pi1*/*Pi2*/*D12*) were 0.33 and 0.72, respectively, which are significantly lower than the scores for Jin 23B and Jinyou 402 (4.01 and 4.02, respectively). Moreover, the hybrid rice from Jin 23B (*Pi1*/*Pi2*/*D12*) under the disease condition had the same yield as under the normal condition (Jiang et al. 2012). In wheat, eight QTLs/genes were pyramided into a popular elite wheat cultivar PBW343 through MAS, endowing the improved PBW343 four grain quality traits and the resistance against three rusts (Tyagi et al. 2014).

Abiotic stress, including drought and salinity, has become more serious for crops in recent years. Stress resistance has been repeatedly achieved in crops through the transformation of a single gene. For example, Monsanto had developed a drought-resistance GM maize variety carrying a cold shock protein B gene *cspB*, Genuity® DroughtGard™ (GM Approval Database of ISAAA). However, the gene pyramiding is still necessary for further enhancement of resistance to abiotic stress. The transgenic maize carrying stacked *betA* (encoding choline dehydrogenase from *Escherichia coli*) and *TsVP* (encoding V-H^+-PPase from *Thellungiella halophila*) had been produced by cross-pollination. It contained higher relative water content (RWC), greater solute accumulation and lower cell damage under drought stress treatment, and grew more vigorously with less growth retardation, shorter anthesis-silking intervals and higher yields than their parental lines, which had either betA or TsVP, and the wild-type (Wei et al. 2011). Nguyen et al. (2013) developed the transgenic maize carrying the stacked *mtlD* (bacterial mannitol-1-phosphate dehydrogenase encoding gene) and *HVA1* (*Hordeum vulgare*) through co-transformation of two vectors. The transgenic plants (*mtlD*+*HVA1*) showed higher leaf RWC and greater plant survival as compared with the transgenic plants of a single transgene and with their control plants under drought stress, and also showed higher fresh and dry shoot matter and dry root matter as compared with the transgenic plants of a single transgene and with their control plants under salt stress.

13.5 Conclusions and Prospects

With the coming of the post-genome era, crop breeding has gradually developed from conventional breeding based on phenotype selection to *molecular design breeding* based on gene selection. As a directed improvement technology, transgenic pyramiding breeding will provide a powerful technological support for crop molecular design breeding. However, transgenic pyramiding breeding also comes up against many technology challenges for meeting the molecular design breeding requirements. In contrast to conventional breeding, transgenic crops should not only be evaluated in agronomic characters such as yield and resistance, but also in their biosafety. Biosafety assessment is a time-consuming and laborious process, and also increases the product cost and prolongs the breeding process. More genes involved in transgenic pyramiding will result in a more complex biosafety assessment. Therefore, to simplify the process of biosafety assessment regarding breeding strategy and technology within the allowable range of biosafety management is important.

Current transgenic crops involve few effect genes only, and all these genes constitutively overexpress in transgenic crops. However, with the development of transgenic pyramiding breeding, the number of transgenes will increase, and their expression patterns, including expression level, tissue-specificity, developmental stage-specificity, etc., will become more complicated. Therefore, it is necessary to ensure every transgene truly expressing according to the breeding objective. Although the current multiple-gene transformation technology can pyramid eight to ten genes or transformed the fragments of 20 kb or more in length, the technology should be further developed to become a conventional breeding technology.

Crop variety is a special commodity, which has commodity properties, such as market demand and cost. Moreover, biological product properties, such as ecological adaptability, stress resistance, yield traits, quality traits, and biosafety, can also be found in the variety. As the second-generation transgenic crops, the commercialization of transgenic crops with stacked pest and herbicide resistance has started, and will be in a fast developing period. A few metabolic engineering crops of carotenoids and VLCPUFAs have achieved the requirement for commercialization. Moreover, by elucidating the metabolic pathways in plants, more metabolic engineering technologies will mature. In recent years, with the rapid development of plant genomics research, a number of genes controlling important agronomy characters have been identified and cloned, and their functions have been illuminated. These achievements can provide the new directions for the transgenic pyramiding breeding of crops.

References

Abbadi A, Domergue F, Bauer J et al (2004) Biosynthesis of very-long-chain polyunsaturated fatty acids in transgenic oilseeds: constraints on their accumulation. Plant Cell 16(10):2734–2748

Aluru M, Xu Y, Guo R et al (2008) Generation of transgenic maize with enhanced provitamin A content. J Exp Bot 59(13):3551–3562

Azadi P, Otang NV, Chin DP et al (2010) Metabolic engineering of *Lilium* × *formolongi* using multiple genes of the carotenoid biosynthesis pathway. Plant Biotechnol Rep 4(4):269–280
Bates SL, Zhao JZ, Roush RT et al (2005) Insect resistance management in GM crops:past, present and future. Nat Biotechnol 23(1):57–62
Bohmert-Tatarev K, McAvoy S, Daughtry S et al (2011) High levels of bioplastic are produced in fertile transplastomic tobacco plants engineered with a synthetic operon for the production of polyhydroxybutyrate. Plant Physiol 155(4):1690–1708
Buntru M, Gärtner S, Staib L et al (2013) Delivery of multiple transgenes to plant cells by an improved version of MultiRound Gateway technology. Transgenic Res 22(1):153–167
Carciofi M, Blennow A, Jensen SL et al (2012) Concerted suppression of all starch branching enzyme genes in barley produces amylose-only starch granules. BMC Plant Biol 12(1):223
Chang YL, Henriquez X, Preuss D et al (2003) A plant-transformation-competent BIBAC library from the *Arabidopsis thaliana* Landsberg ecotype for functional and comparative genomics. Theor Appl Genet 106(2):269–276
Chen QJ, Zhou HM, Chen J et al (2006) A Gateway-based platform for multigene plant transformation. Plant Mol Biol 62(6):927–936
Chen QJ, Xie M, Ma XX et al (2010) MISSA is a highly efficient in vivo DNA assembly method for plant multiple-gene transformation. Plant Physiol 153(1):41–51
Chen H, He H, Zou Y et al (2011) Development and application of a set of breeder-friendly SNP markers for genetic analyses and molecular breeding of rice (*Oryza sativa* L.). Theor Appl Genet 123(6):869–879
Cheng B, Wu G, Vrinten P et al (2010) Towards the production of high levels of eicosapentaenoic acid in transgenic plants: the effects of different host species, genes and promoters. Transgenic Res 19(2):221–229
Cui C, Song F, Tan Y et al (2011) Stable chloroplast transformation of immature scutella and inflorescences in wheat (*Triticum aestivum* L.). Acta Biochim Biophys Sin 43(4):284–291
Daniell H, Lee SB, Pahchal T et al (2001) Expression of the native cholera toxin B subunit gene and assembly as functional oligomers in transgenic tobacco chloroplasts. J Mol Biol 311(5):1001–1009
Davey JW, Hohenlohe PA, Etter PD et al (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12(7):499–510
de Felipe P, Hughes LE, Ryan MD et al (2003) Co-translational, intraribosomal cleavage of polypeptides by the foot-and-mouth disease virus 2A peptide. J Biol Chem 278(13):11441–11448
Diretto G, Al-Babili S, Tavazza R et al (2007) Metabolic engineering of potato carotenoid content through tubers-specific overexpression of a bacterial mini-pathway. PLoS One 2(4), e350
Dokku P, Das KM, Rao GJN (2013) Pyramiding of four resistance genes of bacterial blight in Tapaswini, an elite rice cultivar, through marker-assisted selection. Euphytica 192(1):87–96
el Amrani A, Barakate A, Askari BM et al (2004) Coordinate expression and independent subcellular targeting of multiple proteins from a single transgene. Plant Physiol 135(1):16–24
Fan C, Yu S, Wang C et al (2009) A causal C–A mutation in the second exon of *GS3* highly associated with rice grain length and validated as a functional marker. Theor Appl Genet 118(3):465–472
Fang J, Qian JJ, Yi S et al (2005) Stable antibody expression at therapeutic levels using the 2A peptide. Nat Biotechnol 23(5):584–590
Farre G, Naqvi S, Sanahuja G et al (2012) Combinatorial genetic transformation of cereals and the creation of metabolic libraries for the carotenoid pathway. Methods Mol Biol 847:419–435
Ferguson ME, Hearne SJ, Close TJ et al (2012) Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. Theor Appl Genet 124(4):685–695
Flores T, Karpova O, Su X et al (2008) Silencing of the *GmFAD3* gene by siRNA leads to low a-linolenic acids (18:3) *offad3*-mutant phenotype in soybean *Glycine max* (Merr.). Transgenic Res 17:839–850

François IE, De Bolle MF, Dwyer G et al (2002) Transgenic expression in *Arabidopsis* of a polyprotein construct leading to production of two different antimicrobial proteins. Plant Physiol 128(4):1346–1358
François IE, Van Hemelrijck W, Aerts AM et al (2004) Processing in *Arabidopsis thaliana* of a heterologous polyprotein resulting in differential targeting of the individual plant defensins. Plant Sci 166(1):113–121
Fujisawa M, Takita E, Harada H et al (2009) Pathway engineering of *Brassica napus* seeds using multiple key enzyme genes involved in ketocarotenoid formation. J Exp Bot 60(4):1319–1332
Ghimire MN, Huang F, Leonard R et al (2011) Susceptibility of Cry1Ab-susceptible and -resistant sugarcane borer to transgenic corn plants containing single or pyramided *Bacillus thuringiensis* genes. Crop Prot 30(1):74–81
Gibson DG, Benders GA, Axelrod KC (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. Proc Nat Acad Sci 105(51):20404–20409
Gibson DG, Young L, Chuang RY (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6(5):343–345
GM Approval Database (2004) ISAAA. http://www.isaaa.org/gmapprovaldatabase/. Accessed 12 Feb 2004
Goyal A, Bhowmik PK, Basu SK (2009) Minichromosomes: the second generation genetic engineering tool. Plant Omics J 2(1):1–8
Ha SH, Liang YS, Jung H et al (2010) Application of two bicistronic systems involving 2A and IRES sequences to the biosynthesis of carotenoids in rice endosperm. Plant Biotechnol J 8(8):928–938
Halpin C (2005) Gene stacking in transgenic plants–the challenge for 21st century plant biotechnology. Plant Biotechnol 3(2):141–155
Hamilton CM, Frary A, Xu Y (1999) Construction of tomato genomic DNA libraries in a binary-BAC (BIBAC) vector. Plant J 18(2):223–229
Hou BK, Zhou YH, Wan LH et al (2003) Chloroplast transformation in oilseed rape. Transgenic Res 12(1):111–114
Hu J, Li X, Wu C et al (2012) Pyramiding and evaluation of the brown planthopper resistance genes *Bph14* and *Bph15* in hybrid rice. Mol Breed 29(1):61–69
Huang JC, Zhong YJ, Liu J et al (2013) Metabolic engineering of tomato for high-yield production of astaxanthin. Metab Eng 17(100):59–67
Huang N, Angeles ER, Domingo J et al (1997) Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theor Appl Genet 95(3):313–320
James C (2013) Global status of commercialized Biotech/GM Crops. ISAAA, Ithaca
Jha S, Chattoo BB (2009) Transgene stacking and coordinated expression of plant defensins confer fungal resistance in rice. Rice 2(4):143–154
Jiang H, Feng Y, Bao L et al (2012) Improving blast resistance of Jin 23B and its hybrid rice by marker-assisted gene pyramiding. Mol Breed 30(4):1679–1688
Jiang L, Yu X, Qi X et al (2013) Multigene engineering of starch biosynthesis in maize endosperm increases the total starch content and the proportion of amylose. Transgenic Res 22(6):1133–1142
Kim JH, Lee SR, Li LH et al (2011) High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. PLoS One 6(4), e18556
Kong FN, Jiang SM, Shi LX (2006) Construction and characterization of a transformation-competent artificial chromosome (TAC) library of *Zizania latifolia* (Griseb.). Plant Mol Biol Report 24(2):219–227
Kumar S, Dhingra A, Daniell H (2004) Stable transformation of the cotton plastid genome and maternal inheritance of transgenes. Plant Mol Biol 56(2):203–216
Lange BM, Mahmoud SS, Wildung MR et al (2011) Improving peppermint essential oil yield and composition by metabolic engineering. Proc Nat Acad Sci 108(41):16944–16949
Lee SM, Kang K, Chung H et al (2006) Plastid transformation in the monocotyledonous cereal crop, rice (*Oryza sativa*) and transmission of transgenes to their progeny. Mol Cells 21(3):401–410

Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Meth 4(3):251–256
Li MZ, Elledge SJ (2012) SLIC: a method for sequence-and ligation-independent cloning. Methods Mol Biol 852:51–59
Lin L, Liu YG, Xu X (2003) Efficient linking and transfer of multiple genes by a multigene assembly and transformation vector system. Proc Nat Acad Sci 100:5962–5967
Lu Y, Rijzaani H, Karcher D et al (2013) Efficient metabolic pathway engineering in transgenic tobacco and tomato plastids with synthetic multigene operons. Proc Nat Acad Sci 110(8):E623–E632
Ma C, Mitra A (2002) Expressing multiple genes in a single open reading frame with the 2A region of foot-and-mouth disease virus as a linker. Mol Breed 9(3):191–199
Murata M, Shibata F, Hironaka A et al (2013) Generation of an artificial ring chromosome in *Arabidopsis* by Cre/*Lox*P-mediated recombination. Plant J 74(3):363–371
Naqvi S, Farré G, Zhu C et al (2011) Simultaneous expression of *Arabidopsis* ρ-hydroxyphenylpyruvate dioxygenase and MPBQ methyltransferase in transgenic corn kernels triples the tocopherol content. Transgenic Res 20(1):177–181
Naqvi S, Zhu C, Farre G et al (2009) Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways. Proc Nat Acad Sci 106(19):7762–7767
Nguyen TX, Nguyen T, Alameldin H et al (2013) Transgene pyramiding of the *HVA1* and *mtlD* in T3 maize (*Zea mays* L.) plants confers drought and salt tolerance, along with an increase in crop biomass. Int J Agron, doi:10.1155/2013/598163
Paine JA, Shipton CA, Chaggar S et al (2005) Improving the nutritional value of Golden Rice through increased pro-vitamin A content. Nat Biotechnol 23(4):482–487
Petrie JR, Shrestha P, Zhou XR et al (2012) Metabolic engineering plant seeds with fish oil-like levels of DHA. PLoS One 7(11), e49165
Qi B, Fraser T, Mugford S et al (2004) Production of very long chain polyunsaturated omega-3 and omega-6 fatty acids in plants. Nat Biotechnol 22(6):739–745
Qu S, Coaker G, Francis D (2003) Development of a new transformation-competent artificial chromosome (TAC) vector and construction of tomato and rice TAC libraries. Mol Breed 12(4):297–308
Quan J, Tian J (2009) Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 4(7), e6441
Regina A, Kosar-Hashemi B, Ling S et al (2010) Control of starch branching in barley defined through differential RNAi suppression of starch branching enzyme IIa and IIb. J Exp Bot 61(5):1469–1482
Robert SS, Singh SP, Zhou XR et al (2005) Metabolic engineering of *Arabidopsis* to produce nutritionally important DHA in seed oil. Func Plant Biol 32(6):473–479
Ruf S, Hermann M, Berger I et al (2001) Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. Nat Biotechnol 19(9):870–875
Ruiz-Lopez N, Haslam RP, Napier JA et al (2014) Successful high-level accumulation of fish oil omega-3 long-chain polyunsaturated fatty acids in a transgenic oilseed crop. Plant J 77(2):198–208
Ruiz-Lopez N, Haslam RP, Usher SL et al (2013) Reconstitution of EPA and DHA biosynthesis in *Arabidopsis*: iterative metabolic engineering for the synthesis of n−3 LC-PUFAs in transgenic plants. Metab Eng 17(100):30–41
Ryan MD, Drew J (1994) Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. EMBO J 13(4):928–933
SanGiovanni JP, Chew EY (2005) The role of omega-3 long-chain polyunsaturated fatty acids in health and disease of the retina. Prog Retin Eye Res 24(1):87–138
Sarrion-Perdigones A, Falconi EE, Zandalinas SI et al (2011) GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. PLoS One 6(7), e21622
Sasaki Y, Sone T, Yoshida S et al (2004) Evidence for high specificity and efficiency of multiple recombination signals in mixed DNA cloning by the Multisite Gateway system. J Biotechnol 107(3):233–243

Seeb JE, Carvalho G, Hauser L et al (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. Mol Ecol Res 11(s1):1–8
Sidorov VA, Kasten D, Pang SZ et al (1999) Stable chloroplast transformation in potato: use of green fluorescent protein as a plastid marker. Plant J 19(2):209–216
Skarjinskaia M, Svab Z, Maliga P (2003) Plastid transformation in *Lesquerella fendleri*, an oilseed Brassicacea. Transgenic Res 12(1):115–122
Sun H, Lang Z, Zhu L et al (2012) Acquiring transgenic tobacco plants with insect resistance and glyphosate tolerance by fusion gene transformation. Plant Cell Rep 31(10):1877–1887
Suzuki N, Geletka LM, Nuss DL (2000) Essential and dispensable virus-encoded replication elements revealed by efforts to develop hypoviruses as gene expression vectors. J Virol 74(16):7568–7577
Svab Z, Maliga P (1993) High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. Proc Nat Acad Sci 90(3):913–917
Tan MA, Hutten RC, Visser RG et al (2010) The effect of pyramiding *Phytophthora infestans* resistance genes $R_{Pi\text{-}mcd1}$ and $R_{Pi\text{-}ber}$ in potato. Theor Appl Genet 121(1):117–125
Tan H, Yang X, Zhang F et al (2011) Enhanced seed oil production in canola by conditional expression of *Brassica napus LEAFY COTYLEDON1* and *LEC1-LIKE* in developing seeds. Plant Physiol 156(3):1577–1588
Tang W, Chen H, Xu C et al (2006) Development of insect-resistant transgenic indica rice with a synthetic *cry1C** gene. Mol Breed 18(1):1–10
Tang G, Qin J, Dolnikowski GG et al (2009) Golden Rice is an effective source of vitamin A. Am J Clin Nutr 89(6):1776–1783
Taverniers I, Papazova N, Bertheau Y et al (2008) Gene stacking in transgenic plants: towards compliance between definitions, terminology, and detection within the EU regulatory framework. Envir Biosaf Res 7(4):197–218
Tiwari VK, Wang S, Sehgal S et al (2014) SNP discovery for mapping alien introgressions in wheat. BMC Genomics. doi:10.1186/1471-2164-15-273
Tyagi S, Mir RR, Kaur H et al (2014) Marker-assisted pyramiding of eight QTLs/genes for seven different traits in common wheat (*Triticum aestivum* L.). Mol Breed. doi:10.1007/s11032-014-0027-1
Wan B, Zha Z, Li J et al (2014) Development of elite rice restorer lines in the genetic background of R022 possessing tolerance to brown planthopper, stem borer, leaf folder and herbicide through marker-assisted breeding. Euphytica 195(1):129–142
Wei AY, He CM, Li B et al (2011) The pyramid of transgenes *TsVP* and *BetA* effectively enhances the drought tolerance of maize plants. Plant Biotechnol J 9(2):216–229
WHO (2009) Global prevalence of vitamin A deficiency in populations at risk 1995–2005. WHO global database on vitamin A deficiency. Geneva, World Health Organization, 2009. Available from: http://www.who.int/vmnis/vitamina/en/
Wu G, Truksa M, Datla N et al (2005) Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants. Nat Biotechnol 23(8):1013–1017
Xu C, Cheng Z, Yu W (2012a) Construction of rice mini-chromosomes by telomere-mediated chromosomal truncation. Plant J 70(6):1070–1079
Xu X, Liu X, Ge S et al (2012b) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol 30(1):105–111
Yang H, Tao Y, Zheng Z et al (2012) Application of next-generation sequencing for rapid marker development in molecular plant breeding: a case study on anthracnose disease resistance in *Lupinus angustifolius* L. BMC Genomics. doi:10.1186/1471-2164-13-318
Ye X, Al-Babili S, Klöti A et al (2000) Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287(5451):303–305
Yu W, Lamb JC, Han F et al (2006) Telomere-mediated chromosomal truncation in maize. Proc Nat Acad Sci 103(46):17331–17336
Yu W, Han F, Gao Z (2007) Construction and behavior of engineered minichromosomes in maize. Proc Nat Acad Sci 104(21):8924–8929

Zhai J, Wang Y, Sun C (2013) A plant-transformation-competent BIBAC library of ginseng (*Panax ginseng* C. A. Meyer) for functional genomics research and characterization of genes involved in ginsenoside biosynthesis. Mol Breed 31(3):685–692

Zhang J, Martin JM, Beecher B et al (2010) The ectopic expression of the wheat Puroindoline genes increase germ size and seed oil content in transgenic corn. Plant Mol Biol 74(4–5):353–365

Zhao JZ, Cao J, Li Y et al (2003) Transgenic plants expressing two *Bacillus thuringiensis* toxins delay insect resistance evolution. Nat Biotechnol 21(12):1493–1497

Zhu C, Naqvi S, Breitenbach J et al (2008) Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. Proc Nat Acad Sci 105(47):18232–18237

Chapter 14
RNAi Technology: A Potential Tool in Plant Breeding

Mohammad Abhary and Adel Rezk

Abstract Traditional plant breeding has been very successful in producing agricultural crops with many desirable traits. However, it is a time-consuming process with limited genetic resources, which is restricted to closely related species for crossing. RNA interference (RNAi) technology presents a new potential tool for plant breeding by introducing small non-coding RNA sequences with the ability to switch-off gene expression in a sequence-specific manner. The ability to suppress expression of a specific gene provides an opportunity to acquire a new trait by eliminating or accumulating certain plant traits, leading to biochemical or phenotypic changes that do not exist in non-transgenic plants. RNAi is an ancient evolutionary mechanism adopted by plants as a defense strategy against foreign invading genes but is used today as a tool for generating new quality traits in organisms. In this chapter, we review RNAi applications in plants to acquire new traits, which are difficult to obtain through traditional breeding and present the potential of combining this technology with conventional breeding to overcome the current challenges facing agriculture such as: abiotic stress, biotic stress, nutritional value, allergens, secondary metabolites and flower traits.

Keywords Breeding • Gene silencing • Gene suppression • miRNA • RNAi • Quality traits

M. Abhary,
Department of Biology, Taibah University,
Al-Jamaat Street, Al-Madinah 30002, Saudi Arabia
e-mail: mabhary@taibahu.edu.sa

A. Rezk (✉)
Department of Agricultural Biotechnology, College Agriculture and Food Science,
King Faisal University, Al-Hassa 31982, Saudi Arabia
e-mail: adelrezk@hotmail.com; arazk@kfu.edu.sa

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_14

14.1 Introduction

Plant molecular breeding has become an essential tool in modern agricultural practices, producing and improving sustainable crops to feed the growing world population by introducing desirable traits in a number of crops based upon molecular knowledge of trait loci and genetic markers. The revolution of recent plant breeding techniques has allowed plant breeders to use novel methodologies to import and export quality traits between plants with a high level of confidence (Varshney et al. 2005), speeding up the breeding process and acquiring new traits to confront the current agricultural challenges such as plant diseases and environmental stress. Although genetic mapping is considered a primary tool in the search for quality desirable traits in molecular breeding programs (Varshney et al. 2005), gene functional studies and translational regulations are essential to identify the trait before being transferred between plant lines.

RNAi technology offers a fast and reliable methodology to study gene function through silencing genes based on a sequence-specific manner (Vaucheret et al. 2001). Moreover, RNAi-mediated gene silencing techniques can provide plant researchers with the choice of suppressing genes in full or partially, in time and space, through proper choice of promoters and construct design (Senthil-Kumar and Mysore 2011).

The applicable knowledge of RNAi available today is the result of accumulated observations and research, which began in the last decade of the twentieth century. The discovery of RNAi has revolutionized the research areas of plant breeding and serves as a tool to understand the expression pattern of plant genomes and as a toolbox to control plant gene expression quantitatively and qualitatively (Hirai and Kodama 2008), controlling pathogenicity of plant parasites (Runo 2011) and enhancing resistance against biotic (Wani et al. 2010) and abiotic stresses (Jagtap et al. 2011). The increasing number of research papers on the subject of gene silencing has exceeded 65,000 in the PubMed web-search engine, where almost 7000 plant research papers are listed (http://www.ncbi.nlm.nih.gov/pubmed).

Different forms of gene silencing have been characterized in many organisms, from bacteria to humans, which are all related in the mechanism of reducing the accumulation and the degradation of specific RNA transcripts (Hamilton and Baulcombe 1999). The phenomenon has been given different names according to different organisms, such as co-suppression in plants (Napoli et al. 1990), quelling in fungi (Fagard et al. 2000) and RNA interference (RNAi) in nematodes (Guo and Kemphues 1995), a common evolutionary mechanism of post-transcriptional gene silencing (PTGS). The first observation of this phenomenon came unexpectedly in transgenic petunia experiments in 1990 to improve color pigments (Napoli et al. 1990); scientists were trying to introduce extra copies of pigment coding genes into the petunia, but the resulting transgenic plants exhibited white flowers instead. The endogenous and introduced gene transcripts were turned off or *co-suppressed*. Later, the same phenomenon was observed in roundworm (*Caenorhabditis elegans*) by Guo and Kemphues (1995) when they injected antisense RNA strands of an

endogenous gene into worms to be hybridized with the endogenous mRNA transcript and turn it off. The negative control was the sense strand RNA, which was as effective as antisense in suppressing the endogenous gene expression. The progress of understanding gene silencing started in 1998, when Fire et al. (1998) tested the synergistic effect of sense, antisense and double stranded RNA (dsRNA) (Elbashir et al. 2001). Previously, the idea of using dsRNA was thought impossible because it was hypothesized that dsRNA would be stable and unavailable to pair with target transcripts in the cell. After the astonishing results of the dsRNA effect on gene silencing, the term RNAi was established and the Nobel Prize in Physiology or Medicine was awarded to Fire and Mello in 2006.

The next generation of RNAi experiments focused on its application, with virologists carrying out RNAi research to improve pathogen-derived resistance (PDR) in crops (Abhary et al. 2006) and physiologists were interested in abiotic stress tolerance (Jagtap et al. 2011). The discovery of natural small interfering RNA species enhanced the integration of computer sciences to predict and synthesize artificial silencing triggers for more traits in commercially-important crops (Warthmann et al. 2008).

This chapter aims to illustrate the current knowledge and applications of RNAi technology in plants and presents new prospects for plant breeders. In addition to expanding the available genetic pool, RNAi technology provides the opportunity of integrating new agricultural traits to obtain superior crops with enhanced nutritional value, disease resistance and abiotic stress tolerance as well as new horticultural traits such as flower color.

14.2 microRNA

The microRNAs (miRNA) are naturally existing small non-coding RNA molecules (containing about 22 nucleotides) found in plants, animals and some viruses. The main function of miRNA is to regulate gene expression (Ambros 2004; Bartel 2004). In plants, miRNAs are encoded by nuclear DNA and the viral genome of some DNA viruses. The miRNAs are derived from regions of RNA transcripts that folds back on itself to form a short hairpin, whereas small interfering RNAs (siRNAs) are derived from longer regions of dsRNA (Bartel 2004). In plants, miRNAs are well conserved and have an evolutionary role in genetic regulation (Axtell and Bartel 2005); they usually have near-perfect complementary paring with their mRNA targets, which reduces gene repression through cleavage of the target transcripts (Jones-Rhoades et al. 2006). The first miRNA was discovered by Lee et al. (1993) while studying the expression pattern of the lin-4 gene in roundworm, *Caenorhabditis elegans*. They showed that lin-4 expresses a 20 to 24-nucleotide antisense RNA which repress the lin-14 gene. Several years later, a second miRNA was characterized in *C. elegans* (Reinhart et al. 2000) and was named let-7. Reinhart et al. 2002 isolated and characterized 16 *Arabidopsis* miRNAs that were differentially expressed during development, which control the maturation of the plant in

different stages. By 2005, the number of identified miRNAs from plants reached 872 different sequences belonging to 71 plant species (Zhang et al. 2005). Within 3 years, identified miRNAs reached more than 10,000 in plants (Griffiths-Jones et al. 2008), showing the rapid development of this technology. Several researchers are using computer software programs to mimic the natural microRNAs and produce artificial miRNAs (amiRNAs) for more crop traits and using miRNAs along with amiRNAs in different breeding programs (Kong et al. 2010; Warthmann et al. 2008).

14.3 Mechanism of RNAi

The process of gene expression regulation by RNAi depends on sequence complementarity with the targeted mRNA. Sequence specificity gives a quality control mechanism of gene expression regulation to inhibit gene expression at the stage of translation by cleaving the transcript mRNA into small pieces, which can serve as new templates for more RNAi (Voinnet 2005).

The mechanism of RNAi involves several factors; RNase III enzyme, called Dicer or Dicer-like protein (DCL), cuts the dsRNAs into small pieces ranging between 21 and 25 nucleotides long, named siRNAs. Most plant species have more than one DCL genes; some DCLs are specifically required for the production of the 21-nucleotide long siRNAs. After siRNAs are formed from processed mRNA transcripts or the dsRNA, it is incorporated into a group of proteins in a complex form called RNA-induced silencing complex (RISC). RISC usually exists in small groups inside the cytoplasm, presumably located in the P-body or so-called the RNA quality control centers of the cell (Jouannet et al. 2012); it consists of several proteins including the argonaute (AGO) protein, which acts on slicing and cleaving the complementary mRNA transcripts, preventing gene expression of that specific gene. The RNAi pathway can be divided into three major steps (Fig. 14.1).

First, the initiator step that involves generating the small interfering RNA (siRNA). The initiation of RNAi starts after generating siRNA duplexes by cleaving dsRNA into small pieces, ranging from 21 to 23 nt, with 2 nucleotides overhanging at the 3′-OH and 5′-PO_4. This process is performed by an enzyme called Dicer, a member of the RNase III family, which has a specific endonuclease activity and work in an ATP-dependent manner (Das et al. 2010).

Second, the effector step involving the formation of RNA-induced silencing complex (RISC). After generating the siRNAs in an ATP-dependent manner, the template for RNAi could be established by joining a multi-protein complex that will operate in an ATP-independent manner (Senapedis et al. 2011). The protein complex mediates unwinding the siRNA duplex strands and incorporates one of the strands, a *guide strand,* into the AGO protein component, which is a member of a conserved family of proteins, and then the anti-guide strand is degraded or used to generate more siRNA duplexes by RNA-dependent RNA-polymerase (RDRP). Fagard et al. 2000 identified AGO protein family an *Arabidopsis* mutant called AGO1, which has conserved regions PAZ and PIWI domains, where PAZ is attached

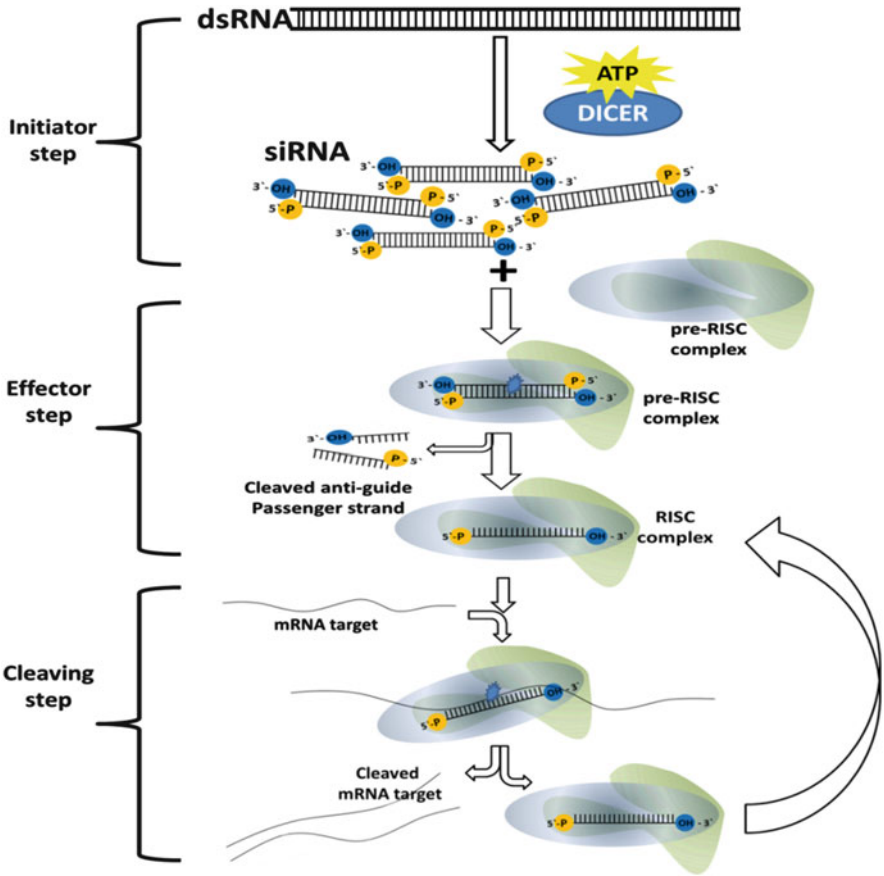

Fig. 14.1 Illustration of the general scheme of RNAi. The three steps of breaking mRNA, initiator step; dsRNA is sliced into siRNA, effector step; siRNA is associated with RISC and cleaving step; mRNA is degraded by RISC on sequence specific manner

to the 3′—OH end of the guide strand and the PIWI domain contains the mRNA cleavage active site (Nykanen et al. 2001). This step is believed to be self-proliferating; the enzyme RDRP generates more siRNA duplexes from the anti-guide strands of the siRNA and provides more guide strands for the AGO protein in the RISC.

Third, the cleaving step involves specific targeting of the mRNA sequence. During this step, the siRNA guide strand of RISC will be in search of its complimentary sequences in the transcript mRNAs pool. When target mRNA is found, the guide strand will bind to it at the PAZ domain site, this action promotes the PIWI domain of the RISC to break one phosphodiester bond on the targeted mRNA strand at a distance of 10 nucleotides. The hydrolyzed mRNA is disrupted at various locations, depending on siRNA complementarity, which will leave the cleaved mRNA un-translatable (Nykanen et al. 2001).

14.4 Methods of Introducing RNAi into Plants

Recent plant transformation technologies have used RNAi as a tool to study gene expression by introducing dsRNA in a similar manner to sense and antisense RNAs, but with higher efficiency, to bring new traits such as disease resistance, stress resistance and nutritional quality traits in various crops. The developed gene delivery techniques has allowed speeding-up the study of gene loss functions in targeted plant tissues by using transient assay systems. RNAi transient delivery systems include: (a) virus-induced gene silencing (VIGS); (b) direct gene bombardment and (c) *Agrobacterium*-mediated transient assays.

(a) Viral vectors have been developed for VIGS in many plant species such as barley (*Hordeum vulgare*) (Liang et al. 2012); tomato (*Lycopersicon lycopersicum*) (Fei et al. 2011); pea (*Pisum sativum*) (Grønlund et al. 2010) and the experimental plants *Arabidopsis* (Burch-Smith et al. 2004) and *Nicotiana benthamiana* (Velásquez et al. 2009). The suitability of viral vectors differs according to the plant species being studied; a limiting factor of using VIGS in some plants besides the biosafety regulations of using some viruses (Shang et al. 2007). Nevertheless, engineered plant viruses have been used for dsRNA production in both transient and stable transformation protocols targeting endogenous plant transcripts and controlling pests during herbivorous insect-plant interactions (Senthil-Kumar and Mysore 2011).

(b) Although particle bombardments (biolistics) have been used as a direct delivery method for dsRNA constructs in stable transformations, the system is being used also in functional transient assays, especially in studying plant pigmentation patterns (Shang et al. 2007). Particle bombardment on mature plants is usually random and limited to the targeted tissue in showing a phenotypical change. Therefore, biolistics are used on specific tissues such as flower petals or developing seeds in transient assays. In addition, microprojectile bombardment is used on monocot plants where *Agrobacterium*-mediated transformation is mostly limited (Schweizer et al. 2000).

(c) The most widely-used method in RNAi transient assays is *Agrobacterium tumerfaciens* infiltration (agroinfiltration). This method is a fast convenient method for delivering silencing constructs to vegetative tissues and obtaining a high level of local phenotypic effect, which can be biochemically analyzed, before proceeding to fully transgenic plants. Agro-infiltration transient assays have been optimized for many plant species, including monocots such as rice (Andrieu et al. 2012).

On the other hand, the delivered RNAi constructs are generally designed to produce a hairpin-RNA; engineered vectors are built with inverted repeats of the targeted sequence downstream to strong or tissue-specific promoters and separated by spacer or intron fragments (Smith et al. 2000). Having such a toolbox, plant gene transcripts can be targeted for silencing in the appropriate time and space. Whether researchers are interested in silencing the genes in full or partially, the choice of

promoter, hairpin and spacer sequences can be adjusted to achieve the demand (Hirai and Kodama 2008).

14.5 Pest and Pathogen Control

The RNAi based gene silencing mechanism is conserved across the plant and animal kingdoms and thus holds promise not only for functional genomics but also in agriculture applications such as engineering for stress tolerance, both biotic and abiotic (Senthil-Kumar and Mysore 2010). There are two strategies in RNAi applications for agricultural crops; the first is silencing a specific targeted gene in plants to improve the agronomic character such as delaying fruit ripening and firmness or increasing the fitness against biotic or abiotic stresses. Methods used in this strategy are referred to as *host gene silencing hairpins* RNAi (HGS-hpRNAi) (Senthil-Kumar and Mysore 2010). The second strategy includes silencing a targeted gene in another organism that feeds on plants, such as insects, plant pathogens and parasites, this process is referred to as *host-delivered* RNAi (hdRNAi). Senthil-Kumar and Mysore (2010) divided this approach into two categories; first, hdRNAi-1 that silence genes in organisms outside the plant that particularly feed and cause direct damage to the plant such as insects, nematodes and parasitic weeds and a second hdRNAi-2 category that silence viral genes inside the host plant cells.

14.5.1 Diseases

The majority of plant diseases cause dramatic crop losses and reduce the artistic and practical values of landscape plants and home gardens. Plant diseases caused by bacteria, virus, fungi, insects and nematodes create several constraints to the production process. Plant disease management aims to reduce economically and aesthetically damages caused by plant diseases in a term called *plant disease control*. RNA silencing-based resistance has been a powerful tool of genetically engineered crops during the last two decades. Engineered plants with RNAi technology are becoming increasingly important and providing more effective strategies in future (Karthikeyan et al. 2013). The advantage of RNAi as a tool for plant gene therapy against, viral, bacterial and fungal infections lies in the fact that it regulates gene expression via mRNA degradation, translation repression and chromatin renovation through small non-coding RNAs and it has no protein product. The applications of RNAi with the use of suitable promoters to silence several genes at the same time should increase researchers' ability to protect crops against many plant diseases.

14.5.2 Viruses

Plant viruses are very difficult to control and their molecular basis for disease development is still under investigation (Jagtap et al. 2011). In stable transgenic plants, such manipulation would require sequential transformations or crosses between plants that could take prolonged periods of time (Fernández-Calvo et al. 2011). The majority of plant viruses have RNA genomes, except geminiviruses, nanoviruses which have ssDNA genomes and replicates through a dsDNA intermediate by a rolling circle mechanism (Laufs et al. 1995) and caulimoviruses, which have a dsDNA genome that replicates through an RNA intermediate using reverse transcription (Hull 1989).

RNAi technology has been used to protect and manage viral infections in plants (Wani et al. 2010; Waterhouse et al. 2001), where post-transcriptional gene silencing (PTGS) acts as a natural antiviral defense system that plays a role in genome maintenance and development (Vanitharani et al. 2005). In virus-infected plants, silencing is initiated by creating a dsRNA that could be viral replication intermediates or ssRNA that becomes dsRNA by host-encoded RNA-dependent RNA polymerase (RdRP) and then acts as a vaccination against viruses (Ahlquist 2002; Vanitharani et al. 2005).

RNAi technology provides the possibility of targeting several viruses in a single transformation so that multiple resistance traits can be generated. Furthermore, dsRNA molecules can be designed to target viral regulatory regions, such as the CaMV 35S promoter, commonly used in transgenic plants (Karthikeyan et al. 2013). The effectiveness of this technology in generating virus resistance was first reported against potato virus Y (PVY) in potato using both sense and antisense transcripts of the helper component proteinase (HC-Pro) gene (Waterhouse et al. 1998). The application of RNAi technology has resulted in inducing immunity reaction against many different RNA and DNA plant viruses such as: tomato leaf curl virus (Abhary et al. 2006; Bian et al. 2006; Rezk et al. 2006; Shelly et al. 2010); potato virus X (PVX) (Soliman et al. 2008); cucumber mosaic virus (CMV) (Kalantidis et al. 2002); barley yellow dwarf virus-PAV (BYDV-PAV) (Wang et al. 2000); bean golden mosaic virus (BGMV) (Bernstein et al. 2001); African cassava mosaic virus (ACMV) (Vanderschuren et al. 2007; 2009); rice hoja blanca virus (Hans et al. 2009); rice dwarf virus (Shimizu et al. 2009); barley stripe mosaic virus (Cakir and Tor 2010); tobacco rattle virus (Andre et al. 2009); citrus tristeza virus (Soler et al. 2011); mungbean yellow mosaic India virus (Haq et al. 2010); rice tungro bacilliform virus (Himani et al. 2008) and many other examples for RNAi application in plant viruses are demonstrated in Table 14.1.

As plant viruses have to deal with RNAi-based defense mechanisms, different silencing processes mediate effective defense mechanisms against viruses in plants and function as the major inducible defense reactions against plant viral pathogens. On the other hand, some viruses can generate a counteracting activity to overcome this mechanism; plant viruses produce silencing suppressors that allow them to proliferate in their specific hosts. Suppressors of gene silencing are often identified in

Table 14.1 Effects of targeted regions of RNAi in various plant virus systems

Virus	Plant host	Targeted region	References
DNA viruses			
Tomato leaf curl ToLCV	*Nicotiana benthamiana* and *Solanum lycopersicum*	PCNA	Pandey et al. (2009)
Tomato yellow leaf curl (TYLCV)	*N. benthamiana* *L. esculentum*	AV1, AV2, AC1	Rezk et al. (2006) Abhary et al. (2006)
African cassava virus	*N. benthamiana* *N. esculenta*	CYP79D1, CYP79D2	Fofana et al. (2004)
Abutilon mosaic virus (AbMV)	*N. benthamiana*	AC1	Krenz et al. (2010)
Beet curly top virus (BCTV)	*Spinacea oleracea* *S. lycopersicum*	rbcS, ChlI	Golenberg et al. (2009)
Cabbage leaf curl virus (CaLCuV)	*N. benthamiana* *Arabidopsis thaliana*	sulfur desaturase (Su), GUS, NtEDS1, GFP, CH42, PDS	Huang et al. (2011); Turnage et al. (2002)
Cotton leaf crumple virus (CLCrV)	*Gossypium* spp.	Magnesium chelatase subunit I gene (ChlI) or the phytoene desaturase gene (PDS)	Tuttle et al. (2008)
Grapevine virus A (GVA)	*N. benthamiana* and *Vitis vinifera*	PDS	Muruganantham et al. (2009)
Pepper huasteco yellow vein virus (PHYVV)	*Capsicum* spp.	Comt (encoding a caVeic acid O-methyltransferase), pAmt, and Kas (a-keto-acyl-[acylcarrier-protein] synthase)	del Rosario Abraham-Juarez et al. (2008)
Rice tungro bacilliform virus (RTBV)	*O. sativa*	(pds)	Purkayastha et al. (2010)
Tomato golden mosaic virus (TGMV)	*N. benthamiana*	Su and luc	Carrillo-Tripp et al. (2006)
Tomato yellow leaf curl China virus (TYLCCNV) Betasatellite	*N. benthamiana* *N. glutinosa*, *N. tabacum* *Lycopersicon esculentum*	PCNA, PDS, SU, GFP	Tao and Zhou, (2004), He et al. (2008)
Tobacco curly shoot virus (TbCSV) Betasatellite	*N. benthamiana*	GFP, SU	Qian et al. (2006)

(continued)

Table 14.1 (continued)

Virus	Plant host	Targeted region	References
DNA viruses			
RNA viruses			
Tobacco rattle virus (TRV)	*N. benthamiana*, *Arabidopsis*, tomato, *Solanum* species, chili pepper, opium poppy, *Aquilegia*	Rar1, EDS1, NPR1/NIM1, pds, rbcS, gfp	Ratcliff et al. (2001), Gould and Kramer (2007), Di Stilio et al. (2010)
Tomato bushy stunt virus (TBSV)	*N. benthamiana*	Magnesium chelatase (ChlH) and phytoene desaturase (PDS) GFP	Hou and Qiu, (2003)
Barley yellow dwarf virus (BYDV)	*Hordeum vulgare* and *Oryza* spp.	BYDV-PAV	Wang et al. (2000)
Turnip yellow mosaic virus (TYMV)	*A. thaliana*	PDS	Pflieger et al. (2008)
Plum pox virus (PPV)	*Nicotiana benthamiana*	(RNA-depending RNA polymerase). RDR6	Vaistij and Jones (2009)
Potato virus X (PVX)	*N. benthamiana* *S. tuberosum*	RDR6, PDS, CP, GFP	Vaistij and Jones (2009) Faivre-Rampant et al. (2004) Soliman et al. (2008)
Banana bract mosaic virus (BBrMV)	Banana	CP	Rodoni et al. (1999)
Cucumber mosaic virus (CMV)	*Glycine max*	chalcone synthase (CHS)	Nagamatsu et al. (2007)
Tobacco mosaic virus (TMV)	*N. benthamiana*	PDS, PSY, ChlH	Kumagai et al. (1995) Hiriart et al. (2003)
Poplar mosaic virus (PopMV)	*Lathyrus odorata* *Medicago truncatula* *N. benthamiana*	GFP	Naylor et al. (2005)
Pea early browning virus (PEBV)	*P. sativum* *Lathyrus odorata* *Medicago truncatula*	pspds, uni, kor, PDS, Nodule inception gene (Nin)	Grønlund et al. (2008); Constantin et al. (2008); Constantin et al. (2004)
Brome mosaic virus (BMV)	*Oryza sativa, Zea mays*, *Hordeum vulgare*, *Festuca arundinacea*, *N. benthamiana*	pds, actin 1, rubisco activase	Ding et al. (2006); Scofield and Nelson (2009); van der Linde et al. (2011)

silencing reversal assay by using green fluorescence protein (GFP) transgenic plants (Goldbach et al. 2003; Voinnet et al. 1999). Suppression of helper component protease (HCPro) that encoded by potyviruses and cucumoviral 2b protein in CMV virus are one of the best-studied suppressors of RNA silencing. Some examples for RNA-silencing suppressors encoded by plant viruses are shown in Table 14.2.

The importance of these suppressors is reflected by the fact that they have been previously identified as pathogenicity factors, or as viral cell-to-cell or long-distance movement proteins, which they are essential for infectivity in indicated hosts. The suppressor proteins encoded by unrelated RNA and DNA viruses bear no similarity to each other in either coding sequence or protein structure, suggesting separate origins and variable functional mechanisms for each suppressor type. Therefore, suppressor proteins can be used as tools to dissect or disturb silencing pathways at different points, and to analyze the effects of these disturbances to host gene expression patterns, development and phenotype (Pumplin and Voinnet 2013).

14.5.3 Bacteria

To date, very few publications have appeared on the use of RNAi against plant fungal or bacterial infections. One of the conspicuous examples of plant bacterial disease management is where RNAi showed a significant type of gene regulation (Escobar et al. 2001). They are reported to control crown gall disease targeting the process of gall formation (tumorigenesis) in plants by initiating RNAi against the iaaM and ipt oncogenes. The transformed plants *Arabidopsis thaliana* and *Lycopersicon esculentum* with RNAi constructs, targeting iaaM and ipt genes, both showed resistance to crown gall disease. These transgenic plants contained a dsRNA version of these two bacterial genes, infecting bacteria were unable to synthesize hormones needed to cause tumors and wild-type plants were hyper-susceptible to *Agrobacterium tumefaciens*. This procedure can be used to develop resistance in ornamental and horticultural plants against crown gall tumorigenesis caused by plant pathogens such as *Albugo candida*, *Synchytrium endobioticum* and *Erwinia amylovora* (Wani et al. 2010).

Katiyar-Agarwal et al. (2006) used RNAi and found endogenous siRNA, nat-siRNA ATGB2 that was strongly induced by *Pseudomonas syringae* in *Arabidopsis*. They demonstrated that the biogenesis of this siRNA requires DCL1, HYL1, HEN1, RDR6, NRPD1A and SGS3, and its induction depends on the cognate host disease resistance genes RPS2 and NDR1. This siRNA contributes to RPS2-mediated race-specific disease resistance by repressing PPRL, a putative negative regulator of the RPS2 resistance pathway against *P. syringae* pv. *tomato* in *Arabidopsis*. Katiyar-Agarwal et al. (2007) also found another class of siRNA, 30–40 nucleotides in length, termed long-siRNAs (lsiRNAs), associated with *P. syringae*. One of these lsiRNAs, AtlsiRNA-1, plays a role in plant bacterial resistance by silencing AtRAP, which is a negative regulator of plant defense.

Table 14.2 RNA-silencing suppressors encoded by plant viruses

Virus	Genus	Suppressor protein	Suppression	References
DNA viruses				
African cassava mosaic virus (ACMV)	Begomovirus	AC2	Compete	Voinnet et al. (1999); Van wazel et al. (2002); Jada et al. (2013)
Mungbean yellow mosaic virus (MYMV)	Begomovirus	AC2	Compete	Voinnet et al. (1999); Van wazel et al. (2002)
Tomato yellow leaf curl virus (TYLCV)	Begomovirus	AC2	Compete	Voinnet et al. (1999); Van wazel et al. (2002)
Tomato yellow leaf curl China virus (TYLCCNV)	Begomovirus	RDR6	complete	Li et al. (2014)
Cauliflower mosaic virus (CaMV)	Caulimovirus	P6	Complete	Pumplin and Voinnet (2013)
RNA virus				
Potato virus X (PVX)	Potexvirus	P25	Partial	Jada et al. (2013); Voinnet et al. (2000)
Potato virus Y	Potyvirus	HC-Pro	Complete	Jada et al. (2013); Brigneti et al. (1998)
Barley yellow mosaic virus (BaYMV)	Bymovirus	Gamma B	Complete	Kubota et al. (2003)
Barley stripe mosaic virus (BSMV)	Hordeivirus	Gamma B	Complete	Yelina et al. (2002)
Cowpea mosaic virus (CPMV)	Comovirus	S protein	Partial	Liu et al. (2004); Voinnet et al. (1999)
Peanut clump virus (PCV)	Pecluvirus	P15	Complete	Dunoyer et al. (2002);
Cucumber mosaic virus (CMV)	Cucumovirus	2b	Partial	Brigneti et al. (1998); Pumplin and Voinnet (2013)
Turnip yellow mosaic virus (TYMV)	Tymovirus	P69	Complete	Chen et al. (2004)
Turnip crinkle virus (TCV)	Carmovirus	P38	Complete	Pumplin and Voinnet (2013)
Rice hoja blanca virus (RHBV)	Tenuivirus	NS3	Complete	Bucher et al. (2003)
Rice yellow mosaic virus (RYMV)	Sobemovirus	P1	Complete	Voinnet et al. (1999); Pumplin and Voinnet (2013)
Tobacco mosaic virus (TMV)	Tobemovirus	P30	Complete	Voinnet et al. (1999)

14.5.4 Fungi

Management of fungal pathogens usually depends on fungicides that are widely used to control phyto-pathogenic fungi. Despite their effect on fungal infections, fungicides have a potentially hazardous effect on the environment and human health. Most pathogenic fungi have evolved various resistant mechanisms to fungicides through genetic adjustment, by which the pathogen becomes less sensitive to agrochemicals (Yang et al. 2008), placing the agricultural industry in search of novel and alternate strategies to manage fungal pathogens. Gene silencing offers an ideal strategy for control of fungal diseases and RNAi offers an emerging strategy by silencing the vital genes that associate with such pathogens. There are many successful examples describing the application of RNAi against fungal pathogens such as; *Aspergillus nidulans*, *Blumeria graminis*, *Cladosporium fulvum*, *Cochliobolus sativus*, *Fusarium graminearum*, *F. verticillioides, Magnaporthae oryzae*, *Neurospora crassa, Phytophthora infestans*; *Puccinia striiformis* f. sp. *tritici* and *Venturia inaequalis*. (Karthikeyan et al. 2013; Wani et al. 2010). The hypermorphic mechanism of RNAi suggests that this technology can be applicable to all those plant pathogenic fungi which are polyploid and polykaryotic in nature, and provides a solution for the problem where frequent lack of multiple marker genes in fungi is practiced. Nowara et al. (2010) reported expression of dsRNA and antisense transcripts that are specific to powdery mildew fungi *Blumeria graminis* in wheat and barley. Yin et al. (2011) used the VIGS method to introduce gene fragments from the rust fungi *Puccinia striiformis* f. sp. *tritici* to plant cells and reduce the expression of the corresponding genes in the rust fungus during infection. Transgenic potato plants expressing RNAi construct against the plasma membrane-localized SYNTAXIN-RELATED 1 (StSYR1) gene showed increased resistance against *Phytophthora infestans* (Eschen-Lippold et al. 2012). Enhanced resistance of StSYR1-RNAi in transgenic potato plants was correlated with enhanced levels of salicylic acid. Recent reports also showed the effect of synthetic dsRNA against *Fusarium oxysporum* f. sp. *cubense* and *Mycosphaerella fijiensis* infecting banana (Mumbanza et al. 2013). In another spectrum, scientists are using host-induced gene silencing (HIGS) to target fungal genes; fungal sterol 14α-demethylase (CYP51) was used in one report to restrict the infection caused by *Fusarium* spp., scientists capitalized on the knowledge that demethylation inhibitor fungicides targets cytochrome P450 lanosterol C-14α- demethylase, this enzyme is encoded by three paralogous genes in *Fusarium graminearum*; therefore, they developed transgenic *Arabidopsis* and barley plants expressing dsRNA targeting all three CYP51 genes and exhibited complete immunity against *F. graminearum*.

14.5.5 Insects

Food production has always faced challenges to keep crops free from pests. Plant breeding and transgenetics are approaches that have invested enormous efforts to develop pest-resistant crops to minimize the cost and environmental hazards of chemical pesticides. Pests, including insects and nematodes, share a common RNAi pathway with most eukaryotes when dsRNA is expressed to trigger vital genes within the pests. The expression of dsRNA in transgenic plants, through hd-RNAi technology, which is directed against the feeding pest, presents a new dimension to plant transgenic and breeding programs (Zha et al. 2011).

Exogenous RNAi technology targeting vital pest genes has offered a powerful solution to down-regulate crucial physiological pest genes to control nematodes and insects (Xiong et al. 2013). Direct injection experiments of dsRNA into insects and nematodes, targeting vital genes, demonstrated the successful effectiveness of RNAi technology in controlling agricultural pests (Mutti et al. 2006; 2008). The delivery of dsRNA into insects and nematodes by expressing dsRNA in host plants showed a significant reduction of feeding damage in vitro. Baum et al. (2007) controlled western corn rootworm (WCR) by feeding this diet. Furthermore, Mutti et al. (2008), when RNAi technology was performed on aphids through direct injection of dsRNA to silence the C002 gene in salivary glands, results showed lethality of aphids on plants but not on artificial diet, suggesting that C002 has a function in aphid interaction with the host plant. Other experiments on aphids have developed an hd-RNAi technology by targeting the Rack-1 gene, which is predominantly expressed in the gut, and C002 gene which is predominantly expressed in the salivary glands. Aphids were fed *Nicotiana benthamiana* leaf disks, transiently producing dsRNA, and on *Arabidopsis thaliana* plants, stably producing the dsRNAs; results showed that C002 and Rack-1 expression were eliminated by up to 60 % on transgenic *A. thaliana* (Pitino et al. 2011). Moreover, pest transcriptional factor genes has been targeted by hd-RNAi strategies against cotton bollworm (*Helicoverpa armigera*) and showed a significant decrease in mRNA and protein levels during larval development stages, resulting in deformity and larval lethality (Xiong et al. 2013).

The hd-RNAi strategy is now widely used in pest control management, especially against nematodes (Karthikeyan et al. 2013); successful studies have highlighted a significant variation of RNAi that results from dsRNA ingestion and the down-regulation of target genes through hd-dsRNA in nematodes (Sindhu et al. 2009). Various plant species are being genetically modified to express dsRNA targeted against nematodes through different developmental stages (Karthikeyan et al. 2007) with future plans to integrate these transgenics into plant breeding programs. Prentice et al. 2015; demonstrated the functionality of the RNAi pathway in the African sweetpotato weevil (SPW) *Cylas puncticollis* by applying dsRNA nano-injection targeting laccase2 gene, which involved in the insect cuticle tanning for sclerotization of insect exoskeleton. The body of treated insects showed inhibition of sclerotization, leading ultimately to death of the insect. The effective down

regulation of laccase2 gene may indicate to the potential of *C. puncticollis* to initiate a systemic RNAi response.

14.6 Improvement of Nutritional Quality

Recent advances in plant breeding techniques have focused on improving yield, disease resistance, tolerance to biotic and abiotic stress and quality traits of the crop. The success of plant breeding depends on the genetic variation of crossed plant species and the spectrum of the genetic pool available, which can restrict the breeding process to limited traits. RNAi technologies add a new tool for plant breeders interested in obtaining new quality traits by down-regulating a certain gene or set of genes in targeted plants and overcoming technical challenges in metabolic engineering programs. The sequence specific RNAi technology, achieved by expressing non-coding RNAs, has shown a high degree of stability in plant biochemical pathways (Baykal and Zhang 2010) and provides a new genetic source of metabolic engineering in plant breeding programs. The success of using RNAi technology is not limited to only overcoming viral, bacterial or fungal diseases, and crop pest damages, but also to improve quality traits in many plant species.

14.6.1 Amino Acid Content

The amino acid content of food and feed is considered one of the most important agronomical traits, where plant breeding and transgenesis programs are racing to provide crops with enhanced levels of essential amino acids and storage proteins. Increasing the content of essential amino acids requires detailed study of its biosynthetic pathways, which have already been successful used to engineer model plants (Ufaz and Galili 2008). The study of amino acid pathways started soon after the discovery of gene silencing; for example, tobacco, a model research plant in gene silencing studies, was engineered with *Arabidopsis* S-Adenosyl-$_{L}$-methionine synthase gene (SAM-S) and suppressed by RNAi; mature transgenic plants showed an increase in $_{L}$-methionine levels (Boerjan et al. 1994). Maize has become a model for lysine modification studies through RNAi technology; successful experiments of lysine modification studies in maize came after an extensive study of its biosynthetic pathway. High lysine content can affect the germination of seeds and the general phenotype of the plant (Segal et al. 2003; Tang and Galili 2004). More recently, RNAi has been used to modify the lysine content in maize by expressing a lysine synthesis enzyme and silencing a lysine degradation enzyme (Frizzi et al. 2008); the resulting transgenic maize exhibited a significant increase in lysine amino acid content. Other reports have demonstrated the effect of silencing, using VIGS, on manipulating the levels of aspartate, glutamate and lysine amino acids through the suppression of the plastid NbAsp5 and NbPAT genes, respectively (de

la Torre et al. 2014). Using RNAi technology to suppress the regulation and degradation of amino acids biosynthetic pathways is considered an advantage for plant breeders, where the genetic pool is expanding and providing new alternatives.

14.6.2 Allergens

Food allergy is a serious health problem affecting children and adults; it is a form of immune response against certain food proteins that are identified as harmful or toxic to the body. Allergic reactions range from mild to severe, depending on the digestibility of those proteins; un-digested proteins are usually tagged by the immunoglobulin E (IgE) and attacked by white blood cells causing various symptoms and physiological disorders in the body. Allergens and toxins are found in various plant species, especially wild-types (Morandini 2010); with the latter, the breeding process through crossing with wild-types is a prolonged and difficult task to achieve and reach domestication in such hybrids. The technology of inactivating specific individual genes, allergens or toxins, is available through RNAi, when such allergenic plant proteins or toxins are not involved in the germination or maturation of these plants.

Soybean (*Glycine max*) is a widely used crop in processed food and known to possess various quantities of P34 proteins, recognized allergens to soybean-sensitive individuals. Dealing with this problem represents a successful example of using gene silencing to prevent the accumulation of such allergenic proteins, without affecting the germination or maturation of the transgenic plant (Herman et al. 2003). Potential allergens in rice, α-globulin and β-glyoxalase, were also silenced using RNAi constructs in a mutated α-amylase/trypsin inhibitor rice line, producing transgenic rice seeds which are free from the three potential allergens that could be used in breeding programs (Wakasa et al. 2011). Other examples of reducing potential allergens in food crops include Ara h2 protein in peanuts (Dodo et al. 2008), Mal d 1 allergen in apple (Gilissen et al. 2005) and Lyc e 1.01 and Lyc e 1.02 allergens in tomatoes (Le et al. 2006). On the other hand, pollen grain allergen such as the Lol P 5 protein, the major allergen in ryegrass pollen, was suppressed by using the antisense strand of the gene to reduce the allergen without affecting the fertility of the pollen grains (Bhalla et al. 1999).

14.6.3 Metabolic Pathways

Plants have a vast network of metabolic pathways, an astonishing diversity of products that are unique to some plant species, producing primary and secondary metabolites that are used in alternative medicine and cannot be found in other organisms (Zorrilla-López et al. 2013). Metabolic engineering in plants requires extensive investigations and study to understand the biochemical pathway networks and the

function of enzymes involved in certain pathways. One way to study metabolic pathways in a plant is reverse genetics (Alhagdow et al. 2007), where losing the function of a specific enzyme can lead to understanding the job performed by the enzyme and the physiological effects on the whole plant. Gene silencing is considered the best candidate tool to study the function of enzymes through reverse genetics and to engineer plant metabolic pathways to overproduce secondary products (Flores et al. 2008), by silencing regulatory enzymes that are economically important such as sugars, fatty acids, aromatic fragrances, steroids and drugs.

To study the effect of $_{L}$-Galactono-1,4-lactone dehydrogenase enzyme in tomatoes, a mitochondrial enzyme catalyzing the last step of vitamin C pathway, Alhagdow et al. (2007) used RNAi to silence SlGalLDH genes, producing this enzyme in tomatoes and found that many secondary metabolites, related to stress response, were affected by the reduction of GalLDH and plant structure, physiology and growth related processes were all compromised. The usage of RNAi technology to improve a quality trait in plants demands a change without affecting the general growth and seed germination of the plant; successful examples of using RNAi in engineering the metabolic pathways without affecting the yield or the plant usually come only after extensive studies on mutants. Studying the Ipa1 mutant maize has inspired scientists to silence the expression of ATP-binding cassette (ABC) transporter, which resulted in reduction of phytic acid and an increase of phosphate in transgenic maize, without affecting the germination or growth of the plant (Shi et al. 2007). Another example of modifying the food quality is transgenic wheat with increased amylose content, where the amylopectin synthesis pathway was modified through inactivated SBEIIa genes by RNAi technology, producing high amylose starch that is beneficial to human digestion (Sestili et al. 2010). Other reports have focused on the applications of RNAi to modify the content of secondary metabolites, which can be removed by chemical or physical processes, but will affect the flavor or characteristics of the plant. Caffeine, for example, was targeted in coffee beans by repressing the theobromine synthase (CaMXMT1) gene using RNAi technology, resulting in 70 % reduction in leaves (Ogita et al. 2003). Furthermore, low caffeine tea was produced in another study by silencing the caffeine synthase (CS), which resulted in the reduction of caffeine in the range of 44–61 % between transgenic tea lines (Mohanpuria et al. 2011). Lowering the α-linolenic acids in soybeans, to avoid hydrogenation processes, was accomplished by using RNAi to suppress the FAD3 genes. Researchers used a conserved region between the three FAD3 genes to build an hpRNAi construct, which they transformed into soybeans, resulting in 7–10 % reduction of α-linolenic acids (18:3) (Flores et al. 2008). Many other examples of modifying secondary metabolites in plants by RNAi have demonstrated the stability and inheritance of the trait, which could be used for breeding (Baykal and Zhang 2010).

14.7 Tolerance to Abiotic Stress

Abiotic stress tolerance in plants is developed as a responsive behavior to overcome sudden changes in the environment, complex physiological and biochemical adaptation mechanisms towards temperature, water availability, light and soil nutrients. Molecular biology tools have revealed that abiotic stress tolerance involves a network of biochemical events and many genes which operate under stress conditions (Senthil-Kumar and Udayakumar 2010), where RNAi applications are useful to discover and study the genes involved in abiotic stress-resistance pathways. In a general perspective, Vanderauwera et al. (2007) found that the poly-(ADP-ribose) polymerase (PAPR) represent a network connection-key between stress-resistance pathways. They repressed the *PAPR* gene in *Arabidopsis thaliana* and *Brassica napus* by RNAi constructs, and found that the transgenic plants had a broad spectrum of abiotic stress-resistant phenotypes and huge transcriptional changes within the stress-resistant genes, analyzed by DNA microarray.

Transcriptome analysis of model plants, such as *Arabidopsis*, revealed that a large amount of non-coding RNAs (ncRNA), except for housekeeping ncRNA, are highly expressed during abiotic stress conditions. Identified ncRNAs are involved in chromatin remodeling and translational repression (Matsui et al. 2013). The miRNAs were also identified in action during abiotic stress, playing major roles in up-regulating or repressing gene expression. In a study, two miRNAs precursors, miR441 and miR446, were found to be positive regulators of abscisic acid (ABA) signaling and tolerance to abiotic stress (Yan et al. 2010). Current and future studies in the subject of abiotic stress are more focused on expressing miRNAs that are related to abiotic stress-resistance in crop plants.

14.8 Flowers

Flower fragrances have long been a source of attraction to people, but for the plant itself helps in attracting pollinators for reproduction. As a source of attraction, breeding programs and market value depend on coloration, scent and flowering time. Therefore, flowering related pathways have been studied with special emphasis on color pigments (To and Wang 2006). The production of colored flowers and controlling the fragrances is a recent field of interest where RNAi approaches are being applied. Gene silencing strategies, using hpRNA or VIGS, are also being studied to control the timing of the transition phase from vegetative to reproductive floral tissues in plants.

14.8.1 Flowering Pigments

The relationship between flower pigments and RNAi began to be understood in the early 1990s, with the first discovery of the RNAi phenomenon involving petunia flowers and their pigmentation, in an experiment designed to change the flower color by overexpressing a component of the flavonoid pathway (Napoli et al. 1990). Flavonoids (anthocyanins) are the main pigments in flowers that produce attractive color patterns from yellow to purple (Fig. 14.2). The biosynthesis and regulation of the anthocyanin pathway compounds is one of the best studied pathways in plants, being a target for biotechnological studies to alter and manipulate this pathway, creating new coloration patterns (Grotewold 2006; He et al. 2010). The overexpression or suppression of one component in the flavonoid pathway results in a color change in the flowers; dsRNA is constantly used to suppress enzymes of the flavonoid pathway to eliminate or to accumulate a certain secondary metabolite in flowers (Tanaka et al. 2008). For example, chalcone isomerase (CHI) enzyme in the flavonoid biosynthesis pathway, down-regulated in tobacco, resulted in reduced pigmentation of petals and accumulation of chalcone in pollen grains, which turn yellow (Nishihara et al. 2005). Back in the 1840s, horticultural societies in Britain and Belgium offered a grand prize to any breeder who could make a true blue rose, a task that was not then accomplished. Katsumoto et al. (2007) manipulated the flavonoid pathway in *Rosa hybrida*. They overexpressed the dihydroflavonol 4-reductase (DFR) gene from *Iris* x *hollandica* and the viola flavonoid 3′,5′ hydroxylase (F3′5′H) gene and suppressed the endogenous DFR gene with RNAi; results showed true blue roses due to the accumulation of delphinidin in the petals. In another study, delphinidin levels were successfully increased up to 80 % in chrysanthemum flowers by silencing the endogenous (F3′5′H) gene and expressing a rose F3′5′H gene, resulting in novel bluish petal colors impossible to obtain by standard breeding (Brugliera et al. 2013).

Nevertheless, delphinidin is not the only metabolite resulting in blue-colored petals; Aida et al. (2000) managed to down-regulate DFR by RNAi and overexpress CHS gene, resulting in accumulation of flavones in the wishbone flower (*Torenia fournieri*) petals and the co-pigmentation resulted in bluer flowers.

On the other end of the spectrum, orange-to-red colored flowers were successfully engineered using RNAi in model tobacco plants. Nakatsuka et al. (2007) suppressed both FLS and F3′H genes in a chimeric RNAi construct and overexpressed the DFR gene in order to accumulate pelargonidin. Transgenic tobacco flowers contained low amounts of flavonol content and high pelargonidin, resulting in orange-to-red colored tobacco petals. Moreover, manipulating the red color intensity of the orchid *Phalaenopsis* cultivar flowers was achieved by suppressing the last enzyme of the pathway (Chen et al. 2011), flavonoid 3-O-glucosyltransferase (3GT). Chen et al. (2011) were able to produce faded-red colored *Phalaenopsis* flowers by gradually reducing the anthocyanin content in petals, resulting in transgenic plants with different degrees of faded red coloring.

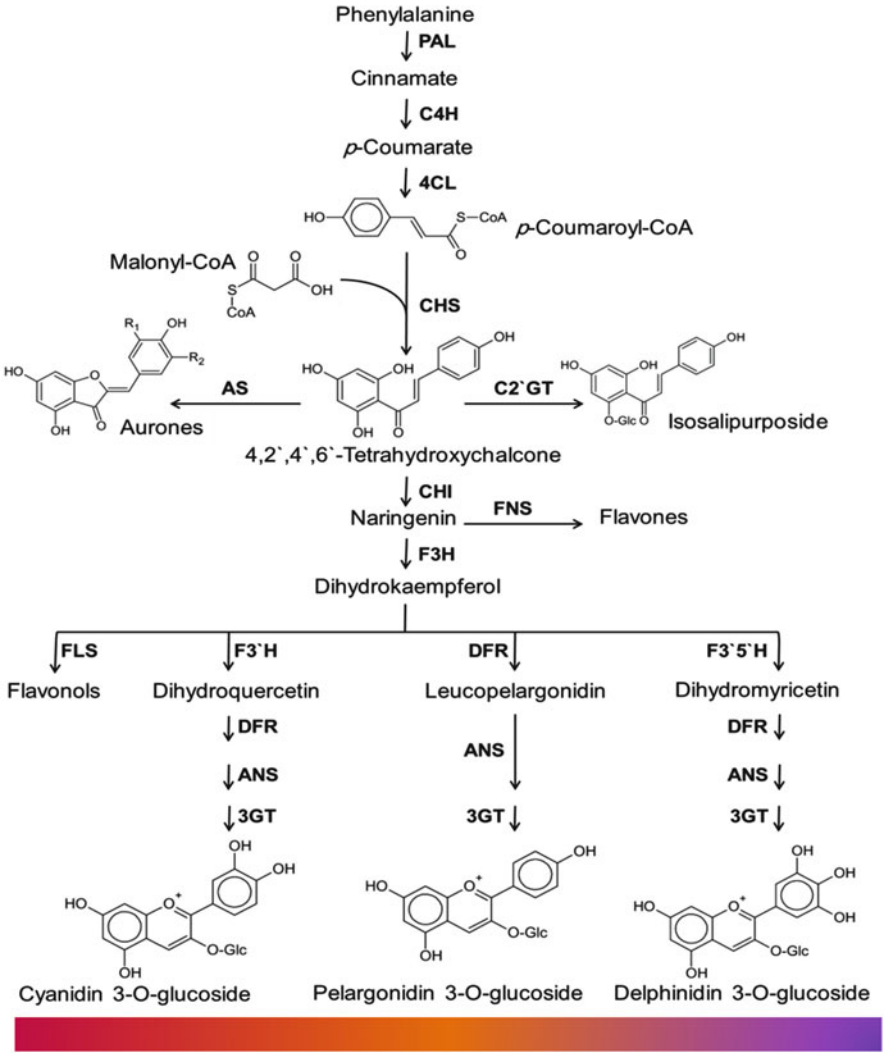

Fig. 14.2 Anthocyanin biosynthetic pathway. The three color pigments of flavonoids in flowers and the catalyzing enzymes. The full biosynthetic pathway can be obtained from http://www.genome.jp/kegg/pathway.html

14.8.2 Flowering Fragrance

The distinctive scent of floral bouquets is the result of volatile compound emissions, a mixture of compounds designed by the plant to recruit pollinators or to repel some florivores during flowering time (Kessler et al. 2008). Volatile compounds are mostly products of the benzenoid and phenylpropanoid pathways, regulated by pollination status and the developmental stage of the plant (Koeduka et al. 2009).

Studies on petunia flowers showed that benzenoid and phenylpropanoid compounds are synthesized and coupled by enzymes from the shikimate pathway. Colquhoun et al. (2010) demonstrated that *Petunia hybrida* chorismate mutase (PhCM1) is one of the early enzymes responsible for coupling of shikimate pathway metabolites, their results showed that silencing the PhCM1 gene in petunia leads to a 60–70 % reduction of benzenoids and phenylpropanoid volatile compounds. In another study, Verdonk et al. (2005) investigated the role of ODORANT1 (ODO1) transcription factor as a circadian regulator of the shikimate pathway and they found that down-regulating ODO1 in transgenic petunia resulted in decreasing volatile emissions and flowers had less fragrance. Suppression of ODO1 also resulted in down-regulating a number of genes in the shikimate pathway. Another transcription regulator, HD20, was studied by Ré et al. (2012) in *Nicotiana attenuata* plants, which opens its flowers and emits fragrance at night. The study revealed that HD20 regulates the expression of several genes related to volatile oil metabolism and suppresses the HD20 gene resulting in increased emission of benzyl acetone (BA) day and night without affecting plant growth or development.

RNAi technology was also shown to be effective in changing the concentrations of volatile products; fragrant volatile oils, such as eugenol and isoeugenol, used in dentistry (Ferracane 2001), are synthesized by EGS1 and IGS1 enzymes, respectively, from a quinone methide-like intermediate (Koeduka et al. 2009). Using RNAi suppression of IGS1, the biosynthesis of isoeugenol was shifted towards eugenol biosynthesis, where high concentrations of eugenol accumulated in petunia petals (Koeduka et al. 2009). Spitzer et al. (2007) used VIGS to suppress the expression of several fragrant-related genes in petunia flowers. When benzioc acid/salicylic acid carboxyl methyltransferases were suppressed; flowers exhibited a 7-fold reduction in methyl-benzoic acid (MeBA) and a 10-fold reduction in methyl-salicylic acid (MeSA). Moreover, silencing phenylacetaldehyde synthase (PAAS) resulted in the reduction of phenylacetaldehyde. On the other hand, the importance of repellent scents has been demonstrated by RNAi technology on many fragrancy-related genes. Kessler et al. (2008, 2012) showed that a complex relationship between attracting pollinators and repelling florivors is sculpted in each floral species to overcome the selective pressure in its environment.

14.8.3 Flowering Time

Beside colors and fragrances, flowering time (anthesis) can also be manipulated by RNAi. The set timing, from vegetative to reproductive growth in plants, is tightly regulated by a set of pathways controlling this transition phase via specific cellular signals that determines the appropriate time of transition (Bäurle and Dean 2006). The FLOWERING LOCUS T 1 (*FT1*) gene in wheat, a homologue to the *Arabidopsis* vernalization (*VRN3*) gene, was reported as a key protein regulating flowering pathways in *Brachypodium* and wheat leaves. Silencing the FT1 gene showed that several flowering pathways and FT-like genes were down-regulated in *Brachypodium*

and wheat, resulting in a late flowering phenotype in wheat and no flowering in *Brachypodium* (Lv et al. 2014). In pineapple, Trusov and Botella (2006) produced late flowering phenotype pineapples by silencing the A1-amino-cyclopropane-1-carboxylate synthase (ACACS2) gene, an interesting trait that generates reproductive parts after generating enough vegetative tissues to support producing fruits, also providing enough adaptation to the surrounding environment before flowering. The flowering regulation pathways are usually branched into other pathways regulating growth and development of the plant (Freiman et al. 2012); this fact was illustrated by Li et al. (2011) when they used RNAi to silence the S-Adenosyl-1-methionine synthase (SAMS) genes 1, 2 and 3 in rice. Transgenic rice with a hairpin OsSAMS1, 2 and 3 showed dwarfism and delayed germination phenotypes besides late flowering. In a chimeric gene silencing strategy, heading date 3a (Hd3a) gene in rice, coding for a flowering transition signal and the RICE FLOWERING LOCUS T1 (RFT1) gene were both silenced using RNAi in rice (Komiya et al. 2008). The double silenced transgenic rice plants exhibited a late flowering phenotype, where silencing each gene alone did not affect flowering in rice.

On the other hand, RNAi approaches have been used to generate early flowering phenotype plants; *Arabidopsis* EARLI1 gene, coding for a proline-rich protein, was down-regulated by RNAi to study its function. Transgenic *Arabidopsis* plants with silenced EARLI1 gene showed early flowering and lower lignin content (Shi et al. 2011). In trees, Szankowski et al. (2009) developed transgenic apple trees with early flowering phenotype. They used RNAi to silence the MdTFL1 gene and showed that transgenic plants were able to generate flower buds and axillary shoots within few months of planting. Early flowering apple trees were also generated using VIGS strategy. Yamagishi et al. (2013) used apple latent spherical virus (ALSV) with a silencing trigger for the MdTFL1 gene to accelerate flowering time in apple trees. Moreover, Frieman et al. (2012) showed that silencing the PcTFL1-1 and PcTFL1-2 genes in pear (*Pyrus communis*) trees produced an early flowering trait without affecting the morphology, pollination or fruit formation in transgenic trees. Early flowering pear trees showed stable and inherited trait, an interesting breeding source which accelerates the breeding period in trees (Freiman et al. 2012).

14.9 Conclusions and Prospects

RNAi stability and trait inheritance in transgenic plant progenies is a very important subject to be accessed in future studies; genetic and phenotypic trait inheritance of RNAi has been confirmed in roundworm (*Caenorhabditis elegans*) by Tavernarakis et al. (2000), but should be further investigated in plants. Naturally-existing miRNAs in plants have evolved and been transferred through generations of inbred and hybrid lines of cultivated crops, such as maize and rice, along with the mechanism of RNAi and showed high stability.

The opportunities presented by RNAi technology to agriculture, by switching-off specific genes, enriches the genetic pool for plant breeders and expands the

spectrum of traits that can be transferred by introgression of stable RNAi transgenic plants with commercial crops, breeding for superior cultivars with disease resistance, stress tolerance, nutritional values, high yield, eliminated allergens, secondary metabolites and many other features that are impossible to obtain in one cultivar by conventional breeding. In trees, breeding takes a very long time to obtain a certain trait and it takes more time to proliferate the desired line. RNAi signals can be transmitted through grafting; this feature can help breeders with a shortcut by grafting stable RNAi transgenic tree cuttings on non-transgenic scions to acquire the trait in a short time.

This review illustrates some opportunities of RNAi technology in plants that can be used in breeding programs to provide better horticultural traits, which are urgently needed to face the changing global climate conditions and the accompanied agricultural challenges.

References

Abhary MK, Anfoka GH, Nakhla MK, Maxwell DP (2006) Post-transcriptional gene silencing in controlling viruses of the tomato yellow leaf curl virus complex. Arch Virol 151(12):2349–2363

Ahlquist P (2002) RNA-dependent RNA polymerases, viruses, and RNA silencing. Science 296(5571):1270–1273

Aida R, Yoshida K, Kondo T et al (2000) Copigmentation gives bluer flowers on transgenic torenia plants with the antisense dihydroflavonol-4-reductase gene. Plant Sci 160:49–56

Alhagdow M, Mounet F, Gilbert L et al (2007) Silencing of mitochondrial ascorbate synthesizing enzyme $_{L}$-Galactono-1.4-Lactone dehydrogenase affects plant and fruit development in tomato. Plant Physiol 145:1408–1422

Ambros V (2004) The functions of animal microRNAs. Nature 431(7006):350–355

Andre CV, Suma C, Gregory BM (2009) Virus-induced gene silencing (VIGS) in *Nicotiana benthamiana* and tomato. J Vis Exp 28:1292–1301

Andrieu A, Breitler J, Sire C et al (2012) An *in planta*, *Agrobacterium*-mediated transient gene expression method for inducing gene silencing in rice (*Oryza sativa L.*) leaves. Rice 5:23

Axtell MJ, Bartel DP (2005) Antiquity of microRNAs and their targets in land plants. Plant Cell 17(6):1658–1673

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116(2):281–297

Baum JA, Bogaert T, Clinton W et al (2007) Control of coleopteran insect pests through RNA interference. Nature Biotechnol 25:1322–1326

Baurle I, Dean C (2006) The timing of developmental transitions in plants. Cell 125:655–664

Baykal U, Zhang Z (2010) Small RNA-mediated gene silencing for plant biotechnology. In: Catalano AJ (ed) Gene silencing: theory, techniques and applications. Nova Sci Publishing Inc, New York, pp 254–269

Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409:363–366

Bhalla PL, Swoboda I, Singh MB (1999) Antisense-mediated silencing of a gene encoding a major ryegrass pollen allergen. Proc Natl Acad Sci U S A 96:11676–11680

Bian XY, Rasheed MS, Seemanpillai MJ, Rezaian MA (2006) Analysis of silencing escape of tomato leaf curl virus: an evaluation of the role of DNA methylation. Mol Plant Microbe 19:614–624

Boerjan W, Bauw G, Van Montagu M, Inze D (1994) Distinct phenotypes generated by overexpression and suppression of S-Adenosyl-$_L$-Methionine Synthase reveal developmental patterns of gene silencing in tobacco. Plant Cell 6:1401–1414

Brigneti G, Voinnet O, Li WX et al (1998) Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. EMBO J 17:6739–6746

Brugliera F, Tao G, Tems U et al (2013) Violet/blue chrysanthemums–metabolic engineering of the anthocyanin biosynthetic pathway results in novel petal colors. Plant Cell Physiol 54(10):1696–1710

Bucher E, Sijen T, De Haan P et al (2003) Negative-strand tospoviruses and tenuiviruses carry a gene for a suppressor of gene silencing at analogous genomic positions. J Virol 77:1329–1336

Burch-Smith TM, Anderson JC, Martin GB, Dinesh-Kumar SP (2004) Applications and advantages of virus-induced gene silencing for gene function studies in plants. Plant J 39:734–746. doi:10.1111/j.1365-313X.2004.02158.x

Cakir C, Tor M (2010) Factors influencing barley stripe mosaic virus-mediated gene silencing in wheat. Phys Mol Plant Pathol 74:246–253

Carrillo-Tripp J, Shimada-Beltran H, Rivera-Bustamante R (2006) Use of geminiviral vectors for functional genomics. Curr Opin Plant Biol 9:209–215

Chen J, Li WX, Xie D et al (2004) Viral virulence protein suppresses RNA silencing-mediated defense but upregulates the role of microRNA in host gene expression. Plant Cell 16:1302–1313

Chen WH, Hsu CY, Cheng HY et al (2011) Downregulation of putative UDP-glucose: flavonoid 3-O-glucosyltransferase gene alters flower coloring in *Phalaenopsis*. Plant Cell Rep 30(6):1007–1017

Colquhoun TA, Schimmel BC, Kim JY et al (2010) A petunia chorismate mutase specialized for the production of floral volatiles. Plant J 61:145–155

Constantin GD, Krath BN, MacFarlane SA et al (2004) Virus-induced gene silencing as a tool for functional genomics in a legume species. Plant J 40:622–631

Constantin GD, Grønlund M, Johansen IE et al (2008) Virus-induced gene silencing (VIGS) as a reverse genetic tool to study development of symbiotic root nodules. MPMI 21(6):720–727

Das D, Moiani D, Axelrod HL et al (2010) Crystal structure of the first eubacterial Mre11 nuclease reveals novel features that may discriminate substrates during DNA repair. J Mol Biol 397:647–663

de la Torre F, El-Azaz J, Avila C, Canovas F (2014) Deciphering the role of aspartate and prephenate aminotransferase activities in plastid nitrogen metabolism. Plant Physiol 164:92–104

del Rosario Abraham-Juárez M, del Carmen Rocha-Granados M, López MG et al (2008) Virus-induced silencing of *Comt, pAmt* and *Kas* genes results in a reduction of capsaicinoid accumulation in chili pepper fruits. Planta 227(3):681–695

Di Stilio VS, Kumar RA, Oddone AM et al (2010) Virus-induced gene silencing as a tool for comparative functional studies in *Thalictrum*. PLoS One 5(8):e12064

Ding XS, Schneider WL, Chaluvadi SR et al (2006) Characterization of a brome mosaic virus strain and its use as a vector for gene silencing in monocotyledonous hosts. Mol Plant Microbe 19:1229–1239

Dodo HW, Konan KN, Chen FC (2008) Alleviating peanut allergy using genetic engineering: the silencing of the immunodominant allergen Ara h2 leads to its significant reduction and a decrease in peanut allergenicity. Plant Biotechnol J 6(2):135–145

Dunoyer P, Pfeffer S, Fritsch C et al (2002) Identification, subcellular localization and some properties of a cysteine-rich suppressor of gene silencing encoded by peanut clump virus. Plant J 29(5):555–567

Elbashir MS, Lendeckel W, Tuschl T (2001) RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev 15:188–200

Eschen-Lippold L, Lübken T, Smolka U, Rosahl S (2012) Characterization of potato plants with reduced StSYR1 expression. Plant Signal Behav 7:559–562

Escobar MA, Civerolo EL, Summerfelt KR, Dander AM (2001) RNAi-mediated oncogene silencing confers resistance to crown gall tumorigenesis. Proc Natl Acad Sci U S A 98:13437–13442

Fagard M, Boutet S, Morel JB et al (2000) AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. Proc Natl Acad Sci U S A 97:11650–11654

Faivre-Rampant O, Gilroy EM, Hrubikova K et al (2004) Potato virus X-induced gene silencing in leaves and tubers of potato. Plant Physiol 134:1308–1316

Fei Z, Joung JG, Tang X et al (2011) Tomato functional genomics database: a comprehensive resource and analysis package for tomato functional genomics. Nucleic Acids Res 39:D1156–D1163

Fernández-Calvo P, Chini A, Fernández-Barbero G et al (2011) The *Arabidopsis* bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. Plant Cell 23:701–715

Ferracane J (2001) Materials in dentistry: principles and applications, 2nd edn. Lippincott Williams & Wilkins, Philadelphia

Fire A, Xu S, Montgomery MK et al (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806–811

Flores T, Karpova O, Su X (2008) Silencing of GmFAD3 gene by siRNA leads to low alpha-linolenic acids (18:3) of *fad3*-mutant phenotype in soybean [*Glycine max* (Merr.)]. Transgenic Res 17:839–850

Fofana IB, Sangare A, Collier R et al (2004) A geminivirus-induced gene silencing system for gene function validation in cassava. Plant Mol Biol 56:613–624

Freiman A, Shlizerman L, Golobovitch S et al (2012) Development of a transgenic early flowering pear (*Pyrus communis* L.) genotype by RNAi silencing of PcTFL1-1 and PcTFL1-2. Planta 235(6):1239–1251

Frizzi A, Huang S, Gilbertson LA (2008) Modifying lysine biosynthesis and catabolism in corn with a single bifunctional expression/silencing transgene cassette. Plant Biotechnol J 6:13–21

Gilissen LJ, Bolhaar ST, Matos CI et al (2005) Silencing the major apple allergen Mal d 1 by using RNA interference approach. J Allergy Clin Immunol 115(2):364–369

Goldbach R, Bucher E, Prins M (2003) Resistance mechanisms to plant viruses: an overview. Virus Res 92:207–212

Golenberg EM, Sather DN, Hancock LC et al (2009) Development of a gene silencing DNA vector derived from a broad host range geminivirus. Plant Methods 5:9

Gould B, Kramer EM (2007) Virus-induced gene silencing as a tool for functional analyses in the emerging model plant *Aquilegia* (columbine, Ranunculaceae). Plant Methods 3:6

Griffiths-Jones S, Saini HK, Van-Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res 36:D154–D158

Grønlund M, Constantin G, Piednoir E et al (2008) Virus-induced gene silencing in *Medicago truncatula* and *Lathyrus odorata*. Virus Res 135:345–349

Grønlund M, Olsen A, Johansen E, Jakobsen I (2010) Protocol: using virus-induced gene silencing to study the arbuscular mycorrhizal symbiosis in *Pisum sativum*. Plant Methods 6:28. http://www.plantmethods.com/content/6/1/28

Grotewold E (2006) The genetics and biochemistry of floral pigments. Annu Rev Plant Physiol Plant Mol Biol 57:761–780

Guo S, Kemphues KJ (1995) par-1, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell 81:611–620

Hamilton A, Baulcombe D (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286:950–952

Hans H, Lucas K, Dick L et al (2009) Binding of small interfering RNA molecules is crucial for RNA interference suppressor activity of rice hoja blanca virus NS3 in plants. J Gen Virol 90:1762–1766

Haq QMI, Arif A, Malathi VG (2010) Engineering resistance against mungbean yellow mosaic India virus using antisense RNA. Indian J Virol 21:82–85

He X, Jin C, Li G et al (2008) Use of the modified viral satellite DNA vector to silence mineral nutrition-related genes in plants: silencing of the tomato ferric chelate reductase gene, FRO1, as an example. Sci China Ser C Life Sci 51:402–409
He F, Mu L, Yan G-L et al. (2010) Biosynthesis of anthocyanins and their regulation in colored grapes. Molecules 15:9057–9091
Herman E, Helm R, Jung R, Kinney A (2003) Genetic modification removes an immunodominant allergen from soybean. Plant Physiol 132(1):36–43
Himani T, Shanmugam R, Manchikatla VR, Dasgupta I (2008) RNA-interference in rice against Rice tungro bacilliform virus results in its decreased accumulation in inoculated rice plants. Transgenic Res 17:897–904
Hirai S, Kodama H (2008) RNAi vectors for manipulation of gene expression in higher plants. Open Plant Sci J 2:21–30
Hiriart JB, Aro EM, Lehto K (2003) Dynamics of the VIGS mediated chimeric silencing of the *Nicotiana benthamiana* ChlH gene and of the tobacco mosaic virus vector. Mol Plant Microbe Interact 16:99–106
Hou H, Qiu W (2003) A novel co-delivery system consisting of a tomato bushy stunt virus and a defective interfering rna for studying gene silencing. J Virol Methods 111:37–42
Huang CJ, Zhang T, Li FF et al (2011) Development and application of an efficient virus-induced gene silencing system in *Nicotiana tabacum* using geminivirus alpha satellite. J Zhejiang Univ Sci B 12:83–92
Hull R (1989) The movement of viruses in plants. Annu Rev Plant Physiol Plant Mol Biol 27:213–240
Jada B, Soitamo AJ, Lehto K (2013) Organ-specific alterations in tobacco transcriptome caused by the PVX-derived P25 silencing suppressor transgene. BMC Plant Biol 13(1):8
Jagtap UB, Gurav RG, Bapat VA (2011) Role of RNA interference in plant improvement. Naturwissenschaften 98:473–492
Jones-Rhoades MW, Bartel DP, Barterl B (2006) MicroRNAs and their regulatory roles in plants. Annu Rev Plant Physiol Plant Mol Biol 57:19–53
Jouannet V, Moreno A, Elmayan T et al (2012) Cytoplasmic Arabidopsis AGO7 accumulates in membrane-associated siRNA bodies and is required for ta-siRNA biogenesis. EMBO J 31:1704–1713
Kalantidis K, Psaradakis S, Tabler M, Tsagris M (2002) The occurrence of CMV-specific short RNAs in transgenic tobacco expressing virus-derived double-stranded RNA is indicative of resistance to the virus. Mol Plant Microbe Interact 15:826–833
Karthikeyan AS, Varadarajan DK, Jain A et al (2007) Phosphate starvation responses are mediated by sugar signaling in *Arabidopsis*. Planta 225:907–918
Karthikeyan A, Deivamani M, Shobhana VG et al (2013) RNA interference: evolutions and applications in plant disease management. Arch Phytopathol Plant Protect 46(12):1430–1441
Katiyar-Agarwal S, Morgan R, Dahlbeck D et al (2006) A pathogen-inducible endogenous siRNA in plant immunity. Proc Natl Acad Sci U S A 103:47–52
Katiyar-Agarwal S, Gao S, Vivian-Smith A, Jin H (2007) A novel class of bacteria-induced small rnas in *Arabidopsis*. Genes Dev 21:3123–3134
Katsumoto Y, Fukuchi-Mizutani M, Fukui Y et al (2007) Engineering of the rose flavonoid biosynthetic pathway successfully generated blue-hued flowers accumulating delphinine. Plant Cell Physiol 48(11):1589–1600
Kessler D, Gase K, Baldwin IT (2008) Field experiments with transformed plants reveal the sense of floral scents. Science 321(5893):1200–1202
Kessler D, Diezel C, Clark D et al (2012) Petunia flowers solve the defense/apparency dilemma of pollinator attraction by deploying complex floral blends. Ecol Lett. doi:10.1111/ele.12038
Koeduka T, Orlova I, Baiga T et al (2009) The lack of floral synthesis and emission of isoeugenol in *Petunia axillaris* subsp. *Parodii* is due to a mutation in the isoeugenol synthase gene. Plant J 58(6):961–969

Komiya R, Ikegami A, Tamaki S et al (2008) Hd3a and RFT1 are essential for flowering in rice. Development 135(4):767–774
Kong YM, Elling AA, Chen B, Deng XW (2010) Differential expression of microRNAs in Maize inbred and hybrid lines during salt and drought stress. Am J Plant Sci 1:69–76
Krenz B, Windeisen V, Wege C et al (2010) A plastid targeted heat shock cognate 70 kDa protein interacts with the abutilon mosaic virus movement protein. Virology 401:6–17
Kubota K, Tsuda S, Tamai A, Meshi T (2003) Tomato mosaic virus replication protein suppresses virus-targeted posttranscriptional gene silencing. J Virol 77:11016–11026
Kumagai MH, Donson J, Della-Cioppa G et al (1995) Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. Proc Natl Acad Sci U S A 92:1679–1683
Laufs J, Traut W, Heyraud F et al (1995) In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. Proc Natl Acad Sci U S A 92:3879–3883
Le LQ, Mahler V, Lorenz Y et al (2006) Reduced allergenicity of tomato fruits harvested from Lyc e 1-silenced transgenic tomato plants. J Allergy Clin Immunol 118:1176–1183
Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75(5):843–854
Li W, Han Y, Tao F, Chong K (2011) Knockdown of SAMS genes encoding S-adenosyl-L-methionine synthases causes methylation alternations of DNAs and histones and leads to late flowering in rice. J Plant Physiol 168:1837–1843
Li F, Huang C, Li Z, Zhou X (2014) Suppression of RNA silencing by a plant DNA virus satellite requires a host calmodulin-like protein to repress RDR6 expression. PLoS Pathog 10(2):e1003921
Liang J, Deng G, Long H, Pan Z, Wang C, Cai P et al (2012) Virus-induced silencing of genes encoding LEA protein in Tibetan hulless barley (Hordeum vulgare ssp. vulgare) and their relationship to drought tolerance. Mol Breed 30:441–451. doi:10.1007/s11032-011-9633-3
Liu L, Grainger J, Canizares MC et al (2004) Cowpea mosaic virus RNA-1 acts as an amplicon whose effects can be counteracted by a RNA-2-encoded suppressor of silencing. Virology 323:37–48
Lv B, Nitcher R, Han X et al (2014) Characterization of flowering locus T1 (FT1) gene in *Brachypodium* and wheat. PLoS One 9(4):e94171. doi:10.1371/journal.pone.0094171
Matsui A, Nguyen AH, Nakaminami K, Seki M (2013) *Arabidopsis* non-coding RNA regulation in abiotic stress responses. Int J Mol Sci 14:22642–22654
Mohanpuria P, Kumar V, Ahuja PS, Yadav SK (2011) Producing low-caffeine tea through post-transcriptional silencing of caffeine synthase mRNA. Plant Mol Biol 76(6):523–524
Morandini P (2010) Inactivation of allergens and toxins. New Biotechnol 27(5):482–493
Mumbanza FM, Kiggundu A, Tusiime G (2013) In vitro antifungal activity of synthetic dsRNA molecules against two pathogens of banana, *Fusarium oxysporum* f sp *cubense* and *Mycosphaerella fijiensis*. Pest Manage Sci 6(10):1155–1162
Muruganantham M, Moskovitz Y, Haviv S et al (2009) Grape vine virus A-mediated gene silencing in *Nicotiana benthamiana* and *Vitis vinifera*. J Virol Methods 155:167–174
Mutti NS, Park Y, Reese JC, Reeck GR (2006) RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrthosiphon pisum*. J Insect Sci 6:1–7
Mutti NS, Louis J, Pappan LK et al (2008) A protein from the salivary glands of the pea aphid, *Acyrthosiphon pisum*, is essential in feeding on a host plant. Proc Natl Acad Sci U S A 105:9965–9969
Nagamatsu A, Masuta C, Senda M et al (2007) Functional analysis of soybean genes involved in flavonoid biosynthesis by virus-induced gene silencing. Plant Biotechnol J 5:778–790
Nakatsuka T, Abe Y, Kakizaki Y et al (2007) Production of red-flowered plants by genetic engineering of multiple flavonoid biosynthetic genes. Plant Cell Rep 26(11):1951–1959
Napoli C, Lemieux C, Jorgensen R (1990) Introduction of chimeric chalcone synthase gene into *Petunia* results in reversible co-suppression of homologous genes in *trans*. Plant Cell 2:279–289
Naylor M, Reeves J, Cooper JI et al (2005) Construction and properties of a gene silencing vector based on poplar mosaic virus (Genus *Carlavirus*). J Virol Methods 124:27–36

Nishihara M, Nakatsuka T, Yamamura S (2005) Flavonoid components and flower color change in transgenic tobacco plants by suppression of chalcone isomerase gene. FEBS Lett 579:6074–6078

Nowara D, Gay A, Lacomme C et al (2010) HIGS: host-induced gene silencing in the obligate biotrophic fungal pathogen *Blumeria graminis*. Plant Cell 22:3130–3141

Nykanen A, Haley B, Zamore PD (2001) ATP requirements and small interfering RNA structure in the RNA interference pathway. Cell 107:309–321

Ogita S, Uefuji H, Yamaguchi Y (2003) Producing decaffeinated coffee plants. Nature 423:823

Pandey P, Choudhury NR, Mukherjee SK (2009) A geminiviral amplicon (VA) derived from tomato leaf curl virus (ToLCV) can replicate in a wide variety of plant species and also acts as a VIGS vector. Virol J 6:152

Pflieger S, Blanchet S, Camborde L (2008) Efficient virus-induced gene silencing in Arabidopsis using a 'one-step' TYMV-derived vector. Plant J 56:678–690

Pitino M, Coleman AD, Maffei ME et al (2011) Silencing of aphid genes by dsRNA feeding from plants. PLoS One 6:e25709. doi:10.1371/journal.pone.0025709

Prentice K, Pertry I, Christiaens O, Bauters L, Bailey A, Niblett C et al (2015) Transcriptome analysis and systemic RNAi response in the African sweetpotato Weevil (Cylas puncticollis, Coleoptera, Brentidae). PLoS One 10(1):e0115336. doi:10.1371/ journal.pone.0115336

Pumplin N, Voinnet O (2013) RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. Microbiology 11:745–760

Purkayastha A, Sharma S, Dasgupta I (2010) A negative element in the downstream region of the rice tungro bacilliform virus promoter is orientation- and position-independent and is active with heterologous promoters. Virus Res 153:166–171

Qian S, McDonough H, Boellmann F, Cyr D, Patterson C (2006) CHIP-mediated stress recovery by sequential ubiquitination of substrates and Hsp70. Nature 440:551–555. doi:10.1038/nature04600

Ratcliff F, Martin-Hernandez AM, Baulcombe DC (2001) Tobacco rattle virus as a vector for analysis of gene function by silencing. Plant J 25:237–245

Ré DA, Raud B, Chan R et al (2012) RNAi-mediated silencing of the HD-Zip gene HD20 in *Nicotiana attenuata* affects benzyl acetone emission from corollas via ABA levels and the expression of metabolic genes. BMC Plant Biol 12:60

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE et al (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. Nature 403(6772):901–906

Reinhart BJ, Weinstein EG, Rhoades MW et al (2002) MicroRNAs in plants. Genes Dev 16:1616–1626

Rezk AA, Abdallah NA, Abdel-Salam AM et al (2006) Transgene-mediated RNA silencing of TYLCV genes affecting the accumulation of viral DNA in plants. Arab J Biotechnol 9(1):143–157

Rodoni BC, Dale JL, Harding RM (1999) Characterization and expression of the coat protein-coding region of the banana bract mosaic potyvirus, development of diagnostic assays and detection of the virus in banana plants from five countries in Southeast Asia. Arch Virol 144:1725–1737

Runo S, Alakonya A, Machuka J, Sinha N (2011) RNA interference as a resistance mechanism against crop parasites in Africa: a 'Trojan horse' approach. Pest Manag Sci 67:129–136. doi:10.1002/Ps.2052

Schweizer P, Pokorny J, Schulze-Lefert P, Dudler R (2000) Double-stranded RNA interferes with gene function at the single-cell level in cereals. Plant J 24:895–903

Scofield SR, Nelson RS (2009) Resources for virus-induced gene silencing in the grasses. Plant Physiol 149:152–157

Segal G, Song R, Messing J (2003) A new opaque variant of maize by a single dominant RNA-interference-inducing transgene. Genetics 165:387–397

Senapedis WT, Kennedy CJ, Boyle PM, Silver PA (2011) Whole genome siRNA cell-based screen links mitochondria to Akt signaling network through uncoupling of electron transport chain. Mol Biol Cell 22:1791–1805

Senthil-Kumar M, Mysore KS (2010) RNAi in plants: recent developments and application in agriculture. In: Gene silencing: theory, techniques and application. Chapter VII. Nova Science publisher, Inc, New York

Senthil-Kumar M, Mysore KS (2011) New dimensions for VIGS in plant functional genomics. Trends Plant Sci 16(12):656–665

Senthil-Kumar M, Udayakumar M (2010) Post transcriptional gene silencing methods for functional characterization of abiotic stress responsive genes in plants. In: Catalano AJ (ed) Gene silencing: theory, techniques and applications. Nova Science Publishing Inc, New York, pp 271–286

Sestili F, Janni M, Doherty A et al (2010) Increasing the amylose content of durum wheat through silencing the *SBEIIa* genes. BMC Plant Biol 10:144–155

Shang Y, Schwinn KE, Bennett MJ et al (2007) Methods for transient assay of gene function in floral tissues. Plant Methods 3:1. doi:10.1186/1746-4811-3-1

Shelly-Praveen S, Ramesh V, Anil KM et al (2010) Silencing potential of viral derived RNAi constructs in tomato leaf curl virus-AC4 gene suppression in tomato. Transgenic Res 19:45–55

Shi J, Wang H, Schellin K et al (2007) Embryo-specific silencing of a transporter reduces phytic acid content of maize and soybean seeds. Nat Biotechnol 25(8):930–937

Shi Y, Zhang X, Xu Z-Y et al (2011) Influence of EARLI1-like genes on flowering time and lignin synthesis of *Arabidopsis thaliana*. Plant Biol 13:731–739

Shimizu T, Yoshii M, Wei T et al (2009) Silencing by RNAi of the gene for Pns12, a viroplasm matrix protein of rice dwarf virus, results in strong resistance of transgenic rice plants to the virus. Plant Biotechnol J 7:24–32

Sindhu AS, Maier TR, Mitchum MG et al (2009) Effective and specific in plant RNAi in cyst nematodes: expression interference of four parasitism genes reduces parasitic success. J Exp Bot 60(1):315–324

Smith NA, Singh SP, Wang MB et al (2000) Total silencing by intron-spliced hairpin RNAs. Nature 407:319–320

Soler N, Fagoaga C, Chiibi S et al (2011) RNAi mediated protection against *Citrus tristeza* virus in transgenic *citrus* plants. In: Erdmann VA, Barciszewski J (eds) Non coding RNAs in plants. Springer, Berlin, pp 447–460

Soliman AM, Barsoum BN, Mohamed GG (2008) siRNA silencing of PVX coat protein gene affects accumulation of viral RNA in potato and tobacco plants. Int J Virol 4(1):14–25

Spitzer B, Zvi M, Ovadis M et al (2007) Reverse genetics of floral scent: application of tobacco rattle virus-based gene silencing in petunia. Plant Physiol 145:1241–1250

Szankowski I, Waidmann S, El-Din Saad O et al (2009) RNAi-silencing of *MdTFL1* induces early flowering in apple. Acta Horticult 839:633–636

Tanaka Y, Nakamura N, Togami J (2008) Altering flower color in transgenic plants by RNAi-mediated engineering of flavonoid biosynthesis pathway. Methods Mol Biol 442:245–257

Tang G, Galili G (2004) Using RNAi to improve plant nutritional value: from mechanism to application. Trends Biotechnol 22(9):463–469

Tao X, Zhou X (2004) A modified viral satellite DNA that suppresses gene expression in plants. Plant J 38:850–860

Tavernarakis N, Wang SL, Dorovkov M et al (2000) Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes. Nat Genet 24:180–183

To KY, Wang CK (2006) Molecular breeding of flower color. In: de Silva JA (ed) Floriculture, ornamental and plant biotechnology: advances in topical issues, vol 1, 1st edn. Global Science Books Ltd, Isleworth, pp 300–310

Trusov Y, Botella JR (2006) Silencing of the ACC synthase gene ACACS2 causes delayed flowering in pineapple [*Ananas comosus* (L.) Merr.]. J Exp Bot 57(14):3953–3960

Turnage MA, Muangsan N, Peele CG, Robertson D (2002) Geminivirus based vectors for gene silencing in *Arabidopsis*. Plant J 30:107–117

Tuttle JR, Idris AM, Brown JK et al (2008) Geminivirus-mediated gene silencing from cotton leaf crumple virus is enhanced by low temperature in cotton. Plant Physiol 148:41–50

Ufaz S, Galili G (2008) Improving the content of essential amino acids in crop plants: goals and opportunities. Plant Physiol 147:954–961
Vaistij FE, Jones L (2009) Compromised virus-induced gene silencing in RDR6-deficient plants. Plant Physiol 149(3):1399–1407
Van der Linde K, Kastner C, Kumlehn J (2011) Systemic virus-induced gene silencing allows functional characterization of maize genes during biotrophic interaction with *Ustilago maydis*. New Phytol 189:471–483
Van Wezel WR, Dong X, Liu H et al (2002) Mutation of three cysteine residues in tomato yellow leaf curl virus-China C2 protein causes dysfunction in pathogenesis and posttranscriptional gene silencing suppression. Mol Plant-Microbe Interact 15:203–208
Vanderauwera S, De Block M, Van de Steene N et al (2007) Silencing of a poly(ADP-ribose) polymerase in plants alters abiotic stress signal transduction. Proc Natl Acad Sci U S A 104(38):15150–15155
Vanderschuren H, Akbergenov R, Pooggin MM (2007) Transgenic cassava resistance to African cassava mosaic virus is enhanced by viral DNA-A bidirectional promoter-derived siRNAs. Plant Mol Biol 64:549–557
Vanderschuren H, Alder A, Zhang P, Gruissem W (2009) Dose-dependent RNAi- mediated geminivirus resistance in the tropical root crop cassava. Plant Mol Biol 70:265–272
Vanitharani R, Chellappan P, Fauquet CM (2005) Geminiviruses and RNA silencing. Trends Plant Sci 10(3):144–151
Varshney RK, Graner A, Sorrells ME (2005) Genomic-assisted breeding for crop improvement. Trends Plant Sci 10:621–630
Vaucheret H, Béclin C, Fagard M (2001) Post-transcriptional gene silencing in plants. J Cell Sci 114:3083–3091
Velásquez AC, Chakravarthy S, Martin GB (2009) Virus-induced gene silencing (VIGS) in *Nicotiana benthamiana* and Tomato. J Vis Exp 28, e1292. doi:10.3791/1292
Verdonk JC, Haring MA, van Tunen AJ, Schuurink RC (2005) ODORANT1 regulates fragrance biosynthesis in petunia flowers. Plant Cell 17(5):1612–1624
Voinnet O (2005) Induction and suppression of RNA silencing: insights from viral infections. Nat Rev Genet 6:206–220
Voinnet O, Pinto YM, Baulcombe DC (1999) Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. Proc Natl Acad Sci U S A 96:14147–14152
Voinnet O, Lederer C, Baulcombe DC (2000) A viral movement protein prevents spread of the gene silencing signal in *Nicotiana benthamiana*. Cell 103:157–167
Wakasa Y, Hirano K, Uriso A et al (2011) Generation of transgenic rice lines with reduced contents of multiple potential allergens using a null mutant in combination with an RNA silencing method. Plant Cell Physiol 52(12):2190–2199
Wang M, Abbott D, Waterhouse PM (2000) A single copy of a virus derived transgene encoding hairpin RNA gives immunity to barley yellow dwarf virus. Mol Plant Pathol 1:401–410
Wani SH, Sanghera GS, Singh NB (2010) Biotechnology and plant disease control-role of RNA interference. Am J Plant Sci 1:55–68
Warthmann N, Chen H, Ossowski S et al (2008) Highly specific gene silencing by artificial miRNAs in rice. PLoS One 3(3):e1829. doi:10.1371/journal.pone.0001829
Waterhouse PM, Graham MW, Wang MB (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. Proc Natl Acad Sci U S A 95:13959–13964
Waterhouse P, Wang M, Lough T (2001) Gene silencing as an adaptive defense against viruses. Nature 411:834–842
Xiong XP, Kurthkoti K, Chang KY et al (2013) Core small nuclear ribonucleoprotein particle splicing factor SmD1 modulates RNA interference in Drosophila. Proc Natl Acad Sci U S A 110(41):16520–16525

Yamagishi N, Kishigama R, Yoshikawa N (2013) Reduced generation time of apple seedlings to within a year by means of plant virus vector: a new plant breeding technique with no transmission of genetic modification to the next generation. Plant Biotechnol J. doi:10.1111/pbi.12116

Yan Y, Zhang Y, Yang K (2010) Small RNAs from MITE-derived stem-loop precursors regulate abscisic acid signaling and abiotic stress responses in rice. Plant J. doi:10.1111/j.1365-313X.2010.04467.x

Yang EC, Chuang YC, Chen L, Chang LH (2008) Abnormal foraging behavior induced by sublethal dosage of imidacloprid in the honey bee (*Hymenoptera: Apidae*). J Econ Entomol 101:1743–1748

Yelina NE, Savenkov EI, Solovyev AG (2002) Long-distance movement, virulence, and RNA silencing suppression controlled by a single protein in hordei- and potyviruses: complementary functions between virus families. J Virol 76(24):12981–12991

Yin C, Jurgenson JE, Hulbert SH (2011) Development of a host-induced RNAi system in the wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici*. Mol Plant Microbe Interact 24:554–561

Zha W, Peng X, Chen R, Du B, Zhu L et al (2011) Knockdown of midgut genes by dsRNA-transgenic plant-mediated RNA interference in the *Hemipteran* insect *Nilaparvata lugens*. PLoS One 6(5):e20504. doi:10.1371/journal.pone.0020504

Zhang B, Pan X, Cobb GP, Anderson TA (2005) Plant microRNA: a small regulatory molecule with big impact. Dev Biol 289:3–16

Zorrilla-López U, Masip G, Arjo G, Bai C, Banakar R et al (2013) Engineering metabolic pathways in plants by multigene transformation. Int J Dev Biol 57:565–576. doi:10.1387/ijdb.130162pc

Part IV
Molecular Breeding

Chapter 15
Molecular Marker-Assisted Breeding: A Plant Breeder's Review

Guo-Liang Jiang

Abstract A plant breeder faces the challenge of how to more effectively and efficiently perform selection and accelerate the breeding progress to satisfy the requirements of changing markets for crop cultivars. Molecular marker-assisted breeding (MAB), the application of molecular biotechnologies (DNA markers) to practical breeding and selection, is a novel strategy and a powerful methodology for plant improvement. It has significant advantages compared with conventional breeding methods. Since the 1990s MAB has received increasingly attention and has been extensively used in different crop species. From a plant breeder's point of view, this chapter addresses the general procedures, theoretical and practical considerations of MAB in plants, including marker-assisted selection (MAS), marker-assisted backcrossing (MABC), marker-assisted gene pyramiding (MAGP), marker-assisted recurrent selection (MARS) and genomewide selection (GWS). Applications of individual MAB methods to practical breeding as well as widely used DNA markers are briefly reviewed, and the challenges and perspectives of MAB are discussed. As a new technology, MAB is not a replacement for but a valued supplement to conventional breeding. Integration of MAB into conventional breeding programs represents an optimistic strategy for future crop improvement.

Keywords Backcrossing • Breeding methodologies • DNA markers • QTL • Gene pyramiding • Molecular mapping • Recurrent selection

15.1 Introduction

Plant breeding has played an important role in the development of agriculture and the improvement of human life. Thousands of years ago, wild plant species were domesticated into crops by artificial selection. Domestication and continual selection of crop species have not only provided humans with essential and diverse food

G.-L. Jiang (✉)
Agricultural Research Station, Virginia State University,
P.O. Box 9061, Petersburg, VA 23806, USA
e-mail: gjiang@vsu.edu; gljiang99@yahoo.com

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_15

and fiber products, but have also established the material foundation of modern breeding for further genetic improvement. Although crossing techniques were used in cultivar development on a small scale before the twentieth century, plant breeding did not develop as a science until the Mendelian principles of inheritance were established. During the early period of plant breeding, success in cultivar development relied considerably on the breeder's experience and subjective judgment, as well as luck. Even within present plant breeding programs aided by well-developed genetic principles and modern techniques, the breeder's experience and unique insight are still an important factor for success in cultivar development. Therefore, plant breeding is generally regarded as a combination of science and art (Acquaah 2007; Fehr 1991). The challenges of plant breeding include the following. How to select parents and form breeding populations? How to identify and select desired variants? How to accurately predict the expression or performance of breeding materials in succeeding generations and variable environments? How to more effectively and efficiently try out selection? How to speed up the breeding progress in cultivar development to satisfy the requirements of changing and increasing markets for crop cultivars? A breeder is always trying to answer these questions, but using traditional breeding methods alone is impractical.

Selection for target traits in practical breeding may be either direct or indirect. Direct selection is based on the performance or phenotypic values of target traits under certain conditions. In many cases, the effects of direct selection for a quantitatively-inherited complex character are limited. Instead, indirect selection may be a better alternative if there is another trait which is inherited in a simpler mode and closely correlated with the target trait as well (Fehr 1991). However, the traits that may be used in indirect selection, particularly for yield, are very limited. A trait is controlled by one or more genes and its phenotypic value is modified by environments. Theoretically, selection for the target gene(s) is the most direct selection and may provide a guarantee for effectiveness. However, it is not practicable to perform selection for target genes for many breeding programs because of the limitation of equipment, facilities and resources. Development of molecular (DNA) markers has created a powerful and practicable tool to perform gene selection in plant breeding, although it is not a real gene selection but the best indirect selection for target genes at the DNA level.

Molecular marker-assisted breeding (MAB), also called *molecular-assisted breeding*, is the application of molecular biotechnologies, specifically DNA markers, in combination with linkage maps and genomics, to alter and improve plant or animal traits on the basis of genotypic assays (Jiang 2013a). This term is used to describe several novel breeding strategies, including marker-assisted selection (MAS), marker-assisted backcrossing (MABC), marker-assisted recurrent selection (MARS), and genomewide selection (GWS) or genomic selection (GS) (Ribaut et al. 2010). MAB is regarded as a novel strategy and a powerful methodology for genetic improvement of crop plants, and up to now it has been extensively used in multiple crop species (Jiang 2013a; Xu 2010). In terms of the resources invested and the expectations presented, however, MAB has not yet been very successful.

There are still many issues to be addressed (Collard and Mackill 2008; Jiang 2013b; Nakaya and Isobe 2012).

The objective of this chapter is to provide a review of the status of MAB in plant improvement and research, from the perspective of a plant breeder's understanding and consideration. Beginning with an introduction to markers widely used in plant breeding, it describes the pre-requisites, general principles and methodologies of marker-assisted breeding in plants, and reviews research progress and applications of different MAB methods in practical breeding, including marker-assisted selection, marker-based backcrossing and marker-based pyramiding of multiple genes. Theoretical and practical issues related to the procedures and applications of this methodology, and the challenges and perspectives are also discussed.

15.2 Markers for Plant Breeding

15.2.1 Morphological and Biochemical Markers

In a broad sense, marker-assisted breeding can be traced back a long time to the use of classical or morphological markers as an assisting tool for selection of plants with desired traits in breeding (Jiang 2013a). During the early history of plant breeding, the markers used mainly included visible morphological traits, such as leaf shape, flower color, pubescence color, pod color, seed color, seed shape, hilum color, awn type and length, fruit shape, rind (exocarp) color and stripes, flesh color, stem length, etc. These morphological markers generally represent genetic polymorphisms which are easily identified and manipulated. Therefore, they are usually used in the construction of linkage maps by classical two- and/or three-point tests (Xu 2010). Some of these markers are linked with other agronomic traits and thus can be used as indirect selection criteria in practical breeding. A successful example of using morphological markers is semi-dwarfism in rice and wheat. During the Green Revolution of the twentieth century, use and selection of semi-dwarfism in rice and wheat was one of the critical steps that lead to the success of high-yielding cultivars. In wheat breeding, the dwarfism governed by gene *Rht10* was introgressed into Taigu nuclear male-sterile wheat by backcrossing, and a tight linkage was generated between *Rht10* and the male-sterility gene *Ta1*. The dwarfism was then used as the marker for identification and selection of the male-sterile plants in breeding populations (Liu 1991). This is particularly helpful for implementation of recurrent selection in wheat. However, morphological markers available are limited, and many of these markers are not associated with important economic traits (e.g. yield and quality) and even have undesirable effects on the development and growth of plants.

Biochemical/protein markers may be regarded as the earliest types of molecular markers, though molecular markers generally refer to DNA markers. Isozymes are alternative forms or structural variants of an enzyme that have different molecular weights and electrophoretic mobility but have the same catalytic activity or function.

Isozymes reflect the products of different alleles rather than different genes because the difference in electrophoretic mobility is caused by point mutation as a result of amino acid substitution (Xu 2010). Therefore, isozyme markers can be genetically mapped onto chromosomes and then used as genetic markers to map other genes. They are also used in seed purity test and occasionally in plant breeding. However, the use of enzyme markers is limited, because there are only a small number of isozymes in most crop species and some of them can be identified only with a specific strain. High molecular weight glutenin subunit (HMW-GS) in wheat is another example of biochemical markers used in plant breeding. Payne et al. (1987) discovered a correlation between the presence of certain HMW-GS and gluten strength, measured by the SDS-sedimentation volume test. They designed a numeric scale to evaluate bread-making quality as a function of the described subunits (Glu-1 quality score) (Payne et al. 1987; Rogers et al. 1989). Assuming the effect of the alleles to be additive, the bread-making quality could be predicted by adding the scores of the alleles present in the particular line. It has been established that the allelic variation at the Glu-D1 locus have a greater influence on bread-making quality than the variation at the other Glu-1 loci. Subunit combination 5 + 10 for locus Glu-D1 (Glu-D1 5 + 10) renders stronger dough than Glu-D1 2 + 12, largely due to the presence of an extra cysteine residue in the Dx-5 subunit compared to the Dx-2 subunit, which would promote the formation of polymers with larger size distribution. Therefore, breeders may enhance the bread-making quality in wheat by selecting subunit combination Glu-D1 5 + 10 instead of Glu-D1 2 + 12. Of course, the variation of bread-making quality among different varieties cannot be explained only by the variation in HMW-GS composition, because the low molecular weight glutenin subunit (LMW-GS) (as well as the gliadins in a smaller proportion) and their interactions with the HMW-GS also play an important role in the gluten strength and bread-making quality (Jiang 2013a).

15.2.2 DNA Markers

DNA markers are defined as fragments of DNA revealing mutations/variations, which can be used to detect polymorphism between different genotypes or alleles of a gene for a particular sequence of DNA in a population or gene pool. Such fragments are associated with a certain location within the genome and may be detected by means of certain molecular technology. Simply stated, a DNA marker is a small region of DNA sequence showing polymorphism (base deletion, insertion and substitution) between different individuals (Jiang 2013a). There are two basic methods to detect the polymorphism: Southern blotting, a nuclear acid hybridization technique (Southern 1975), and PCR, a polymerase chain reaction technique (Mullis 1990). Using PCR and/or molecular hybridization followed by electrophoresis (e.g. PAGE – polyacrylamide gel electrophoresis, AGE – agarose gel electrophoresis, CE – capillary electrophoresis), the variation in DNA samples or polymorphism for a specific region of DNA sequence can be identified based on the product features,

such as band size and mobility. In addition, other detection systems have been also developed. For instance, several new array chip techniques use DNA hybridization combined with labeled nucleotides, and new sequencing techniques detect polymorphism by sequencing. The marker techniques help in selection of multiple desired characters simultaneously using F_2 and backcross populations, near isogenic lines, doubled haploids and recombinant inbred lines. DNA markers are also called molecular markers in many cases and play a major role in molecular breeding. Therefore, molecular markers in this chapter are mainly referred to as DNA markers except specific definitions are given, although isozymes and protein markers are also molecular markers. Depending on the application and species involved, ideal DNA markers for efficient use in marker-assisted breeding should meet the following criteria (Jiang 2013a):

- High level of polymorphism.
- Even distribution across the whole genome (not clustered in certain regions).
- Co-dominance in expression (so that heterozygotes can be distinguished from homozygotes).
- Clear distinct allelic features (so that the different alleles can be easily identified).
- Single copy and no pleiotropic effect.
- Low cost to use (or cost-efficient marker development and genotyping)
- Easy assay/detection and automation.
- High availability (unrestricted use) and suitability to be duplicated/ multiplexed (so that the data can be accumulated and shared between laboratories).
- Genome-specific in nature (especially with polyploids).
- No detrimental effect on phenotype.

Since DNA restriction fragment length polymorphism (RFLP) was first used in human-linkage mapping (Botstein et al. 1980), substantial progress has been achieved in development and improvement of molecular techniques that help to easily find markers of interest on a large-scale. As a result, DNA markers have been extensively and successfully used in human genetics, animal genetics and breeding, plant genetics and breeding, and germplasm characterization and management. The techniques that have been extensively used and are particularly promising for application to plant breeding included: the restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), microsatellites or simple sequence repeat (SSR) and single nucleotide polymorphism (SNP). According to a causal similarity of SNPs with some of these marker systems and fundamental difference with several other marker systems, the molecular markers can also be classified into SNPs (due to sequence variation, e.g. RFLP) and non-SNPs (due to length variation, e.g. SSR) (Gupta et al. 2001). In view of space limitation, only the five marker systems mentioned above are briefly addressed here (Table 15.1). For details about the technical methods of how to develop DNA markers and the procedures how to detect them in practice, recent reviews and books may be consulted (Farooq and Azam 2002a, b; Gupta et al. 2001; Semagn et al. 2006a; Xu 2010).

Table 15.1 Feature description and comparison of most widely used DNA marker systems in plants

Feature	RFLP	RAPD	AFLP	SSR	SNP
Genomic abundance	High	High	High	Moderate to high	Very high
Genomic coverage	Low copy coding region	Whole genome	Whole genome	Whole genome	Whole genome
Expression/inheritance	Co-dominant	Dominant	Dominant / co-dominant	Co-dominant	Co-dominant
Number of loci	Small (<1000)	Small (<1000)	Moderate (1000 s)	High (1000 s – 10,000 s)	Very high (>100,000)
Level of polymorphism	Moderate	High	High	High	High
Type of polymorphism	Single base changes, indels	Single base changes, indels	Single base changes, indels	Changes in length of repeats	Single base changes, indels
Type of probes/primers	Low copy DNA or cDNA clones	10 bp random nucleotides	Specific sequence	Specific sequence	Allele-specific PCR primers
Cloning and/or sequencing	Yes	No	No	Yes	Yes
PCR-based detection	Usually no	Yes	Yes	Yes	Yes
Radioactive detection	Usually yes	No	Yes or no	Usually no	No
Reproducibility/reliability	High	Low	High	High	High
Effective multiplex ratio	Low	Moderate	High	High	Moderate to high
Marker index	Low	Moderate	Moderate to high	High	Moderate
Genotyping throughput	Low	Low	High	High	High
Amount of DNA required	Large (5–50 μg)	Small (0.01–0.1 μg)	Moderate (0.5–1.0 μg)	Small (0.05–0.12 μg)	Small (≥0.05 μg)
Quality of DNA required	High	Moderate	High	Moderate to high	High
Technical demanding	Moderate	Low	Moderate	Low	High
Time demanding	High	Low	Moderate	Low	Low
Ease of use	Not easy	Easy	Moderate	Easy	Easy
Ease of automation	Low	Moderate	Moderate to high	High	High
Development/start-up cost	Moderate to high	Low	Moderate	Moderate to high	High
Cost per analysis	High	Low	Moderate	Low	Low
Polymorphic loci detected per analysis	1.0–3.0	1.5–5.0	20–100	1.0–3.0	1.0
Primary application	Genetics	Diversity	Diversity and genetics	All purposes	All purposes

Adapted from Jiang (2013a); Xu (2010); Semagn et al. (2006a); Collard et al. (2005) and others

A marker system has advantages and disadvantages, relevant to the purposes of research, available genetic resources or databases, equipment and facilities, funding and personnel resources, etc. (Jiang 2013a). The choice and use of DNA markers in research and breeding is still a challenge for plant breeders. In most cases, a breeder needs to take a number of factors into consideration when choosing one or more molecular marker types (Semagn et al. 2006a). An appropriate choice should be one that best meets the requirements under the conditions and resources available for the breeding program.

15.2.2.1 RFLP Marker

RFLP markers are the first generation of DNA markers and one of the important tools for plant genome mapping. They are a type of Southern-Boltting-based markers. In living organisms, mutation events (deletion and insertion) may occur at restriction sites or between adjacent restriction sites in the genome. The gain or loss of restriction sites resulting from base pair changes and insertions or deletions at restriction sites within the restriction fragments may cause differences in the size of restriction fragments. These variations may cause alternation or elimination of the recognition sites for restriction enzymes. Consequently, when homologous chromosomes are subjected to restriction enzyme digestion, different restriction products are produced and can be detected by electrophoresis and DNA probing techniques.

RFLP markers are powerful tools for comparative and synteny mapping. Most RFLP markers are co-dominant and locus-specific. RFLP genotyping is highly reproducible, and the methodology is simple and no special equipment is required. By using an improved RFLP technique, i.e. cleaved amplified polymorphism sequence (CAPS), also known as PCR-RFLP, high-throughput markers can be developed from RFLP probe sequences. Very few CAPS are developed from probe sequences, which are complex to interpret. Most CAPS are developed from SNPs found in other sequences followed by PCR and detection of restriction sites. CAPS technique consists of digesting a PCR-amplified fragment and detecting the polymorphism by the presence/absence of restriction sites (Konieczny and Ausubel 1993). Another advantage of RFLP is that the sequence used as a probe need not be known. All that a researcher needs is a genomic clone that can be used to detect the polymorphism. Very few RFLPs have been sequenced to determine what sequence variation is responsible for the polymorphism. However, it may be problematic to interpret complex RFLP allelic systems in the absence of sequence information. RFLP analysis requires large amounts of high-quality DNA, has low genotyping throughput and is very difficult to automate. Radioactive autography involving in genotyping and physical maintenance of RFLP probes limit its use and sharing between laboratories. RFLP markers were predominantly used in the 1980/1990s, but since the last decade fewer direct uses of RFLP markers in genetic research and plant breeding have been reported. Most plant breeders may think that RFLP is too laborious and demands too much pure DNA to be important for plant breeding, although it was central for various types of scientific studies (Xu 2010).

15.2.2.2 RAPD Markers

RAPD is a PCR-based marker system. In this system, the total genomic DNA of an individual is amplified by PCR using a single, short (usually about ten nucleotides/bases) and random primer. The primer which binds to many different loci is used to amplify random sequences from a complex DNA template that is complementary to it (maybe including a limited number of mismatches). Amplification can occur during the PCR, if two hybridization sites are similar to one another (at least 3000 bp) and in opposite directions. The amplified fragments generated by PCR depend on the length and size of both the primer and the target genome. The PCR products (up to 3 kb) are separated by agarose gel electrophoresis and imaged by ethidium bromide (EB) staining. Polymorphisms resulted from mutations or rearrangements either at or between the primer-binding sites can be detected in the electrophoresis by the presence or absence of a particular RAPD band.

RAPD predominantly provides dominant markers. This system yields high levels of polymorphism and is simple and easy to conduct (Xu 2010). First, neither DNA probes nor sequence information is required for the design of specific primers. Second, the procedure does not involve blotting or hybridization steps, and thus it is a quick, simple and efficient technique. Third, relatively small amounts of DNA (about 10 ng per reaction) are required and the procedure can be automated, and higher levels of polymorphism also can be detected compared with RFLP. Fourth, no marker development is required, and the primers are non-species specific and can be universal. Fifth, the RAPD products of interest can be cloned, sequenced and then converted into or used to develop other types of PCR-based markers, such as sequence characterized amplified region (SCAR), single nucleotide polymorphism (SNP), etc. However, RAPD also has some limitations/disadvantages, such as low reproducibility and incapability to detect allelic differences in heterozygotes.

15.2.2.3 AFLP Markers

AFLPs are PCR-based markers, simply RFLPs visualized by selective PCR amplification of DNA restriction fragments. Technically, AFLP is based on the selective PCR amplification of restriction fragments from a total double-digest of genomic DNA under high stringency conditions, i.e. the combination of polymorphism at restriction sites and hybridization of arbitrary primers (Xu 2010). Because of this, AFLP is also called *selective restriction fragment amplification* (SRFA). An AFLP primer (17–21 nucleotides in length) consists of a synthetic adaptor sequence, the restriction endonuclease recognition sequence and an arbitrary, non-degenerate *selective* sequence (1–3 nucleotides). The primers used in this technique are capable of annealing perfectly to their target sequences (the adapter and restriction sites) as well as a small number of nucleotides adjacent to the restriction sites. The first step in AFLP involves restriction digestion of genomic DNA (about 500 ng) with two restriction enzymes, a rare cutter (6-bp recognition site, EcoRI, PtsI or HindIII) and a frequent cutter (4-bp recognition site, MseI or TaqI). The adaptors are then ligated

to both ends of the fragments to provide known sequences for PCR amplification. The double-stranded oligonucleotide adaptors are designed in such a way that the initial restriction site is not restored after ligation. Therefore, only the fragments which have been cut by the frequent cutter and rare cutter will be amplified. This property of AFLP makes it very reliable, robust and immune to small variations in PCR amplification parameters (e.g. thermal cycles, template concentration), and it also can produce a high marker density. The AFLP products can be separated in high-resolution electrophoresis systems. The fragments in gel-based or capillary DNA sequencers can be detected by dye-labeling primers radioactively or fluorescently. The number of bands produced can be manipulated by the number of selective nucleotides and the nucleotide motifs used.

A typical AFLP fingerprint (restriction fragment patterns generated by the technique) contains 50–100 amplified fragments, of which up to 80 % may serve as genetic markers. In general, AFLP assays can be conducted using relatively small DNA samples (1–100 ng per individual). AFLP has a very high multiplex ratio and genotyping throughput, and is relatively reproducible across laboratories. Another advantage is that it does not require sequence information or probe collection prior to generating the fingerprints, and a set of primers can be used for different species. This feature is especially useful when DNA markers are rare. However, AFLP assays also have some limitations. For instance, polymorphic information content for bi-allelic markers is low (the maximum is 0.5). High quality DNA is required for complete restriction enzyme digestion. AFLP markers usually cluster densely in centromeric regions in some species with large genomes (e.g. barley and sunflower). In addition, marker development is complicated and not cost-efficient, especially for locus-specific markers. The applications of AFLP markers include biodiversity studies, analysis of germplasm collections, genotyping of individuals, identification of closely linked DNA markers, construction of genetic DNA marker maps and physical maps, gene mapping and transcript profiling.

15.2.2.4 SSR Markers

SSRs, also called *microsatellites*, *short tandem repeats* (STRs) or sequence-*tagged microsatellite sites* (STMS), are PCR-based markers. They are randomly tandem repeats of short nucleotide motifs (2–6 bp/nucleotides long). Di-, tri- and tetranucleotide repeats, e.g. (GT)n, (AAT)n and (GATA)n, are plentiful and widely distributed throughout the genomes of plants and animals. The copy number of these repeats varies among individuals and is a source of polymorphism in plants. Because the DNA sequences flanking microsatellite regions are usually conserved, primers specific for these regions can be designed for use in the PCR reaction. One of the most important attributes of microsatellite loci is their high level of allelic variation, thus making them valuable genetic markers. The unique sequences bordering the SSR motifs provide templates for specific primers to amplify the SSR alleles via PCR. SSR loci are individually amplified by PCR using pairs of oligonucleotide primers specific to unique DNA sequences flanking the SSR sequence. The

PCR-amplified products can be separated in high-resolution electrophoresis systems (e.g. AGE and PAGE) and the bands can be visually recorded by fluorescent labeling or silver-staining (Xu 2010).

SSR markers are characterized by their hyper-variability, reproducibility, co-dominant nature, locus-specificity and random genomewide distribution in most cases. The advantages of SSR markers also include that they can be readily analyzed by PCR and easily detected by PAGE or AGE. SSR markers can be multiplexed, have high throughput genotyping and can be automated. SSR assays require only very small DNA samples (~100 ng per individual) and low start-up costs for manual assay methods. Thus SSR makers are regarded as the most breeder-friendly marker system, although the SSR technique requires nucleotide information for primer design, labor-intensive marker development process and high start-up costs for automated detections. Since the 1990s, SSR markers have been extensively used in constructing genetic linkage maps, QTL mapping, marker-assisted selection and germplasm analysis in plants (Jiang 2013a). In many species, numerous breeder-friendly SSR markers have been developed and are available to breeders. In soybean, for instance, over 35,000 SSR markers have been developed and mapped onto all 20 linkage groups (Song et al. 2010).

15.2.2.5 SNP Markers

An SNP is a single nucleotide base difference between two DNA sequences or individuals. SNPs can be categorized according to nucleotide substitutions either as transitions (C/T or G/A) or transversions (C/G, A/T, C/A or T/G). In practice, single base variants in cDNA (mRNA) are considered to be SNPs as are single base insertions and deletions (indels) in the genome (Xu 2010). SNPs provide the ultimate/simplest form of molecular markers as a single nucleotide base is the smallest unit of inheritance, and thus they can provide maximum markers. SNPs occur very commonly in animals and plants. Typically, SNP frequencies are in a range of 1 SNP for every 100–300 bp in plants (Edwards et al. 2007; Xu 2010). SNPs may present within coding sequences of genes, non-coding regions of genes or in the intergenic regions between genes at different frequencies in different chromosome regions.

Based on various methods of allelic discrimination and detection platforms, many SNP genotyping methods have been developed. A convenient method for detecting SNPs is RFLP (SNP-RFLP) or by using the CAPS marker technique. If one allele contains a recognition site for a restriction enzyme while the other does not, digestion of the two alleles will produce fragments different in length. A simple procedure is to analyze the sequence data stored in the major databases and identify SNPs. Four alleles can be identified when the complete base sequence of a segment of DNA is considered and these are represented by A, T, G and C at each SNP locus in that segment. There are several SNP genotyping assays, such as allele-specific hybridization, primer extension, oligonucleotide ligation and invasive cleavage based on the molecular mechanisms (Sobrino et al. 2005), and different detection methods to analyze the products of each type of allelic discrimination reaction, such

as gel electrophoresis, mass spectrophotometry, chromatography, fluorescence polarization, arrays or chips, etc. At present, SNPs are also widely detected by sequencing (Jiang 2013a).

SNPs are co-dominant markers, often linked to genes and present in the simplest/ultimate form for polymorphism, and thus they have become very attractive potential genetic markers in genetic studies and breeding. In addition, SNPs can be very easily automated and quickly detected, with a high efficiency for detection of polymorphism. Therefore, it can be expected that SNPs will be increasingly used for various purposes, particularly as whole DNA sequences become available for more and more species (e.g. rice, soybean, maize, etc.). Recently SNPs have been increasingly used in genome-wide association studies in plants. However, high costs for start-up or marker development, the high-quality DNA required and high technical/equipment demands limit, to some extent, the application of SNPs in some laboratories and practical breeding programs, although commercial services of SNP assays are available for many species. Development of SNP chips, a high-density genotyping array with a huge number (thousands, tens or hundreds of thousands) of SNP markers, has provided a powerful tool for characterizing genetic diversity and linkage disequilibrium, and for constructing high-resolution linkage maps to improve the entire genome sequence assembly (Song et al. 2013).

15.3 Prerequisite and General Procedure of Marker-Assisted Breeding

15.3.1 Prerequisite and Considerations for an Efficient Marker-Assisted Breeding Program

Compared with conventional breeding approaches, DNA marker-assisted breeding requires more complicated equipment and facilities. In general, the prerequisites listed below are essential for marker-assisted breeding (MAB) in plants (Jiang 2013a).

15.3.1.1 Appropriate Marker System and Reliable Markers

For a plant species or crop, a suitable marker system and available reliable markers are critically important to initiate a marker-assisted breeding program. As discussed above, suitable markers should have the following attributes: (a) ease and low-cost of use and analysis; (b) small amount of DNA required; (c) co-dominance; (d) repeatability/reproducibility of results; (e) high levels of polymorphism; and (f) occurrence and even distribution genomewide (Jiang 2013a).

In addition, close association with the target gene(s) is another important desirable attribute for the markers to be used. If the markers are located in close proxim-

ity to the target gene or present within the gene, selection of the markers will ensure the success in selection of the gene. DNA markers for polymorphism are available throughout the genome, and their presence or absence is not affected by environment and usually do not directly affect the phenotype. DNA markers can be detected at any stage of plant growth, but the detection of classical markers is usually limited to certain growth stages. However, each type of markers has advantages and disadvantages for specific purposes. Relatively speaking, SSRs have most of the desirable features and thus are the ideal markers of choice for many crops. SNPs require more detailed knowledge of the specific, single nucleotide DNA changes responsible for genetic variation among individuals. As more and more SNPs have become available in many species, however, they are also considered an important type for marker-assisted breeding.

15.3.1.2 Quick DNA Extraction and High Throughput Marker Detection

For most plant breeding programs, hundreds to thousands of plants/individuals are usually screened for desired marker patterns. In addition, breeders need the results instantly to make selections in a timely manner. Therefore, a quick DNA extraction technique and a high throughput marker detection system are essential to handling a large number of tissue samples and a large-scale screening of multiple markers in breeding programs (Jiang 2013a). Extracting DNA from small tissue samples in 96- or 384-well plates and streamlined operations are adopted in many labs and programs. High throughput PAGE and AGE systems are commonly used for marker detection. Some labs also provide marker detection services using automated detection systems, e.g. SNP chips based on thousands to tens of thousands of markers.

15.3.1.3 Knowledge of Marker-Trait Association

The most crucial factor for marker-assisted breeding is knowledge of the associations between markers and the traits of interest. Only those markers that are closely associated with the target traits or tightly linked to the genes can provide a sufficient guarantee for success in practical breeding. The more closely the markers are associated with the traits, the higher the probability of success and efficiency of use will be. To obtain this information, one can use one or more appropriate methods, such as gene mapping, QTL analysis, association mapping, genomewide association study, classical mutant analysis, linkage or recombination analysis, bulked segregant analysis, etc. In addition, it is also critical to know the linkage situation, i.e. the markers that are linked in cis/trans (coupling or repulsion) with the desired allele of the trait. Even if some markers have been reported to be tightly linked with a QTL, a plant breeder still needs to determine the association of alleles in his own breeding material. The effects of a QTL are not fixed but vary with genetic backgrounds and environments (Wang et al. 2014a, b). This makes QTL information difficult to directly transfer between different materials.

15.3.1.4 Genetic Maps

Linkage maps provide a framework for detecting marker-trait associations and for choosing markers to use in marker-assisted breeding. Therefore, a genetic linkage map, particularly a high-density linkage map is very important for MAB (Jiang 2013a). To use markers and select a desired trait present in a specific germplasm line, a proper population of segregation for the trait of interest is required to construct a linkage map. Once a marker or a few markers are found to be associated with the trait in a given population, a dense molecular marker map in a standard reference population will help identify makers that are close to (or flank) the target gene. If a region is found associated with the desired traits, fine mapping also can be done with additional markers to identify the marker(s) tightly linked to the gene controlling the trait. A favorable genetic map should have an adequate number of evenly-spaced polymorphic markers to accurately locate desired QTLs/genes (Babu et al. 2004).

15.3.1.5 Quick and Efficient Data Processing and Management

Quick and efficient data processing and management may provide timely and useful reports for breeders. In a marker-assisted breeding program, as discussed above, not only are large numbers of samples handled, but multiple markers for each sample also need to be screened at the same time. As such, it is especially important to have a fast and efficient system for labeling, storing, retrieving, processing and analyzing large data sets, and even integrating data sets available from other programs. The development of bioinformatics and statistical software packages provides a useful tool for this purpose (Jiang 2013a).

15.3.1.6 Economic Factors

In addition to the abovementioned prerequisites, the cost of molecular tools and net return of end products are critical for a breeder to make a decision if MAB is an appropriate strategy and deserves to be used. Compared with conventional breeding programs, MAB needs additional cost to establish a MAB laboratory equipped with essential equipment (Fig. 15.1), e.g. DNA extraction instrument, PCR machine, electrophoresis system, gel imaging system, etc. If there is no MAB lab and/or equipment available within the breeding program, MAB may also be conducted by sending DNA samples or plant tissue samples to a specific laboratory which provides such services (marker screening or genotyping). In general, MAB is more suitable for large-scale breeding programs equipped with high throughput genotyping and phenotyping facilities. For breeding programs without high throughput genotyping, MAB may be used for the improvement of some specific traits of interest, such as resistance to diseases/pests, quality traits controlled by major genes. Only if the net return of end products can cover the cost related to the use of

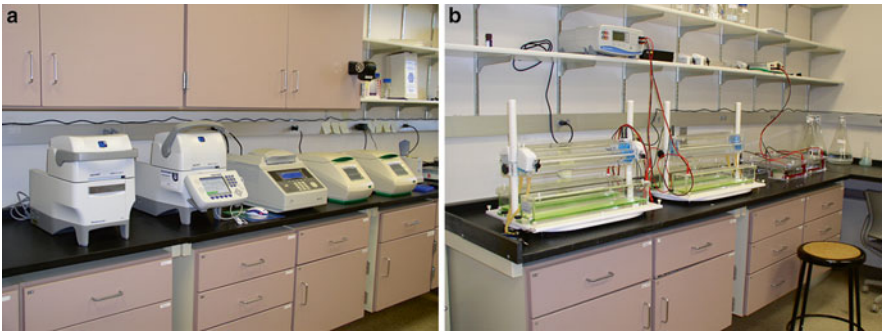

Fig. 15.1 Molecular-assisted breeding (MAB) laboratory: (**a**) thermal cyclers (PCR machines) and (**b**) electrophoresis systems

molecular tools, can MAB be employed by breeders. From the marketing point of view, of course, early release of new cultivars by MAB may be regarded as compensation for the cost invested in MAB.

15.3.2 Procedure of Marker-Assisted Breeding

Marker-assisted breeding involves the following activities provided the prerequisites are well equipped or available:

- Planting the breeding populations with potential segregation for traits of interest or polymorphism for the markers used.
- Sampling plant tissues, usually at early stages of growth, e.g. emergence to young seedling stage.
- Extracting DNA from tissue sample of each individual or family in the populations, and preparing DNA samples for PCR and marker screening.
- Running PCR or other amplifying operation for the molecular markers associated with or linked to the trait of interest (Fig. 15.1a).
- Separating and scoring PCR/amplified products, by means of appropriate separation and detection techniques (Fig. 15.1b), e.g. PAGE, AGE, etc.
- Identifying individuals/families carrying the desired marker alleles.
- Selecting the best individuals/families with both desired marker alleles for target traits and desirable performance/phenotypes of other traits, by jointly using marker data and other selection criteria.
- Repeating the above activities for several generations, depending upon the association between the markers and the traits, as well as the status of marker alleles (homozygous or heterozygous), and advancing the individuals selected in breeding program until stable superior or elite lines with improved traits are developed.

15.4 Marker-Assisted Selection

15.4.1 MAS Procedure and Theoretical and Practical Considerations

Marker-assisted selection (MAS) refers to a breeding procedure in which DNA marker detection and selection are integrated into a traditional breeding program. Taking a single cross as an example, the general procedure can be described as follows (Jiang 2013a):

- Select parents and make the cross, at least one (or both) possess the DNA marker allele(s) for the desired trait of interest.
- Plant F_1 population and detect the presence of the marker alleles to eliminate false hybrids.
- Plant segregating F_2 population, screen individuals for the marker(s), and harvest the individuals carrying the desired marker allele(s).
- Plant $F_{2:3}$ plant rows, and screen individual plants with the marker(s). A bulk of F_3 individuals within a plant row may be used for the marker screening for further confirmation in case needed if the preceding F_2 plant is homozygous for the markers. Select and harvest the individuals with required marker alleles and other desirable traits.
- In the subsequent generations (F_4 and F_5), conduct marker screening and make selection similarly as for $F_{2:3}$ s, but more attention is given to superior individuals within homozygous lines/rows of markers.
- In $F_{5:6}$ or $F_{4:5}$ generations, bulk the best lines according to the phenotypic evaluation of target trait and the performance of other traits, in addition to marker data.
- Plant yield trials and comprehensively evaluate the selected lines for yield, quality, resistance and other characters of interest (Fig. 15.2).

A frequently asked question about marker-assisted selection concerns the number of QTLs that should be selected for MAS. Theoretically, all the QTLs contributing to the trait of interest could be taken into account. For a quantitatively-inherited character like yield, however, numerous QTLs or genes are usually involved. It is almost impossible to select all QTLs or genes simultaneously so that the selected individuals incorporate all the desired QTLs due to the limitation of resources and facilities (Jiang 2013a). More loci involved in MAS might not necessarily lead to higher prediction accuracy or efficiency as expected (Zhang et al. 2015) The number of individuals in the population increases exponentially with the increase of target loci involved. The relative efficiency of MAS decreases as the number of QTLs increases and their heritability decreases (Moreau et al. 1998). In other words, MAS will be less effective for a highly complex character governed by more genes than for a simply inherited character controlled by fewer genes. In addition, the number of genes/QTLs also impacts the breeding design and implementing scheme (details discussed below). Typically no more than three QTLs are regarded as an appropriate and feasible choice (Ribaut and Betran 1999), although five QTLs were

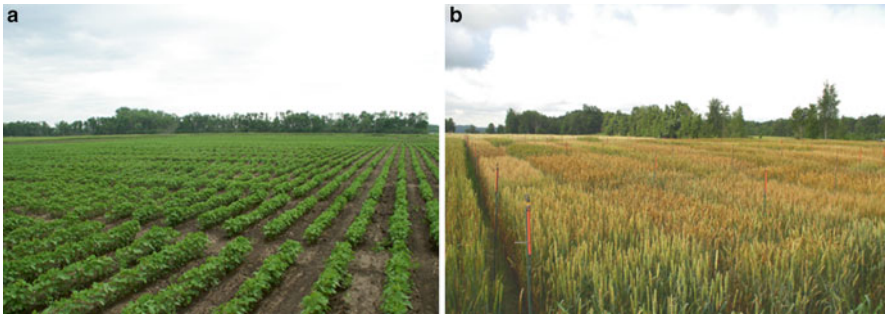

Fig. 15.2 Breeding programs: (**a**) soybean yield trials and (**b**) wheat disease (head scab) nursery

used in improvement of fruit quality traits in tomato via marker-assisted introgression (Lecomte et al. 2004). With development of SNP markers (especially rapid automated detection and genotyping technologies), selection of more QTLs at the same time might be preferred and practicable (Kumpatla et al. 2012).

For MAS for multiple genes/QTLs, Hospital (2003) suggested that the number of genes undergoing selection should be limited to three to four if they are the QTLs selected on the basis of linked markers, and to five to six if they are the known loci selected directly. Only the multi-environmentally verified QTLs that possess medium to large effects are selected. The first priority should be given to the major QTLs that can explain the greatest proportion of phenotypic variation and/or can be consistently detected across a range of environments and different populations. In addition, an index for selection that weights markers differently could be constructed, depending on their relative importance to the breeding objectives. Flint-Garcia et al. (2003) presented an example of such an index used to select for QTLs with different effect magnitudes.

Another commonly asked question concerns the number of markers which should be used in MAS. The more markers associated with a QTL that are used, the greater opportunity of success in selecting the QTL of interest will be ensured (Jiang 2013a). However, efficiency is also important for a breeding program, especially when the resources and facilities are limited. Taking both effectiveness and efficiency into account, for a single QTL it is usually suggested to use two markers (i.e. flanking markers) that are tightly linked to the QTL of interest. The markers to be used should be close enough to the gene/QTL of interest (<5 cM) to ensure that only a minor proportion of the selected individuals will be recombinants. One marker alone should be appropriate if it is found to be located within the region of gene sequence of interest or in such a close proximity to the QTL/gene that no recombination occurs between the marker and the QTL/gene. However, if a marker is not tightly linked to a gene of interest, recombination between the marker and gene may reduce the efficiency of MAS because a single crossover may alternate the linkage association and lead to selection errors. The efficiency of MAS decreases with the increase of the recombination frequency (genetic distance) between the marker and gene. Use of two flanking markers rather than one may decrease the chance of such

errors due to homologous recombination and increase the efficiency of MAS. In this case, only a double crossover (i.e. two single crossovers occurring simultaneously on both sides of the gene/QTL in the region) may result in selection errors, but the occurrence of a double crossover is considerably rare. For instance, if two flanking markers with an interval of 20 cM or so between them are used, there will be higher probability (99 %) for recovery of the target gene than if only one marker is used.

In practical MAS, a breeder is also concerned about how the markers should be detected, how many generations of MAS have to be conducted and how large size of the population is needed (Jiang 2013a). In general, detection of marker polymorphism is performed at early stages of plant growth. This is true especially for marker-assisted backcrossing and marker-assisted recurrent selection, because only the individuals that carry preferred marker alleles are expected to be used in backcrossing to the recurrent parent and/or inter-mating between selected individuals/progenies. The generations of MAS required vary with the number of markers used, the degree of association between the markers and the QTLs/genes of interest, and the status of marker alleles. In many cases, marker screening is performed for two to four consecutive generations in a segregating population. Fewer markers are used and the markers are in a closer proximity to the QTL or gene of interest, fewer generations will be needed. Marker screening may be stopped if homozygous status of marker alleles of interest is detected in two consecutive generations. Bonnett et al. (2005) discussed the strategies for efficient implementation of MAS involving several issues, e.g. breeding systems or schemes, population sizes, number of target loci, etc. Their strategies include F_2 enrichment, backcrossing, and inbreeding.

In MAS, phenotypic evaluation and selection is still very helpful if conditions permit, and even necessary in cases when the QTLs selected for MAS are unstable across environments and the association between the selected markers and QTLs is not so close. Moreover, the impact of genetic background should also be taken into consideration. The presence of a QTL or marker does not necessarily guarantee the expression of the desired trait (Jiang 2013a). QTL data derived from multiple environments and different populations provides a better understanding of the interactions of QTL × environment and QTL × QTL or QTL × genetic background, and thus assures a better use of MAS. In addition to genotypic (markers) and phenotypic data for the trait of interest, a breeder often pays considerable attention to other important traits, because the trait of interest is not the only objective of breeding in most cases.

There are several indications for adoption of molecular markers in the selection for the traits of interest in practical breeding (Collard and Mackill 2008; Jiang 2013a). The situations favorable for MAS include:

- The selected character is expressed late in plant development, like fruit and flower features or adult characters with a juvenile period (so that it is not necessary to wait for the plant to become fully developed before propagation occurs or can be arranged).
- The target gene is recessive (so that individuals which are heterozygous for the recessive allele can be selected and/or crossed to produce some homozygous offspring with the desired trait).

- Special conditions are required in order to invoke expression of the target gene(s), as in the case of breeding for disease and pest resistance (where inoculation with the disease or subjection to pests would otherwise be required), or the expression of target genes is highly variable with the environments.
- The phenotype of a trait is conditioned by two or more unlinked genes. For example, selection for multiple genes or gene pyramiding may be required to develop enhanced or durable resistance against diseases or insect pests.

15.4.2 MAS for Major Genes or Improvement of Qualitative Traits

In crop plants, many economically-important characteristics are controlled by major genes/QTLs, such as resistance to diseases/pests, male sterility, self-incompatibility and others related to shape, color and architecture of whole plants and/or plant parts. These traits are often of mono- or oligogenic inheritance in nature. Even for some quality traits, one or a few major QTLs or genes can account for a very high proportion of the phenotypic variation of the trait (Bilyeu et al. 2006; Pham et al. 2012). Therefore, transfer of such a gene to a specific line can lead to tremendous improvement of the trait in the cultivar under development. The marker loci which are tightly linked to major genes can be used for selection and sometimes might be more efficient than direct selection for the target genes. Such advantages in efficiency may be due to higher expression of the marker mRNA in the cases that the marker is actually within a gene. Alternatively, in such cases that the target gene of interest differs between two alleles by a difficult-to-detect SNP, an external marker of which polymorphism is easier to detect may present as the most realistic option.

Soybean cyst nematode (SCN) (*Heterodera glycines*) may be taken as an example of MAS for major genes. This pathogen is the most economically-significant soybean pest. The principal strategy to reduce or eliminate damage from this pest is the use of resistant cultivars (Cregan et al. 1999). However, it is a difficult and expensive process to identify resistant segregants in breeding populations. A widely used phenotypic assay takes 5 weeks, requires a large greenhouse space, and about 5–10 h of labor for every 100 plant samples processed (Young 1999). Fortunately, the SSR marker Satt309 has been identified to be located only 1–2 cM away from the resistance gene *rhg1* (Cregan et al. 1999), which has laid the basis for many public and commercial breeding efforts. In a direct comparison, genotypic selection with Satt309 was 99 % accurate in predicting lines that were susceptible in subsequent greenhouse assays for two test populations, and 80 % accurate in a third population, each with a different source of SCN resistance (Young 1999). Also in soybean, using molecular markers in a cross J05×V94-5152, Shi et al. (2009) developed five $F_{4:5}$ lines that were homozygous for all eight marker alleles linked to the genes/loci of resistance to soybean mosaic virus (SMV). These lines exhibited resistance to SMV strains G1 and G7 and presumably carried all three resistance

genes (*Rsv1*, *Rsv3* and *Rsv4*) that would potentially provide broad and durable resistance to SMV (Shi et al. 2009).

15.4.3 MAS for Improvement of Quantitative Traits

Most of the important agronomic traits are polygenic or controlled by multiple QTLs. MAS for the improvement of such traits is a complex and difficult task because it involves many genes or QTLs, QTL×E interaction and epistasis (Jiang 2013a). Usually, each of these genes has a small effect on the phenotypic expression of the trait and expression is affected by environmental conditions. The QTL×E interaction reduces the efficiency of MAS and epistasis can result in a skewed QTL effect on the trait. Therefore, phenotyping of quantitative traits is a complex endeavor, and determining marker-phenotype association becomes difficult as well. To accurately characterize the effects of the QTLs and to evaluate the stability across environments, repeated field tests are required.

Despite a tremendous amount of QTL mapping experiment over the past decade, application and utilization of QTL mapping information in plant breeding has been constrained by a number of factors (Collard and Mackill 2008):

- Strong QTL-environmental interaction which makes phenotyping difficult since expression may vary from one location/year to another.
- Lack of universally-valid QTL-marker associations applicable across populations. A new project of QTL mapping is usually required to identify new QTL markers whenever new germplasm is used, putting some breeders off and decreasing their interest in MAS.
- Deficiencies in QTL statistical analysis which lead to either overestimation or underestimation of the number of QTLs involved and their effect on the trait.
- Often times, there are no QTLs with major effects on the trait, and this means a large number of QTLs have to be identified. In many cases this becomes a difficult goal to achieve and further complicates identification of marker-QTL association.

In order to improve the efficiency of MAS for quantitative traits, appropriate field experimental designs and approaches have to be employed. Attention should be given to replications both over time and space, consistency in experimental techniques, samplings and evaluations, robust data processing and statistical analysis (Jiang 2013a). For example, composite interval mapping (CIM) allows the integration of data from different locations for joint analysis to estimate QTL-environment interaction so that stable QTLs across environments can be identified. A saturated linkage map enables accurate identification of both targeted QTLs as well as linked QTLs in coupling and repulsion linkage phases. In practical breeding for improvement of a quantitative trait, usually not many minor QTLs are considered but only a few major QTLs are used in MAS (Jiang 2013a). In case many QTLs, especially

minor-effect QTLs, are involved, a breeder would prefer to consider the strategy of gene pyramiding (see the later section).

Fusarium head blight (FHB) or scab caused by *Fusarium* species is one of the most destructive worldwide disease in wheat and barley. To combat this disease, a great effort worldwide from multiple fields, including plant breeding and genetics, molecular genetics and genomics, plant pathology, and integrated management, has been dedicated since the 1990s. Resistance to FHB in both wheat and barley is quantitatively inherited, and many QTLs have been identified from different germplasm resources (Buerstmayr et al. 2009; Jiang et al. 2007a, b). Use of MAS to improve the resistance has become a choice for many breeding programs. In wheat, a major QTL designated as *Fhb1* was consistently detected across multiple environments and populations, explaining 20–40 % of phenotypic variation in most cases (Buerstmayr et al. 2009; Jiang et al. 2007a, b). Thus wheat breeders would especially prefer to use this major QTL to develop new cultivars with FHB resistance. Pumphrey et al. (2007) compared 19 pairs of NIL for *Fhb1* derived from an ongoing breeding program and found that the average reduction in disease severity between NIL pairs was 23 % for disease severity and 27 % for kernel infection. Later investigations from the group also demonstrated successful implementation of MAS for this QTL (Anderson et al. 2007).

In addition, researchers also tried to incorporate multiple QTLs by MAS to improve resistance to FHB. Miedaner et al. (2006) demonstrated that MAS for three FHB resistance QTLs simultaneously was highly effective in enhancing FHB resistance in German spring wheat. FHB resistance was the highest in recombinant lines with multiple QTLs combined, especially 3B plus 5A. Jiang et al. (2007a, b) made a comparison of multiple-locus combinations in a RIL population derived from the cross Veery × CJ 9306. For three loci, the average levels of resistance, from low to high, in genotypes were: no favorable allele – one favorable allele – two favorable alleles – three favorable alleles, except for the non-reciprocal comparisons. When four or five loci carrying favorable alleles from the resistant parent CJ 9306 were considered simultaneously, the coefficients of determination between the accumulated effects of alleles for different combinations and the averages of number or percentage of diseased spikelets for the corresponding RILs were 0.33–0.41 ($P<0.01$) (Jiang et al. 2007a). Therefore, the authors concluded that the effects of FHB-resistance QTLs could be accumulated and the resistance could be feasibly enhanced by selection of favorable marker alleles for multiple loci in breeding programs.

Asoro et al. (2013) recently compared the responses to two cycles of selection for β-glucan concentration in oat with marker-based selection methods and the best linear unbiased prediction (BLUP) – based phenotypic selection. They found that the averages of marker-based selection methods in Cycle 2 were significantly greater than those of phenotypic selection, and progenies with the highest β-glucan level came from the marker-based selection methods. In addition, greater genetic variance was maintained by marker-based selection methods than BLUP phenotypic selection, suggesting that marker-based selection methods could enable greater future selection gains (Asoro et al. 2013). However, MAS for higher β-glucan

concentration resulted in a later heading date. Since 2004, the USDA-NIFA has awarded many Coordinated Agricultural Projects (CAPs) to encourage collaborative efforts in applied plant genomics and molecular research in several crops, such as rice, wheat, barley, beans, potato, tomato, etc. An important strategy of these CAPs is applying marker-assisted selection to plant breeding and efficiently using genetic resources and facilities available, including thousands and tens of thousands of DNA markers and plant introductions, to facilitate development of crop cultivars with improved yield, resistance and quality. It can be expected that MAS will be more extensively applied to improvement of quantitative traits in crops.

15.5 Marker-Assisted Backcrossing

15.5.1 MABC Procedure and Theoretical and Practical Considerations

Marker-assisted or marker-based backcrossing (MABC) is regarded as the simplest form of marker-assisted selection, and at present it is the most widely and successfully used method in practical molecular breeding (Jiang 2013a). MABC aims to transfer one or a few genes/QTLs of interest from one genetic source (serving as the donor parent and may be inferior agronomically or fall short in comprehensive performance in many cases) into a superior cultivar or elite breeding line (serving as the recurrent parent) to improve the targeted trait. Unlike traditional backcrossing, MABC is based on the alleles of markers associated with or linked to gene(s)/QTL(s) of interest instead of phenotypic performance of target trait. The general procedure of MABC is as follows, regardless of dominant or recessive nature of the target trait in inheritance (Jiang 2013a):

- Select parents and make the cross, one parent is superior in comprehensive performance and serves as recurrent parent (RP), and the other one used as donor parent (DP) should possess the desired trait and the DNA marker's allele(s) associated with or linked to the gene for the trait.
- Plant F_1 population and detect the presence of the marker allele(s) at early stages of growth to eliminate false hybrids, and cross the true F_1 plants back to the RP.
- Plant BCF_1 population, screen individuals for the marker(s) at early growth stages, and cross the individuals carrying the desired marker allele(s) (in heterozygous status) back to the RP. Repeat this step in subsequent seasons for two to four generations, depending upon the practical requirements and operation situations as discussed below.
- Plant the final backcrossing population (e.g. BC_4F_1), and screen individual plants with the desired marker(s) for the target trait and discard the individuals carrying homozygous RP marker alleles for the target trait. Have the individuals with required marker allele(s) selfed and harvest them.

- Plant the progenies of backcrossing-selfing (e.g. BC_4F_2), detect the markers and harvest individuals carrying homozygous DP marker allele(s) of target trait for further evaluation and release.

Theoretically, in a population large enough in size (or with adequate individuals) and provided no selection is made during backcrossing (i.e. *blind* backcrossing only), the proportion of the RP genome after n generations of backcrossing is given by $1-(1/2)^{n+1}$ for a single locus and $[1-(1/2)^{n+1}]^k$ for k loci, respectively. The percentage of the RP genome is the average of the population, with some individuals possessing more and others possessing less of the RP genome. To fully recover the genome of the RP, 6–8 generations of backcrossing is needed typically in case no selection is made for the RP. However, this process is usually slower than expected for the target gene-carrier chromosome, i.e. linkage drag, especially in case a linkage exists between the target gene and other undesirable traits (Jiang 2013a). On the other hand, the process of introgression of QTLs/genes and recovery of the RP genome may be accelerated by selection using markers flanking the QTLs and evenly spaced markers from other chromosomes (i.e. unlinked to the QTLs) of the RP (Collard et al. 2005) or selection for the performance of the RP conducted simultaneously. For MABC program, therefore, there are two types of selection recognized: foreground selection and background selection (Hospital 2003).

In foreground selection, the selection is made only for the marker allele(s) of the donor parent at the target locus to maintain the target locus in a heterozygous state until the final backcrossing is completed. Then the selected plants are selfed and the progeny plants with homozygous DP allele(s) of selected markers are harvested for further evaluation and release. As described above, this is the general procedure of MABC. The effectiveness of foreground selection depends on the number of genes/loci involved in the selection, the marker-gene/QTL association or linkage distance and the undesirable linkage to the target gene/QTL (Jiang 2013a).

In background selection, however, the selection is made for the marker alleles of recurrent parent in all genomic regions of desirable traits except the target locus, or selection against the undesirable genome of donor parent. The objective is to hasten the restoration of the RP genome and eliminate undesirable genes brought in by the DP. The progress in recovery of the RP genome depends on the number of markers used in background selection. The more evenly are makers located on all the chromosomes are selected for the RP alleles, the faster the recovery of the RP genome will be achieved, but larger population size and more genotyping will be required as well. In addition, the linkage drag also can be efficiently addressed by background selection using DNA markers, although it is difficult to overcome in a traditional backcrossing program.

Foreground selection and background selection are two respective aspects of MABC with different foci of selection. In practice, however, both foreground and background selections are usually conducted in the same program, either simultaneously or successively (Jiang 2013a). In many cases, they can be performed alternatively even in the same generation. The individuals that have the desired marker alleles for target trait are selected first (foreground selection). Then the selected

individuals are screened for other marker alleles again for the RP genome (background selection). It is understandable and appropriate to do so because selection of the target gene/QTL is the essential and only critical point for the backcrossing program, and the individuals that do not have the allele of the target gene will be discarded and thus it is not necessary to genotype them for other traits.

The efficiency of MABC depends upon several factors, such as the population size for each generation of backcrossing, marker-gene association or the distance of markers from the target locus, number of markers used for target trait and RP background, as well as undesirable linkage drag (Jiang 2013a). Based on simulations of 1000 replicates, Hospital (2003) presented the expected results of a typical MABC program, in which heterozygotes were selected at the target locus in each generation, and RP alleles were selected for two flanking markers on target chromosome each located 2 cM apart from the target locus and for three markers on non-target chromosomes. As shown in Table 15.2, a faster recovery of the RP genome could be achieved by MABC with combined foreground and background selection, compared to traditional backcrossing. Therefore, MABC can lead to considerable time savings compared to conventional backcrossing (Collard et al. 2005; Frisch et al. 1999).

In a MABC program, the population to be analyzed should contain at least one genotype that has all favorable alleles for a particular QTL. Later on, the number of QTLs may be increased progressively, but not beyond 6 QTLs in most cases because of prohibitive difficulty in handling all QTLs (Hospital 2003). In addition, the more QTLs/genes are transferred, the larger the proportion of unwanted genes would be due to linkage drag. In general, most of the unwanted genes are located on non-target chromosomes in early BC generations, and are rapidly removed in subsequent BC generations. On the contrary, the quantity of DP genes on the target chromosome decreases much more slowly, and even after generation BC_6 many of the unwanted donor genes are still located on the target chromosome in a segregating state (Newbury 2003). Given a total genome length is 3000 cM, 1 % donor DNA fragments after 6 backcrosses represents a 30 cM chromosomal segment or region, which may host many unwanted genes, especially if the DP is a wild genetic resource. Young and Tanksley (1989) genotyped a collection of tomato varieties in

Table 15.2 Expected results of a MABC program with combined foreground and background selection used

Backcross generation	Number of individuals	Homozygosity (%) of recurrent parent alleles at selected markers		Recurrent (%) parent genome	
		Chromosome with target locus	All other chromosomes	Marker-assisted backcross	Conventional backcross
BC_1	70	38.4	60.6	79.0	75.0
BC_2	100	73.6	87.4	92.2	87.5
BC_3	150	93.0	98.8	98.0	93.7
BC_4	300	100.0	100.0	99.0	96.9

Adapted from Hospital (2003)

which the resistance gene was previously transferred at the *Tm-2* locus with RLFP markers. Their data indicated that the size of chromosomal segment retained around the *Tm-2* locus during backcross breeding was very variable, with one line exhibiting a donor segment of 50 cM after 11 backcrosses and the other possessing 36 cM donor segment after 21 backcrosses. This clearly demonstrates the need for background selection.

Linkage drag unfavorably impacts the application and efficiency of backcrossing. As discussed above, linkage drag can be reduced by performing background selection. Typically, two markers flanking the target gene are used, and the individuals (or double recombinants) that are heterozygous at the target locus and homozygous for the recipient (RP) alleles at both flanking markers are selected (Jiang 2013a). Use of closer flanking markers leads to more effective and faster reduction of linkage drag compared to distant markers. However, less distance between two flanking markers implies less probability of double recombination, and thus larger populations and more genotyping are needed. In order to optimize the genotyping effort (i.e. the cost of the program), therefore, it is important to determine the minimal population sizes necessary to ensure the desired genotypes can be obtained. Hospital and Decoux (2002) developed statistical software for determining the minimum population size required in BC program to identify at least one individual that is double-recombinant with heterozygosity at the target locus and homozygosity for recurrent parent alleles at the flanking marker loci. In addition, another point should be taken into consideration. For closely-linked flanking markers, it is unlikely to obtain double recombinant genotypes through only one generation of backcrossing. Therefore, additional backcrossing should be conducted. For instance, in one BC generation (e.g. BC_1) single recombination on one side of the target gene is selection, and single recombination on the other side may be selected in another BC generation (e.g. BC_2) (Young and Tanksley 1989). In this way, individuals with desired RP alleles at two flanking markers and donor allele at target locus can be finally obtained.

To accelerate the recovery of RP genome on non-target chromosomes, scientists suggested using markers in backcrossing and discussed how many makers should be used (Hospital et al. 1992; Tanksley et al. 1989; Visscher et al. 1996). In background selection, the approaches involve selecting individuals that are of homozygous recipient type at a collection of markers located on non-carrier chromosomes. From a point of both effectiveness and efficiency, it is important to determine an appropriate number of markers to be used. More markers do not necessarily lead to better benefits in practice (Jiang 2013a). Generally, several markers are involved and MABC should be performed over two or more generations. It is unlikely that the selection objective can be realized in a single BC generation.

Dense marker coverage of non-target chromosomes is not mandatory to increase the overall proportion of recurrent parent genome, unless fine-mapping of specific chromosome regions is highly important. An appropriate number of markers and optimal position on chromosomes are important (Jiang 2013a). Computer simulation suggested that for a chromosome of 100 cM, two to four markers are sufficient, and selection based on markers would be most efficient if the markers are optimally

positioned along the chromosomes (Servin and Hospital 2002). In practice, at least two or three markers per chromosome are needed, and every chromosome should be involved. In such a MABC scheme, three to four generations of backcrossing is generally enough to achieve more than 99 % of the recurrent parent genome. With respect to the time necessary to release new varieties, the gain due to background selection can be economically valuable. In addition, background selection in late BC generations is more efficient than that conducted in early BC generations. For example, if a BC breeding scheme is conducted over three successive BC generations and yet the preference is to genotype individuals only once, then it is more efficient to genotype and select the individuals in BC_3 generation rather than in the BC_1 generation (Hospital et al. 1992; Ribaut et al. 2002).

15.5.2 Application of MABC

Success in integrating MABC as a breeding approach lies in identifying situations in which markers offer noticeable advantages over conventional backcrossing or valuable complements to conventional breeding efforts (Jiang 2013a). MABC is essential and advantageous when: (a) phenotyping is difficult and/or expensive or impossible; (b) heritability of the target trait is low; (c) the trait is expressed in late stages of plant development and growth, such as flowers, fruits, seeds; (d) the traits are controlled by genes that require special conditions to express; (e) the traits are controlled by recessive genes; and (f) gene pyramiding is needed for one or more traits.

MABC has been the most widely and successfully used molecular breeding method in plants to date. It has been applied to different types of traits (e.g. disease/pest resistance, drought tolerance and quality) in many species, e.g. rice, wheat, maize, barley, pear millet, soybean, tomato, etc. (Collard et al. 2005; Dwivedi et al. 2007; Xu 2010). In maize, for example, *Bacillus thuringiens* is a bacterium that produces insecticidal toxins, which can kill corn borer larvae when they ingest the toxins in corn cells (Ragot et al. 1995). The integration of the *Bt* transgene into various maize genetic backgrounds has been achieved by using MABC. Aroma in rice is controlled by a recessive gene which is due to an eight base-pair deletion and three single nucleotide polymorphism in a gene that codes for betaine aldehyde dehydrogenase 2 (Bradbury et al. 2005a). This discovery allows identification of the aromatic and non-aromatic rice varieties and discriminates homozygous recessive and dominant as well as heterozygous individuals in segregating populations for the trait. MABC has been used to select for aroma in rice (Bradbury et al. 2005b). High lysine opaque-2 gene in maize was incorporated using MABC (Babu et al. 2005). However, the rate of success decreases when large numbers of QTLs are targeted for introgression. Sebolt et al. (2000) used MABC for two QTL for seed protein content in soybean. However, only one QTL was confirmed in $BC_3F_{4:5}$. When that QTL was introduced in three different genetic backgrounds, it had no effect in one background. In tomato, Tanksley and Nelson (1996) proposed a MABC strategy,

advanced backcross-QTL (AB-QTL), to transfer resistance genes from wild relative/unadapted genotype into elite germplasm. The strategy has proven effective for various agronomically-important traits in tomato, including fruit quality and black mold resistance (Bernacchi et al. 1998; Fulton et al. 2002; Tanksley and Nelson 1996). In addition, AB-QTL has been used in other crop species, such as rice, barley, wheat, maize, cotton and soybean, collectively demonstrating that this strategy is effective in transferring favorable alleles from the wild/unadapted germplasm to elite germplasm (Concibido et al. 2003; Wang and Chee 2010).

In barley, a marker linked (0.7 cM) to the *Yd2* gene for resistance to barley yellow dwarf virus was successfully used to select for resistance in a backcrossing scheme (Jefferies et al. 2003). Compared to lines without the marker, the BC_2F_2-derived lines carrying the linked marker had lighter leaf symptoms and higher yield when infected by the virus. In maize, marker-facilitated backcrossing was also successfully employed to improve complex traits such as grain yield. Using MABC, six chromosomal segments each in two elite lines, Tx303 and Oh43, were transferred into two widely used inbred lines, B73 and Mo17, through three generations of backcrossing followed by two selfing generations. Then the enhanced lines with better performance were selected based on initial evaluations of testcross hybrids. The single-cross hybrids of enhanced B73 × enhanced Mo17 out-yielded the check hybrids by 12–15 % (Stuber et al. 1999). Zhao et al. (2012) reported that a major quantitative trait locus (named *qHSR1*) for resistance to head smut in maize was successfully integrated into ten high-yielding inbred lines (susceptible to head smut). Each of the ten high-yielding lines was crossed with a donor parent Ji 1037 that contains *qHSR1* and is completely resistant to head smut, followed by five generations of backcrossing to the respective recurrent parents. In BC_1 through BC_3 only phenotypic selection was conducted to identify highly resistant individuals after artificial inoculation. In BC_4 phenotypic selection, foreground selection and recombinant selection were conducted to screen for resistant individuals with the shortest *qHSR1* donor regions. In BC_5, phenotypic selection, foreground selection and background selection were performed to identify resistant individuals with the highest proportion of the recurrent parent genome, followed by one generation of self-pollination to obtain homozygous genotypes at the *qHSR1* locus. The ten improved inbred lines all showed substantial resistance to head smut, and the hybrids derived from these lines also showed a significant increase in the resistance. With the aid of DNA markers, Wan et al. (2014) developed a new restorer line of rice (KR022) possessing stacked BPH-resistance genes *Bph14* and *Bph15*, *Bt* gene *cry1C* and glufosinate-resistance gene *bar*. Semagn et al. (2006b) provided a detail review on the progress and prospects of MABC in crop breeding.

Recently, a cooperative marker-based backcrossing project for high-oleic acid in soybean has been initiated among multiple U.S. land-grant universities and USDA-ARS (Jiang 2013a). Backcrossing and selection are performed using the markers tightly linked to the high-oleic genes/loci. The high-oleic (80 % or higher) traits are being transferred from mutant lines or derived lines into other locally superior cultivars/lines, or combined with other unique traits like low linolenic acid (1–2 %) (Pham et al. 2012).

15.6 Marker-Assisted Gene Pyramiding and Marker-Assisted Recurrent Selection

One of the most important applications of DNA markers to plant breeding is gene pyramiding. Marker-assisted gene pyramiding (MAGP) has been proposed and applied to enhance resistance to diseases and insects by selecting for two or more genes at a time (Jiang 2013a). For example, in rice such pyramids have been developed against bacterial blight and blast (Huang et al. 1997; Luo et al. 2012; Singh et al. 2001). Castro et al. (2003) reported success in pyramiding qualitative gene and QTLs for resistance to stripe rust in barley. The advantage of using markers in this case allows selecting for QTL-allele-linked markers that have the same phenotypic effect. To enhance or improve a quantitatively inherited trait in plant breeding, pyramiding of multiple genes or QTLs has been recommended as a potential strategy (Richardson et al. 2006). The cumulative effects of multiple-QTL pyramiding have been proven in crop species like wheat, barley and soybean (Jiang et al. 2007a, b; Li et al. 2010; Richardson et al. 2006; Wang et al. 2012, 2014a, 2015). Pyramiding of multiple genes/QTLs may be achieved through different approaches: multiple-parent crossing or complex crossing, backcrossing, and recurrent selection. A suitable breeding scheme for MAGP depends on the number of genes/QTLs required for improvement of traits, the number of parents that contain the required genes/QTLs, the heritability of traits of interest, and other factors (e.g. marker-gene association, expected duration to complete the plan and relative cost) (Jiang 2013a). Assuming three or four desired genes/QTLs exist separately in three or four lines, pyramiding of them can be realized by three-way, four-way or double crossing. They may also be integrated by convergent backcrossing or stepwise backcrossing (See below for detail). However, if there are more than four genes/QTLs to be pyramided, a breeder often prefers to complex or multiple crossing and/or recurrent selection.

15.6.1 Marker-Assisted Gene Pyramiding

For MABC-based gene pyramiding, in general, there may be three strategies or breeding schemes: stepwise, simultaneous/synchronized and convergent backcrossing or transfer (Jiang 2013a). Supposing cultivar W is superior in comprehensive performance but lack of a trait of interest, and four different genes/QTLs (desired gene/marker alleles designed as A, B, C and D) contributing to the trait have been identified in four germplasm lines (e.g. P1, P2, P3 and P4). Three MABC schemes for pyramiding the genes/QTLs are described in Figs. 15.3, 15.4 and 15.5, respectively.

In the stepwise MABC (Fig. 15.3), four target genes/QTLs are transferred into the recurrent parent W in order. In one step of backcrossing, one gene/QTL is targeted and selected, followed by the next step of backcrossing for another gene/QTL,

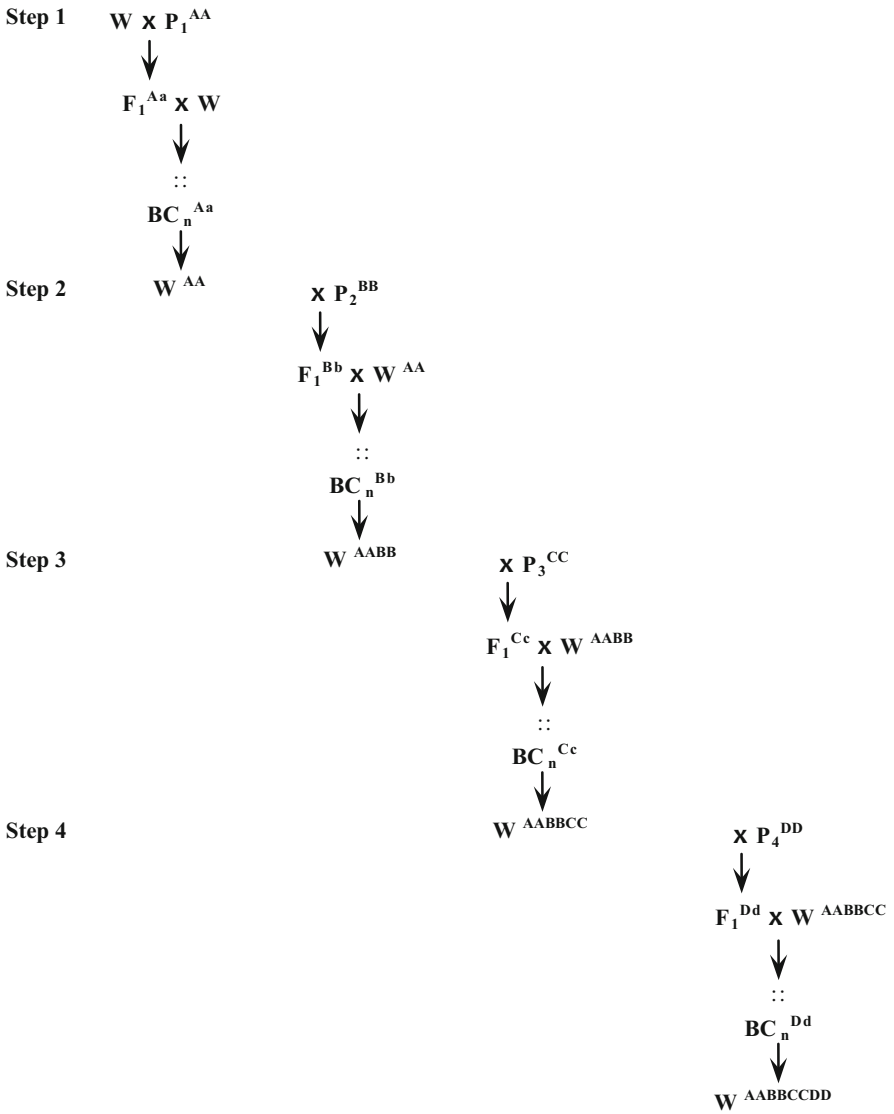

Fig. 15.3 Procedure of stepwise marker-assisted backcrossing (MABC) (Jiang 2013a)

until all target genes/QTLs have been introgressed into the RP. The advantage of this scheme is that gene pyramiding is more precise and easier to implement as it involves only one gene/QTL at one time and thus the population size and genotyping amount will be small. The improved recurrent parent may be released before the final step as long as the integrated genes/QTLs (e.g. two or three) meet the requirement at that time (Jiang 2013a). The disadvantage is that it takes a longer time to complete. In the simultaneous or synchronized MABC (Fig. 15.4) the recurrent

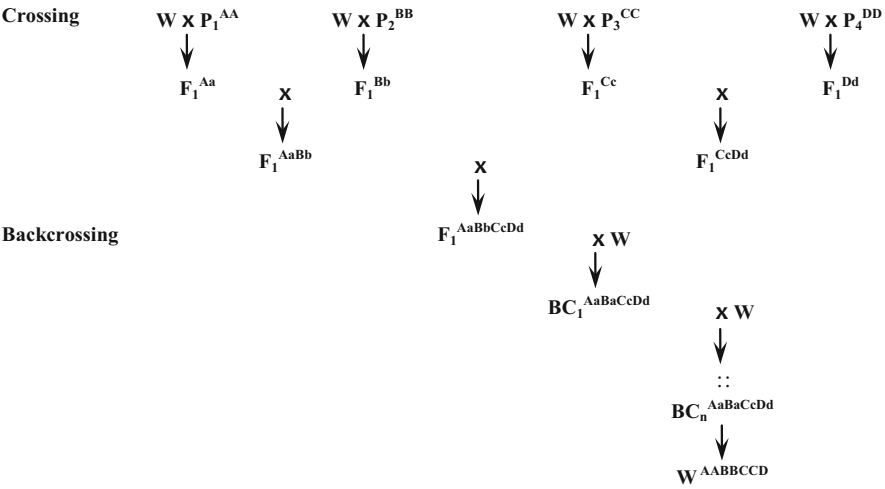

Fig. 15.4 Procedure of simultaneous/synchronized marker-assisted backcrossing (MABC) (Jiang 2013a)

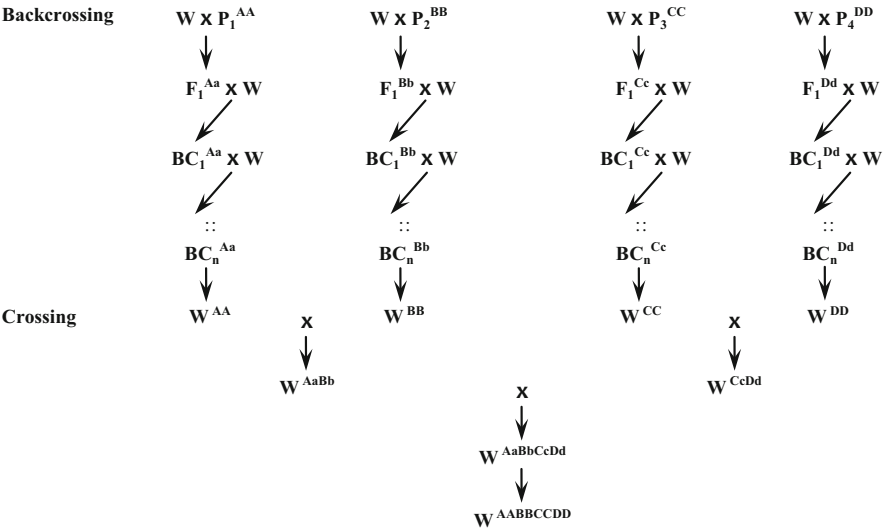

Fig. 15.5 Procedure of convergent marker-assisted backcrossing (MABC) (Jiang 2013a)

parent W is first crossed to each of four donor parents to produce four single-cross F_1s. Two of the four single-cross F_1s are crossed with each other to produce two double-cross F_1s, and these two double-cross F_1s are crossed again to produce a hybrid which integrates all four target genes/QTLs in heterozygous state. The hybrid and/or progeny with heterozygous markers for all four target genes/QTLs is subsequently crossed back to the RP W until a satisfactory recovery of the RP genome,

and finalized by one generation of selfing (Jiang 2013a). The advantage of this method is that it takes the shortest time to complete. However, in the backcrossing all target genes/QTLs are involved at the same time and thus it requires a large population and more genotyping. Relatively speaking, convergent MABC (Fig. 15.5) is a strategy that combines the advantages of both stepwise and synchronized backcrossing. First the four target gene/QTLs are transferred separately from the donors into the recurrent parent W by single crossing followed by backcrossing based on markers linked to the target genes/QTLs, to produce four improved lines (W^{AA}, W^{BB}, W^{CC}, and W^{DD}). Two of the improved lines are crossed with each other and the two hybrids are then intercrossed to integrate all four genes/QTLs together and develop the final improved line with all four genes/QTLs pyramided (e.g. $W^{AABBCCDD}$). Therefore, convergent MABC is more acceptable because in this scheme not only is time reduced (compared to stepwise transfer) but gene fixation and/or pyramiding are also more easily assured (compared to simultaneous transfer) (Jiang 2013a).

Theoretical issues and efficiency of MABC for gene pyramiding have been investigated through computer simulations (Ribaut et al. 2002; Servin et al. 2004; Ye and Smith 2008). Practical application of MABC to gene pyramiding has been reported in many crops, including rice, wheat, barley, cotton, soybean, common bean and pea, especially for developing durable resistance to stresses in crops (Jiang 2013a). However, there is very limited information available about the release of commercial cultivars resulted from this strategy. Joshi and Nayak (2010) and Xu (2010) reviewed the techniques and practical cases in marker-based gene pyramiding. Somers et al. (2005) implemented a molecular breeding strategy to introduce multiple pest resistance genes into Canadian wheat. They used high throughput SSR genotyping and half-seed analysis to process backcrossing and selection for six FHB resistance QTLs, plus orange blossom wheat midge resistance gene *Sm1* and leaf rust resistance gene *Lr21*. They also used 45–76 SSR markers to perform background selection in backcrossing populations to accelerate the restoration of the RP genetic background. This strategy resulted in 87 % fixation of the elite genetic background at the BC_2F_1 on average and successfully introduced all (up to 4) of the chromosome segments containing FHB, *Sm1* and *Lr21* resistance genes in four separate crosses (Somers et al. 2005). More recently, Lou and Yin (2013) employed convergent MABC aiming at pyramiding of semi-dwarf gene *sd1*, submergence tolerance gene *Sub1A*, blast resistance gene *Pi9* and bacterial blight resistance genes *Xa21* and *Xa27* in KDML 105. Through this strategy, they successfully developed an improved line (T5105) which combines all these desired traits to boost the production of high-quality aromatic rice in tropical regions.

Similar to the simultaneous/synchronized backcrossing scheme, marker-assisted complex or convergent crossing (MACC) can be undertaken to pyramid multiple genes/QTLs. In particular, MACC is a proper option of breeding schemes for gene pyramiding if all the parents are improved cultivars or lines with good comprehensive performance and have different or complementary genes or favorable alleles for the traits of interest (Jiang 2013a). The difference from simultaneous backcrossing is that selfing hybrid and progenies replaces backcrossing hybrid to the recurrent parent. In MACC, the hybrid of convergent crossing is subsequently self-pollinated

and marker-based selection for target traits is performed for several consecutive generations until genetically stable lines with desired marker alleles and traits have been developed. In order to reduce population size and to avoid loss of most important genes/QTLs, different markers may be used and selected in different generations, depending on their relative importance (Jiang 2013a). The markers for the most important genes/QTLs can be detected and selected first in early generations and the markers of less importance later. Once homozygous alleles of the markers for a gene/locus are detected, they may not be necessarily detected again in the subsequent generations. Instead, phenotypic evaluation should be conducted if conditions permit.

15.6.2 Marker-Assisted Recurrent Selection

Using markers to select or pyramid for multiple genes/QTLs is more complex and less proven. Recurrent selection is widely regarded as an effective strategy for the improvement of polygenic traits (Fehr 1991; Hallauer 1992; Jiang et al. 1994). However, the effectiveness and efficiency of selection are not very satisfactory in some cases because phenotypic selection is highly dependent upon environments and genotypic selection takes a longer time (2–3 crop seasons at least for 1 cycle of selection) (Jiang 2013a). Marker-assisted recurrent selection (MARS) is a scheme which allows performing genotypic selection and intercrossing in the same crop season for one cycle of selection (Fig. 15.6). Therefore, MARS could enhance the efficiency of recurrent selection and accelerate the progress of the procedure (Jiang et al. 2007a), particularly helps in integrating multiple favorable genes/QTLs from different sources through recurrent selection based on a multiple-parental population.

For complex traits such as grain yield, biotic and abiotic resistance, MARS has been proposed for *forward breeding* of native genes and pyramiding multiple QTLs

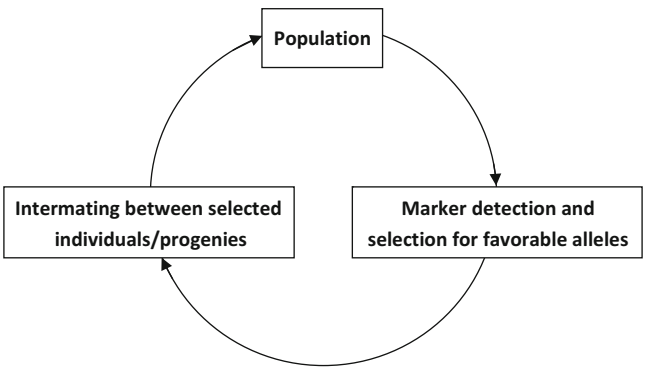

Fig. 15.6 General procedure of marker-assisted recurrent selection (MARS) (Jiang 2013a)

(Crosbie et al. 2006; Eathington 2005; Ragot et al. 2000; Ribaut et al. 2000, 2010). MARS is a recurrent selection scheme using molecular markers for the identification and selection of multiple genomic regions involved in the expression of complex traits to assemble the best-performing genotype within a single or across related populations (Ribaut et al. 2010). Johnson (2004) presented an example to demonstrate the efficiency of MARS for quantitative traits. In their maize MARS programs, a large-scale use of markers in bi-parental populations, first for QTL detection and then for MARS on yield (i.e. rapid cycles of recombination and selection based on associated markers for yield), could allow increased efficiency of long-term selection by increasing the frequency of favorable alleles. Eathington (2005) and Crosbie et al. (2006) also indicated that the genetic gain achieved through MARS in maize was about twice that of phenotypic selection (PS) in some reference populations. In upland cotton, Yi et al. (2004) also demonstrated the effectiveness of MARS for resistance to *Helicoverpa armigera*. Their data indicated that the mean levels of resistance in improved populations after recurrent selection were significantly higher than those of preceding populations.

15.7 Genomic Selection

Genomic selection (GS) or genome-wide selection (GWS) is a form of marker-based selection, referring to the simultaneous selection for many (tens or hundreds of thousands of) markers, which cover the entire genome in a dense manner so that all genes are expected to be in linkage disequilibrium with at least some of the markers (Meuwissen 2007). In GS genotypic data (genetic markers) across the whole genome are used to predict complex traits with accuracy sufficient to allow selection on that prediction alone. Selection of desirable individuals is based on genomic estimated breeding value (GEBV) (Nakaya and Isobe 2012), which is a predicted breeding value calculated using an innovative method based on genome-wide dense DNA markers (Meuwissen et al. 2001). GS does not need significant testing and identifying a subset of markers associated with the trait (Meuwissen et al. 2001), i.e., GS can remove the need to search for significant QTL-marker loci associations individually (Desta and Ortiz 2014). In other words, QTL mapping with populations derived from specific crosses can be avoided in GS. However, it does first need to develop GS models, i.e. the formulae for GEBV prediction (Nakaya and Isobe 2012). In this process (training phase), phenotypes and genome-wide genotypes are investigated in the training population (a subset of a population) to predict significant relationships between phenotypes and genotypes using statistical approaches. Subsequently, GEBVs are used for the selection of desirable individuals in the breeding phase, instead of the genotypes of markers used in traditional MAS (Jiang 2013a). For accuracy of GEBV and GS, genomewide genotype data are necessary and require high marker density in which all quantitative trait loci (QTLs) are in linkage disequilibrium with at least one marker.

The use of high-density markers is one of the fundamental features of GS (Desta and Ortiz 2014). GS is possible only when high-throughput marker technologies, high-performance computing and appropriate new statistical methods are available. This approach has become feasible due to the discovery and development of a large number of single nucleotide polymorphisms (SNPs) by genome sequencing and new methods to efficiently genotype large number of SNP markers (Jiang 2013a). The ideal method to estimate the breeding value from genomic data is to calculate the conditional mean of the breeding value given the genotype at each QTL (Goddard and Hayes 2007). This conditional mean can only be calculated by using a prior distribution of QTL effects, and thus this should be part of the research to implement GS. In practice, this method of estimating breeding values is approximated by using the marker genotypes instead of the QTL genotypes, but the ideal method is likely to be approached more closely as more sequence and SNP data are obtained (Goddard and Hayes 2007). In a recent review, Desta and Oritz (2014) discussed in detail the estimation of GEBV, the accuracy and gain of selection of GS, and other related issues.

Since the application of GS was proposed by Meuwissen et al. (2001) to breeding populations, theoretical, simulation and empirical studies have been conducted, mostly in animals (Goddard and Hayes 2007; Jannink et al. 2010). Relatively speaking, GS in plants was less studied and large-scale empirical studies are not available in public sectors for plant breeding (Desta and Ortiz 2014; Jannink et al. 2010), but it has attracted more and more attention in recent years (Bernardo 2010; Bernardo and Yu 2007; Guo et al. 2011; Heffner et al. 2010, 2011; Lorenzana and Bernardo 2009; Wong and Bernardo 2008; Zhong et al. 2009). Studies indicated that in all cases, accuracies provided by GS were greater than might be achieved on the basis of pedigree information alone (Jannink et al. 2010). In oil palm, for a realistic yet relatively small population, GS was superior to MARS and PS in terms of gain per unit cost and time (Wong and Bernardo 2008). The studies have demonstrated the advantages of GS, suggesting that it would be a potential method for plant breeding and it could be performed with realistic sizes of populations and markers when the populations used are carefully chosen (Nakaya and Isobe 2012). Heffner et al. (2010) assessed relative annual genetic gains from GS, MAS, and PS plus MAS, and their data showed that a GS accuracy of 0.53 could lead to threefold and twofold annual genetic gains relative to MAS and PS in maize and wheat, respectively. Recently, Massman et al. (2013) compared the realized gains from selection with GWS and marker-assisted recurrent selection (MARS). Multi-location trials indicated that gains for the Stover Index and Yield + Stover Index were 14–50 % larger (significant at $P = 0.05$) with GWS than with MARS. Gains in individual traits were mostly non-significant. They concluded that using all available markers for predicting genotypic value could lead to greater gain than using a subset of markers with significant effects (Massman et al. 2013). Experiments with oats suggest that genomic selection is a superior method for selecting a polygenic complex trait like β-glucan concentration (Asoro et al. 2013).

The superiority of GS in terms of genetic gain over PS and/or MAS is usually proven by using multiple cycles per year and intensifying selection gains per unit time and cost, particularly for complex traits with low heritability (Desta and Ortiz 2014). However, the gain of GS per cycle is not necessarily better than other selec-

tion methods, e.g. PS and MAS. Using empirical data from *Arabidopsis*, corn and barley, Lorenzana and Bernardo (2009) suggested that the gain from GS per cycle was half of the gain from PS. We recently performed a GWAS as well as GS and MAS for seed weight in soybean. The results suggested that GS exhibited higher prediction accuracy than MAS either for various cross-validations within the association panel or when unrelated panels were used in validation (Zhang et al. 2015). However, the number of loci/markers involved in MAS is usually much smaller than that needed in GS. MAS might have advantages in lowering genotyping cost for a relatively high prediction accuracy.

GS has been highlighted as a new approach for MAS in recent years and is regarded as a powerful, attractive and valuable tool for plant breeding. Kumpatla et al. (2012) recently presented an overall review on the GS for plant breeding. Desta and Oritz (2014) suggested that with the advent of cutting-edge next-generation sequencing (NGS) and high-throughput phenotyping techniques, GS would revolutionize the applications of plant improvement programs. However, GS has not become a popular methodology in plant breeding, and there may be a long way to go before the extensive use of GS in plant breeding programs (Jiang 2013a). The major reason might be the unavailability of sufficient knowledge of GS for practical use (Nakaya and Isobe 2012). Statistics and simulation discussed in terms of formulae in GS studies are most likely too specific and difficult for plant breeders to understand and to use in practical breeding programs. In addition, GS relies on the degree of genetic similarity between training population and breeding population in the LD between marker and trait loci (Desta and Ortiz 2014), but in practice the breeding populations breeders are working on are considerably different from the training population studied. Therefore, one should not get too excited about GS in plants, in particular it directly applying to breeding programs. The comprehensive nature of population structure, especially in inbreeding or self-pollinated species, is a major barrier to implementing GS in plant breeding (Desta and Ortiz 2014). From a plant breeder's point of view, GS can be practicable for a few breeding populations with a specific purpose, but may be impractical for an entire breeding program dealing with hundreds and thousands of crosses/populations at the same time (Jiang 2013a). Therefore, GS must shift from theory to practice, and its accuracy and cost effectiveness must be evaluated in practical breeding programs to provide convincing empirical evidence and warrant a practicable addition of GS to the plant breeder's toolbox (Heffner et al. 2009). Development of easily understandable formulae for GEBVs and user-friendly software packages for GS analysis will be helpful in facilitating and enhancing the application of GS in plant breeding.

15.8 Conclusions and Prospects

Plant breeding has a long history of development beginning with the domestication of crop species. Modern plant breeding based on the fundamental principles of inheritance has become one of the most important components of agricultural

science and technology (Jiang 2013b). It has features of both science and art. Conventional breeding methodologies have extensively proven successful in development of plant cultivars and germplasm. The semi-dwarf high-yielding cultivars of cereals developed during the Green Revolution and hybrid rice developed in 1970s are the most renowned examples. However, conventional breeding is still dependent to a considerable extent on subjective evaluation and empirical selection. Scientific breeding needs less subjectiveness and more science, i.e. practical and accurate evaluation, and effective and efficient selection (Jiang 2013b). Molecular marker-assisted breeding (MAB) has brought great challenges, opportunities and prospects for conventional breeding.

Along with progress in molecular biotechnology, various types of molecular markers in crop plants were developed during the 1980s and 1990s (Xu 2010). The rapid development of molecular markers (particularly DNA markers) and continuous improvement of molecular assays has led to a significant revolution of breeding technology in plants, i.e. the development and application of molecular marker-assisted breeding (MAB). The extensive use of molecular markers in various fields of plant science, e.g. germplasm evaluation, genetic mapping, map-based gene discovery, characterization of traits and crop improvement, has demonstrated that molecular technology is a powerful and reliable tool in genetic manipulation of agronomically-important traits in crop plants (Jiang 2013a; Xu 2010). Compared with conventional breeding methods, MAB has significant advantages (Jiang 2013b):

- MAB can allow selection for all kinds of traits to be carried out at the seedling stage and thus reduces the time required before the phenotype of an individual plant is identified. For the traits that are expressed at later developmental stages, undesirable genotypes can be quickly eliminated by marker-assisted selection (MAS). This feature is particularly important and useful for some breeding schemes such as backcrossing and recurrent selection, in which crossing with or between selected individuals is required.
- MAB is not affected by environment and thus allows the selection to be performed under any environmental conditions (e.g. greenhouse and off-season nurseries). This is very helpful for improvement of certain traits that are expressed only when favorable environmental conditions present, e.g. disease/pest resistance and stress tolerance. For the low-heritability traits that are easily affected by the environment, MAS based on reliable markers tightly linked to the quantitative trait loci (QTLs) for traits of interest can be more effective and efficient than phenotypic selection (PS).
- MAB using co-dominance markers (e.g. SSR and SNP) can allow effective selection of recessive alleles of desired traits in the heterozygous status. No selfing or test-crossing is needed to detect the traits controlled by recessive alleles, and thus MAB may saves time and accelerate breeding progress.
- For the traits controlled by multiple genes/QTLs, individual genes/QTLs in the same individuals can be identified and selected simultaneously in MAB, and thus MAB is particularly suitable for gene pyramiding. In traditional phenotypic

selection, however, it is problematic to distinguish individual genes/loci because one gene may mask the effect of others.
- Genotypic assays based on molecular markers may be faster, cheaper and more accurate than conventional phenotypic assays, depending on the traits and conditions, and thus MAB may have higher effectiveness and efficiency in terms of time, resources and saved effort.

As such, the research and use of MAB in plants has continued to increase in the public and private sectors, particularly since the 2000s (Babu et al. 2004; Jiang 2013a; Xu 2010). In a sense, MAB represents a new direction of future development in plant breeding. However, as a new strategy and methodology of plant breeding, MAB has not been perfect and it has some defects (Jiang 2013b). Marker-assisted selection (MAS) and/or marker-assisted backcrossing (MABC) have been primarily applied to simply-inherited traits, such as monogenic or oligogenic resistance to diseases/pests, although quantitative traits were also involved (Collard and Mackill 2008; Segmagn et al. 2006b; Wang and Chee 2010). MAB in plants has not achieved the results previously expected in terms of extent and success (e.g. release of commercial cultivars). Collard and Mackill (2008) listed ten reasons for the low impact of MAS and MAB in general. Improvement of economically-important agronomic traits like yield and quality that are complicatedly inherited is still a great challenge for MAB, including the newly developed genome-wide selection (GWS) or genomic selection (GS) (Jiang 2013a; Nakaya and Isobe 2012). From the standpoint of a plant breeder, therefore, MAB is not universally or necessarily advantageous (Jiang 2013a). The application of molecular technologies to plant breeding is still facing the following drawbacks and/or challenges (Jiang 2013a, b):

- Not all markers are breeder-friendly. This problem may be solved by converting non-breeder-friendly markers to other types of breeder-friendly markers (e.g. RFLP to STS, sequence tagged site, and RAPD to SCAR, sequence characterized amplified region).
- Not all markers can be applicable across populations due to lack of marker polymorphism or reliable marker-trait association. Multiple mapping populations are helpful for a better understanding of marker allelic diversity and genetic background effects. In addition, QTL positions and effects also need to be validated and re-estimated by breeders in their own germplasm of interest (Heffner et al. 2009).
- False selection may occur due to recombination between the markers and the genes/QTLs of interest. Use of flanking markers or more markers for the target gene/QTL can help to perform reliable selection.
- Inaccurate estimations of QTL locations and effects result in slower progress than expected. The efficiency of QTL detection is attributed to multiple factors, such as algorithms, mapping methods, number of polymorphic markers, and population type and size (Wang et al. 2012, 2014a, b). Fine mapping with high marker density and in large populations and well-designed phenotyping across multiple environments may provide more accurate estimates of QTL locations and effects.

- The methods and schemes of MAB must be easily understandable, acceptable and implementable for plant breeders. A MAB scheme designed on a computer without practical considerations is unlikely used on a large scale in practical breeding programs.
- A plant breeder often handles hundreds of segregation populations each with hundreds or thousands of individuals. With limited resources, it is very difficult to timely complete tissue sampling, DNA extraction, marker screening and data analysis.
- A large number of breeding programs have not been equipped with adequate facilities and conditions for the large-scale adoption of MAB in practice.
- Higher startup expenses and labor costs are still a concern in many cases.

Therefore, as other new methods of plant breeding like transgenic breeding or genetic manipulation do, MAB cannot replace conventional breeding but is only a supplementary addition to conventional breeding (Jiang 2013a, b). High costs and technical or equipment demands of MAB will continue to be a major obstacle for its large-scale use in the near future, especially in developing countries (Collard and Mackill 2008; Ribaut et al. 2010). Therefore, integration of MAB into conventional breeding programs will be an optimistic strategy for crop improvement in the future (Jiang 2013a, b). It can be expected that the drawbacks of MAB will be gradually overcome, as its theory, technology and application are further developed and improved. This should lead to the wide adoption and use of MAB in practical breeding programs for more crop species and in more countries as well.

References

Acquaah G (2007) Principles of plant genetics and breeding. Blackwell Publishing, USA

Anderson JA, Chao S, Liu S (2007) Molecular breeding using a major QTL for Fusarium head blight resistance in wheat. Crop Sci 47S:112–119

Asoro FG, Newell MA, Beavis WD et al (2013) Genomic, marker-assisted, and pedigree-BLUP selection methods for β-glucan concentration in elite oat. Crop Sci 53:1894–1906

Babu R, Nair SK, Prasanna BM, Gupta HS (2004) Integrating marker-assisted selection in crop breeding – Prospects and challenges. Curr Sci 87:607–619

Babu R, Nair SK, Kumar A et al (2005) Two-generation marker-aided backcrossing for rapid conversion of normal maize lines to quality protein maize (QPM). Theor Appl Genet 111:888–897

Bernacchi D, Beck-Bunn T, Emmatty D et al (1998) Advanced backcross QTL analysis of tomato: II. Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derived from *Lycopersicon hirsutum* and *L. pimpinellifolium*. Theor Appl Genet 97:170–180

Bernardo R (2010) Genomewide selection with minimal crossing in self-pollinated crops. Crop Sci 50:624–627

Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative traits in maize. Crop Sci 47:1082–1090

Bilyeu K, Palavalli L, Sleper DA, Beuselinck P (2006) Molecular genetic resources for development of 1 % linolenic acid soybeans. Crop Sci 46:1913–1918

Bonnett DG, Rebetzke GJ, Spielmeyer W (2005) Strategies for efficient implementation of molecular markers in wheat breeding. Mol Breed 15:75–85
Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331
Bradbury LMT, Fitzgerald TL, Henry RJ et al (2005a) The gene for fragrance in rice. Plant Biotechnol J 3:363–370
Bradbury LMT, Henry RJ, Jin Q et al (2005b) A perfect marker for fragrance genotyping in rice. Mol Breed 16:279–283
Buerstmayr H, Ban T, Anderson JA (2009) QTL mapping and marker-assisted selection for Fusarium head blight resistance in wheat: a review. Plant Breed 128:1–26
Castro AJ, Capettini F, Corey AE et al (2003) Mapping and pyramiding of qualitative and quantitative resistance to stripe rust in barley. Theor Appl Genet 107:922–930
Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc B 363:557–572
Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169–196
Concibido VC, La Vallee B, Mclaird P et al (2003) Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. Theor Appl Genet 106:575–582
Cregan PB, Mudge J, Fickus EW et al (1999) Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by the *rhg1* locus. Theor Appl Genet 99:811–818
Crosbie TM, Eathington SR, Johnson GR et al (2006) Plant breeding: past, present and future. In: Lamkey KR, Lee M (eds) Plant breeding: the Arnel R. Hallauer international symposium. Blackwell Publishing, Oxford, pp 3–50
Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci 19:592–601
Dwivedi SL, Crouch JH, Mackill DJ et al (2007) The molecularization of public sector crop breeding: progress, problems and prospects. Adv Agron 95:163–318
Eathington SR (2005) Practical applications of molecular technology in the development of commercial maize hybrids. In: Proceedings of the 60th Annual Corn and Sorghum Seed Research Conferences, American Seed Trade Association, Washington, DC
Edwards D, Forster JW, Chagne D, Batley J (2007) What is SNPs? In: Oraguzie NC, Rikkerink EHA, Gardiner SE, de Silva HN (eds) Association mapping in plants. Springer, Berlin, pp 41–52
Farooq S, Azam F (2002a) Molecular markers in plant breeding-I: Concepts and characterization. Pak J Biol Sci 5:1135–1140
Farooq S, Azam F (2002b) Molecular markers in plant breeding-II: Some pre-requisites for use. Pak J Biol Sci 5:1141–1147
Fehr WR (1991) Principles of cultivar development vol. 1: Theory and technique. MacMillan Publishing Company, USA
Flint-Garcia SA, Darrah LL, McMullen MD, Hibbard BE (2003) Phenotypic versus marker-assisted selection for stalk strength and second-generation European corn borer resistance in maize. Theor Appl Genet 107:1331–1336
Frisch M, Bohn M, Melchinger AE (1999) Comparison of selection strategies for marker-assisted backcrossing of a gene. Crop Sci 39:1295–1301 (See also Errata, Frisch et al. 1999. Crop Sci 39:1903)
Fulton TM, Bucheli P, Voirol E et al (2002) Quantitative trait loci (QTL) affecting sugars, organic acids and other biochemical properties possibly contributing to flavor, identified in four advanced backcross populations of tomato. Euphytica 127:163–177
Goddard ME, Hayes BJ (2007) Genomic selection. J Anim Breed Genet 124:323–330
Guo Z, Tucker DM, Liu J et al (2011) Evaluation of genomewide selection efficiency in maize nested association mapping populations. Theor Appl Genet 124:261–275

Gupta PK, Roy JK, Prasad M (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr Sci 80:524–535

Hallauer AR (1992) Recurrent selection in maize. Plant Breed Rev 9:115–179

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. Crop Sci 49:1–12

Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. Crop Sci 50:1681–1690

Heffner EL, Jannink J-L, Iwata H et al (2011) Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Sci 51:2597–2606

Hospital F (2003) Marker-assisted breeding. In: Newbury HJ (ed) Plant molecular breeding. Blackwell Publishing and CRC Press, Oxford/Boca Raton, pp 30–59

Hospital F, Decoux G (2002) Popmin: a program for the numerical optimization of population sizes in marker-assisted backcross breeding programs. J Hered 93:383–384

Hospital F, Chevalet C, Mulsant P (1992) Using markers in gene introgression breeding programs. Genetics 231:1199–1210

Huang N, Angeles ER, Domingo J et al (1997) Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theor Appl Genet 95:313–320

Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9:166–177

Jefferies SP, King BJ, Barr AR et al (2003) Marker-assisted backcross introgression of the *Yd2* gene conferring resistance to barley yellow dwarf virus in barley. Plant Breed 122:52–56

Jiang G-L (2013a) Molecular markers and marker-assisted breeding in plants. In: Andersen SB (ed) Plant breeding from laboratories to fields. InTech, Croatia, pp 45–83

Jiang G-L (2013b) Plant marker-assisted breeding and conventional breeding: challenges and perspectives. Adv Crop Sci Tech 1, e106. doi:10.4172/acst.1000e106

Jiang G-L, Wu Z-S, Huang D-C (1994) Effects of recurrent selection for resistance to scab (*Gibberella zeae*) in wheat. Euphytica 72:107–113

Jiang G-L, Dong Y, Shi J, Ward RW (2007a) QTL analysis of resistance to Fusarium head blight in the novel wheat germplasm CJ 9306. II. Resistance to deoxynivalenol accumulation and grain yield loss. Theor Appl Genet 115:1043–1052

Jiang G-L, Shi J, Ward RW (2007b) QTL analysis of resistance to Fusarium head blight in the novel wheat germplasm CJ 9306. I. Resistance to fungal spread. Theor Appl Genet 116:3–13

Johnson R (2004) Marker assisted selection. In: Jannick J (ed) Plant breeding reviews, vol 24, Part 1., pp 293–309

Joshi RK, Nayak S (2010) Gene pyramiding – a broad spectrum technique for developing durable stress resistance in crops. Biotech Mol Biol Rev 5:51–60

Konieczny A, Ausubel F (1993) A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR based markers. Plant J 4:403–410

Kumpatla SP, Buyyarapu R, Abdurakhmonov IY, Mammadov JA (2012) Genomics-assisted plant breeding in the 21st century: technological advances and progress. In: Abdurakhmonov IY (ed) Plant breeding. InTech, Rijeka, Croatia, pp 131–184

Lecomte L, Duffé P, Buret M et al (2004) Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds. Theor Appl Genet 109:568–668

Li X, Han Y, Teng W et al (2010) Pyramided QTL underlying tolerance to Phytophthora root rot in mega-environment from soybean cultivar 'Conrad' and 'Hefeng 25'. Theor Appl Genet 121:651–658

Liu B-H (1991) Development and prospects of dwarf male-sterile wheat. Chin Sci Bull 36(4):306

Lorenzana RF, Bernardo R (2009) Accuracy of genotypic predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120:151–161

Luo Y, Yin Z (2013) Marker-assisted breeding of Thai fragrance rice for semi-dwarf phenotype, submergence tolerance and disease resistance to rice blast and bacterial blight. Mol Breed 32:709–721

Luo Y, Sangha JS, Wang S et al (2012) Marker-assisted breeding of *Xa4*, *Xa21* and *Xa27* in the restorer lines of hybrid rice for broad-spectrum and enhanced disease resistance to bacterial blight. Mol Breed. doi:10.1007/s11032-012-9742-7

Massman JM, Jung H-JG, Bernardo R (2013) Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. Crop Sci 53:58–66

Meuwissen T (2007) Genomic selection: marker assisted selection on a genomewide scale. J Anim Breed Genet 124:321–322

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Miedaner T, Wilde F, Steiner B et al (2006) Stacking quantitative trait loci (QTL) for Fusarium head blight resistance from non-adapted sources in an European elite spring wheat background and assessing their effects on deoxynivalenol (DON) content and disease severity. Theor Appl Genet 112:562–569

Moreau L, Charcosset A, Hospital F, Gallais A (1998) Marker-assisted selection efficiency in populations of finite size. Genetics 148:1353–1365

Mullis K (1990) The unusual origin of the polymerase chain reaction. Sci Am 262(4):56–61, 64–65

Nakaya A, Isobe SN (2012) Will genomic selection be a practical method for plant breeding? Ann Bot. doi:10.1093/aob/mcs109

Newbury HJ (2003) Plant molecular breeding. Blackwell Publishing/CRC Press, Birmingham

Payne PI, Nigtingale MA, Krattiger AF, Holt LM (1987) The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat varieties. J Sci Food Agric 40:51–65

Pham A-T, Shannon JG, Bilyeu KD (2012) Combinations of mutant FAD2 and FAD3 genes to produce high oleic acid and low linolenic acid soybean oil. Theor Appl Genet 125:503–515

Pumphrey MO, Bernardo R, Anderson JA (2007) Validating the *Fhb1* QTL for Fusarium head blight resistance in near-isogenic wheat lines developed from breeding populations. Crop Sci 47:200–206

Ragot M, Biasiolli M, Dekbut MF et al (1995) Marker-assisted backcrossing: a practical example. In: Berville A, Tersac M (eds) Les Colloques, vol 72. INRA, Paris, pp 45–46

Ragot M, Gay G, Muller JP, Durovray J (2000) Efficient selection for the adaptation to the environment through QTL mapping and manipulation in maize. In: Ribaut JM, Poland D (eds) Molecular approaches for the genetic improvement of cereals for stable production in water-limited environments. CIMMYT, Mexico, pp 128–130

Ribaut JM, Betran J (1999) Single large-scale marker-assisted selection (SLS-MAS). Mol Breed 5:531–541

Ribaut JM, Edmeades G, Perotti E, Hoisington D (2000) QTL analysis, MAS results and perspectives for drought-tolerance improvement in tropical maize. In: Ribaut JM, Poland D (eds) Molecular approaches for the genetic improvement of cereals for stable production in water-limited environments. CIMMYT, Mexico, pp 131–136

Ribaut JM, Jiang C, Hoisington D (2002) Simulation experiments on efficiencies of gene introgression by backcrossing. Crop Sci 42:557–565

Ribaut JM, de Vicente MC, Delannay X (2010) Molecular breeding in developing countries: challenges and perspectives. Curr Opin Plant Biol 13:1–6

Richardson KL, Vales MI, Kling JG et al (2006) Pyramiding and dissecting disease resistance QTL to barley stripe rust. Theor Appl Genet 113:485–495

Rogers WJ, Payne PI, Harinder K (1989) The HMW glutenin subunit and gliadin compositions of German-Grown wheat varieties and their relationship with bread-making quality. Plant Breed 103:89–100

Sebolt AM, Shoemaker RC, Diers BW (2000) Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Sci 40:1438–1444

Semagn K, Bjornstad A, Ndjiondjop MN (2006a) An overview of molecular marker methods for plants. Afr J Biotechnol 5:2540–2568
Semagn K, Bjornstad A, Ndjiondjop MN (2006b) Progress and prospects of marker assisted backcrossing as a tool in crop breeding programs. Afr J Biotechnol 5:2588–2603
Servin B, Hospital F (2002) Optimal positioning of markers to control genetic background in marker-assisted backcrossing. J Hered 93:214–217
Servin B, Martin OC, Mezard M, Hospital F (2004) Toward a theory of marker-assisted gene pyramiding. Genetics 168:513–523
Shi A, Chen P, Li D et al (2009) Pyramiding multiple genes for resistance to soybean mosaic virus in soybean using molecular markers. Mol Breed 23:113–124
Singh S, Sidhu JS, Huang N et al (2001) Pyramiding three bacterial blight resistance genes (*xa5*, *xa13* and *Xa21*) using marker-assisted selection into indica rice cultivar PR106. Theor Appl Genet 102:1011–1015
Sobrino B, Briona M, Carracedoa A (2005) SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Sci Int 154:181–194
Somers DJ, Thomas J, DePauw R et al (2005) Assembling complex genotypes to resist *Fusarium* in wheat (*Triticum aestivum* L.). Theor Appl Genet 111:1623–1631
Song Q, Jia G, Zhu Y et al (2010) Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. Crop Sci 50:1950–1960
Song Q, Hyten DL, Jia G et al (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS One 8(1), e54985
Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517
Stuber CW, Polacco M, Senior ML (1999) Synergy of empirical breeding, marker-assisted selection and genomics to increase crop yield potential. Crop Sci 39:1571–1583
Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191–203
Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Bio Technology 7:257–263
Visscher PM, Thompson R, Haley CS (1996) Confidence intervals in QTL mapping by bootstrapping. Genetics 143:1013–1020
Wan B, Zha Z, Li J et al (2014) Development of elite rice restorer lines in the genetic background of R022 possessing tolerance to brown planthopper, stem borer, leaf folder and herbicide through marker-assisted breeding. Euphytica 195:129–142
Wang B, Chee PW (2010) Application of advanced backcross quantitative trait locus (QTL) analysis in crop improvement. J Plant Breed Crop Sci 2:221–232
Wang X, Jiang G-L, Green M et al (2012) Quantitative trait locus analysis of saturated fatty acids in a population of recombinant inbred lines of soybean. Mol Breed 30:1163–1179
Wang X, Jiang G-L, Green M et al (2014a) Quantitative trait locus analysis of unsaturated fatty acids in a recombinant inbred population of soybean. Mol Breed 33:281–296
Wang X, Jiang G-L, Green M et al (2014b) Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. Mol Genet Genomics 289:935–949. doi:10.1007/s00438-014-0865-x
Wang X, Jiang G-L, Song Q et al (2015) Quantitative trait locus analysis of seed sulfur-containing amino acids in two recombinant inbred line populations of soybean. Euphytica 201:293–305. doi:10.1007/s10681-014-1223-0
Wong CK, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theor Appl Genet 116:815–824
Xu Y (2010) Molecular plant breeding. CAB International, Wallingford, UK/Cambridge, MA
Ye G, Smith KF (2008) Marker-assisted gene pyramiding for inbred line development: basic principles and practical guidelines. Int J Plant Breed 2:1–10

Yi C, Guo W, Zhu X et al (2004) Pyramiding breeding by marker-assisted recurrent selection in upland cotton II. Selection effects on resistance to *Helicoverpa armigera*. Sci Agric Sin 37:801–807

Young ND (1999) A cautiously optimistic vision for marker-assisted breeding. Mol Breed 5:505–510

Young ND, Tanksley SD (1989) RFLP analysis of the size of chromosomal segments retained around the *Tm-2* locus of tomato during backcross breeding. Theor Appl Genet 77:353–359

Zhang J, Song Q, Cregan PB, Jiang G-L (2015) Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (Glycine max). Theor Appl Genet. doi:10.1007/s00122-015-2614-x

Zhao X, Tan G, Xing Y et al (2012) Marker-assisted introgression of *qHSR1* to improve maize resistance to head smut. Mol Breed. doi:10.1007/s11032-011-9694-3

Zhong S, Dekkers JCM, Fernando RL, Jannink J-L (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182:355–364

Chapter 16
Mutant-Based Reverse Genetics for Functional Genomics of Non-model Crops

Toyoaki Anai

Abstract In the past decade, innovations in high-throughput sequencing (next-generation sequencing) technologies have accelerated whole-genome sequencing of various non-model crop species. Taking advantage of huge polymorphic sequence data provided by the results of whole-genome sequencing, we can easily develop novel molecular markers. It may boost the use of forward genetics approach to isolating the corresponding genes for QTLs in non-model crops. Furthermore, this forward genetics approach is a steady and robust method but it is still difficult to increase its throughput. The sequenced genes have been annotated on the basis of sequence similarity; however, the functions of most genes (and the resulting phenotypes) are still obscure. Although we can easily obtain multiple crop genomic sequences from public databases, it is necessary to increase the throughput of functional genomics. Reverse genetics, which uses mutants or transgenic lines for the genes of interest, is an attractive approach to determine gene function. Mutant-based reverse genetics has several advantages over transgene-based reverse genetics: (a) its higher throughput, (b) the absence of restrictions for growing non-transgenic mutants in the field and (c) the possibility to use the mutants directly for traditional cross-breeding programs as valuable genetic resources of non-model crops. This chapter describes recent advances in functional genomics research on non-model crops, with a focus on mutant-based reverse genetics approaches.

Keywords HRM • Mutants • Reverse genetics • TILLING • Whole-genome sequencing

T. Anai (✉)
Laboratory of Plant Genetics and Breeding, Faculty of Agriculture, Saga University, Honjyo-machi 1, Saga 840-8502, Japan
e-mail: anai@cc.saga-u.ac.jp

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_16

16.1 Introduction

A sustainable food supply is required to support the increasing world population. Crop improvement requires the elucidation and use of the diversity of functions of specific genes in vivo. Over the past decade, continuous improvement of next-generation sequencing (NGS) technologies (Bolger et al. 2014) has provided unprecedented opportunities for high-throughput whole-genome sequencing of non-model crops. The field of functional genomics originated from predicting the functions of genes by using sequence similarities and gene expression profiles. However, precise gene functions cannot always be determined by using these approaches alone.

An established research strategy for functional genomics of economically-important crop species has been eagerly awaited by plant breeding researchers for a long time. Reverse genetics is an attractive strategy to determine the functions of particular genes by studying phenotypes resulting from alterations in these genes. This approach has been mainly developed using transgenic technology in the model plant *Arabidopsis thaliana*, the genome sequence of which was completed in 2000 (AGI Initiative 2000); this strategy was then extended to several crop species (Bolger et al. 2014). T-DNA tagging and transposon tagging are well-known tools of transgene-based reverse genetics of model plant species. The establishment of large insertion-mutant collections has supported reverse genetics in *Arabidopsis* (Alonso et al. 2003) and rice (Krishnan et al. 2009). RNAi-based technology has been also used for target-specific gene silencing (Waterhouse and Helliwell 2003). Even though transgenic technology is attractive, several factors (e.g. plant regeneration rate and growing space) limit its application in many economically-important crop plants that have a genome larger in size than *Arabidopsis*; thus, technologies not based on transgenes are more suitable for non-model crops. The rice Tos17-mediated system, which employed an endogenous retrotransposon (Tos17) to develop insertion mutants, was an attractive alternative (Miyao et al. 2003). However, active endogenous transposons are not available in most crop species. Traditional chemical or physical mutagenesis combined with high-throughput mutation screening is a promising alternative for reverse genetics of various crop species. This chapter describes recent progress in functional genomics and its application in non-model crops, especially the development of mutant-based reverse genetics tools.

16.2 Advances in High-Throughput Sequencing Technologies and Whole-Genome Sequencing

Whole-genome sequencing was first completed for the model plant *Arabidopsis* (AGI Initiative 2000), and was subsequently reported for several crops (Table 16.1).

Table 16.1 Sequenced crop genomes

Year	Crop species	Genome size	Sequencing method[a]	References
2002	*Oryza sativa*	389 Mb	WGS (dideoxy) BAC by BAC (dideoxy)	Goff et al. (2002) Yu et al. (2002)
2006	*Populus trichocarpa*	485 Mb	WGS (dideoxy)	Tuskan et al. (2006)
2007	*Vitis vinifera*	487 Mb	WGS (dideoxy)	Jaillon et al. (2007) Velasco et al. (2007)
2008	*Carica papaya*	372 Mb	WGS (dideoxy)	Ming et al. (2008)
2009	*Sorghum bicolor*	730 Mb	WGS (dideoxy)	Paterson et al. (2009)
	Zea mays	2.3 Gb	BAC by BAC (dideoxy)	Schnable et al. (2009)
	Cucumis sativus	350 Mb	WGS (hybrid)	Huang et al. (2009)
2010	*Glycine max*	1.1 Gb	WGS (dideoxy)	Schmutz et al. (2010)
	Malus × domestica	742 Mb	WGS (hybrid)	Velasco et al. (2010)
2011	*Brassica rapa*	284 Mb	WGS (dideoxy)	Wang et al. (2011)
	Theobroma cacao	430 Mb	WGS (hybrid)	Argout et al. (2011)
	Fragaria vesca	240 Mb	WGS (NGS)	Shulaev et al. (2011)
	Solanum tuberosum	727 Mb	WGS (hybrid)	Potato Genome Consortium (2011)
2012	*Hordeum vulgare*	5.1 Gb	BAC by BAC (dideoxy) WGS (hybrid)	Mayer et al. (2011) IBGS Consortium (2012)
	Linum usitatissimum	739 Mb	WGS (hybrid)	Tomato Genome Consortium (2012)
	Musa acuminate	523 Mb	WGS (hybrid)	D'Hont et al. (2012)
	Cucumis melo	375 Mb	WGS (NGS)	Garcia-Mas et al. (2012)
	Citrullus lanatus	353 Mb	WGS (NGS)	Guo et al. (2013)
2013	*Cicer arietinum*	738 Mb	WGS (NGS)	Varshney et al. (2013)
	Citrus sinensis	367 Mb	WGS (NGS)	Xu et al. (2013)
	Prunus persica	265 Mb	WGS (dideoxy)	International Peach Genome Initiative (2013)

[a]BAC by BAC, using BAC contigs for sequencing
WGS, using sheared whole genome for sequencing and assembling
Hybrid, a combination of dideoxy and NGS methods

Initially, whole-genome sequencing projects were time- and labor-consuming, because sequencing reactions had to be carried out with overlapping BAC contigs using dideoxy method-based capillary array sequencers. This approach was used, for example, for rice, maize, and barley (Table 16.1). However, recently the strategy has drastically shifted towards whole-genome shotgun sequencing (WGS) using NGS. This approach was used, for example, for sorghum and soybean. Because NGS does not require contig construction, it has a much higher throughput and is less time- and labor-consuming than the dideoxy method-based sequencing.

The first NGS platform, 454, was based on parallel pyrosequencing and had relatively low data yield and short reads. More recent Illumina and SOLiD platforms use different sequencing chemistries. The Illumina platform recently became the predominant NGS platform for whole-genome sequencing (Bolger et al. 2014). At an early stage, the performance of NGS was not sufficient for whole-genome sequencing of crop genomes and the hybrid method, a combination of NGS and dideoxy method-based sequencing, was adopted for several crop species (e.g. apple and potato). Recently, due to improvements in platforms and assembly algorithms, WGS of several crops (e.g. chickpea, orange and wild strawberry) was completed by using NGS alone, but WGS of highly-complex crop genomes (e.g. wheat) needs further tool improvement. A possible solution would be to extend the length of each read without sacrificing total sequence throughput. In this respect, the most attractive candidate at present is the Pacific Biosciences (PacBio) platform, a single-molecule DNA sequencer.

The first crop genome, rice, completed by using WGS (Goff et al. 2002; Yu et al. 2002), became a reference for genomic research on cereal crops (Jung et al. 2008). This success accelerated whole-genome sequencing of other crops, and nucleotide sequence data with molecular marker locations and homology-based annotations became available from public databases. These databases are useful for map-based cloning and marker-assisted selection (MAS) in conventional crop breeding programs. Several genes for important agricultural traits have been successfully identified by using whole-genome sequencing data (Ashikari et al. 2005; Powell et al. 2012; Riedelsheimer et al. 2012). However, the functions of individual genes still remained unclear because of gene redundancy in many crops. Thus, development of efficient and robust functional genomics tools for non-model crops was needed.

16.3 Strategies for Functional Genomics Studies on Non-model Crops

Functional genomics uses two different strategies, forward genetics (phenotype to gene) and reverse genetics (gene to phenotype) (Fig. 16.1). A traditional functional study of an individual gene starts from a specific phenotype, and then the gene responsible for this phenotype is identified, usually by linkage analysis. During domestication, many crops evolved highly complex genomes containing functionally redundant genes. Quantitative trait locus (QTL) mapping is the most effective method to discover more complex genotype-phenotype relationships at a whole-genome scale. Whole-genome sequencing can provide many polymorphic markers (e.g. SSRs, SNPs and IN-DELs) to assist high-resolution QTL mapping (Varshney et al. 2014). Over the past decade, many genes for important agronomic traits have been identified in and cloned from various crops, but QTL mapping may be difficult for functionally redundant genes; genes with weaker phenotypes are also difficult to analyze efficiently using the forward genetics approach.

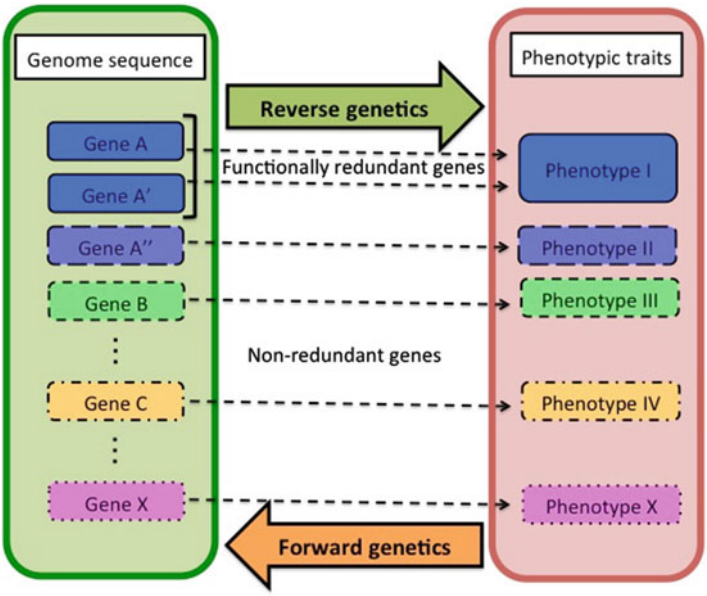

Fig. 16.1 Two different strategies for functional genomics

In contrast to forward genetics, in reverse genetics individual gene functions are elucidated by the sequence of each target gene. Until recently, the reverse genetics approach was not applicable to non-model crop species. Whereas *Arabidopsis* T-DNA and Ac/Ds insertion lines are available from the Arabidopsis Biological Resource Center (ABRC) and rice T-DNA and Tos17 insertion lines are also distributed by several institutes, there are only a few comparable mutant resources for crop species other than rice. Several reported transgenic approaches for reverse genetics (Table 16.2) are very attractive, but they cannot be easily applied to crop species cultivated in the field. Because non-transgenic mutant populations can be easily scaled up, the mutant-based reverse genetics approach can be applied for functional genomics studies on a whole-genome scale.

16.4 Mutant Resources for Functional Genomics of Non-model Crops via the Reverse Genetics Approach

T-DNA tagging and RNAi-mediated gene silencing are very popular for producing *Arabidopsis* gene-specific mutants, but these methods cannot be easily adopted for economically-important crops because of the large plant genome size and low transformation efficiency. Furthermore, consumers do not fully accept GM crops as food. These challenges created a great demand to develop reverse genetics methodologies for functional genomics studies on non-model crops without using

Table 16.2 Transgene-based methods of reverse genetics

Method	Mutation specificity	Mutation stability	Allelic variations	Screening throughput	References
Overexpression	Uncontrollable	Unstable	Uncontrollable	Depends on the host	Ichikawa et al. (2006)
Gene silencing	Uncontrollable	Unstable	Uncontrollable	Depends on the host	Schwab et al. (2006)
T-DNA tagging	Random	Stable	Limited	Depends on the host	Sessions et al. (2002)
Transposon tagging[a]	Random	Stable	Limited	Depends on the host	Tissier et al. (1999)
Homologous recombination	Specific	Stable	Possible	Very low	Shaked et al. (2005)
Zinc-finger nuclease	Specific	Stable	Possible	Low	Lloyd et al. (2005)
TALENs	Specific	Stable	Possible	Low	Li et al. (2012)
CRISPR/Cas	Specific	Stable	Possible	Low	Cong et al. (2013)

TALENs, sequence specific artificial nuclease mediated gene editing system
CRISPR/Cas, type II prokaryotic CRISPR (clustered regularly interspaced short palindromic repeats)/Cas mediated RNA-guided site-specific gene editing system
[a]Except for the rice *Tos17* system, which uses endogenous retrotransposons (Miyao et al. 2003)

transgenic plants. The combination of mutagenesis and a high-throughput mutation detection system is one of the most attractive alternatives to satisfy this demand. Efficient chemical mutagens (e.g. alkylation reagents and crosslinking reagents) and physical mutagens (e.g. X-rays, gamma rays and fast neutrons) have been already established for various crop species (Maluszynski et al. 1995). The mutation type (which mainly depends on the mutagen type) and mutation frequency (which also depends on the mutagen dose and treatment period) are important factors in developing mutant populations (Table 16.3).

In general, radiation and crosslinking reagents such as diepoxybutane (DEB) and psoralen induce deletions of various sizes (Wu et al. 2005), whereas alkylation reagents such as ethyl methane sulfonate (EMS) (Greene et al. 2003) and *N-methyl-N*-nitrosourea (MNU) (Satoh et al. 2010) induce nucleotide substitutions. Large deletions may eliminate multiple-functional loci, whereas small deletions frequently induce frame-shifts, protein truncations, or mis-splicing and lead to loss of function. Nucleotide substitutions induce nonsense, missense, and mis-splicing mutations, which may also result in loss of function. The mutation rate of deletions was relatively lower than that of nucleotide substitutions (Anai 2012). Nucleotide substitutions need smaller population sizes and their handling is easier than that of deletions; therefore, nucleotide-substitution mutant populations developed with alkylation reagents are predominantly used for reverse genetics analysis.

Table 16.3 Popular mutagens and their properties

Mutagen	Mode of action	Mutation type	Mutation frequency	Handling	References
X-rays	Induction of DNA strand breaks	Deletion	Low ~0.1/Mb	Special equipment needed	Anai (2012)
Gamma rays	Induction of DNA strand breaks	Deletion	Low ~0.2/Mb	Special equipment needed	Sato et al. (2006)
Fast neutrons	Induction of DNA strand breaks	Deletion	Low ~0.045/Mb	Special equipment needed	Li et al. (2001)
Ion beam	Induction of DNA strand breaks	Deletion	Moderate 0.03–0.08% (Phenotipic)	Special equipment needed	Kazama et al. (2011)
EMS	Alkylation of nucleotide base	Base substitution	High ~41.7/Mb	Easy	Slade et al. (2005)
MNU	Alkylation of nucleotide base	Base substitution	High ~7.4/Mb	Easy	Suzuki et al. (2008)
DEB	DNA crosslinking	Deletion, base substitution	Moderate 0.07–2.79 % (Phenotypic)	Easy	Wu et al. (2005)
Psoralen	DNA crosslinking	Deletion	Moderate ~0.028% (Phenotipic)	Special equipment needed	Gengyo-Ando and Mitani (2000)
NaN_3	Unclear	Base substitution	High ~2.6/Mb	Easy	Talamè et al. (2008)

16.5 High-Throughput Mutant Screening Systems for Non-model Crops

High-throughput mutation screening optimized for each mutant population is a key for the advancement of functional genomics in non-model crops. This section describes typical high-throughput screening methods for different types of mutant resources (Table 16.4).

Two potent screening methods, Deletagene (Fig. 16.2) and comparative genomic hybridization (CGH) (Fig. 16.3), are predominant choices for obtaining mutants with large deletions. Deletagene, the most simple and cost-effective method, uses PCR on genomic DNA *mega pools* (~2500 lines/pool) derived from a fast neutron–treated mutant population (Li et al. 2001). In this method, detection of shorter amplicons indicates deletion of the target sequence.

The CGH method is a traditional dot-blot-like method, which uses a high-density microarray with many probes to identify deletions in mutants generated by irradiation with fast neutrons and X-ray (Bolon et al. 2011); at present, this method is

Table 16.4 List of high-throughput mutant screening systems

Screening method	Mutation type	Cost and skills required	References
Deletagene	Large deletion (>10 kbp)	Low cost, easy to handle	Li et al. (2001)
CGH	Large deletion (>100 kbp)	Expensive, easy to handle	Bolon et al. (2011)
TILLING	Base substitution, IN-DELs	Low cost[a], easy to handle	McCallum et al. (2000) Colbert et al. (2001) Hoshino et al. (2010)
HRM	Base substitution, IN-DELs	Moderate cost, easy to handle	Botticella et al. (2010)
KeyPoint	All types	Expensive, special techniques required	Rigola et al. (2009)

[a]The cost of TILLING was calculated for a modified method (Hoshino et al. 2010)

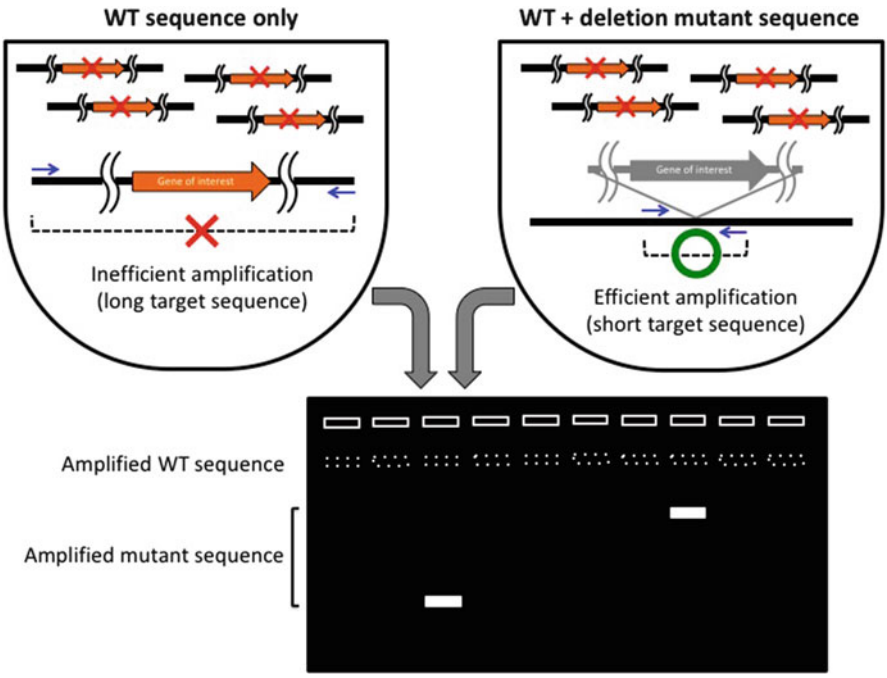

Fig. 16.2 Principals of Deletagene method

difficult to apply to screen for mutations in specific nucleotide sequences on a whole-genome scale because of the high cost of high-density microarrays.

Targeting-Induced Local Lesions IN Genomes (TILLING) (Fig. 16.4) and high-resolution melting (HRM) (Fig. 16.5) are currently the most attractive methods to screen for novel mutant alleles in nucleotide-substitution mutant populations.

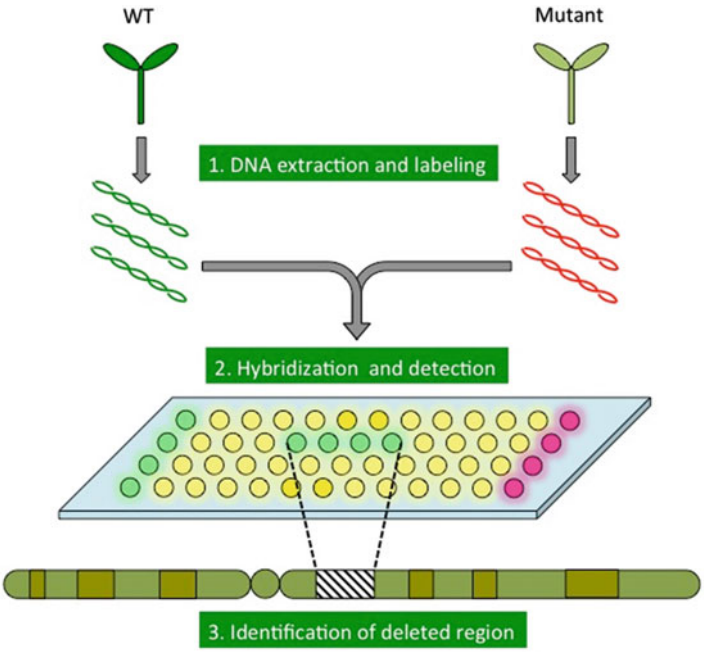

Fig. 16.3 Principals of comparative genomic hybridization (CGH) method

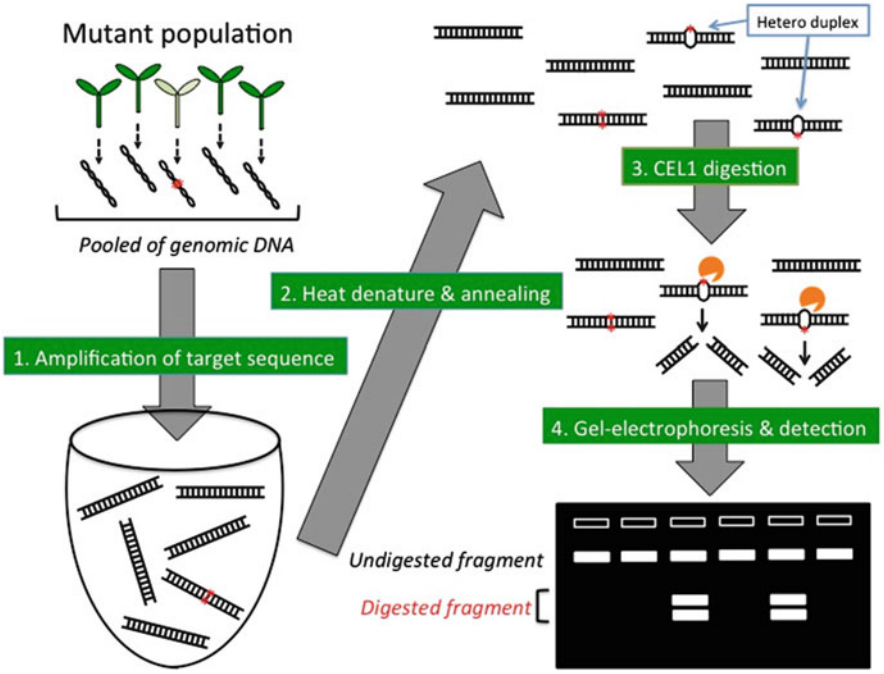

Fig. 16.4 Principals of Targeting-Induced Local Lesions IN Genomes (TILLING) method

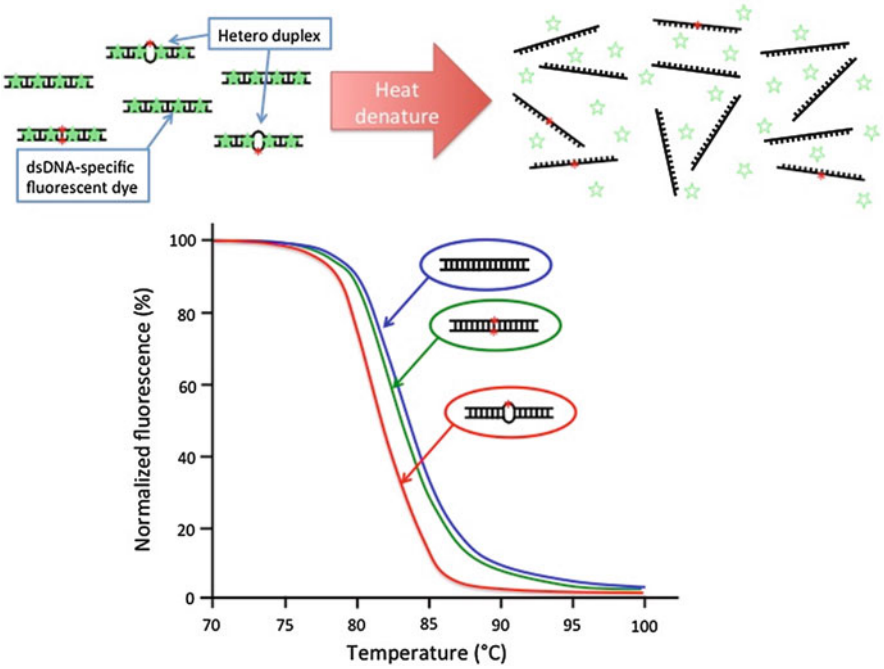

Fig. 16.5 Principals of high resolution melting (HRM) method

TILLING was originally developed as a method to detect mismatches in heteroduplex DNA molecules with high-performance liquid chromatography under denaturing conditions (McCallum et al. 2000). Subsequently, a mismatch-specific nuclease (CEL I) and the LI-COR gel imaging system with fluorescent dye-labeled primers were introduced to improve its throughput (Colbert et al. 2001). To achieve higher throughput at a lower cost, we routinely use a combination of non-labeled primers, high-sensitivity GelRed dye, and agarose gel electrophoresis (Hoshino et al. 2010; Watanabe et al. 2009, 2011; Xia et al. 2012). Many mutants have been isolated from various crop species by using TILLING (Table 16.5).

HRM compares the melting points of samples during a temperature shift by using a real-time fluorescent monitoring system with a double-stranded DNA-specific fluorescent dye (Botticella et al. 2010). This method is attractive because automation can easily improve its throughput, but the target region that can be detected in one reaction is shorter than that of TILLING. The mutation recognition spectrum may slightly differ between TILLING and HRM (our unpublished data), hence both methods are important for developing multiple mutant alleles of the target sequence in functional genomics. Furthermore, TILLING and HRM are frequently used for EMS- treated mutant populations, but it is also able to detect short nucleotide deletions in ionizing radiation-treated populations.

Table 16.5 List of major crops for which mutant alleles have been isolated using TILLING

Crop species	Target gene	Mutagen	References
Arachis hypogaea	Arah1.01, Arah1.02, Arah2.01, Arah2.02, AhFAD2A, AhFAD2B	EMS	Knoll et al. (2011)
Brassica napus	FAE1	EMS	Wang et al. (2008)
Brassica rapa	BraA.RPL.a, BraA.RPL.b, BraA.RPL.c, BraA.IND.a, BraA.MET1.a, BraA.MET1.b	EMS	Stephenson et al. (2010)
Cucumis sativus	CsACS1, CsACS2, CsWIP1, CsRMS4, CsRMS3	EMS	Boualem et al. (2014)
Glycine max	gmclav, gmclavb, gmnark, gmppck4, gmrhg1b, gmrhg4b, gmsacpd2, FT3	EMS X-ray	Cooper et al. (2008) Watanabe et al. (2009)
Helianthus annuus	FatA, SAD	EMS	Kumar et al. (2013)
Hordeum vulgare	HvCO1, Rpg1, eIF4E, NR	NaN_3	Talamè et al. (2008)
Linum usitatissimum	CAD, C3H	EMS	Chantreau et al. (2013)
Oryza sativa	Os1433, OsBZIP, OsCALS8R, OsDREB, OsEXTE, OsMAPK, OsPITA, OsR1A, OsRPLD1, OsTPS1, OsAHP1, OsSAD1, PLA1	EMS NMU	Till et al. (2007) Suzuki et al. (2008)
Solanum lycopersicum	Rab11a, PG, Exp1, RIN, Lcy-b, Lcy-e	EMS	Minoia et al. (2010)
Sorghum bicolor	ACO1, COMT, MIK1, PHYA	EMS	Xin et al. (2008)
Triticum durum	Waxy	EMS	Slade et al. (2005)
Triticum aestivum	SEBIIa, SEBIIb, WKS1, WKS2	EMS	Uauy et al. (2009)
Zea mays	DMT101, DMT102, DMT103, DMT106, HAC110, HDA105	EMS	Till et al. (2004)

16.6 Conclusions and Prospects

The processes of functional genomics have been developed in parallel with innovations in NGS and reverse genetics techniques. The reverse genetics techniques were first developed for model plants; yet, non-model crops have recently become the main target of functional genomics. This is an interesting and important development, which suggests that the distinction between model plants and non-model crops has become less clear because of technical innovations. In the near future, this trend is likely to further expand.

The most important aim of functional genomics in non-model crops is to develop genetic resources carrying novel functional alleles that can be directly used in conventional breeding programs. In this respect, the non-transgenic mutant-based reverse genetics approach is highly desirable. Because the NGS throughput continues to increase, NGS-based mutation screening systems like KeyPoint technology

(Rigola et al. 2009) will become more popular in the near future. KeyPoint technology employs many unique sequence-tags and multi-dimensional pooled library to identify mutation in target sequence. Targeting-mutagenesis methods (e.g. zinc finger nucleases, TALENs, CRISPR/Cas) are now categorized in transgenic technologies (Cong et al. 2013; Li et al. 2012; Lloyd et al. 2005), but they may become feasible alternatives of TILLING and HRM, because their products contain no foreign sequences. However, these technologies need to become more efficient to enable their use in non-model crops. Functional genomics of non-model crops will allow not only basic exploration of gene functions but also the development of novel genetic resources containing innovative functional alleles, leading to the next step in crop improvement.

References

AGI Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Alonso JM, Stepanova AN, Leisse TJ et al (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301:653–657

Anai T (2012) Potential of a mutant-based reverse genetic approach for functional genomics and molecular breeding in soybean. Breed Sci 61:462–467

Argout X, Salse J, Aury JM et al (2011) The genome of *Theobroma cacao*. Nat Genet 43:101–108

Ashikari M, Sakakibara H, Lin S et al (2005) Cytokinin oxidase regulates rice grain production. Science 309:741–745

Bolger ME, Weisshaar B, Scholz U et al (2014) Plant genome sequencing – applications for crop improvement. Curr Opin Biotechnol 26:31–37

Bolon Y-T, Haun WJ, Xu WW et al (2011) Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. Plant Physiol 156:240–253

Botticella E, Sestili F, Hernandez-Lopez A et al (2010) High resolution melting analysis for the detection of EMS induced mutations in wheat *SbeIIa* genes. BMC Plant Biol 11:156

Boualem A, Fleurier S, Troadec C et al (2014) Development of a *Cucumis sativus* TILLinG platform for forward and reverse genetics. PLoS One 9:e97963

Chantreau M, Grec S, Gutierrez L et al (2013) PT-Flax (phenotyping and TILLinG of flax): development of a flax (*Linum usitatissimum* L.) mutant population and TILLinG platform for forward and reverse genetics. BMC Plant Biol 13:159

Colbert T, Till BJ, Tompa R et al (2001) High-throughput screening for induced point mutations. Plant Physiol 126:480–484

Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/cas systems. Science 339:819–823

Cooper JL, Till BJ, Laport RG et al (2008) TILLING to detect induced mutations in soybean. BMC Plant Biol 8:9

D'Hont A, Denoeud F, Aury JM et al (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature 488:213–217

Garcia-Mas J, Benjak A, Sanseverino W et al (2012) The genome of melon (*Cucumis melo* L.). Proc Natl Acad Sci U S A 109:11872–11877

Gengyo-Ando K, Mitani S (2000) Characterization of mutations induced by ethyl methanesulfonate, UV, and trimethylpsoralen in the nematode *Caenorhabditis elegans*. Biochem Biophys Res Comm 269:64–69

Goff SA, Ricke D, Lan TH, Presting G et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100
Greene EA, Codomo CA, Taylor NE et al (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. Genetics 164:731–740
Guo S, Zhang J, Sun H et al (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. Nat Genet 45:51–58
Hoshino T, Takagi Y, Anai T (2010) Novel *GmFAD2-1b* mutant alleles created by reverse genetics induce marked elevation of oleic acid content in soybean seed in combination with *GmFAD2-1a* mutant alleles. Breed Sci 60:419–425
Huang S, Li R, Zhang Z et al (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet 41:1275–1281
IBGS Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716
Ichikawa T, Nakazawa M, Kawashima M et al (2006) The FOX hunting system: an alternative gain-of-function gene hunting technique. Plant J 48:974–985
International Peach Genome Initiative (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet 45:487–494
Jaillon O, Aury JM, Noel B et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467
Jung K-H, An G, Ronald PC (2008) Towards a better bowl of rice: assigning function to tens of thousands of rice genes. Nat Rev Genet 9:91–101
Kazama Y, Hirano T, Saito H, Liu Y et al (2011) Characterization of highly efficient heavy-ion mutagenesis in *Arabidopsis thaliana*. BMC Plant Biol 11:161
Knoll JE, Ramos ML, Zeng Y et al (2011) TILLING for allergen reduction and improvement of quality traits in peanut (*Arachis hypogaea* L.). BMC Plant Biol 11:81
Krishnan A, Guiderdoni E, An G et al (2009) Mutant resources in rice for functional genomics of the grasses. Plant Physiol 149:165–170
Kumar APK, Boualem A, Bhattacharya A et al (2013) SMART – Sunflower mutant population and reverse genetic tool for crop improvement. BMC Plant Biol 13:38
Li X, Song Y, Century K et al (2001) A fast neutron deletion mutagenesis-based reverse genetics system for plants. Plant J 27:235–242
Li T, Liu B, Spalding MH et al (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat Biotechnol 30:390–392
Lloyd A, Plaisier CL, Carroll D, Drews GN (2005) Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis*. Proc Natl Acad Sci U S A 102:2232–2237
Maluszynski M, Ahloowalia BS, Sigurbjörnsson B (1995) Application of *in vivo* and *in vitro* mutation techniques for crop improvement. Euphytica 85:303–315
Mayer KFX, Martis M, Hedley PE et al (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263
McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeted screening for induced mutations. Nat Biotechnol 18:455–457
Ming R, Hou S, Feng Y, Yu Q (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). Nature 452:991–996
Minoia S, Petrozza A, D'Onofrio O et al (2010) A new mutant genetic resource for tomato crop improvement by TILLING technology. BMC Res Notes 3:69
Miyao A, Tanaka K, Murata K et al (2003) Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. Plant Cell 15:1771–1780
Paterson AH, Bowers JE, Bruggmann R et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556
Potato Genome Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195

Powell ALT, Nguyen CV, Hill T et al (2012) Uniform ripening encodes a golden 2-like transcription factor regulating tomato fruit chloroplast development. Science 336:1711–1715
Riedelsheimer C, Czedik-Eysenberg A, Grieder C et al (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44:217–220
Rigola D, van Oeveren J, Janssen A et al (2009) High-throughput detection of induced mutations and natural variation using KeyPoint™ technology. PLoS One 4:e4761
Sato Y, Shirasawa K, Takashashi Y et al. (2006) Mutant selection from progeny of gamma-ray-irradiated rice by DNA heteroduplex cleavage using Brassica petiole extract. Breed Sci 56:179–183.
Satoh H, Matsusaka H, Kumamaru T (2010) Use of *N*-methyl-*N*-nitrosourea treatment of fertilized egg cells for saturation mutagenesis of rice. Breed Sci 60:475–485
Schmutz J, Cannon SB, Schlueter J et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183
Schnable PS, Ware D, Fulton RS et al (2009) The b73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115
Schwab R, Ossowski S, Riester M et al (2006) Highly specific gene silencing by artificial microRNAs in *Arabidopsis*. Plant Cell 18:1121–1133
Sessions A, Burke E, Presting G et al (2002) A high-throughput *Arabidopsis* reverse genetics system. Plant Cell 14:2985–2994
Shaked H, Melamed-Bessudo C, Levy AA (2005) High-frequency gene targeting in *Arabidopsis* plants expressing the yeast *RAD54* gene. Proc Natl Acad Sci U S A 102:2232–2237
Shulaev V, Sargent DJ, Crowhurst RN et al (2011) The genome of woodland strawberry (*Fragaria vesca*). Nat Genet 43:109–116
Slade AJ, Fuerstenberg SI, Loeffler D et al (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat Biotechnol 23:75–81
Stephenson P, Baker D, Girin T et al (2010) A rich TILLING resource for studying gene function in *Brassica rapa*. BMC Plant Biol 10:62
Suzuki T, Eiguchi M, Kumamaru T, Satoh H (2008) MNU-induced mutant pools and high performance TILLING enable finding of any gene mutation in rice. Mol Genet Genomics 279:213–223
Talamè V, Bovina R, Sanguineti MC et al (2008) TILLMore, a resource for the discovery of chemically induced mutants in barley. Plant Biotechnol J 6:477–485
Till BJ, Reynolds SH, Weil C et al (2004) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4:12
Till BJ, Cooper J, Tai TH, Colowit P (2007) Discovery of chemically induced mutations in rice by TILLING. BMC Plant Biol 7:19
Tissier AF, Marillonnet S, Klimyuk V et al (1999) Multiple independent defective suppressor-mutator transposon insertions in *Arabidopsis*: a tool for functional genomics. Plant Cell 11:1841–1852
Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641
Tuskan GA, Difazio S, Jansson S et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604
Uauy C, Paraiso F, Colasuonno P et al (2009) A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. BMC Plant Biol 9:115
Varshney RK, Song C, Saxena RK et al (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nat Biotechnol 31:240–246
Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoS Biol 12:e1001883
Velasco R, Zharkikh A, Troggio M et al (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS One 2:e1326
Velasco R, Zharkikh A, Affourtit J et al (2010) The genome of the domesticated apple (*Malus*×*domestica* Borkh.). Nat Genet 42:833–839

Wang N, Wang Y, Tian F et al (2008) Functional genomics resource for *Brassica napus*: development of an EMS mutagenized population and discovery of *FAE1* point mutations by TILLING. New Phytol 180:751–765

Wang X, Wang H, Wang J et al (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet 43:1035–1039

Watanabe S, Hideshima R, Xia Z et al (2009) Map-based cloning of the gene associated with the soybean maturity locus E3. Genetics 182:1251–1262

Watanabe S, Xia Z, Hideshima R et al (2011) A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering. Genetics 188:395–407

Waterhouse PM, Helliwell CA (2003) Exploring plant genomes by RNA-induced gene silencing. Nat Rev Genet 4:29–38

Wu J-L, Wu C, Lei C et al (2005) Chemical- and irradiation-induced mutants of Indica rice IR64 for forward and reverse genetics. Plant Mol Biol 59:85–97

Xia Z, Watanabe S, Yamada T et al (2012) Positional cloning and characterization reveal the molecular basis for soybean maturity locus e1 that regulates photoperiodic flowering. Proc Natl Acad Sci U S A 109:E2155–E2164

Xin Z, Wang ML, Barkley NA et al (2008) Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. BMC Plant Biol 8:103

Xu Q, Chen LL, Ruan X et al (2013) The draft genome of sweet orange (*Citrus sinensis*). Nat Genet 45:59–66

Yu J, Hu S, Wang J et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79–92

Chapter 17
Prospects of Functional Genomics in Sugarcane

Ram Kushal Singh and Sujeet Pratap Singh

Abstract Sugarcane is an important commercial crop cultivated for sugar and energy. However, cultivar improvement may be limited due to the narrow genetic base of desired genes. In this sense, functional genomics is a promising tool to assist in the process of developing improved cultivars. Genetic maps linking DNA markers and traits have been developed, but marker-assisted breeding is in its infancy in sugarcane, and genome sequencing has just recently commenced. Substantial resources are available for the sugarcane transcriptome, and both specific and overlapping gene expression patterns for many traits have been established. Gene silencing and over-expression show promise as tests for gene function in sugarcane, and progress has been made in dissecting sucrose accumulation pathways. To broadly assign functions to unknown genes, different fast and multiple parallel approaches are currently used and developed. Such methods allow analysis of the different constituents of the cell that help to deduce gene function, namely the transcripts, proteins and metabolites. Similarly the phenotypic variations of entire mutant collections can now be analyzed in a much faster and more efficient manner than earlier. The different methodologies have developed to form their own fields within the functional genomics technological platform and are termed transcriptomics, proteomics, metabolomics and phenomics. This chapter focuses on recent technological developments and their impact on the field of sugarcane functional genomics. Understanding the unique biological attributes of sugarcane through functional genomics will provide innovative improvement applications that can underpin future, bio-energy and biomaterial industries.

Keywords Conventional breeding • Functional genomics • Marker-assisted selection • Phenomics • Sugarcane • Transcriptomics • Transgenics

R.K. Singh (✉) • S.P. Singh
Centre for Sugarcane Biotechnology, Sugarcane Research Institute (U.P. Council of Sugarcane Research), Shahjahanpur-242 001, India
e-mail: rks.upcsr@gmail.com; sujeetsugarcane@rediffmail.com

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_17

17.1 Introduction

Sugarcane (*Saccharum* spp.) is cultivated on 24 million ha which corresponds to 0.5 % of global agricultural area and produces two-thirds of the world's sugar (Lembke et al. 2012). Ethanol derived from the fermentation of sucrose is an increasingly important product that can be used as fuel, either directly or mixed with refined petroleum (Menossi et al. 2008). It has been estimated that ethanol from sugarcane may replace up to 10 % of the world's refined petroleum consumption in 15–20 years (Goldemberg 2007). Sugarcane bagasse (the major waste product generated by sugar mills after extraction of the sucrose from cane juice) is largely used for energy cogeneration at the mill or for the production of animal feed increasing the overall efficiency of the crop system. Recently, there has been increased interest in using bagasse for processes such as paper production, as a dietary fiber in bread, as a wood substitute in the production of wood composite, and in the synthesis of carbon fibers. Sugarcane, *Saccharum* L., belongs to the Poaceae family, Andropogoneae tribe. There are many species such as *S. officinarum* with juicy cane and sweet softness; Chinese cane, *S. sinense*; Indian canes, *S. barberi*, *S. robustum*; and *S. spontaneum,* wild species that are sugar free but are disease resistant and have abiotic tolerant ability (Roach 1972). Modern sugarcane cultivars (*Saccharum* spp.) derive from introgressions into the highly polyploid domesticated sugar-producing species *S. officinarum*, (x = 10 and 8x = 80) of the wild *S. spontaneum* species characterized by different cytotype levels (x = 8 and 2n = 5x – 16x = 40 – 128) (D'Hont et al. 1998; Sreenivasan et al. 1987).

Sugarcane is a genetically complex polyaneuploid crop, which makes difficult the identification of associations between genes and traits. Because of its multi-specific origin, sugarcane is thought to have one of the most complex plant genomes, carrying variable chromosome numbers (generally 2n = 8× or 10 × 100–130) with a commensurately large DNA content (Lu et al. 1994). The basic genome size ranges from 760 to 926 Mbp, which is twice the size of the rice genome (389 Mbp) and similar to that of sorghum (760 Mbp) (D'Hont and Glaszmann 2001). As a result, sugarcane cultivars have a large and complex genome of about 110–130 chromosomes (10 Gb) corresponding to about 12 homologous sets of a monoploid genome of 1500 cM. However, its large (10 Gb), polyploid, complex genome has hindered genome-based breeding efforts. This is the main cause of slow rate of developing high sugar, high yielding, disease and pest tolerant varieties. Improvement in sugar content is more desirable because more sucrose in less biomass produced would result in lower cost of sugarcane production (Singh and Singh 2004; Singh et al. 2011, 2013). The results of natural hybridization and conventional breeding programs have given the modern sugarcane genome more complexity. Nevertheless sugarcane transgenics are still lagging behind. Weed-resistant and viral-resistant transgenic sugarcane have been reported but so far there has been no genetically modified (GM) sugarcane for commercial cultivation. This is probably due to institutional/intellectual property and regulatory biosafety issues, but may also be related to the fact that for highly complex traits, such as sucrose

content, the genes to be used have not yet been proved ideal for better agronomic performance. Gene discovery and identification is essential for breeding programs, either for transgenic sugarcane development or for marker-assisted breeding (Singh et al. 2013). Genomics science facilitates characterization of entire eukaryote genomes at the DNA sequence level, but for crop plants with complex genomes such as sugarcane, gene characterization is currently best achieved via expressed sequence tag (EST) analysis where sequence information is restricted to genes that are actually functioning in a particular tissue or situation. The development of DNA markers has irretrievably changed the sugarcane breeding approaches. Applications of DNA markers in breeding for cultivar development are called *marker-assisted selection* (MAS). By determining the allele of a DNA marker, sugarcane that possesses genes of economic interest, quantitative trait loci (QTLs) may be identified based on their genotype rather than their phenotype (Collard et al. 2005).

Genetic maps linking DNA markers and traits have been developed, but marker-assisted breeding is in its infancy in sugarcane, and genome sequencing has just commenced (Singh et al. 2013). Linkage maps have been utilized for identifying chromosomal regions that contain genes controlling simple traits (controlled by a single gene) and quantitative traits using QTL analysis (Collard et al. 2005). Proteomics of sugarcane is also at an early stage, especially when dealing with stalk tissue, where there is meager study to date. A systematic proteome analysis of stalk tissue remains to be investigated in sugarcane, wherein the stalk tissue is well-known for its rigidity, fibrous nature, and the presence of oxidative enzymes, phenolic compounds and extreme levels of carbohydrates, thus making the protein extraction complicated (Amalraj et al. 2010). The complete genome sequence of a sugarcane cultivar is not yet available. DNA microarrays allow simultaneous expression analysis of thousands of genes. Gene discovery has been limited in sugarcane. Substantial resources are available for the sugarcane transcriptome, and both specific and overlapping gene expression patterns for many traits have been established. Gene silencing and over-expression show promise as tests for gene function in sugarcane, and progress has been made in dissecting sucrose accumulation pathways.

Functional genomics includes function-related aspects of the genome itself such as mutation and polymorphism (such as single nucleotide polymorphism (SNP) analysis), as well as measurement of molecular activities. The latter comprises a number of *-omics* such as transcriptomics (gene expression), proteomics (protein expression), and metabolomics (interaction among the metabolites). Functional genomics uses mostly multiplex techniques to measure the abundance of many or all gene products such as mRNAs or proteins within a biological sample. Together these measurement modalities endeavor to quantify the various biological processes and improve our understanding of gene and protein functions and interactions. Functional genomics has been extensively used for crop improvements, typically in conjunction with advanced breeding techniques (Kurowska et al. 2011; Langridge and Fleury 2011; Mir et al. 2012; Tran and Mochida 2010). The advent of genome and RNA sequencing is likely, however, to significantly facilitate the breeding process replacing many of the marker-assisted breeding techniques with cheaper whole genome or transcriptome studies.

Sugarcane is a crop with great potential for metabolic engineering, but progress has been limited by highly efficient transgene silencing. The potential exists to utilize efficient gene silencing in molecular improvement through down-regulation of sugarcane genes. However, sugarcane is highly polyploid and heterozygous, which may complicate efforts to employ transgene-mediated silencing of endogenous genes. Gene silencing technology is especially attractive for polyploid organisms because it can be applied to reduce the expression of all homo(eo)logous transcripts of a target gene followed by phenotypic, biochemical and molecular analysis to determine gene function (Manners and Casu 2011). Hairpin-mediated gene silencing (RNAi) should therefore be a powerful tool for the molecular improvement of this important crop (Osabe et al. 2009). These could also be applied to different varieties and ecotypes of crops enhancing the availability of potential target genes for breeding (Mittler and Shulaev 2013).

The objectives of sugarcane functional genomics are to understand the expansion of biological investigation from studying single genes or proteins to studying all genes or proteins at once in a systematic fashion. Understanding the unique biological attributes of sugarcane through functional genomics will provide innovative sugarcane improvement applications for sugar and bio-energy that can underpin future, carbon-neutral, bio-energy and biomaterial industries.

This chapter focuses on the progress made in the dynamic aspects such as the gene at the genomic level (genomics), mRNA during transcription (transcriptomics), as a result formation of the protein (proteomics) which start functions with the help of metabolites (metabolomics) and finally phenotypic expression of sugarcane plant (phenomics), as compared to the static aspects of the genomic information such as DNA sequence or structures.

17.2 Functional Genomics

Functional genomics is often used broadly to refer to the many possible approaches to understanding the properties and function of the entirety of an organism's genes and gene products. This definition is somewhat variable; Gibson and Muse define it as *approaches under development to ascertain the biochemical, cellular, and/or physiological properties of each and every gene product*, while Pevsner includes the study of nongenic elements in his definition: *the genome-wide study of the function of DNA (including genes and nongenic elements), as well as the nucleic acid and protein products encoded by DNA*. It is a field of molecular biology that attempts to make use of the vast wealth of data produced by genomic projects (such as genome sequencing projects) to describe gene (and protein) functions and interactions. Unlike genomics, functional genomics focuses on the dynamic aspects such as gene transcription, translation, and protein-protein interactions, as compared to the static aspects of the genomic information such as DNA sequence or structures.

The functional genomics/systems biology platform is an extremely complex and powerful approach to determine the function of individual genes, pathways,

networks and ultimately entire genomes. The principal premise behind this approach is to evaluate and study the entire cell or organism as a system and understand how different biological processes occur within this system, how they are controlled and how they are executed (Mittler and Shulaev 2013). Functional genomics attempts to answer questions about the function of DNA at the levels of genes, RNA transcripts and protein products. Functional genomics involves studies of natural variation in genes (structural genomics), RNA, and proteins over time (such as an organism's development) or space (such as its body regions), as well as studies of natural or experimental functional disruptions affecting genes, chromosomes, RNA, or proteins (functional genomics) and comparative genomics/integrative genomics (bioinformatics). Thus it works at various levels of genomic research viz; the gene at the genomic level (genomics), mRNA in transcriptome (transcriptomics) as a result formation of the protein (proteomics) which start functions with the help of metabolites (metabolomics) and finally phenotypic expression of sugarcane plant (phenomics) can be observed (Fig. 17.1).

Structural genomics deals with DNA sequencing, sequence assembly, sequence organisation and management. Basically it is the starting stage of genome analysis i.e. construction of genetic, physical or sequence maps of high resolution of the organism. The complete DNA sequence of an organism is its ultimate physical map. Due to rapid advancement in DNA technology and completion of several genome sequencing projects over the last few years, the concept of structural genomics has come to a stage of transition. Based on the information of structural genomics, the next step is to reconstruct genome sequences and to find out the function that the genes perform. This information also lends support to design experimentation to find out the functions that specific genome does. The strategy of functional genomics has widened the scope of biological investigations. This strategy is based on systematic study of single gene/protein to all genes/proteins. Hence, functional genomics provide the novel information about the genome. This facilitates the understanding of genes and the function of proteins, and protein interactions. The wealth of knowledge about this untold story is being unraveled by the scientists with the development of microarray technology and proteomics. These two technologies

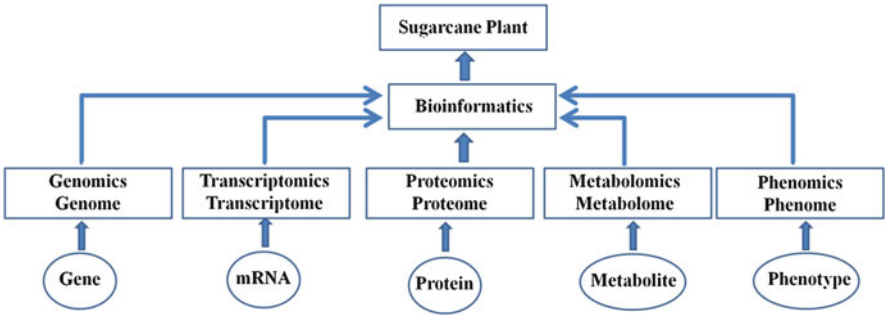

Fig. 17.1 Arrow diagram indicating the integration of results from different technological levels of functional genomics leading to the construction of a sugarcane genome

have helped to explore the instantaneous events of all the genes expressed in a cell/tissue present at varying environmental conditions like temperature and pH.

The sugarcane genome is a complex store of biological information, but on its own it is unable to release that information to the cell. The transcriptome of sugarcane is complex and includes transcripts of homo(eo)logous reflecting the highly polyploid genome of commercial sugarcane hybrids derived recently from two *Saccharum* species. Utilization of the biological information contained in the sugarcane genome requires the coordinated activity of metabolic enzymes and other proteins, which participate in a complex series of biochemical reactions referred to as genome expression (Fig. 17.2). The initial product of genome expression is the transcriptome, a collection of RNA molecules derived from those protein-coding genes whose biological information is required by the cell at a particular time. The characterization of sugarcane genes and their association with biological traits such as sugar accumulation, biomass yield and stress tolerance has so far primarily relied on studies of the sugarcane transcriptome (Manners and Casu 2011). The transcriptome is maintained by the process called transcription, in which individual genes are copied into RNA molecules. The second product of genome expression is the proteome, the cell's repertoire of proteins, which specifies the nature of the biochemical reactions that the cell is able to carry out. The proteins that make up the proteome are synthesized by translation of the individual RNA molecules present in the transcriptome (Casu et al. 2005; de Setta et al. 2014; Manners and Casu 2011).

The advancement in modern technology in recent years has included the development of Next generation DNA sequencing platforms. These have increased our ability to sequence entire genome over the course of time. More than 100 plant genomes have already been fully sequenced, and have reduced the cost of whole genome and whole transcriptome sequencing to a level that is highly accessible for individual researchers (Appleby et al. 2009; Edwards et al. 2013; Mittler and Shulaev 2013). The implications of this advancement are immense and far-reaching

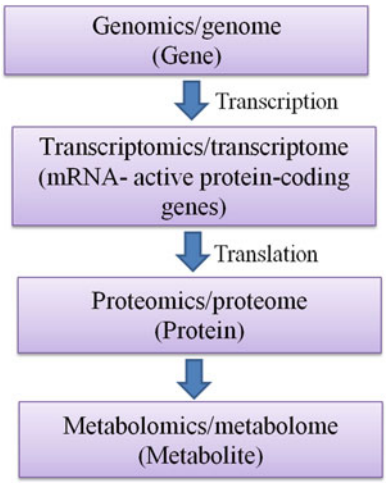

Fig. 17.2 Three way irreversible structure and functions of genome, transcriptome and proteome

in many fields including sugarcane agriculture, disease and pest incidence, medicine and public health. In addition to the sequence data that generates a physical map of the genome, Next generation sequencers can provide information on the methylation state of the entire genome (methylome), thereby enabling us to determine the epigenetic control of different genes and their functions within the genome. Bio-ethanol development is a hot topic these days and functional genomics is already being applied to identify and engineer different pathways into the sugarcane genome. When sugarcane achieves functional genomics, even the sky is not the limit. The massive effort to sequence sugarcane genomes coupled with the application of further functional genomics tools should eventually yield an enormously high level of understanding of sugar accumulation functions, development and its regulation. In the near future there are a number of different research avenues that would significantly benefit from the application of functional genomics.

17.2.1 Goals of Functional Genomics

Sugarcane productivity in the twenty-first century will face unprecedented challenges due to the changing global climate, including unstable patterns of precipitation and temperature, intensified biotic and abiotic stressors, and soil erosion (Walthall et al. 2012). Growing enough sugar, fuel and fiber to accommodate a growing world population in a changing climate without further compromising the environment presents complications in the form of increased fertilizer use and subsequent runoff, increased pesticide/herbicide use and loss of biodiversity. These issues are not independent of each other. Trends toward decreasing crop yields due to increased drought, heat and pathogen stress may motivate the use of mitigating strategies that include higher inputs of increasingly scarce resources and pesticides. A more complete understanding of sugarcane response to the surrounding environment, particularly the molecular and genetic mechanisms that sense and regulate this response, is critically needed to accelerate development of widely-adaptable crops. With increased drought tolerance, nutrient-use efficiency, and resistance/tolerance to pathogens and insects, these improved varieties, ideally, will tolerate and even thrive under suboptimal conditions and require few inputs of functional genomics (Kasuga et al. 2004).

The goal of functional genomics is to understand the relationship between an organism's genome and its phenotype. The past decade has seen an exponential growth in genome-scale measurements of diverse cellular parameters, such as transcription, translation, protein abundance, protein-protein interactions, metabolites and genetic phenotypic interaction. In order to gain a systems-level understanding of a given process, cell or organism, the current challenge is to convert these static qualitative maps into dynamic quantitative models of cellular processes. This rather daunting task can only be achieved through a multidisciplinary approach, which requires intensive integration of technology and thinking from basic biology, genomics, computational biology, mathematics, engineering and physics.

17.2.2 Sugarcane ESTs in Functional Genomics

The information carried by an expressed sequence tags (ESTs) collection is a significant path to determine an organism's genome content but more pragmatically, when considering key crops, it can directly point to genes which may contribute to agronomical trait development (e.g. tolerance to abiotic and biotic stresses, and sugar content among others). Significant progress has been noted recently with the development of tools such as ESTs. To date 284353 ESTs have become available to explore the large polyploid sugarcane genome and consequently renewed the interest in sugarcane genetics (NCBI 2014). ESTs represent tags of the expressed portion of a genome and therefore potentially identify genes encoding proteins, natural antisense transcripts (Carson et al. 2002; Ma et al. 2006), miRNA, transacting siRNA precursors (Zhang et al. 2005), and more generally non-coding RNA (Menossi et al. 2008). Several sugarcane ESTs collections have been developed (Ma et al. 2006; Parida et al. 2009; Singh et al. 2013; Vettore et al. 2003). The publicly-available sugarcane ESTs were assembled into tentative consensus sequences (virtual transcripts), singletons, and mature transcripts, referred to as the Sugarcane Gene Index (SGI; http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=*s.officinarum*). The Brazilian sugarcane ESTs project collection has generated 237954 ESTs, which were organized into 43141 putative unique sugarcane transcripts (26803 contigs and 16338 single tons) referred to as sugarcane assembled sequences (SASs) on SUCEST, http://sucestfun.org. An internal redundancy analysis suggested that this collection of SASs represented 33000 sugarcane genes, but this estimation was likely an overestimation, since a two-fold redundancy among SASs that presented significant similarity with *Arabidopsis* proteins (60 % of the SASs) was detected (Menossi et al. 2008). A detailed organization of sugarcane genes into functional categories (i.e. signal transduction components, regulation of gene expression, development, biotic and abiotic stresses, transposable elements, metabolism, etc.) had completed and represents the basis to develop functional genomics approaches (Vettore et al. 2003). The availability of ESTs allows for large-scale gene expression analysis using a variety of tools (Papini-Terzi et al. 2005).

The SSRs developed from ESTs are derived from expressed genes. The advantage of using ESTs is that they can facilitate the mapping of genes with known function pathways (Cato et al. 2001; Singh et al. 2013). They are also located in the transcribed portion of the genome (Marconi et al. 2011), which allows for a direct association between genes and important agronomic traits. SSR markers include genomic SSRs and EST-derived SSRs. EST-SSR markers identify variation in the expressed portion of the genome, so that gene tagging should give *perfect* marker-trait associations. EST-SSRs, unlike genomic SSRs, which showed a high percentage of cross transferability, these are utilized across a number of related genera/species (Singh et al. 2011; Varshney et al. 2005). They have shown to be useful for comparative mapping across species, comparative genomics, and evolutionary studies and they have been shown to possess a higher potential for inter-specific transferability than genomic SSRs (Gupta et al. 2010). On the other hand, they are expected to be less polymorphic within the species due to their conserved nature

(Varshney et al. 2005). Thus, EST-SSRs have provided a valuable source of new PCR-based molecular markers in many crops. A total of ~285000 ESTs, generated from 30 tissues and many developmental stage-specific cDNA libraries, are available in the SUCEST web page (sugarcane EST project in Brazil) (Bugos and Thom 1993; Vettore et al. 2003). These sugarcane ESTs have aided in the identification and transcription profiling of signal transduction-related genes in sugarcane tissues (Papini-Terzi et al. 2005), low-temperature responsive genes (Nogueira et al. 2003) and ABA – and MeJA activated sugarcane bZIP genes (Schlog et al. 2008).

EST-derived microsatellites are the sequence redundancy that yields multiple sets of markers at the same locus. However, SSRs developed from assembled ESTs or unigenes, popularly known as unigene-derived microsatellite (UGMS) markers, have the advantage of assaying variation in the expressed component of the genome with unique identity and positions (Parida et al. 2006). With the increasing emphasis on functional genomics, a large number of EST sequences are being developed and assembled into unigene databases (http://www.ncbi.nlm.nih.gov/unigene) (Lopez et al. 2004). Furthermore, with evolving bioinformatics tools, it is now possible to identify and develop UGMS markers efficiently on a large scale in much less time (Parida et al. 2006). Because of the above advantages of UGMS markers, and relatively easy accessibility of large unigene resources, increasing numbers of UGMS markers, are being identified and used for a variety of applications in a number of plant species like the gerbera daisy (Gong and Deng 2010), and cereals such as wheat, barley, maize, sorghum, rice and *Arabidopsis* (Parida et al. 2006). The UGMS markers can also be used to accurately assay functional diversity in the natural populations and available germplasm collections as well as for comparative mapping and evolutionary studies as anchor markers (Reddy et al. 2011).

17.2.3 Functional Genomics for Economic Traits

EST sequencing has significantly contributed to gene discovery and expression studies used to associate function with sugarcane genes. A significant amount of data exists on regulatory events controlling responses to herbivory, drought and phosphate deficiency, which cause important constraints on yield and on endophytic bacteria, which are highly beneficial (Riera et al. 2005). The means to reduce drought, phosphate deficiency and herbivory by the sugarcane borer have a negative impact on the environment. Improved tolerance for these constraints is being sought. Sugarcane's ability to accumulate sucrose up to 16 % of its culm dry weight is a challenge for genetic manipulation. Genome-based technology such as cDNA microarray data indicate genes associated with sugar content that may be used to develop new varieties improved for sucrose content or for traits that restrict the expansion of the cultivated land. The genes can also be used as molecular markers of agronomic traits in traditional breeding programs. Gene discovery through the SUCEST sequencing program has been a major breakthrough for the breeding programs throughout the world, and functional studies based on cDNA arrays are uncovering pathways of plant adaptation and responses to the environment. EST-simple sequence repeats (SSRs) have been

successfully used for genetic relationship analysis, extending the knowledge of the genetic diversity of sugarcane to a functional level. Development of new markers based on ESTs and their integration into genetic maps will renew breeding programs and help MAB technology speed up breeding programs (Menossi et al. 2008). Genes can be silenced or over-expressed to study their function and to produce new phenotypes not possible through conventional breeding. Metabolic profiling associated with gene expression studies are certainly the future tools of the sugarcane industry. Also, the analysis of the transcriptome in transgenic plants altered for genes of interest would certainly prove to be an excellent tool to unravel sugarcane regulatory networks associated with important traits (Leibbrandt and Snyman 2003; Menossi et al. 2008).

17.2.4 Functional Annotations for Genes

Putative genes can be identified by scanning a genome for regions likely to encode proteins, based on characteristics such as long open reading frames, transcriptional initiation sequences and poly adenylation sites. A sequence identified as a putative gene must be confirmed by further evidence, such as similarity to cDNA or EST sequences from the same organism, similarity of the predicted protein sequence to known proteins, association with promoter sequences, or evidence that mutating the sequence produces an observable phenotype (Buchel et al. 1996). The identification of signal transduction components and transcription factors that might regulate sugar accumulation is highly desirable if we are to improve this characteristic of sugarcane. Terzi et al. (2009) evaluated 30 genotypes that have different Brix (sugar) levels and identified genes differentially expressed in internodes using cDNA microarrays. These genes were compared to existing gene expression data for sugarcane plants subjected to diverse stress and hormone treatments. The comparisons revealed a strong overlap between the drought and sucrose-content datasets and a limited overlap with ABA signaling (Riera et al. 2005). Genes associated with sucrose content were extensively validated by qRT-PCR, which highlighted several protein kinases and transcription factors that are likely to be regulators of sucrose accumulation (Boudsocq and Lauriere 2005). Moreover, sucrose-associated genes were shown to be directly responsive to short-term sucrose stimuli, confirming their role in sugar-related pathways. Gene expression analysis of contrasting sugarcane populations for sucrose content indicated a possible overlap with drought and cell wall metabolism processes and suggested signaling and transcriptional regulators to be used as molecular markers in breeding programs.

17.3 Transcriptome Profile Analysis of Sugarcane

The transcriptome is the total mRNA in a cell or organism and the template for protein synthesis in a process called *translation*. The transcriptome reflects the genes that are actively expressed at any given moment. Gene expression

microarrays measure packaged mRNA (mRNA with the introns spliced out) as a summary of gene activity. While advances in microarray technology have resulted in progress in genomics and transcriptomics (and the resultant literature), it is important to highlight some limitations. Specifically, gene expression microarrays measure changes in mRNA abundance, not protein, and thus there is a lack of consensus around the interpretation of microarray data (Manners and Casu 2011). Regulation of gene expression is critical for a variety of essential processes in plants, such as growth, development, differentiation, metabolic regulation and stress tolerance. The transcriptome, which can vary with external environmental conditions, is the set of all RNA molecules, including mRNA, rRNA, tRNA, and noncoding RNAs produced in one or a population of cells (Wei et al. 2011). Transcription, the first step in gene expression, plays a central role in the regulation of the expression of genes. Transcription factors are a family of proteins that control the physiological and biochemical process of a plant. The MYB family of proteins is the largest family of transcription factors (TF) that plays an important role in controlling growth and development and stress responses. These genes have been extensively studied in many dicot and monocot plants (Geethalakshmi et al. 2014)

Among the grasses, sugarcane has a complex polyploidy genome and only a little information is available about the MYB TF family. Therefore, the analysis of the TF is essential for an understanding of the mechanisms of gene expression and their genome organization. With increasing importance of this biofuel crop, 57 R2R3-MYB, 20 MYB-like and 1 3R1-MYB genes were computationally predicted and 51 ScMYBs were isolated by RT-PCR. Subgroup-specific conserved motifs outside the MYB domain demonstrated their phylogeny and functional conservation in sugarcane (Geethalakshmi et al. 2014). Next generation sequencing has been widely employed to obtain transcriptome data in species without a sequenced genome. Transcriptome analysis using high-throughput short-read sequencing technology, such as Solexa sequencing, is straight forward, and does not have to be restricted to the genome of model organisms (Zenoni et al. 2010). This analysis can provide information on gene expression and regulation and thus is essential to interpret the functional elements of the genome and reveal molecular mechanisms. Solexa sequencing is a high-throughput, short-read, massively parallel sequencing platform, of which the read length is relatively short (21 bp), and bioinformatics analysis of the corresponding differentially expressed genes has to rely only on sugarcane EST databases (Wu et al. 2013). Therefore, in the absence of corresponding sequenced genome information as a reference, many of the differentially-expressed genes can not be functionally annotated. However, the genomes of sorghum, maize and rice can act as reference information (Singh et al. 2011). According to sugarcane unigenes identified and annotated by RNA-seq and sorghum, maize and rice reference genome, we hope to establish a platform for future genetic and functional genomic research in sugarcane. In conclusion, the usefulness of the Solexa sequencing in identifying genes related to sugarcane smut defense has been successfully demonstrated by Wu et al. (2013). However, most of the molecular mechanisms of sugarcane-smut interaction are as yet unknown. More genes related to sugarcane defense and their expression profiles in response to smut infection should be ana-

lyzed further. Wu et al. (2013) provided a Solexa sequencing platform for gene expression research on sugarcane and also a reference for studying the molecular mechanism in non-model organisms like sugarcane.

17.3.1 Role of RNAs in Functional Genomics

In recent years, the opinion that RNA molecules only serve as a genetic information carrier has dramatically changed. Today, it is widely accepted that most of the genome is actively transcribed and that non-coding RNA molecules would have key functional roles in several fundamental cell processes such as cell division, growing and differentiation (Alexander et al. 2010; Emidio et al. 2011). Indeed, RNA molecules are now known to carry diverse functions, such as catalysis, transcription regulation and chromosome repair, among others (Wei et al. 2011).

Non-coding RNAs (ncRNAs) have been found to have roles in a great variety of processes, including transcriptional regulation, chromosome replication, RNA processing and modification, messenger RNA stability and translation, and even protein degradation and translocation. Recent studies indicate that ncRNAs are far more abundant and important than initially imagined (Huttenhofer and Vogel 2006; Storz 2002). The development of new methods for assessing the accuracy of various RNAs structure models can benefit from previous experience gathered in the field of protein structure prediction and assessment (Fig. 17.3) (Ken 2010; Wei et al. 2011).

17.3.2 DNA Microarray in Sugarcane

DNA microarray technology has been used successfully to detect the expression of many thousands of genes, to detect DNA polymorphisms and to map genomic DNA clones. It permits quantitative analysis of RNAs transcribed from both known and unknown genes and allows comparison of gene expression patterns in normal and pathological cells and tissues (Heller et al. 1997). DNA microarrays have been used extensively in recent years to study mRNA expression profiles of different cell types under various growth conditions. These steady-state mRNA profiles provide a wealth of information about cellular functions and responses. However, they do not necessarily reflect the ultimate gene expression profile of a cell since the step of translation might lead to discrepancies between the mRNA profile and the profile of the actual functioning unit in the cell, the protein. To investigate transcriptome changes in response to environmental inputs that alter yield we used cDNA microarrays to profile expression of 1545 genes in plants submitted to drought, phosphate starvation, herbivory and N_2-fixing endophytic bacteria (Xiang and Brownstein 2003). The arrayed elements correspond mostly to genes involved in signal transduction, hormone biosynthesis, transcription factors, novel genes and genes corresponding to unknown proteins (Papini-Terzi et al. 2005). Microarrays

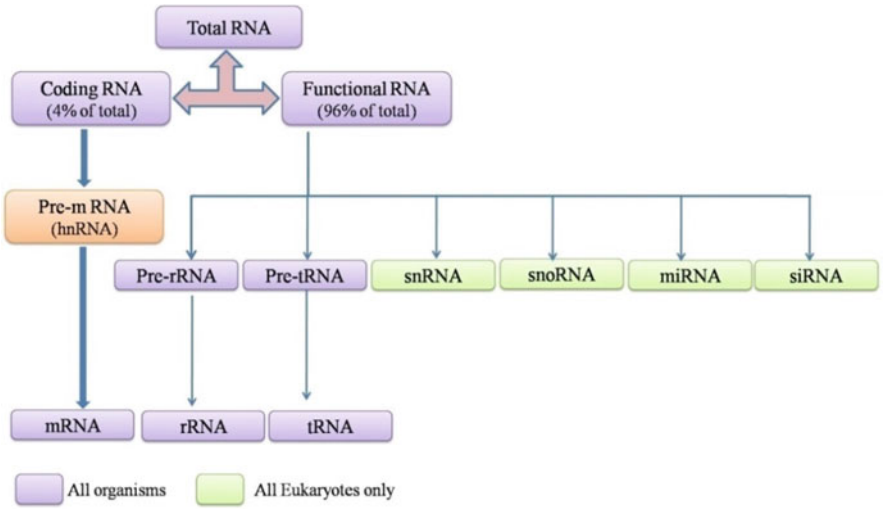

Fig. 17.3 Structural RNAs working in all organisms

measure the amount of mRNA in a sample that corresponds to a given gene or probe DNA sequence. Probe sequences are immobilized on a solid surface and allowed to hybridize with fluorescently labeled *target* mRNA. The intensity of fluorescence of a spot is proportional to the amount of target sequence that has hybridized to that spot, and therefore to the abundance of that mRNA sequence in the sample. Microarrays allow for identification of candidate genes involved in a given process based on variation between transcript levels for different conditions and shared expression patterns with genes of known function. An extensive study on the sugarcane transcriptome was performed by Sharma and Mishra (2014). Sugarcane genes responsive to phytohormones and to challenges sugarcane commonly deals with in the field were identified. Additionally, the protein kinases were annotated based on a phylogenetic approach. The experimental design and statistical analysis applied proved robust to unravel genes associated with a diverse array of conditions attributing novel functions to previously unknown or undefined genes. The data consolidated in the SUCAST database resource can guide further studies and be useful for the development of improved sugarcane varieties (Xiang and Brownstein 2003).

17.4 Serial Analysis of Gene Expression (SAGE)

SAGE is an alternate method of gene expression analysis based on RNA sequencing rather than hybridization. It relies on the sequencing of 10–17 base pair tags which are unique to each gene. These tags are produced from poly-A mRNA and ligated end-to-end before sequencing. SAGE gives an unbiased measurement of the number of transcripts per cell, since it does not depend on prior knowledge of what

transcripts to study (as microarrays do). Sugarcane can be considered a potential candidate for phytoremediation because of its outstanding biomass production (commercial average 300 mt/ha), but its prospective metal tolerance had not been previously evaluated (Calsa and Figueira 2007).

Identification of differentially-expressed genes in sugarcane under various stresses can give clues as to what defense mechanisms and biochemical pathways are regulated during different types of stress. To date, differentially-expressed genes can be studied using a multitude of methods but mainly involving representational difference analysis of cDNA (RDA), serial analysis of gene expression (SAGE) (Yamamoto et al. 2001), suppression subtractive hybridization (SSH) (Aharoni and Vorst 2002), cDNA microarray analysis, and cDNA-amplified fragment length polymorphism (AFLP). Among them, cDNA-AFLP has proven to be an excellent tool to identify novel genes related to plant resistance to pathogens. The cDNA-AFLP technique incorporates AFLP for the analysis of the differential expression of mRNA (You-Xiong et al. 2011).

17.5 Proteomics and Photosynthetic Genes in Sugarcane

Proteomics is the study of the proteome, the protein complement of the genome. Genes are being discovered by the millions, including many that encode proteins with no known function. To describe the entire protein complement encoded by an organism's DNA, researchers have coined the term *proteome* (Fields 2001). Proteomics are complementing work carried out on cellular intermediary metabolism and nucleic acid metabolism to provide a new and increasingly complete picture of biochemistry at the level of cells and even organisms. Sugarcane proteomic analyses have examined general protein polymorphisms, protein expression changes after dedifferentiation of leaf tissue in callus culture and drought-stress responsive proteins (Sugihario et al. 2002). An increase of sucrose content in elite sugarcane cultivars may be a main point to be addressed by using genetic transformation, and is directly dependent of increasing photosynthetic efficiency. The vast majority of photosynthetic proteins is nucleus-encoded and requires N-terminal pre sequences, named chloroplast transit peptides, to target them to the chloroplast. About 2100–3600 distinct chloroplast proteins are nuclear-encoded, while about 100–120 are encoded by the organelle genome. The complete nucleotide sequences of the chloroplast genome of sugarcane (*Saccharum officinarum*) are determined. It consists of 141182 base-pairs (bp), containing a pair of inverted repeat regions (IRA, IRB) of 22794 bp each. The IRA and IRB sequences separate a small single copy region (12546 bp) and a large single copy (83048 bp) region. The gene content and relative arrangement of the 116 identified genes (82 peptide-encoding genes, 4 ribosomal RNA genes, 30 tRNA genes), with the 16 ycf genes, are highly similar to maize (Calsa et al. 2004). The sugarcane photosynthetic efficiency is expected to be improved upon manipulation of photosynthetic genes (i.e. ribulose-1,5-bisphosphate carboxylase/oxygenase, phosphoenolpyruvate carboxylase, carbonic anhydrase)

generating novel knowledge in this research field as well as leading to increased synthesis of triose phosphates, and ultimately, increased sucrose content in the transgenic cultivars. Editing events, defined as C-to-U transitions in the mRNA sequences, comparable with those observed in maize, rice and wheat. The conservation of gene organization and mRNA editing suggests a common ancestor for the sugarcane and maize plastomes. These data provide the basis for functional analysis of plastid genes and plastid metabolism within the Poaceae (Calsa et al. 2004).

17.5.1 RNAi

RNA interference (RNAi) is an important process, used by organisms to regulate the activity of genes at the stage of translation or by hindering the transcription of specific genes (Fire et al. 1998). It is a biological course of action in which RNA molecules reduce gene expression, typically by causing the demolition of specific mRNA molecules (Jing et al. 2005). The use of the RNA interference (RNAi) pathway to eliminate gene products has greatly facilitated the understanding of gene function. Behind this remarkable pathway is an intricate network of proteins that ensures the degradation of the target mRNA (Fig. 17.4a, b). RNA interference (RNAi) methods can be used to transiently silence or knock down gene expression using ~20 base-pair double-stranded RNA typically delivered by transfection of synthetic ~20-mer short-interfering RNA molecules (siRNAs) or by virally encoded short-hairpin RNAs (shRNAs). RNAi screens, typically performed in cell culture-based assays or experimental organisms (such as the nematode *Caenohrhabditis elegans*) can be used to systematically disrupt nearly every gene in a genome or subsets of genes (sub-genomes); possible functions of disrupted genes can be assigned based on observed phenotypes.

17.6 Protein-Protein Interactions in Sugarcane

Plants respond to pathogens and insect attacks by inducing and accumulating a large set of defense-related proteins. During the process of plant development, sugarcane is frequently infested by pathogens (including bacteria, fungi and viruses) which are major biotic stresses. Regardless of whether the interaction between plant and pathogen is a disease-resistant (non-affinity interaction) or susceptible reaction (affinity interaction), it is due to the interaction between the disease-resistance gene of the host plant and the corresponding a virulence gene of pathogen, which could induce the result of the coordination expression of a series of defense genes in host plant. Two homologous of a barley wound-inducible protein (BARWIN) have been characterized in sugarcane, SUGARWIN1 and SUGARWIN2 (sugarcane wound-inducible proteins). SUGARWIN2 specificity to pathogenic fungi associated with red rot suggests an unfavorable interaction between the sugarcane borer (*Diatraea saccharalis*)

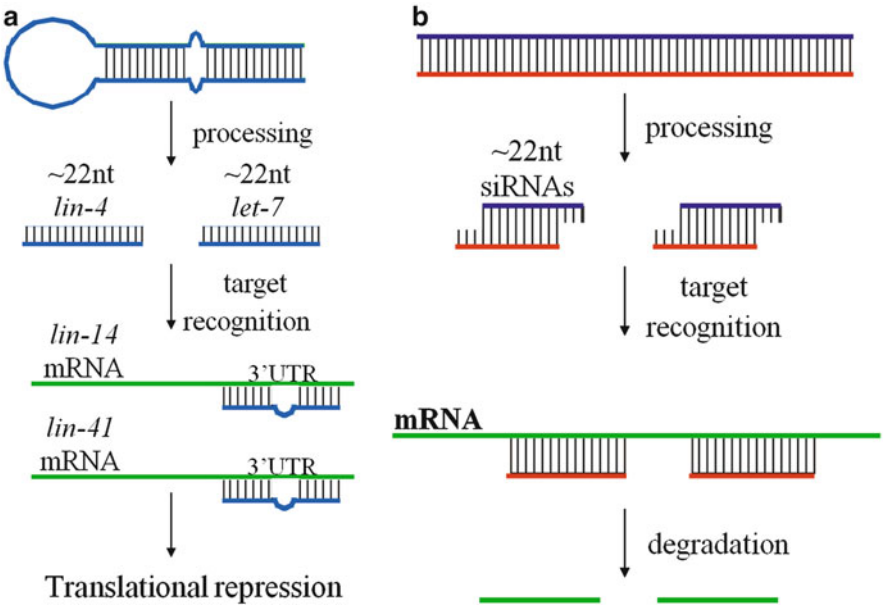

Fig. 17.4 (**a**) Developmental regulation by stRNAs (μRNAs), (**b**) RNAi by siRNAs

and red rot (*Colletotrichum falcatum*) because the protein is expressed due to the borer attack and has deleterious effects only on the fungus (Franco et al. 2014). In addition, the protein itself does not show any effect on insect development; instead, it has antimicrobial activities toward *Fusarium verticillioides*, an opportunistic fungus that usually occurs after the *D. saccharalis* borer attacks sugarcane. However, in the reactions of affinity interaction or non-affinity interaction, the differential genes had significant differences in spatial distribution, expression rate and expression intensity (Que et al. 2009). Therefore, these differences at the level of gene expression directly appeared in the various generation rate, intensity and spatial distribution of proteins in disease-resistant and susceptible varieties after the interaction between plant and pathogen. Que et al. (2011) showed that there were significant differences in protein 2-DE atlas between resistant and susceptible varieties, and also between the inoculated and the control sugarcane. In total, 23 proteins, including 11 up regulated, 9 down regulated, and 3 newly induced after infection, were identified by MALDI-TOF-TOF/MS. Bioinformatic analysis revealed that the functions of these 23 differential proteins were related to photosynthesis, signal transduction and disease resistance. Sugarcane provides a biosecure platform for transgene containment and we have used it to produce a biologically-active high value protein. However, protein accumulation was too low to be economically viable due to post-transcriptional gene silencing (PTGS). The long-term goal of this study is to find a suppressor to prevent PTGS of transgenes without interfering with sugarcane growth and development. Using sugarcane plants to produce high-value proteins could significantly increase the economic competitiveness of the Hawaiian sugar industry. For example, recombinant GM-CSF for human therapeutic use is produced in yeast and sold under

the name Leukine by Berlex (www.berlex.com). Because of the biosecurity and other advantages that sugarcane provides, there are several significant efforts underway in other countries to develop sugarcane as a bio factory system. In Australia, the Bureau of Sugar Experiment Stations and the University of Queensland have developed transgenic sugarcane that produces a thermoplastic poly hydroxyl butyrate, and they are also working on other high value products.

17.7 Metabolomics

The study of global metabolite profiles in a system (cell, tissue or organism) under a given set of conditions is known as *metabolomics*. It has a number of theoretical advantages over the other omic approaches. It is the systematic study of the unique chemical fingerprints, metabolites, that specify cellular processes of an organism at different developmental stages and under different environmental conditions (Broadhurst and Kell 2006). Metabolomics embraces several strategies that aim to increase our understanding of how metabolite levels and their interactions influence phenotypes. The metabolome is the final downstream product of gene transcription, and therefore, changes in the metabolome are amplified relative to changes in the transcriptome and the proteome (Sharma and Mishra 2014). Sucrose content increases with internode development down the stem of sugarcane. In an attempt to determine which other changes in metabolites may be linked to sucrose accumulation, gas chromatography and mass spectrometry were used to obtain metabolic profiles from methanol/water extracts of four samples of different age down the stem of cultivar Q117 (Glassop et al. 2007).

Metabolomics has proven to be very rapid and superior to any other post-genomics technology for pattern-recognition analyses of biological samples. One of the key advantages of metabolomics is that a sequenced genome is not required to make sense of the data (Weckwerth and Morgenthal 2005). Basically, this technique can elucidate differences in genomes without using any DNA. Metabolomics studies mainly focus on the observed changes in metabolite concentration to changes in gene expression and perhaps even in the genes themselves. Metabolomics can be used in analyzing diversity, linking genotypic to phenotypic diversity and analyzing plant-pathogen interactions, in determining the nutritional difference between traditional and genetically modified crops. It should be useful in characterizing specific metabolic networks associated with several plant metabolisms viz; nitrogen metabolism, lipid metabolism, glucosinolate metabolism, starch metabolism, and flavonoids metabolism in understanding the adaptive response of biotic and abiotic stress tolerance in crop plants. Metabolomics may also help to identify uncommon and valuable phytochemicals using the integration of transcript and protein profiling in the identification of regulated key sites in metabolic networks (Tiessen et al. 2002). Metabolomics is considered an efficient tool for addressing future needs in agriculture and human nutrition for assessing food safety. Recent studies on plant biology have focused on several applications of metabolomics including the connectivity of different pathways, in the engineering of biodegradation and mineralization processes for the removal of toxic

substances from atmosphere, in discovering novel pathways, in uncovering silent phenotypes of mutations and in the nutraceutical breeding of crop plants. Identification and quantification of specific metabolites in complex mixtures and their data analysis and integration remains the major problems of metabolomics studies. As the scientific knowledge of metabolic pathways increases (gene, proteins and metabolites interactions), this can be expected to improve in the near future (Sharma and Mishra 2014).

17.8 Phenomics: Genotype to Phenotype

Recent advances in DNA sequencing and phenotyping technologies, in concert with analysis of large datasets, have spawned *phenomics*, defined as the use of large-scale approaches to study how genetic instructions from a single gene or the whole genome translate into the full set of phenotypic traits of an organism. It has focused on analyzing phenotype, because it is frequently slower and more expensive than genomics due to the difficulties of measuring molecular, cellular or organismal traits with sufficient throughput, resolution and precision. Thus, phenomics has broad importance in applied and basic biology and is equally relevant to goals as disparate as yield improvement in food and energy crops, environmental remediation using microbes and plants, and understanding complex networks that control fundamental life processes. Examples of phenotypic parameters in sugarcane include gross morphological measures such as stalk diameter, height, number, internodes, cane weight, sugarcane yield, dynamic measures such as rate of cell division, stomatal diffusive resistance, etc. However, prediction of phenotype from genotype is generally a difficult problem due to the large number of genes and gene products that contribute to most phenotypes in concert with complex and changeable environmental influences (NIFA-NSF 2011). The study of phenotypes is quite different. Unlike a genotype, the phenotype of an organism can be described at many levels, from specific molecules to dynamic metabolic networks to complex cellular developmental and physiological systems, all the way to the aggregate or social behaviors of complex populations. Interactions with symbionts, pathogens or competing organisms create additional levels of phenotypic complexity. Moreover, phenotypes are dynamic and the time scales in which they change vary tremendously. Consider, for instance, the rapid responses of a bacterium to nutrient changes or the dynamic changes in photosynthesis of a leaf as a single cloud passes shading the sun, compared to the slow morphological changes in long-lived sugarcane plants or even the lifelong changes in the outward appearance of a human being (Heffner et al. 2011; Lu et al. 2011; Murchie and Niyogi 2011).

17.9 Functional Genomics and Bioinformatics

Modern bioinformatics came of age with the development of genomic technologies, specifically the ability to produce large amounts of sequence information at an ever-decreasing cost. Because of the large quantity of data produced by these techniques

and the desire to find biologically meaningful patterns, bioinformatics is crucial to analysis of functional genomics data. Examples of techniques in this class are data clustering or principal component analysis for unsupervised machine learning (class detection) as well as artificial neural networks or support vector machines for supervised machine learning (class prediction, classification). Functional enrichment analysis is used to determine the extent of over- or under-expression (positive- or negative-regulators in case of RNAi screens) of functional categories relative to background sets. Gene ontology based enrichment analysis are provided by DAVID and Gene Set Enrichment Analysis (GSEA), pathway-based analysis by Ingenuity and Pathway studio and protein complex based analysis by COMPLEAT. Bioinformatic tools were developed that allowed the assembling and sequence analysis of the sugarcane ESTs (Edwards and Batley 2004). A total of 49 groups headed by researchers with diverse expertise in many different areas of biological science contributed for a detailed annotation of the sugarcane EST project (SUCEST) database (http://sucest.lad.ic.unicamp.br/en) carried out by the Brazilian consortium. The vast quantities of diverse biological data generated by recent biotechnological advances have led to the development and evolution in the field of bioinformatics. This relatively new field facilitates both the analysis of genomic and post genomic data and the integration of information from the related fields of transcriptomics, proteomics, metabolomics and phenomics (Edwards and Batley 2004).

17.10 Conclusions and Prospects

Commercial sugarcane cultivars are poly-aneuploid, interspecific hybrids and contain 110–130 chromosomes. However, its large (10 Gb), complex genome has hindered genome-based breeding efforts. Gene discovery was a necessary first step for commencing genomics research in sugarcane. The strength of functional genomics is that it enables us to synergistically bring together complementary approaches. Commercial arrays for large-scale expression profiling hold promise for the development of a databank of expression profiling for further data-mining. Candidate genes from large-scale expression profiling as well as other studies have been a source for targeting peptide discovery, promoter discovery, transgenic testing for trait modification, including use of RNAi/DNA marker development. The world sugarcane ESTs collection is a valuable resource to assist in the production of the sugarcane genome. Most of the functional genomics studies conducted to date are performed on laboratory or transgenic house-grown GM sugarcane. The natural environment is nevertheless significantly different from the artificial conditions used in vivo or in vitro and the development of improved sugarcane would be significantly enhanced if the functional genetics platform would be applied to sugarcane grown under natural conditions at a commercial scale.

The characterization of sugarcane genes and their association with biological traits, that is sugar accumulation, biomass yield and stress tolerance should be well documented for further studies. The impact of this genomic complexity on transcription will be greatly informed by data emerging from the International Sugarcane

Genome Sequencing Project. Reverse genetics approaches would be possible in sugarcane but only very few sugarcane genes have had their specific functions examined in transgenic sugarcane plants, thus this study should be conducted in transgenic experiments. The development and design of sugarcane for sugar, bio-fuel, feed-stalk and energy by co-generation could be enhanced by the application of functional genomics tools to identify beneficial pathways and engineer them into or out of specific sugarcane cultivars. This would result in the enhancement of sugar and bio-fuel quality without the need to modify sugarcane at the post harvest stage. An integration of genome, transcriptome and metabolome data will emerge and inform the molecular breeding of this important sugar and energy crop to meet the requirements of a growing population in the near future.

References

Aharoni A, Vorst O (2002) DNA microarrays for functional plant genomics. Plant Mol Biol 48:99–118

Alexander RD, Steven A, Barrass JD et al (2010) Splicing-dependent RNA polymerase pausing in yeast. Mol Cell 40(4):582–593

Amalraj RS, Selvaraj N, Veluswamy GK et al (2010) Sugarcane proteomics: establishment of a protein extraction method for 2-DE in stalk tissues and initiation of sugarcane proteome reference map. Electrophoresis 31(12):1959–1974

Appleby N, Edwards D, Batley J (2009) New technologies for ultra-high throughput genotyping in plants. Methods Mol Biol 513:19–39

Boudsocq M, Lauriere C (2005) Osmotic signaling in plants: multiple pathways mediated by emerging kinase families. Plant Physiol 138(3):1185–1194

Broadhurst DI, Kell DB (2006) Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2:171–196

Buchel CS, Dry IB, Robinson SP (1996) Isolation of a full-length cDNA encoding polyphenol oxidase from sugarcane, a C_4 grass. Plant Mol Biol 31:1233–1238

Bugos RC, Thom M (1993) Glucose transporter cDNAs from sugarcane. Plant Physiol 103:1469–1470

Calsa T Jr, Figueira A (2007) Serial analysis of gene expression in sugarcane (*Saccharum* spp.) leaves revealed alternative C4 metabolism and putative antisense transcripts. Plant Mol Biol 63(6):745–762

Calsa T Jr, Carraro DM, Benatti MR et al (2004) Structural features and transcript-editing analysis of sugarcane (*Saccharum officinarum* L.) chloroplast genome. Curr Genet 46(6):366–373

Carson DL, Huckett BI, Botha FC (2002) Sugarcane ESTs differentially expressed in immature and maturing intermodal tissue. Plant Sci 162(2):289–300

Casu RE, Mannersa JM, Bonnett GD et al (2005) Genomics approaches for the identification of genes determining important traits in sugarcane. Field Crop Res 92(2–3):137–147

Cato SA, Gardner RC, Kent J, Richardson TE (2001) A rapid PCR-based method for genetically mapping ESTs. Theor Appl Genet 102:296–306

Collard BCY, Jahufer MZZ, Brouwer JB et al (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169–196

D'Hont A, Glaszmann JC (2001) Sugarcane genome analysis with molecular markers, a first decade of research. Proc Int Soc Sugarcane Technol 24:556–559

D'Hont A, Ison D, Alix K et al (1998) Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of RNA genes. Genome 41:221–225

de Setta N, Monteiro-Vitorello CB, Metcalfe CJ et al (2014) Building the sugarcane genome for biotechnology and identifying evolutionary trends. BMC Genome 15:540
Edwards D, Batley J (2004) Plant bioinformatics: from genome to phenome. Trends Biotechnol 22(5):222–237
Edwards D, Batley J, Snowdon RJ (2013) Accessing complex crop genomes with next-generation sequencing. Theor Appl Genet 126:1–11
Emidio C, Norambuena T, Marti-Renom MA, Melo F (2011) All-atom knowledge-based potential for RNA structure prediction and assessment. Bioinformatics 27(8):1086–1093
Fields S (2001) Proteomics: proteomics in genomeland. Science 291:1221–1224
Fire A, Xu S, Montgomery MK et al (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806–811
Franco FP, Santiago AC, Henrique-Silva F et al (2014) The sugarcane defense protein SUGARWIN2 causes cell death in *Colletotrichum falcatum* but not in non-pathogenic fungi. PLoS One 9(3), e91159. doi:10.1371/journal.pone.0091159
Geethalakshmi S, Barathkumar S et al (2014) The MYB transcription factor family genes in sugarcane (*Saccharum* sp). Plant Mol Biol Rep. doi:10.1007/s11105-014-0768-3
Glassop D, Roessner U, Bacic A, Bonnett GD (2007) Changes in the sugarcane metabolome with stem development. Are they related to sucrose accumulation? Plant Cell Physiol 48(4):573–584
Goldemberg J (2007) Ethanol for a sustainable energy future. Science 315:808–810
Gong L, Deng Z (2010) EST-SSR markers for gerbera (*Gerbera hybrida*). Mol Breed 26:125–132
Gupta V, Raghuvanshi S, Gupta A et al (2010) The water-deficit stress- and red-rot-related genes in sugarcane. Funct Integr Genomics 10:207–214
Heffner EL, Jannink JL, Sorrells ME (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome 4:65–75
Heller RA, Schena M, Chai A et al (1997) Discovery and analysis of inflammatory disease-related genes using cDNA microarrary. Proc Natl Acad Sci U S A 94:2150–2155
Huttenhofer A, Vogel J (2006) Experimental approaches to identify non-coding RNAs. Nucleic Acid Res 34(2):635–646
Jing Q, Huang S, Guth S et al (2005) Involvement of microRNA in AU-rich element-mediated mRNA instability. Cell 120:623–634
Kasuga M, Miura S, Shinozaki K, Yamaguchi-Shinozaki K (2004) A combination of the *Arabidopsis* DREB1A gene and stress-inducible *rd29A* promoter improved drought- and low-temperature stress tolerance in tobacco by gene transfer. Plant Cell Physiol 45(3):346–350
Ken NP (2010) The functional genomics of inbreeding depression: a new approach to an old problem. Bio Sci 60(4):267–277
Kurowska M, Daszkowska-Golec A, Gruszka D et al (2011) Tilling: a shortcut in functional genomics. J Appl Genet 52:371–390
Langridge P, Fleury D (2011) Making the most of 'omics' for crop breeding. Trends Biotechnol 29:33–40
Leibbrandt NB, Snyman SJ (2003) Stability of gene expression and agronomic performance of a transgenic herbicide resistant sugarcane line in South Africa. Crop Sci 43(2):671–677
Lembke CG, Nishiyama MY Jr, Sato PM et al (2012) Identification of sense and antisense transcripts regulated by drought in sugarcane. Plant Mol Biol 79:461–477
Lopez C, Jorge V, Piegu B et al (2004) A unigene catalogue of 5700 expressed genes in cassava. Plant Mol Biol 56:541–554
Lu YH, D'Hont A, Paulet F et al (1994) Molecular diversity and genome structure in modern sugarcane varieties. Euphytica 78:217–226
Lu Y, Savage LJ, Larson MD et al (2011) Chloroplast: a database for large-scale phenotypic screening of *Arabidopsis* mutants. Plant Physiol 155:1589–1600
Ma J, Morrow DJ, Fernandes J, Walbot V (2006) Comparative profiling of the sense and antisense transcriptome of maize lines. Genome Biol 7(3):R22
Manners JM, Casu RE (2011) Transcriptome analysis and functional genomics of sugarcane. Trop Plant Biol 4(1):9–21

Marconi TG, Costa EA, Miranda HR et al (2011) Functional markers for gene mapping and genetic diversity studies in sugarcane. BMC Res Notes 4:264. doi:10.1186/1756-0500-4-264
Menossi M, Silva-Filho MC, Vincentz M et al (2008) Sugarcane functional genomics: gene discovery for agronomic trait development. Int J Plant Genom 458732. doi:10.1155/2008/458732
Mir RR, Zaman-Allah M, Sreenivasulu N et al (2012) Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops. Theor Appl Genet 125:625–645
Mittler R, Shulaev V (2013) Functional genomics, challenges and perspectives for the future. Physiol Plant 148:317–321
Murchie EH, Niyogi KK (2011) Manipulation of photoprotection to improve plant photosynthesis. Plant Physiol 155:86–92
NCBI (2014) http://www.ncbi.nlm.nih.gov/nucest/?term=sugarcane
NIFA-NSF PW Report (2011) Phenomics: genotype to phenotype, a report of the phenomics workshop sponsored by the USDA and NSF
Nogueira FTS, Rosa VE Jr, Menossi M et al (2003) Expression profiles and data mining of sugarcane response to low temperature. Plant Physiol 132:181–1824
Osabe M, Sugatani J, Takemura A, Kurosawa M, Yamazaki Y, Ikari A, Miwa M (2009) Up-regulation of CAR expression through Elk-1 in HepG2 and SW480 cells by serum starvation stress. FEBS Lett 583:885–889
Papini-Terzi FS, Rocha FR, Vêncio RZ et al (2005) Transcription profiling of signal transduction-related genes in sugarcane tissues. DNA Res 12(1):27–38
Parida SK, Rajkumar KA, Dalal V et al (2006) Unigene derived microsatellite markers for the cereal genomes. Theor Appl Genet 112:808–817
Parida SK, Kalia SK, Kaul S et al (2009) Informative genomic microsatellite markers for efficient genotyping applications in sugarcane. Theor Appl Genet 118:327–338
Que YX, Yang ZX, Xu LP, Chen RK (2009) Isolation and identification of differentially expressed genes in sugarcane infected by *Ustilago scitaminea*. Acta Agron Sin 35(3):452–458
Que Y, Xu L, Lin J et al (2011) Differential protein expression in sugarcane during sugarcane-Sporisorium *scitamineum* interaction revealed by 2-DE and MALDI-TOF-TOF/MS. Compar Funct Genomics Article ID 989016. doi:org/10.1155/2011/989016
Reddy RN, Madhusudhana R, Mohan SM et al (2011) Characterization, development and mapping of unigene-derived microsatellite markers in sorghum [*Sorghum bicolor* (L.) Moench]. Mol Breed. doi:10.1007/s11032-011-9571-0
Riera M, Valon C, Fenzi F et al (2005) The genetics of adaptive responses to drought stress: abscisic acid dependent and abscisic acid-independent signalling components. Physiol Plant 123(2):111–119
Roach BT (1972) Nobilisation of sugarcane. Proc Int Soc Sugar Cane Technol 14:206–216
Schlog PS, Nogueira FTS, Drummond R et al (2008) Identification of new ABA- and MEJA-activated sugarcane bZIP genes by data mining in the SUCEST database. Plant Cell Rep 27:335–345
Sharma MK, Mishra SK (2014) Applications of metabolomics in plant biology. Agrotech 2:4
Singh RK, Singh SB (2004) Breeding Strategies for commercially elite early maturing varieties of sugarcane (*Saccharum* species complex). Sugar Tech 6(1&2):89–92
Singh RK, Singh RB, Singh SP, Sharma ML (2011) Identification of sugarcane microsatellites associated to sugar content in sugarcane and transferability to other cereal genomes. Euphytica 182:335–354. doi:10.1007/s10681-011-0484-0
Singh RK, Singh SP, Tiwari DK et al (2013) Genetic mapping and QTL analysis for sugar yield-related traits in sugarcane. Euphytica 191:333–353. doi:10.1007/s10681-012-0841-7
Sreenivasan TV, Ahloowalia BS, Heinz DJ (1987) Cytogenetics. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier, Amsterdam, pp 211–254
Storz G (2002) An expanding universe of noncoding RNAs. Science 296(5571):1260–1263
Sugihario B, Ermawati N, Mori N et al (2002) Identification and characterization of a gene encoding drought-inducible protein localized in the bundle sheath cell of sugarcane. Plant Cell Physiol 43:350–354

Terzi FSP, Flávia R, Rocha FR et al (2009) Sugarcane genes associated with sucrose content. BMC Genome 10:120. doi:10.1186/1471-2164-10-120

Tiessen A, Hendriks JHM, Stitt M et al (2002) Starch synthesis in potato tubers is regulated by post-translational redox modification of ADP-glucose pyrophosphorylase: a novel regulatory mechanism linking starch synthesis to the sucrose supply. Plant Cell 14:2191–2213

Tran LS, Mochida K (2010) Functional genomics of soybean for improvement of productivity in adverse conditions. Funct Integr Genomics 10:447–462. doi:10.1007/s10142-010-0178-z

Varshney RK, Garner A, Sorells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Vettore AL, da Silva FR, Kemper EL et al (2003) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res 13(12):2725–2735

Walthall CL, Hatfield J, Backlund P et al (2012) Climate change and agriculture in the United States: effects and adaptation, USDA Technical Bulletin. www.usda.gov/oce/climate_change/effects_2012/effects_agriculture.htm

Weckwerth W, Morgenthal K (2005) Abiotic stress in plants and metabolic responses agricultural and biological sciences. Drug Discov Today 10(22):1551–1558

Wei W, Qi X, Wang L et al (2011) Characterization of the Sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. BMC Genom 12:451. doi:10.1186/1471-2164-12-451

Wu Q, Xu L, Guo J et al (2013) Transcriptome profile analysis of sugarcane responses to *Sporisorium scitaminea* infection using solexa sequencing technology. Biomed Res Int. doi.org/10.1155/2013/298920

Xiang C, Brownstein M (2003) Fabrication of cDNA microarrays. In: Brownstein MJ, Khodursky A (eds) Methods in molecular biology: vol. 224: functional genomics: methods and protocols. Humana Press, Totowa, pp 1–7

Yamamoto M, Wakatsuki T, Hada A, Ryo A (2001) Use of serial analysis of gene expression (SAGE) technology. J Immunol Methods 250(1–2):45–66

You-Xiong Q, Jian-Wei L, Xian-Xian S et al (2011) Differential gene expression in sugarcane in response to challenge by fungal pathogen *Ustilago scitaminea* revealed by cDNA-AFLP. J Biomed Biotechnol. doi.org/10.1155/2011/160934

Zenoni S, Ferrarini A, Giacomelli E (2010) Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. Plant Physiol 152(4):1787–1795

Zhang BH, Pan XP, Wang QL et al (2005) Identification and characterization of new plant microRNAs using EST analysis. Cell Res 15(5):336–360

Chapter 18
Potentials of Proteomics in Crop Breeding

Aayudh Das, Bimal Paudel, and Jai S. Rohila

Abstract Recent incorporation of new molecular technologies, for example genomics and molecular markers, to the conventional plant breeding methods has resulted in a paradigm shift in crop improvement strategies. Proteomics is a comparatively new tool which is gradually becoming essential for the new generation of plant breeders because it is capable of providing new insights at cellular levels. Proteome, a translational version of a genome, is vital to expose molecular mechanisms essential for plant growth, development, and their interactions with physical and biological environment. A specific advantage of proteomics over all other -omics methods in crop breeding advancement is its ability to consider post-translational modifications that reflect the functional impressions of protein modifications on crop plant productivity. In this chapter we discuss various proteomic techniques, with examples of their applications for the advancement of conventional crop breeding programs. Proteomic methods can be used to measure subtle changes in protein expression levels in response to selective breeding, and for biotic and abiotic stress tolerance studies among different germplasm or cultivars. We demonstrate that the prospects of possible inclusion of proteomic techniques will lead to stronger crop breeding programs and a solid food security in the coming years.

Keywords Abiotic • Biotic • Genomics • Plant breeding • Proteomics

18.1 Introduction

There are almost one billion people in the world who are severely malnourished, (FAO 2014; Salekdeh and Komatsu 2007) and in next 40 years, this number is expected to increase by two billion people (Pretty 2000). To provide food for this hungry world, actions should be directed towards increased food production in a way that utilizes comparatively less resources. To reduce the number of people at risk of hunger, efforts to augment agricultural productivity will be a key element

A. Das • B. Paudel • J.S. Rohila (✉)
Department of Biology & Microbiology, College of Agriculture and Biological Sciences, South Dakota State University, 252, Biostress Building, 57007 Brookings, SD, USA
e-mail: aayudhdas@gmail.com; bimal.paudel@sdstate.edu; jai.rohila@sdstate.edu

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_18

(Tester and Langridge 2010). This effort requires an integrated methodology in which infrastructure improvement, resource management and crop breeding, are essential elements. Among these elements, crop breeding can subsidize this worldwide food scarcity because it is an economical tool for increasing the yields and nutritional value of crops, and is as an essential tool for shaping future agriculture; however, at the same time some modifications to traditional approaches are necessary to perform at a more advanced level (Bänziger 2000).

Traditional crop breeding practices rely on homologous recombination and use inter-breeding or crossing of closely or distantly-related crops to introduce new traits (genes) from one variety of plant to produce new plant varieties with desirable characteristics. Progeny plants from this inter-breeding are then backcrossed with the parental plants to maintain the desired genetic pool (Duvick 1996). The essential purposes of a plant breeder are to increase (1) the amount and quality of the yield, (2) the tolerance for abiotic stress (e.g. heat, salinity, drought) (Mittler 2006) and (3) the resistance against various diseases.

Among many, one crucial task that can help crop breeding programs is identification of the novel genes that are responsible for desirable traits with a potential for utilization in commercial productions of field crops such as rice, wheat, maize, etc. Numerous completed genome sequences (e.g. *Arabidopsis*, rice, soybeans) provide access to a greater number of genes to plant breeders (Eldakak et al. 2013; Kim et al. 2014). Fortuitously, challenges also ascend at a time when plant breeders are observing significant advancements in understanding fundamental mechanisms involved in plant growth and development (Delmer 2005). The vital tools for enhancing molecular breeding include physical gene mapping on the chromosomes, genetic mapping, gene sequencing, microarrays, marker-aided or -assisted selection (MAS), and validating candidate genes and most importantly proteomics for high throughput analysis of gene expression. Whereas genomic studies are examining the blueprints theoretically, proteomic studies are focused on studying the functional players; moreover, post-translational modifications (PTMs) reveal the functional aspects for facilitating the definite biological functions (Kav et al. 2007). Proteomic studies together with functional genomics knowledge permit an understanding of cellular processes and also aid plant breeders in overcoming certain limitations of conventional crop breeding. New methods for construction and screening of breeding populations ultimately increase the productivity of plant selection and accelerate the rates of genetic gain of the trait of interest (Langridge and Fleury 2011).

As the central dogma states that genes are eventually expressed as proteins; the application of protein technologies to identify new genes is prevailing. Proteomic investigations of a crop of interest offer a new alternative for the determination of the genes that are crucial for crop breeding advancements. Over the last two decades, technical progress in the extraction, separation, quantification, and identification of plant proteins have made the high-throughput analysis of crop proteins feasible, and the reproducibility of these technologies has reduced errors in proteomic assays (Chen and Harmon 2006).

Advancements in genomics and proteomic technologies over the last two decades have been beneficial for the success of crop breeding programs and will help in making substantial progress towards variety development in the near future (Ahmed et al. 2013).

18.2 Advancement of Conventional Crop Breeding: The Link between Genomics and Proteomics

Phenotypic selection is the primary tool of traditional crop breeding to improve genetic traits. Although phenotypic selection programs have several limitations; they are most predominantly focused on genes with major effects that produce a clear phenotype (Langridge and Fleury 2011). In the case of additive gene effects, large populations can be used to provide a functional possibility of combining required alleles at multiple loci into a single line (Gamble 1962; Langridge and Fleury 2011). However, when the traits are complex with no prior genetic understanding, breeders consider the germplasm as a population and try to alter the mean population reaction in a specific direction for the desirable trait. With intensively bred species, a plant breeder has to evaluate millions of lines each year to maintain acceptable rates of genetic improvement (Richards et al. 1993). To solve these difficulties, the introduction of molecular markers has provided an opening for genotypic selection in which the genetic position of main loci is known. As marker densities increase the use of markers to monitor the genetic background of lines becomes more effective and this advancement has allowed the prospect for accelerated backcrossing (Langridge and Fleury 2011). Since crop improvement is based on genetic variation, genomics can provide the ability to discover novel genes and pathways that affect traits and provide platforms to practice effective selection strategies (Bouchez and Höfte 1998).

Consistency is an important advantage gained by using modern genomic tools as the phenotyping assay to detect the gene or quantitative trait locus (QTL). This consistency is possible because of genetic linkage analysis based on the principle of genetic recombination during meiosis (Tanksley and Nelson 1996). Genetic background determination is essential for large populations that represent many recombination cycles, and these markers are used to detect recombinations in the region of the target QTL. In this analysis, the recombinants are first phenotyped to narrow down the QTL by resolving a portion of the locus that identifies new markers. This process continues, and the populations are then screened by high-throughput genotyping technologies (Langridge and Fleury 2011). Genotyping by sequencing the QTL region can offer a prompt means to construct high-resolution maps in a single cycle. However, because whole-genome sequencing of enormous and repetitive genomes remains expensive, shotgun sequencing should be the alternative strategy to resolve this problem (Collard and Mackill 2008). A quick method to discover expressed genes is by determining partial sequences of cDNAs using a method

known as *expressed sequenced tags* (ESTs), in which thousands of sequences are determined with limited investment and in a shorter time interval. Compared with the speculative molecular mass prediction from genomic sequences, comprehensive organism-definite protein databases allow for the attribution of a sequence to its protein fragments (Bouchez and Höfte 1998).

The identification of QTLs using DNA markers was a key invention in the representation of quantitative traits, but we need the assistance of proteomic technologies to identify novel gene for plant breeders (Collard and Mackill 2008). MAS with the use of these DNA markers in crop breeding programs became essential for the selection of desired genes/QTLs for the development of superior breeding lines (Collard and Mackill 2008; Eldakak et al. 2013). In 1994, French researchers invented an approach that combined proteomics and MAS components (Damerval et al. 1994). These researchers identified protein quantity loci (PQL) that elucidated approximate spot intensity variations. This type of approach is exclusively beneficial for crop breeding because through rigorous breeding selection, lines can be obtainable with differing phenotypic degrees that can complete the correlation between responsive genes and observed stress tolerance phenotypes. This correlation can be validated by analyzing progressive mapping populations such as near isogenic lines (NILs), recombinant inbred lines (RILs) and double haploid (DH) lines (Salekdeh and Komatsu 2007). In this way, plant breeders should be able to integrate the selected genes in marker-assisted breeding programs to improve the genetic trait under investigation for the desirable outcome (Salekdeh and Komatsu 2007).

To monitor mRNA levels hybridization techniques i.e. a reverse northern technique is commonly used, in which DNA fragments corresponding to different cDNAs (or genes) are immobilized onto a solid substance and then hybridized with probes (that are prepared from total mRNA pools and converted to cDNA). The advantage of this approach is that it allows a large number of gene fragments to be analyzed within a short interval (Bouchez and Höfte 1998; DeRisi et al. 1997). A more advanced method to monitor mRNA levels is the DNA microarray that uses fluorescent probes, and hybridization signals are perceived using a scanner (Schena et al. 1995). However, information on mRNA levels is not sufficient to reach a complete picture of gene expression. Protein expression data are more informative, but are much more difficult to acquire in a parallel fashion. The proteome represents the total protein complement expressed by the genome. Thus, proteomic approaches are essential to investigate the protein abundance and PTM of several thousands of proteins in a parallel fashion (Humphery-Smith et al. 1997).

There are many potential advantages of proteomics over genomics. Because genes are transcribed into mRNA with the use of alternative splicing, there is no one-to-one association between the genome and the transcriptome. Transcripts are translated into proteins and undergo PTMs. Consequently, one gene can produce several different protein isoforms and protein stoichiometry are influenced by environmental factors, proteins interact with other proteins, and proteins may undergo degradations. Because the stoichiometry of the mature proteins ultimately determines its function in the cell, high-throughput proteomic screening methods for

detecting deviations in protein expression may be more appropriate for the identification of biomarkers, rather than the use of genomic technologies (Sabel et al. 2011).

Proteomics supplements many functional genomics approaches that include microarray-based expression profiles and systematic phenotypic profiles at the organism level. Assimilation of these data with bioinformatics approaches will reveal the functional aspects of the genes that will eventually reflect the protein properties and functions that will be useful for researchers and plant breeders to comprehensively understand the genetic trait of interest (Tyers and Mann 2003).

18.3 Proteomic Approaches for Advancements in Crop Breeding

Proteomics is becoming progressively more important for several investigations, including signaling and biochemical pathways in plants. The studies on protein levels are becoming vital in exposing molecular mechanisms that are essential for plant growth, development and environmental interactions. Present proteomic technologies have provided enormous opportunities for high-throughput proteomic studies followed by protein identification to analyze various functional aspects, predominantly quantification, protein-protein interactions, subcellular localization and PTMs (Chen and Harmon 2006; Patterson and Aebersold 2003; Rohila et al. 2004, 2006, 2009; Roy et al. 2011). The plant proteome is complex and dynamic in nature, so a single diagnostic technique for protein studies will not represent comprehensive analysis of the entire proteome. To enhance proteome resolution, coverage and understanding researchers use multiple techniques for complementary results (Whitelegge 2004).

Over the last two decades, two-dimensional gel electrophoresis (2-DE) has become the most predominant technique for protein separation (Komatsu et al. 2012). Liquid chromatography was a common practice in many laboratories, but with a standard 2-DE analysis, protein modification and degradation can be more clearly visualized. Apart from 2-DE, many other technologies have also evolved for protein separation and protein profiling. Crop proteomics have been limited by inadequate accessibility of genomic data, but with the effective progress of next-generation sequencing (NGS) platforms, the identification of proteins and their isoforms of a particular plant species has become much more achievable (Komatsu et al. 2013a).

A specific advantage of proteomics over all other -omics methods in crop breeding advancement is its ability to investigate PTMs that reflect the functional impressions of protein modifications on crop plant productivities. The application of proteomics to investigate function in plants will benefit from further advancement in plant phenotyping. Advanced, automated proteomic techniques will improve the selection of suitable genotypes by proteome-based analyses that can characterize important traits in crop breeding programs (Komatsu et al. 2013a). A few of the popular proteomic techniques are briefly discussed below.

18.3.1 Two-Dimensional Gel Electrophoresis (2-DE)

Two-dimensional gel electrophoresis (2-DE) with immobilized pH gradients (IPGs) together with protein detection by mass spectrometry (MS) is presently a method of choice for many in proteomics field. This 2-DE is the most classical method for investigating the active changes of plant proteome patterns in response to environmental stimuli such as heat, cold and drought (Roy et al. 2011). It has become the most common tool for profiling the complex proteomes of various plant species. Isoelectric focusing coupled with molecular weight basis separation (SDS-PAGE) of proteins is usually referred to as 2-DE. Protein separation based on these two unrelated properties can produce notable, composite maps of proteins that can specifically detect deviations in protein expression level, isoforms and post-translational modifications (Görg et al. 2004; Thelen and Peck 2007)).

The 2-DE separates proteins in two independent dimensions: isoelectric point (pI) in the first dimension and molecular mass (M_r) in the second dimension (Fig. 18.1). These separated proteins are then analyzed as spots on the gel using one of the several staining methods such as, anionic dye (Coomassie brilliant blue), negative staining with metal cations (zinc imidazole), silver-staining (silver nitrate), fluorescence staining or labeling, radioactive isotopes, autoradiography, fluorography or phosphor-imaging (Görg et al. 2004; Westermeier and Marouga 2005).

Modern day 2-DE technology with IPG strips has overwhelmed the prior limits of carrier ampholyte-based 2-DE; new technologies have progressed in terms of reproducibility, handling ease, resolution and separation of optimum basic or acidic proteins (Boguth et al. 2000; O'Farrell et al. 1977). Proteomes with extreme alkaline proteins can be analyzed with the help of modified 2-DE with most recently developed IPGs between pH 2.5–12 (Gorg et al. 2009). Narrow-overlapping IPGs offer greater resolution to detect less abundant proteins. With the improved 2-DE technology we can resolve thousands and thousands of proteins, as lowly expressed as 1 ng of protein per spot (Görg et al. 2004, 2009).

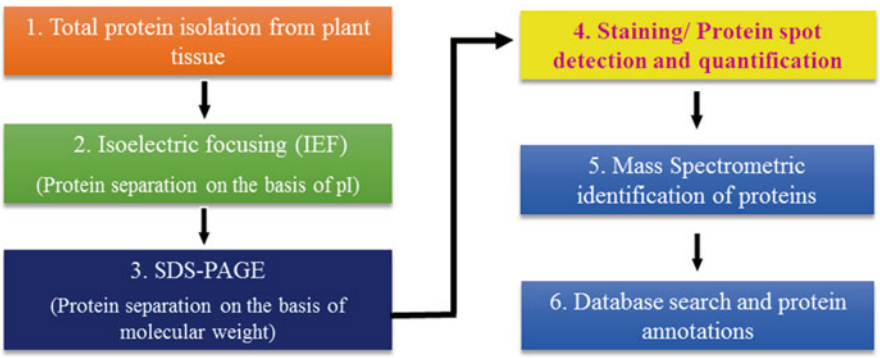

Fig. 18.1 A general 2-DE approach for developing protein profile databases

18.3.2 2D-Difference in Gel Electrophoresis (2D-DIGE)

Comparing different proteomes by classical 2D electrophoresis is perplexing and complicated because of extensive gel-to-gel variation (Van den Bergh and Arckens 2004). Two-dimensional difference gel electrophoresis (2D-DIGE) is the improved form of 2-DE that permits researchers to relate the same protein on two or more gels (Viswanathan et al. 2006). The 2D-DIGE technique robustly overcomes the high experimental variations of 2-DE and increases the reproducibility of protein profiling (Casasoli et al. 2008; Robbins et al. 2013; Ünlü et al. 1997) 2D-DIGE uses a differential staining technique in which each protein sample is covalently labeled with different fluorophores (i.e. Cy2, Cy3, Cy5) at a lysine residue (Viswanathan et al. 2006). The protein samples are then mixed and separated on the same gel, permitting high-sensitivity identification with a suitable imaging system that reveals that the proteins that are common to all samples appear as *spots* with a constant ratio of fluorescent signals, but proteins that vary between the samples (differential expression at protein levels) have different fluorescence ratios (Van den Bergh and Arckens 2004; Viswanathan et al. 2006). The 2D-DIGE method can be used to measure subtle changes in protein expression levels in response to selective breeding, biotic and abiotic stress, and genetic manipulation to analyze natural variation among different cultivars, and to compare the expression levels of allergen proteins (Keeler et al. 2007; Teshima et al. 2010).

18.3.3 Mass Spectrometry

Mass-spectrometry is used for the large-scale analysis of proteins and has been performed routinely in many plant biology laboratories (Schulze and Usadel 2010). For precise quantitation using ion intensities and mass-to-charge ratios (m/z), comparisons between different samples containing large numbers of unknown proteins can only be achieved with LC-MS/MS experiments (Hu et al. 2005). The major achievements in this field started with the development of soft protein ionization methods, such as electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) (Schulze and Usadel 2010). MALDI is normally used to analyze relatively simple peptide mixtures, and ESI-MS system is favored for the analysis of complex peptide mixtures (Aebersold and Mann 2003). Statistical treatment of the data is necessary and, despite the complexity of the data, many tools are available for data normalization and detection of differential expression (Hu et al. 2005).

Proteins usually employ functions by interacting with other proteins under tight regulation. To investigate the members of the protein complex that play vital roles in their interactions, MS is the most convenient method for their identifications after pulling-down the whole complex via tandem affinity purification (TAP), because in TAP one of the proteins itself is used as an affinity reagent to find its binding partners (Rohila and Fromm 2008; Rubio et al. 2005). Because many biologically-relevant

interactions are low affinity in nature, MS-based methods are useful to detect a subset of the protein interactions (Aebersold and Mann 2003). By studying the major functions of groups of proteins that may control specific traits of a cultivar, MS-based detection aids researchers and plant breeders to identify the fate of particular proteins that may employ specific functions.

18.3.4 Isotope-Coded Affinity Tagging (ICAT)

Although classical gel-based proteomic methods are the most frequently used techniques for detecting protein expression level, gel-free alternatives, such as isotope-coded affinity tagging (ICAT), have been established to quantitatively compare proteomes; ICAT is primarily used to circumvent the difficulty of contamination in assigning proteins to subcellular organelles (Dunkley et al. 2004). Because ICAT is not based on SDS-PAGE and relies on MS for protein quantification, it can be used to comparatively more efficiently compare membrane proteomes (Rombouts et al. 2013).

This approach is based on the ability of free cysteine thiols to be labeled by the iodoacetamide-based ICAT, and the two labeled protein samples are then combined and proteolyzed to peptides with trypsin digestion (Sethuraman et al. 2004). Afterwards, the labeled peptides are fractionated by cation exchange chromatography and then refined by avidin affinity chromatography. Ultimately, mass spectrometry can be used to detect the labeled free thiols (Gygi et al. 1999; Shiio and Aebersold 2006). ICAT provides a major contribution for identifying unknown proteins from a mixture of proteins and will be particularly useful in cases of crop breeding in which researchers and breeders are searching for desired proteins controlling a major biological function in a cultivar.

18.3.5 Isobaric Tagging for Relative and Absolute Quantification (iTRAQ)

To understand the complex biological systems of plants and determine the functional relationships between plant-based proteins and protein-protein interactions, proteomic techniques are vital to characterize proteins from complex protein mixtures. This type of proteomic analysis requires some type of differential comparison tool to compare a given experimental protein in a test sample to some type of a control sample. Isobaric tags for relative and absolute quantitation (iTRAQ) is a gel-free technique that uses a set of isobaric tags which are amine specific, to quantify proteins from diverse sources in a single test. The amine specificity of these isobaric tags confirms that there will be no loss of information from samples involving post-translational modifications. iTRAQ reagents permit LC-MS or MS-MS experimental systems that enhance statistical validation in proteome profiles. With iTRAQ, researchers and plant breeders can improve the temporal analysis of protein expression and can elucidate various markers, e.g. for drought and diseases, that can

be useful to make appropriately genetically modified crops or by crop breeding techniques (Zieske 2006).

18.3.6 Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC)

SILAC is a potent technique for relative and quantitative proteomics, and has recently been applied to many different plant protein analyses (Lewandowska et al. 2013). A major advantage of SILAC is that it explains large-scale kinetics of proteomes and directly aids to identifying important proteins in signaling pathways that control biological functions (Mann 2006).

Efficient labelling of entire *Arabidopsis* seedlings by SILAC method opens new prospects to exploit the genetic assets of *Arabidopsis* and quantitatively analyze the impact of the mutations in vivo (Lewandowska et al. 2013). A protein research group from Denmark demonstrated that metabolic labeling by amino acid [$^{13}C_6$] arginine made it possible to explaining comparative study of protein expression in *Arabidopsis* and adjustment-dependent interaction screen using affinity pull-downs (Gruhler et al. 2005).

18.3.7 Multidimensional Protein Identification Technology (MudPIT)

For large-scale protein identification, the representative shotgun proteomic method is the multidimensional protein identification technology (MudPIT) (Lee et al. 2011). The work flow of MudPIT is: tryptic digestion of complex proteins, separation of digested proteins by liquid chromatography and loading eluents onto a mass spectrometer for protein profiling (Bhalerao et al. 1999). MudPIT is a potential technique for surveying protein expression patterns.

One of the major achievements in crop breeding in recent years was the reduction in crop plant heights resulting in the Green Revolution. After this achievement, reducing the tiller numbers has been an imperative focus in crop breeding studies (Khush 2000). Current molecular breeding studies have aimed to reveal the mechanisms responsible for controlling tiller numbers using the MudPIT technology (Lee et al. 2011).

18.3.8 Proteome Mapping and Proteome Database

At present, the proteome databases are in a developing stage from which proteome research can address specific biological queries. The information from proteome databases enables the identification of alterations in protein expression level under diverse growth conditions (Salekdeh and Komatsu 2007). Available databases are

present for plant models such as *Arabidopsis* and significant crop species such as rice or maize. These databases include comparisons of polypeptide patterns in various plant organs or tissues that aid in a better understanding the genome expression (Thiellement et al. 1999).

One exceptional benefit of proteomic investigations is that they permit the analysis of genome expression at the subcellular level compared with other genome-wide approaches, as demonstrated previously (Masson and Rossignol 1995). In the case of *Arabidopsis*, a related approach directed with the help of databases revealed that many membrane peripheral proteins are present in various subcellular locations (Santoni et al. 1998). An additional significant feature of annotated protein databases is that they link with genomic data (Thiellement et al. 1999). In the case of the *Arabidopsis* plasma membrane database, numerous unknown proteins were detected with the help of ESTs, reflecting vital information on the subcellular expression of the encoded genes. A similar type of application applied with 2-D gel data compared the accumulation of transcripts and their corresponding proteins in *Arabidopsis* (Thiellement et al. 1999).

18.4 Application of Proteomics in Molecular Crop Breeding: Today We Are One Step Closer than Yesterday…But Still Many Miles to Go

The improvement of crop breeding is predominantly focused on the betterment of overall crop quality, reduction of yield loss, improved resistance against diseases and, most importantly, protection from various biotic and abiotic stresses. The application of proteomics to appropriately assess crop plants, particularly in crop breeding, has rapidly improved within the last decade. Proteomics is more advantageous than other -omics techniques because a fraction of the plant proteome provides a complete representation of cellular proteins (Komatsu et al. 2013a). A quick understanding of different phases where the proteomics and crop improvement programs blend together can be made from Fig. 18.2.

18.4.1 Improvement of Crop Yields and Quality

Proteomics has identified numerous proteins that play vital roles in plant growth and development (Gong and Wang 2013). From an agricultural perspective, seed viability is one of the important factors related to crop yields. Proteomic analyses of seed protein content and the roles of enzymes in the starch biosynthesis pathway aids plant breeders and plant biologists to analyze genes responsible for seed quality and predictive hypotheses. For example, applied proteomics to investigate the regulation of rice seed germination revealed the detailed mechanism of starch

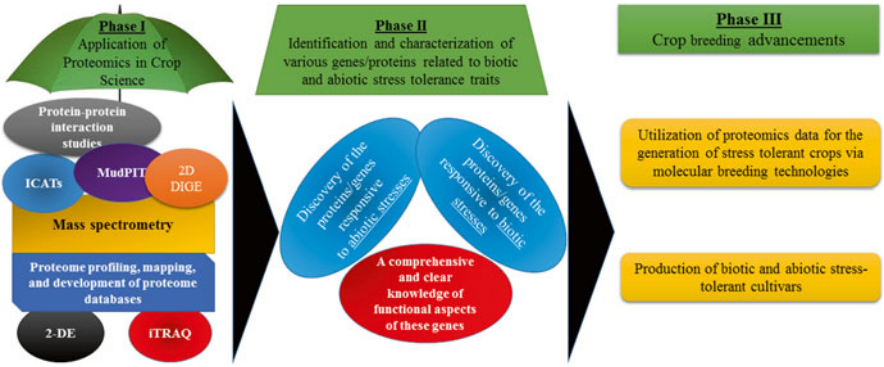

Fig. 18.2 A proposed model for proteomics and its key role in crop improvement programs

degradation in endosperm and starch biosynthesis in the embryo during germination (Komatsu et al. 2013a; Wang et al. 2013).

In crop breeding, heterosis has so far been the most popular mechanism and has been used widely in crop improvement programs in which a sterile male line is required for hybrid breeding (Wang et al. 2013). The application of crop proteomic studies identified the proteins involved in the regulation of male sterility, may become a major breakthrough that could benefit plant breeders. The application of transgenic techniques is becoming an essential method to obtain germplasm with desired gene expression compared to the traditional breeding, but evaluation of these genetically-modified crops require various proteomic studies to verify protein expressions at desired levels (Gong et al. 2012). With the recent advancements in proteomic technologies, plant breeders and researchers can now solve the challenges faced during the creation of male sterile lines, and heterosis is being re-investigated with proteomics incorporated, systems-level approaches to understand fundamental molecular functions. A vibrant concept of differential expression of a variety of genes and a brief overview of metabolic pathways with the application of proteomics manifested the difficulties of heterosis Baranwal et al. (2012).

18.4.2 Combating Abiotic Stresses

Plants are repeatedly forced to recognize and respond to adverse environmental changes to avoid unfavorable effects on their growth and development (Barkla et al. 2013b). Understanding the mechanisms by which crop plants exhibit resistance to abiotic stress is of significant interest for plant breeders to create agriculture breeding tactics to safeguard sustainable efficiencies in crop production. Abiotic stress includes deviations in temperature, osmotic balance, light, water supply and nutrient levels that negatively influence plant performance. Currently, the major concern of plant breeders in agriculture is the stress-related alterations in plant development,

because they ultimately restrict the yield and can result in enormous economic losses (Barkla et al. 2013b). To meet the demand for increased crop production, improving plant fitness over a wider range of adverse environmental conditions is very useful to plant breeders (Gisladottir and Stocking 2005). Proteomics allows the proper analysis of stressed plants at the protein level. Widespread, advanced proteomic technologies, such as 2D-DIGE, MudPIT, iTRAQ, and label free quantification MS, has increased the pace of discovery of low abundant proteins (e.g. transcriptional factors, kinases, channels and transporters) and novel regulatory mechanisms in salt stress signaling and metabolic pathways (Zhang et al. 2012).

18.4.2.1 Heat (High Temperature) Stress

High temperature (HT) is a major abiotic stress limiting plant growth, metabolism and crop productivity worldwide (Hasanuzzaman et al. 2013). According to the International Panel on Climate Change, the increased temperatures (2–4 °C) of the past two decades have triggered a loss of approximately five billion mt of major food crops, such as rice, wheat, maize and soybeans (Peng et al. 2004). Heat stress increases membrane damage and impairs metabolic functions. Plants mainly activate some defense systems upon sensing heat stress to survive under those unfavorable conditions, and one of the key factors for heat tolerance is the generation of heat shock proteins (HSPs) (Hashiguchi et al. 2010). A plant breeder needs to develop a proper protection strategy for the survival of the crop plant under such heat stress conditions. Heat stress tolerance mechanisms of plants are complex and controlled by multiple genes, proteins, ion transporters, osmo-protectants, antioxidants, and other factors involving numerous physiological and biochemical modifications in the cell, such as adjustments in the plasma membrane structure and function, protein composition and primary and secondary metabolites (Huang and Xu 2008).

Because proteins are directly involved in heat stress response, proteomic analyses can considerably contribute to unravel the likely interactions between protein abundances and plant stress acclimations (Kosová et al. 2011). Comprehensive proteomic profiling initiatives are beneficial techniques for increasing the understanding of heat stress responses and improving the breeding of heat-tolerant crops to confront the global temperature escalation (Zou et al. 2011). The identification of protein responsive to heat stress is an essential first step toward understanding the molecular mechanisms, to eventually produce heat-tolerant cultivars with enhanced resistance to heat stress (Kang et al. 2010). Recently, researchers identified a total of 48 heat shock proteins (HSPs), those play many vital roles in energy and carbohydrate metabolism and redox homeostasis, while investigating the rice leaf proteome in response to high temperature of 42 °C by using 2-DE coupled with MALDI-TOF analysis (Lee et al. 2007). Their research suggests that HSPs may play a central role in preventing cellular damage under preeminent heat stress in addition to the disruption of major metabolic pathways upon elevated temperature. Another proteomic study comparing various wheat cultivars with different heat tolerance capacities revealed that 16 kDa heat shock protein (HSPs) and transcription factors upstream

of these HSPs had crucial roles in providing thermo-tolerance to the crop (Majoul et al. 2004). Proteomic analysis of HSP proteins in tomato revealed that inhibiting the function of HSP100 by incorporating an antisense construct resulted in poor survival of those plants under heat stress conditions (Yang et al. 2009).

Refining the grain biology of cereal crops has been a major interest, for example, in rice breeding understanding the mechanism of rice kernel development has a special objective. In plant breeding, improving crop yields and plant survival are equally important, keeping pace with these two aspects at the same time is difficult for plant breeders. Because cereals such as wheat and rice plants are most sensitive during their flowering periods more specific care is required for heat tolerance analysis (Satake and Yoshida 1978). Proteomic analyses have been performed for protein content analyses of rice grain and caryopsis development of rice under severe heat stress, and researchers observed that there was a sustainable relationship between elevated temperature and grain quality (Lin et al. 2005). Heat stress responses in rice included both protein up- and down-regulations, so biologically there is both an increase and decrease in the abundance of some proteins. Subsequently, because a large numbers of proteins were down-regulated during heat response; these down-regulated proteins may have significant effects on rice breeding for heat-tolerance.

18.4.2.2 Salinity Stress

Salinity stress is a major impediment in crop breeding and one of the chief types of abiotic stress in plant agriculture that negatively influences crop yield worldwide (Yokoi et al. 2002). Because salinity stress is an obstacle to achieve required food productions, researchers are concentrating their research on the development of salt tolerant plants to improve crops (Zhu 2001). Crop breeding is a classical method to produce salt tolerant lines that has been used for a long time, but with little success because of the multi-genic nature of salt resistance and also the occurrence of low genetic variation available in major crop plants (Turan et al. 2012). A proteomic study provided plant breeders new insights into salt stress response in plant roots and accurately demonstrated the functional aspects of individual proteins responsible for the salt-tolerant response. The incorporation of proteomics results with outcomes from other -omics and bioinformatic approaches will enable the creation of molecular networks fundamental for salt stress response and tolerance (Zhang et al. 2012).

Over the last decade, proteomic analyses of salt tolerant plants have gained popularity in stress biology, and proteomic studies have depended mostly on 2-DE (whole tissue protein separation). However, with developments in recent years, MS-based proteomic approaches have also become common methods to apply complex mixtures of proteins and to pre-fractionate extracts to identify post-translationally modified peptides (Barkla et al. 2013a; Matros et al. 2011). For example, a plant proteome analysis identified more than 560 salt-responsive proteins in plant bodies of 34 different plant species (Zhang et al. 2011). An overview of salinity tolerance mechanisms with proteomic studies revealed many functional aspects, such as photosynthesis, signal transduction, osmotic homeostasis, ROS

scavenging system, ion homeostasis, cross-membrane transport, cytoskeleton dynamics, protein fate and cross-tolerance to multiple stresses (Zhang et al. 2011).

Currently, proteomic methodologies are providing extraordinary insights into plant biology, crop breeding, plant pathology and stress biology. Many proteomic approaches are applied to understand the molecular basis of the salt stress response. Recently, a proteomic approach was used to identify the salt stress-responsive proteins in an elite wheat cultivar. A total of 2358 protein spots were identified from NaCl treated (1.0, 1.5, 2.0 and 2.5 %) wheat leaves by 2D-DIGE, among which 125 spots showed a substantial change in protein abundance. Then using Q-TOF mass spectrometry a total 52 salt-responsive spots were identified (26 were up-regulated and 21 were down-regulated, 5 have multi-expression patterns) which were categorized into six functional groups that included transport-associated proteins, ATP synthase, protein folding, detoxifying enzymes, carbon metabolism and proteins with unfamiliar biological roles. In that study it was observed that some particular proteins namely H^+-ATPases, glutathione S-transferase, ferritin and triose-phosphate isomerase might have key roles under salt stress condition (Gao et al. 2011). Yan et al. (2005) reported a systematic proteomic analysis of rice root proteins under high salinity stress and identified 10 different salt stress-responsive proteins that are involved in many vital cellular processes, including energy metabolism, mRNA and protein processing, ROS scavenging and cytoskeleton stability.

18.4.2.3 Drought Stress

Drought stress is one of the major constraints to agriculture where limited water availability for crop irrigation causes a reduction in carbon fixation by the photosynthetic apparatus that result in net yield losses (Eldakak et al. 2013). Plant breeders desire novel and more accurate tools to fast-track breeding programs required for the increasing demands of food and fodder because farmers are facing a changing climate in which drought stress has major impacts on crop damages globally (Abreu et al. 2013). Crop loss, because of drought stress, has a huge impact on agriculture industry and the economy. Therefore, to prevent economic loss and to make the crop plants more tolerant to drought stress (so that crop can thrive even in drought-prone situations), a variety of approaches such as conventional breeding and/or genetic engineering are being practiced.

Studying the response of a crop to drought is the first step in the crop breeding strategy of tolerant genotypes. Drought prompts changes in the leaf proteome of maize that were studied by 2-DE and iTRAQ that yielded consistent results. The analysis demonstrated that drought resulted in the up-regulation (overexpression) of protective, stress-related proteins and changes in the activities of antioxidant enzymes, indicating an abridged level of photosynthesis. Thus, this result established that early stomatal closure in the susceptible genotype leads to the inhibition of photosynthesis and activates a less-efficient synthesis of the detoxification proteins which are linked with drought tolerance mechanisms (Benesova et al. 2012). Drought-induced changes in the leaf proteome of sugar beet (*Beta vulgaris*) were

studied by 2-DE and quantitatively analyzed using image analysis software that highlighted crucial genes. Of more than 500 proteins, 11 exhibited an association with drought signal transduction, oxidative stress, redox regulation and chaperone activities when analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Some of these proteins could contribute a physiological advantage under drought, making them potential targets for selection strategies in crop breeding (Hajheidari et al. 2005). The same research group applied a proteomic approach to investigate changes in wheat grain in response to drought using MALDI-TOF/TOF that led to the identification of some crucial proteins that were thioredoxin targets and revealed an association between drought and oxidative stress. In addition, that study proposed the application of recognized applicant genetic marker in wheat breeding for improved drought tolerance (Hajheidari et al. 2007). Proteomic analysis of drought-affected barley suggests that chloroplast metabolism and energy associated proteins influence the adaptation progression of barley seedlings under drought (Kausar et al. 2013). To improve the understanding of the intricate mechanisms involved in the common bean (*Phaseolus vulgaris*) in response to drought stress, 2D-DIGE was used to compare changes in protein abundances between stressed and control plants, and the majority of identified proteins were shown to play crucial roles in photosynthesis, energy metabolism, ATP inter-conversion, protein synthesis and proteolysis (Zadraznik et al. 2013). This knowledge has helped bean breeders to understand the drought stress response more completely.

Although plant breeders have made significant improvements of plants through conventional breeding by developing drought-tolerant lines/cultivars of many selected crops, this approach contains several limitations, in that it is time-consuming, laborious and not very cost-effective. Thus, plant breeders have leveraged insights from recent proteomic studies that will be advantageous to the conventional breeding process in the near future (Ashraf 2010).

18.4.2.4 Cold (Chilling) Stress

The world's population is increasing at an alarming rate, whereas crop production is not able to keep a pace with increasing demands because of hindrances from various environmental factors. Specifically, cold stress (low temperature, 1–10 °C) is one of the largest concerns of plant breeders because it is a major threat to the sustainability of crop yields in some geographical areas (Janská et al. 2010; Yadav 2010). Cold stress unfavorably upsets plant growth and development and various phenotypic symptoms include yellowing of leaves (chlorosis), poor germination, reduced leaf expansion, stunted seedlings, wilting and sometimes lead to necrosis (Yadav 2010). Most plants acquire moderate freezing tolerance by cold acclimation (Chinnusamy et al. 2007). One of the most vital adaptation strategies of plants from freezing is the adjustment of plasma membrane compositions and functions (Takahashi et al. 2013). Therefore, to investigate these tolerance mechanisms and find out the several of cold stress responses, plant breeders include the assistance of proteomic studies. Comparative proteomics have helped researchers to recognize the overall response

of cold tolerant plants and retrieval mechanism against chilling stress (Ghosh and Xu 2014). Thus, the collection of proteomic data in response to the cold stress is of importance for application to crop breeding efforts, mainly to increase cold tolerance in crops (Takahashi et al. 2013).

Because we know that under cold stress the plasma membrane will be the first site of primary injury, analysis has been performed to explore plasma membrane proteomics. This helps researchers to understand how plants adapt to stressful environments and how we can improve crop yields in this type of severe environment. Several plasma membrane protein expressions have been reported to change in response to a chilling environment, which cause an increase in a plant's cold tolerance. For example, in *Arabidopsis*, a plasma membrane-associated protein called phospholipase Dδ, which hydrolyzes membrane phospholipids, has been confirmed to increase in amount during cold acclimation (Minami et al. 2009; Takahashi et al. 2013). Quantitative analysis of plasma membrane proteome of rice roots grown under cold stress situation has shown that proteins associated with signal transduction and membrane permeability are higher in expressions (Hashimoto et al. 2009). Using proteomic analyses, an investigation was performed to understand the molecular adaptation mechanisms of cold stress in rice (*Oryza sativa*) seedlings. Proteins were extracted from the leaves of the plants with stressed seedlings and then, by fractionation, nearly 1700 protein spots were separated and visualized on 2-D gels. Of those proteins, 41 were identified using MALDI-TOF MS or ESI/MS/MS that play important roles in the biosynthesis of proteins, cell wall components and proteins linked to energy pathway and signal transduction (Cui et al. 2005).

Winter damage in Norway by freezing of strawberry plants is one of the greatest factors in reducing crop yield (average crop loss of 20 % each year) and quality in temperate regions. Therefore, construction of cultivars with better freezing durability is one of major challenge for strawberry breeders. Improvement of cold hardiness can be achieved by utilizing proteomic analysis with 2-DE and label-free quantitative proteomics. Many proteins were newly identified as associated with cold tolerance, including metabolic enzymes, antioxidants/detoxifying enzymes, pathogenesis-related proteins and the flavonoid pathway. These results reflect a more comprehensive interpretation of antioxidation, detoxification and disease resistance in reference to cold stress (Koehler et al. 2012). For peach breeders, it is a major challenge to store harvested fruit properly at low temperatures as peach undergoes a rapid softening process that leads to undesirable changes such as mealiness and browning, which lowers fruit quality and value. Using a comparative proteomic method with 2-D DIGE helped researchers to classify proteins that showed phase-precise fluctuations in their accumulation pattern and a noteworthy proportion of the proteins were identified as associated with softening, chilling stress (Nilo et al. 2010). Quantitative proteomic analyses of an Iranian spring wheat identified down-regulated proteins after cold stress and some vital Krebs cycle enzymes (malate dehydrogenase, isocitrate dehydrogenase) together with many photosynthesis-related proteins (Calvin cycle enzymes, oxygen-evolving complex proteins, ATP synthase subunits) (Rinalducci et al. 2011).

Proteomic findings allow present molecular breeding research to be more constructive, contributing to the current knowledge on these long-term responses to

cold stress, which may be crucial to cold stress adaptation under unavoidable temperature condition. Thus, comparative proteomics has established itself as an appreciated tool for plant breeders to understand cold stress and plant chilling tolerance.

18.4.2.5 Flooding Stress

Flooding is another major abiotic stress for many natural and artificial ecosystems worldwide that hampers agriculture, and affects crop productions every year (Bailey-Serres and Voesenek 2008; Striker 2012). Soil inundation induces various plant physiological dysfunctions, leading to deterioration in plant growth and survival capacity. Some of the most crucial effects of flooding include decrease in metabolism, and reduction in water and nutrient uptake. Consequently prolonged soil flooding will also eventually lead to anoxic conditions with severe effects on plants respiratory metabolism (Dat et al. 2004). Complete submergence causes a marked reduction in cellular O_2 content, severely suppresses production of energy required for plant growth and ultimately leads to plant death (Komatsu et al. 2012). Currently it is major challenge for the plant breeders to create flooding-tolerant germplasm so that they can reduce the crop loss related to flooding conditions. To develop a flooding-tolerant cultivar, genetic and molecular knowledge can help the breeder tremendously. The application of quantitative proteomics, as shown in the next example, is initiating the venture for the breeders towards flood-tolerant varieties development, and to improve agricultural productivity. Flooding stress is a serious concern for the soybean breeders because it reduces growth and ultimately loss of yield. To investigate the flooding tolerance mechanism of soybean, initially a flooding-tolerant mutant was isolated and studied using proteomics. The results suggest that activation of the alcohol fermentation system in the primary stages of flooding may be a key factor for getting flood-tolerance in soybean (Komatsu et al. 2013b). In another study, cell wall proteins from roots and hypocotyls of soybeans flooded for 2 days were analyzed by 2-DE. Results revealed that 16 of 204 cell wall proteins (e.g. lipoxygenases, glycoprotein precursors, germin-like protein precursors, superoxide dismutase were down-regulated) showed responses to flooding stress. These outcomes suggest that the roots and hypocotyls of soybean instigated the suppression of lignification through reduction of these proteins by down regulation of ROS and jasmonate biosynthesis under flooding stress (Komatsu et al. 2010). Breeders can use such science-based information for pyramiding the genes in the development of new germplasm for flood-prone areas.

18.4.3 Combating Biotic Stresses

Crop losses caused by plant pathogens (e.g. fungi, bacteria, viruses, nematodes) has become one of the most serious constraints to world agriculture, in the struggle to increase crop yield and productivity (Teng et al. 1984). The enormous number of plant pathogens can be detrimental to the crops and their effects range from mild

symptoms to epidemic, followed by the destruction of food crops planted over large areas. At least 10 % of global food production is lost every year because of plant disease. It has been estimated that regardless of the prevention measures practiced, a crop loss worth around USD 9.1 billion occurs each year in United States because of disease (Agrios 2005; James 1998; Strange and Scott 2005). Thus, crop management and yield loss assessment requires multidisciplinary approaches because pathogens interact with other biotic and abiotic factors and also with one another to affect the yield more than predicted (Teng et al. 1984). Because pathogenic organisms pose such a threat to modern day agriculture, plant breeders are attempting to find proper ways to analyze pathogenic detriments with the help of proteomic analyses to combat them and to make crops disease-resistant to prevent yield losses.

Crop losses from many of the viral diseases have been abridged considerably by operative use of conventional breeding activities with the application of proteomic techniques (Ali and Yan 2012). For example, leaf rust in barley (*Hordeum vulgare*) by fungal pathogen *Puccinia hordei* is one of the most destructive foliar diseases. The leaf rust resistance gene Rph15 is of unresolved interest for resistance breeding because it confers resistance to *Puccinia*. Thus the aim of the breeders to investigate the Rph15-based defense responses using a proteomic study and apply this knowledge in trait development may help to develop disease resistant varieties of crops. Protein arrangement variations in response to pathogen infections in barley were examined in two near isogenic lines (NILs): Bowman (leaf rust susceptible line) and Bowman-Rph15 (leaf rust resistant line). Almost all of the pathogen-responsive proteins were recognized in the Bowman-Rph15 resistant and were characterized by LC MS/MS analysis. The results revealed the involvement of proteins in photosynthesis and carbohydrate metabolism, energy metabolism, protein degradation and defense responses (Bernardo et al. 2012). In a similar study of spike disease in wheat caused by *Fusarium graminearum*, the main aim was to identify proteins in resistant wheat cultivar Wangshuibai induced by *F. graminearum* infection. Extracted proteins from infected wheat spikes were separated by 2-DE, and 30 protein spots showed massive change in abundance compared with the control and these were characterized by MALDI-TOF MS. Many of the proteins identified related to disease resistance in plants, indicating that proteins related to the defense reactions were triggered shortly after inoculation (Wang et al. 2005). Incorporation of proteomic experiments on pathogens grown in vitro has been a successful tool in crop breeding for pathological analysis. For example, comparative proteomic application based on multi-dimensional peptide separation (MudPIT) on *Phytophthora infestans*, one of the most destructive plant pathogens causing potato blight, gave insight into the energy utilization strategy of this pathogen throughout cyst germination and host colonization, and identified several candidate host-range proteins (Rampitsch and Bykova 2012; Savidor et al. 2008). In recent research with *Botrytis cinerea* (a necrotrophic fungus) using PAGE followed by reverse-phase HPLC and Edman N-terminal sequencing, revealed the representation of the first avirulence gene product (Avr9) (González-Fernández and Jorrin-Novo 2010).

Thus, proteomics has become a powerful tool for providing important information regarding pathogenic and virulent factors, thereby introducing plant breeders to new possibilities for crop disease diagnosis and crop protection from pathogens.

18.5 Conclusions and Prospects

The recent incorporation of advances in proteomics, genomics and molecular marker applications with conventional crop breeding approaches has shaped the foundation for molecular crop breeding and is revolutionizing twenty-first century crop improvement programs for better yields, improved nutritional qualities and additional traits of commercial importance (Moose and Mumm 2008).

One of the major challenges for proteomics applications in crop breeding has been the incapability of proteomics to examine all protein components, which might correlate in part to the technology itself, protein extraction inefficiencies, proteome complexities and biological variances (Chen and Harmon 2006). The application of proteomics technology in the advancement of crop breeding is based on its two major complementary approaches, gel-based and gel-free (Deswal et al. 2013). Although proteomics applications for breeding have been dominated by 2-DE approaches, gel-free quantitative proteomics methods are gaining widespread acceptance, even though these methodologies have been adapted to a lesser extent by breeders (Matros et al. 2011). The 2-DE approaches have several limitations and one of the crucial drawback is that it is inadequate for the analysis of very complex mixtures and the protein detections are strongly biased toward the more abundant ones (Salekdeh and Komatsu 2007). Thus, in future gel-free quantitative proteomic techniques can contribute to a greater extent in helping crop breeding programs by incorporating proteome level analysis (Barkla et al. 2013b; Schaff et al. 2008). The recent introduction of MRM MS (Multiple Reaction Monitoring mass spectrometry), FT-ICR MS (Fourier transform ion cyclotron resonance mass spectrometry) and HPLC ion mobility MS into proteomics is expected to channel plant proteomics into an adequate scale of dynamic resolution (Allwood et al. 2012; Kopka et al. 2004; Zabrouskov et al. 2003). Protein arrays offer a different key explanation and have the likely potential for high-throughput applications to identify molecular pathways and novel protein markers. Although protein chips remain under development, they have demonstrated their utility for investigating protein functions and expression levels. However, more optimization of these practices is required before they can be broadly used in crop proteome analyses (Salekdeh and Komatsu 2007).

Quantitative proteomics will better define the vicissitudes in protein abundance within the cell and analysis of PTMs which can affect the binding and activity of a protein (e.g. phospho-proteomes), will improve our understanding of the molecular mechanisms concomitant with stress responses. Currently, it is challenging to draw constant assumptions from different crop proteomics studies because of the use of different plant genotypes and sample preparation, along with the different proteomic techniques being applied (Balbuena et al. 2011). A higher conformity of proteomic studies between different groups of researchers, breeders and the application of standard operation protocol would expedite the comparability of proteomic outcomes (Salekdeh and Komatsu 2007).

In conclusion, the knowledge derived from proteomic studies will help modify the conventional crop breeding architecture and enable the production of crop plants

with improved qualities (stress resistant, disease tolerant, high yielding), and thereby accomplishing our goal of enhancing crop productivity.

Acknowledgements The authors acknowledge financial support from USDA/SunGrant, SD Agriculture Experiment Station, SD Wheat Commission and SD Soybean Research and Promotion Council.

References

Abreu IA, Farinha AP, Negrao S et al (2013) Coping with abiotic stress: proteome changes for crop improvement. J Proteomics 93:145–168. doi:10.1016/j.jprot.2013.07.014

Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422(6928):198–207

Agrios G (2005) Plant pathology, 5th edn. Academic Press, New York

Ahmed F, Rafii MY, Ismail MR et al (2013) Waterlogging tolerance of crops: breeding, mechanism of tolerance, molecular approaches, and future prospects. Biomed Res Int 2013:963525. doi:10.1155/2013/963525

Ali F, Yan J (2012) Disease resistance in maize and the role of molecular breeding in defending against global threat. J Integr Plant Biol 54(3):134–151. doi:10.1111/j.1744-7909.2012.01105.x

Allwood JW, Parker D, Beckmann M et al (2012) Fourier transform ion cyclotron resonance mass spectrometry for plant metabolite profiling and metabolite identification. In: Plant metabolomics. Springer, New York, pp 157–176

Ashraf M (2010) Inducing drought tolerance in plants: recent advances. Biotechnol Adv 28(1):169–183. doi:10.1016/j.biotechadv.2009.11.005

Bailey-Serres J, Voesenek L (2008) Flooding stress: acclimations and genetic diversity. Annu Rev Plant Biol 59:313–339

Balbuena TS, Dias LL, Martins ML et al (2011) Challenges in proteome analyses of tropical plants. Braz J Plant Physiol 23(2):91–104

Bänziger M (2000) Breeding for drought and nitrogen stress tolerance in maize: from theory to practice. CIMMYT, Mexico

Baranwal VK, Mikkilineni V, Zehr UB et al (2012) Heterosis: emerging ideas about hybrid vigour. J Exp Bot 63(18):6309–6314. doi:10.1093/jxb/ers291

Barkla BJ, Castellanos-Cervantes T, de Leon JL et al (2013a) Elucidation of salt stress defense and tolerance mechanisms of crop plants using proteomics—current achievements and perspectives. Proteomics 13(12–13):1885–1900. doi:10.1002/pmic.201200399

Barkla BJ, Vera-Estrella R, Pantoja O (2013b) Progress and challenges for abiotic stress proteomics of crop plants. Proteomics 13(12–13):1801–1815. doi:10.1002/pmic.201200401

Benesova M, Hola D, Fischer L et al (2012) The physiology and proteomics of drought tolerance in maize: early stomatal closure as a cause of lower tolerance to short-term dehydration? PLoS One 7(6), e38017. doi:10.1371/journal.pone.0038017

Bernardo L, Prinsi B, Negri AS et al (2012) Proteomic characterization of the rph15 barley resistance gene-mediated defence responses to leaf rust. BMC Genomics 13:642. doi:10.1186/1471-2164-13-642

Bhalerao RP, Salchert K, Bakó L et al (1999) Regulatory interaction of prl1 wd protein with arabidopsis snf1-like protein kinases. Proc Natl Acad Sci 96(9):5322–5327

Boguth G, Harder A, Scheibe B et al (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 21(6):1037–1053

Bouchez D, Höfte H (1998) Functional genomics in plants. Plant Physiol 118(3):725–732

Casasoli M, Spadoni S, Lilley KS et al (2008) Identification by 2-D DIGE of apoplastic proteins regulated by oligogalacturonides in arabidopsis thaliana. Proteomics 8(5):1042–1054

Chen S, Harmon AC (2006) Advances in plant proteomics. Proteomics 6(20):5504–5516

Chinnusamy V, Zhu J, Zhu J-K (2007) Cold stress regulation of gene expression in plants. Trends Plant Sci 12(10):444–451
Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc B Biol Sci 363(1491):557–572
Cui S, Huang F, Wang J et al (2005) A proteomic analysis of cold stress responses in rice seedlings. Proteomics 5(12):3162–3172. doi:10.1002/pmic.200401148
Damerval C, Maurice A, Josse J et al (1994) Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. Genetics 137(1):289–301
Dat JF, Capelli N, Folzer H et al (2004) Sensing and signalling during plant flooding. Plant Physiol Biochem 42(4):273–282
Delmer DP (2005) Agriculture in the developing world: connecting innovations in plant research to downstream applications. Proc Natl Acad Sci U S A 102(44):15739–15746
DeRisi JL, Iyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278(5338):680–686
Deswal R, Gupta R, Dogra V et al (2013) Plant proteomics in India and Nepal: current status and challenges ahead. Physiol Mol Biol Plants 19(4):461–477. doi:10.1007/s12298-013-0198-y
Dunkley TP, Dupree P, Watson RB et al (2004) The use of isotope-coded affinity tags (ICAT) to study organelle proteomes in arabidopsis thaliana. Biochem Soc Trans 32(Pt3):520–523. doi:10.1042/BST0320520
Duvick DN (1996) Plant breeding, an evolutionary concept. Crop Sci 36(3):539–548
Eldakak M, Milad SI, Nawar AI et al (2013) Proteomics: a biotechnology tool for crop improvement. Front Plant Sci 4:35
FAO (2014) Food and agriculture organization of the United Nations. http://www.fao.org/hunger/en/. Accessed 29 May 2014
Gamble E (1962) Gene effects in corn (Zea mays l.): I. Separation and relative importance of gene effects for yield. Can J Plant Sci 42(2):339–348
Gao L, Yan X, Li X et al (2011) Proteome analysis of wheat leaf under salt stress by two-dimensional difference gel electrophoresis (2D-DIGE). Phytochemistry 72(10):1180–1191
Ghosh D, Xu J (2014) Abiotic stress responses in plant roots: a proteomics perspective. Front Plant Sci 5:6. doi:10.3389/fpls.2014.00006
Gisladottir G, Stocking M (2005) Land degradation control and its global environmental benefits. Land Degrad Dev 16(2):99–112
Gong CY, Wang T (2013) Proteomic evaluation of genetically modified crops: current status and challenges. Front Plant Sci 4:1–8. doi:10.3389/fpls.2013.00041
Gong CY, Li Q, Yu HT et al (2012) Proteomics insight into the biological safety of transgenic modification of rice as compared with conventional genetic breeding and spontaneous genotypic variation. J Proteome Res 11(5):3019–3029. doi:10.1021/pr300148w
González-Fernández R, Jorrin-Novo JV (2010) Proteomics of fungal plant pathogens: the case of Botrytis cinerea. In: Current research technology and education topics in applied microbiology and microbial biotechnology. Formatex Research Center, Badajoz, pp 205–217
Görg A, Weiss W, Dunn MJ (2004) Current two-dimensional electrophoresis technology for proteomics. Proteomics 4(12):3665–3685
Gorg A, Drews O, Luck C et al (2009) 2-de with IPGs. Electrophoresis 30(Suppl 1):S122–S132. doi:10.1002/elps.200900051
Gruhler A, Schulze WX, Matthiesen R et al (2005) Stable isotope labeling of arabidopsis thaliana cells and quantitative proteomics by mass spectrometry. Mol Cell Proteomics 4(11):1697–1709. doi:10.1074/mcp.M500190-MCP200
Gygi SP, Rist B, Gerber SA et al (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 17(10):994–999
Hajheidari M, Abdollahian-Noghabi M, Askari H et al (2005) Proteome analysis of sugar beet leaves under drought stress. Proteomics 5(4):950–960. doi:10.1002/pmic.200401101
Hajheidari M, Eivazi A, Buchanan BB et al (2007) Proteomics uncovers a role for redox in drought tolerance in wheat. J Proteome Res 6(4):1451–1460. doi:10.1021/pr060570j

Hasanuzzaman M, Nahar K, Alam MM et al (2013) Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. Int J Mol Sci 14(5):9643–9684
Hashiguchi A, Ahsan N, Komatsu S (2010) Proteomics application of crops in the context of climatic changes. Food Res Int 43(7):1803–1813
Hashimoto M, Toorchi M, Matsushita K et al (2009) Proteome analysis of rice root plasma membrane and detection of cold stress responsive proteins. Protein Pept Lett 16(6):685–697
Hu J, Coombes KR, Morris JS et al (2005) The importance of experimental design in proteomic mass spectrometry experiments: some cautionary tales. Brief Funct Genomic Proteomic 3(4):322–331
Huang B, Xu C (2008) Identification and characterization of proteins associated with plant tolerance to heat stress. J Integr Plant Biol 50(10):1230–1237
Humphery-Smith I, Cordwell SJ, Blackstock WP (1997) Proteome research: complementarity and limitations with respect to the RNA and DNA worlds. Electrophoresis 18(8):1217–1242
James C (1998) Global review of commercialized transgenic crops: 1998, vol 8. ISAAA, Ithaca
Janská A, Maršík P, Zelenková S et al (2010) Cold stress and acclimation–what is important for metabolic adjustment? Plant Biol 12(3):395–405
Kang S, Chen S, Dai S (2010) Proteomics characteristics of rice leaves in response to environmental factors. Front Biol 5(3):246–254
Kausar R, Arshad M, Shahzad A et al (2013) Proteomics analysis of sensitive and tolerant barley genotypes under drought stress. Amino Acids 44(2):345–359. doi:10.1007/s00726-012-1338-3
Kav NN, Srivastava S, Yajima W et al (2007) Application of proteomics to investigate plant-microbe interactions. Curr Proteomics 4(1):28–43
Keeler M, Letarte J, Hattrup E et al (2007) Two-dimensional differential in-gel electrophoresis (DIGE) of leaf and roots of lycopersicon esculentum. Methods Mol Biol 355:157–174. doi:10.1385/1-59745-227-0:157
Khush GS (2000) New plant type of rice for increasing the genetic yield potential. In: Nanda JS (ed) Rice breeding and genetics, research priorities and challenges. Science Publishers, Enfield, pp 99–108
Kim DW, Agrawal GK, Rakwal R et al (2014) Genomic methods for improving abiotic stress tolerance in crops. In: Ricroch A, Chopra S, Fleischer S (eds) Plant biotechnology: experience and future prospects. Springer International Publishing, Cham. doi:10.1007/978-3-319-06892-3__4
Koehler G, Wilson RC, Goodpaster JV et al (2012) Proteomic study of low-temperature responses in strawberry cultivars (fragaria x ananassa) that differ in cold tolerance. Plant Physiol 159(4):1787–1805. doi:10.1104/pp. 112.198267
Komatsu S, Kobayashi Y, Nishizawa K et al (2010) Comparative proteomics analysis of differentially expressed proteins in soybean cell wall during flooding stress. Amino Acids 39(5):1435–1449. doi:10.1007/s00726-010-0608-1
Komatsu S, Hiraga S, Yanagawa Y (2012) Proteomics techniques for the development of flood tolerant crops. J Proteome Res 11(1):68–78. doi:10.1021/pr2008863
Komatsu S, Mock H-P, Yang P et al (2013a) Application of proteomics for improving crop protection/artificial regulation. Front Plant Sci 4:1–3. doi:10.3389/fpls.2013.00522
Komatsu S, Nanjo Y, Nishimura M (2013b) Proteomic analysis of the flooding tolerance mechanism in mutant soybean. J Proteomics 79:231–250. doi:10.1016/j.jprot.2012.12.023
Kopka J, Fernie A, Weckwerth W et al (2004) Metabolite profiling in plant biology: platforms and destinations. Genome Biol 5(6):109
Kosová K, Vítámvás P, Prášil IT et al (2011) Plant proteome changes under abiotic stress—contribution of proteomics studies to understanding plant stress response. J Proteomics 74(8):1301–1322
Langridge P, Fleury D (2011) Making the most of 'omics' for crop breeding. Trends Biotechnol 29(1):33–40
Lee DG, Ahsan N, Lee SH et al (2007) A proteomic approach in analyzing heat-responsive proteins in rice leaves. Proteomics 7(18):3369–3383

Lee J, Jiang W, Qiao Y et al (2011) Shotgun proteomic analysis for detecting differentially expressed proteins in the reduced culm number rice. Proteomics 11(3):455–468. doi:10.1002/pmic.201000077
Lewandowska D, ten Have S, Hodge K et al (2013) Plant silac: stable-isotope labelling with amino acids of arabidopsis seedlings for quantitative proteomics. PLoS One 8(8), e72207. doi:10.1371/journal.pone.0072207
Lin SK, Chang MC, Tsai YG et al (2005) Proteomic analysis of the expression of proteins related to rice quality during caryopsis development and the effect of high temperature on expression. Proteomics 5(8):2140–2156
Majoul T, Bancel E, Triboï E et al (2004) Proteomic analysis of the effect of heat stress on hexaploid wheat grain: characterization of heat-responsive proteins from non-prolamins fraction. Proteomics 4(2):505–513
Mann M (2006) Functional and quantitative proteomics using silac. Nat Rev Mol Cell Biol 7(12):952–958
Masson F, Rossignol M (1995) Basic plasticity of protein expression in tobacco leaf plasma membrane. Plant J 8(1):77–85
Matros A, Kaspar S, Witzel K et al (2011) Recent progress in liquid chromatography-based separation and label-free quantitative plant proteomics. Phytochemistry 72(10):963–974
Minami A, Kawamura Y, Yamazaki T et al (2009) Plasma membrane and plant freezing tolerance: possible involvement of plasma membrane microdomains in cold acclimation. In: Plant cold hardiness: from the laboratory to the field. CAB International, Wallingford, pp 62–71
Mittler R (2006) Abiotic stress, the field environment and stress combination. Trends Plant Sci 11(1):15–19
Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol 147(3):969–977
Nilo R, Saffie C, Lilley K et al (2010) Proteomic analysis of peach fruit mesocarp softening and chilling injury using difference gel electrophoresis (DIGE). BMC Genomics 11:43. doi:10.1186/1471-2164-11-43
O'Farrell PZ, Goodman HM, O'Farrell PH (1977) High resolution two-dimensional electrophoresis of basic as well as acidic proteins. Cell 12(4):1133–1142
Patterson SD, Aebersold RH (2003) Proteomics: the first decade and beyond. Nat Genet 33:311–323
Peng S, Huang J, Sheehy JE et al (2004) Rice yields decline with higher night temperature from global warming. Proc Natl Acad Sci U S A 101(27):9971–9975
Pretty J (2000) Food security through sustainable agriculture. In: Novartis foundation for sustainable development symposium. Nutrition and Development
Rampitsch C, Bykova NV (2012) Proteomics and plant disease: advances in combating a major threat to the global food supply. Proteomics 12(4–5):673–690
Richards R, López-Castañeda C, Gomez-Macpherson H et al (1993) Improving the efficiency of water use by plant breeding and molecular biology. Irrig Sci 14(2):93–104
Rinalducci S, Egidi MG, Karimzadeh G et al (2011) Proteomic analysis of a spring wheat cultivar in response to prolonged cold stress. Electrophoresis 32(14):1807–1818. doi:10.1002/elps.201000663
Robbins ML, Roy A, Wang PH et al (2013) Comparative proteomics analysis by DIGE and iTRAQ provides insight into the regulation of phenylpropanoids in maize. J Proteomics 93:254–275. doi:10.1016/j.jprot.2013.06.018
Rohila JS, Fromm ME (2008) Tap-tagging system in rice for protein complex isolation. In: Plant proteomics. Wiley, Hoboken, pp 525–542. doi:10.1002/9780470369630.ch36
Rohila JS, Chen M, Cerny R et al (2004) Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. Plant J 38(1):172–181. doi:10.1111/j.1365-313X.2004.02031.x
Rohila JS, Chen M, Chen S et al (2006) Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice. Plant J 46(1):1–13. doi:10.1111/j.1365-313X.2006.02671.x

Rohila JS, Chen M, Chen S et al (2009) Protein-protein interactions of tandem affinity purified protein kinases from rice. PLoS One 4(8), e6685. doi:10.1371/journal.pone.0006685
Rombouts I, Lagrain B, Brunnbauer M et al (2013) Improved identification of wheat gluten proteins through alkylation of cysteine residues and peptide-based mass spectrometry. Sci Rep 3:2279. doi:10.1038/srep02279
Roy A, Rushton PJ, Rohila JS (2011) The potential of proteomics technologies for crop improvement under drought conditions. Crit Rev Plant Sci 30(5):471–490
Rubio V, Shen Y, Saijo Y et al (2005) An alternative tandem affinity purification strategy applied to arabidopsisproteincomplexisolation.PlantJ41(5):767–778.doi:10.1111/j.1365-313X.2004.02328.x
Sabel MS, Liu Y, Lubman DM (2011) Proteomics in melanoma biomarker discovery: great potential, many obstacles. Int J Proteomics 2011:1–8
Salekdeh GH, Komatsu S (2007) Crop proteomics: aim at sustainable agriculture of tomorrow. Proteomics 7(16):2976–2996
Santoni V, Rouquié D, Doumas P et al (1998) Use of a proteome strategy for tagging proteins present at the plasma membrane. Plant J 16(5):633–641
Satake T, Yoshida S (1978) High temperature-induced sterility in indica rices at flowering. Proc Crop Sci Soc Jpn 47(1):6–17
Savidor A, Donahoo RS, Hurtado-Gonzales O et al (2008) Cross-species global proteomics reveals conserved and unique processes in phytophthora sojae and phytophthora ramorum. Mol Cell Proteomics 7(8):1501–1516
Schaff JE, Mbeunkui F, Blackburn K et al (2008) Silip: a novel stable isotope labeling method for in planta quantitative proteomic analysis. Plant J 56(5):840–854
Schena M, Shalon D, Davis RW et al (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270(5235):467–470
Schulze WX, Usadel B (2010) Quantitation in mass-spectrometry-based proteomics. Annu Rev Plant Biol 61:491–516
Sethuraman M, McComb ME, Huang H et al (2004) Isotope-coded affinity tag (ICAT) approach to redox proteomics: identification and quantitation of oxidant-sensitive cysteine thiols in complex protein mixtures. J Proteome Res 3(6):1228–1233
Shiio Y, Aebersold R (2006) Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. Nat Protoc 1(1):139–145
Strange RN, Scott PR (2005) Plant disease: a threat to global food security. Phytopathology 43:83–116. doi:10.1146/annurev.phyto.43.113004.133839
Striker GG (2012) Flooding stress on plants: anatomical, morphological and physiological responses. Botany InTech, Rijeka, pp 3–28. doi:10.5772/2245
Takahashi D, Li B, Nakayama T et al (2013) Plant plasma membrane proteomics for improving cold tolerance. Front Plant Sci 4:90. doi:10.3389/fpls.2013.00090
Tanksley S, Nelson J (1996) Advanced backcross qtl analysis: a method for the simultaneous discovery and transfer of valuable qtls from unadapted germplasm into elite breeding lines. Theor Appl Genet 92(2):191–203
Teng P, Shane W, MacKenzie DR (1984) Crop losses due to plant pathogens. Crit Rev Plant Sci 2(1):21–47
Teshima R, Nakamura R, Satoh R et al (2010) 2D-DIGE analysis of rice proteins from different cultivars. Regul Toxicol Pharmacol 58(3 Suppl):S30–S35. doi:10.1016/j.yrtph.2010.05.010
Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science 327(5967):818–822
Thelen JJ, Peck SC (2007) Quantitative proteomics in plants: choices in abundance. Plant Cell 19(11):3339–3346
Thiellement H, Bahrman N, Damerval C et al (1999) Proteomics for genetic and physiological studies in plants. Electrophoresis 20(10):2013–2026
Turan S, Cornish K, Kumar S (2012) Salinity tolerance in plants: breeding and genetic engineering. Aust J Crop Sci 6(9):1337
Tyers M, Mann M (2003) From genomics to proteomics. Nature 422(6928):193–197

Ünlü M, Morgan ME, Minden JS (1997) Difference gel electrophoresis. A single gel method for detecting changes in protein extracts. Electrophoresis 18(11):2071–2077
Van den Bergh G, Arckens L (2004) Fluorescent two-dimensional difference gel electrophoresis unveils the potential of gel-based proteomics. Curr Opin Biotechnol 15(1):38–43
Viswanathan S, Ünlü M, Minden JS (2006) Two-dimensional difference gel electrophoresis. Nat Protoc 1(3):1351–1358
Wang Y, Yang L, Xu H et al (2005) Differential proteomic analysis of proteins in wheat spikes induced by Fusarium graminearum. Proteomics 5(17):4496–4503. doi:10.1002/pmic.200401317
Wang K, Peng X, Ji Y et al (2013) Gene, protein, and network of male sterility in rice. Front Plant Sci 4:1–10. doi:10.3389/fpls.2013.00092
Westermeier R, Marouga R (2005) Protein detection methods in proteomics research. Biosci Rep 25:19–32
Whitelegge JP (2004) Mass spectrometry for high throughput quantitative proteomics in plant research: lessons from thylakoid membranes. Plant Physiol Biochem 42(12):919–927
Yadav SK (2010) Cold stress tolerance mechanisms in plants. A review. Agron Sustain Dev 30(3):515–527
Yan S, Tang Z, Su W et al (2005) Proteomic analysis of salt stress-responsive proteins in rice root. Proteomics 5(1):235–244
Yang M-F, Liu Y-J, Liu Y et al (2009) Proteomic analysis of oil mobilization in seed germination and postgermination development of Jatropha curcas. J Proteome Res 8(3):1441–1451
Yokoi S, Bressan RA, Hasegawa PM (2002) Salt stress tolerance of plants. JIRCAS Work Rep 23(01):25–33
Zabrouskov V, Giacomelli L, van Wijk KJ et al (2003) A new approach for plant proteomics characterization of chloroplast proteins of arabidopsis thaliana by top-down mass spectrometry. Mol Cell Proteomics 2(12):1253–1260
Zadraznik T, Hollung K, Egge-Jacobsen W et al (2013) Differential proteomic analysis of drought stress response in leaves of common bean (*Phaseolus vulgaris* l.). J Proteomics 78:254–272. doi:10.1016/j.jprot.2012.09.021
Zhang H, Han B, Wang T et al (2011) Mechanisms of plant salt response: insights from proteomics. J Proteome Res 11(1):49–67
Zhang H, Han B, Wang T et al (2012) Mechanisms of plant salt response: insights from proteomics. J Proteome Res 11(1):49–67. doi:10.1021/pr200861w
Zhu J-K (2001) Plant salt tolerance. Trends Plant Sci 6(2):66–71
Zieske LR (2006) A perspective on the use of iTRAQ™ reagent technology for protein complex and profiling studies. J Exp Bot 57(7):1501–1508
Zou J, Liu C, Chen X (2011) Proteomics of rice in response to heat stress and advances in genetic engineering for heat tolerance in rice. Plant Cell Rep 30(12):2155–2165

Chapter 19
Utilization of Diverse Sequencing Panels for Future Plant Breeding

Aysen Yumurtaci

Abstract Discovery of the DNA double helix represents one of the great milestones in biology. For plant researchers, a set of complex sequence composition of four different nucleotides arrayed on DNA present important clues for solving the role of genes. Along with the Human Genome Project, unveiling the genome sequence of many organisms contributed to the development of next-generation technologies and their applications have increasingly gained momentum. Initial sequencing efforts have been performed on viruses, bacteria and small plant genomes. Drafting of the entire genome with high sensitivity made possible high-throughput analysis of nucleotide strings. Today, molecular markers have pivotal role in the development of resistant plants of high quality. In plant breeding, choosing the most appropriate genotype via modern and cost-effective tools has become an urgent need. With the introduction of Sanger sequencing, reference genome sequences of some plants have begun to be released publicly with different coverage levels. Mining of raw data, originally generated by different sequencing platforms, has caused a shift toward ongoing genome sequencing projects. Also, polyploidy and repetitive DNA sequences are the two main challenging factors of sequencing which prevent seeing the big picture. Above all, development of long-read length based and single molecule real-time sequencing methods have began to take place in the commercial market. Thus, it will be possible to sequence larger parts of a genome at one time without loss. Extending the sequencing revolution will reduce cost and give high-throughput from the laboratory to the field and upgrade the understanding of plant genome organization.

Keywords DNA • Genome • High-throughput • Plant breeding • Sequencing

A. Yumurtaci (✉)
Faculty of Science and Letters, Department of Biology, University of Marmara, Istanbul, Turkey
e-mail: yumurtaciay@yahoo.com; aysen.yumurtaci@gmail.com

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_19

19.1 Introduction

Sequencing is an important tool to interpret the structural and functional genomic mechanisms in all organisms. Almost four decades ago, Sanger and Coulson (1975) developed a sequencing-by-synthesis (SBS) method that was based on chain termination reaction during amplification of specific DNA fragments. The most significant point during Sanger sequencing is the competition between normal nucleotides (dNTPs) and dideoxynucleotides (ddNTPs) for binding to DNA during the amplification reaction. Each amplified molecule contains radioactively or fluorescently labeled ddNTP at the 3' primed site, and reaction products are separated according to their fragment sizes via capillary electrophoresis (CE). A manual separation method was used in the beginning with four different reactions performed to identify the strings of the nucleotides on the DNA molecule. Later, a combination of four different fluorescent dyes allowed carrying out the amplification reaction in a single tube, which also reduced the cost and time, and increased the accuracy of the sequencing (Fig. 19.1).

In addition, advancements in polyacrylamide polymer biochemistry have been devised to reuse the gels, and multichannel automation increased the sample number per run. In the Sanger method, average length of error free reads is 600–800 bps. However, automated Sanger sequencing applications facilitate optical sequence detection and allow reading high-resolution CE separation. The method had some limitations such as expensive chemicals, bubble formation in the capillary injection system and unequal peak heights; however, this type of sequencing paved the way for development of newer systems (Stranneheim and Lundeberg 2012).

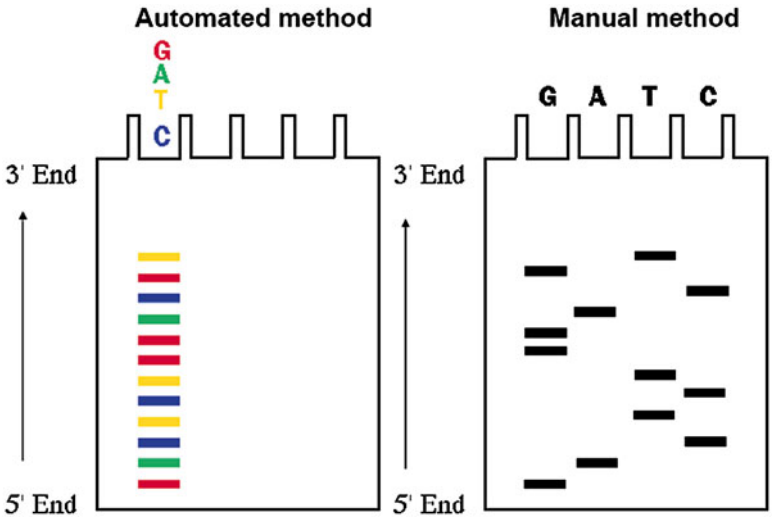

Fig. 19.1 Basic overview for gel screening of fluorescently and radioactively labeled products after the Sanger sequencing reaction

Past efforts have focused on a whole genome sequencing approach to make large-scale comparisons between genomes. In recent years, there has been a rapid increase in the range of sequencing applications, and their infrastructure has been supported with high-throughput technologies exhibiting greater improvements in meeting demands. Basically, all next-generation sequencing methods use either clonally amplified DNA templates or single molecule DNA. Additionally, the length of fragment size is one of the key factors for selection of a sequencing method. For example, so-called *primer walking* is used for a DNA molecule length of 2 Kb. For much longer molecules, such as 20 Kb, first, a restriction enzyme digested template DNA is cloned into BAC and YAC vectors, and then *shotgun sequencing* is performed by using these clones. As a last step, insert sequences, each carrying out individual clones, are assembled into a single linear DNA called *contigs*. This type of sequencing has served as a tool for median-sized crop genomes such as *Oryza* and *Sorghum* in the recent past (Jackson et al. 2011). In addition, crops having more complex and relatively larger genomes than other species such as wheat, changed the nature of genome sequencing to offer high quality whole genome sequence data. However, repetitive sequences, transposable element density and predisposition to high ploidy in plants have caused some difficulties during sequencing (Feuillet et al. 2011).

In accordance with the demand for high coverage, quality and long read lengths for complicated genome analysis, automated platforms with high capability have raised sequencing standards. From the historical perspective, new types of sequencing instruments were brought together in order to make technical advancements that were embodied with fluorescent DNA sequencers. Notably, the newest type of massively parallel sequencing systems permit collecting data even from a single cell (Mardis 2013). Massive parallelization of sequencing has allowed the entire flow cell to be imaged before new nucleotide incorporation (Laurie et al. 2013).

Different sequencing applications have been initiated to show several variation patterns according to the target molecule type that originated from DNA or RNA. In parallel to the increase in the number of plant genome sequence projects, sequencing data belonging to different sources, such as RNA sequencing (RNA-seq), were also promoted into a leading position (Fig. 19.2). Whole-genome sequencing, targeted resequencing, the discovery of transcription factor binding sites, noncoding RNA expression profiling, global and targeted resequencing for SNPs, de novo sequencing of whole genomes or genes, copy number variation (CNV) characterization, amplicon sequencing, chromatin immunoprecipitation sequencing (ChIP-seq), transcriptome sequencing and epigenome sequencing have been applied to remodel the plant genomes. Especially in de novo sequencing, a high quality of long reads with a low error rate is necessary before assembling the large amount of data. In contrast, short-read technologies can be easily adapted into resequencing studies to provide highly qualified raw data.

The most complicated part of sequencing is that of reliable read overlap on the genome. In other words, weaknesses of some existing methods mostly indicate an assembly problem. At this stage, long reads with high coverage can confirm generation of a sufficient number of gap fillings between sequence reads. Also, assembly

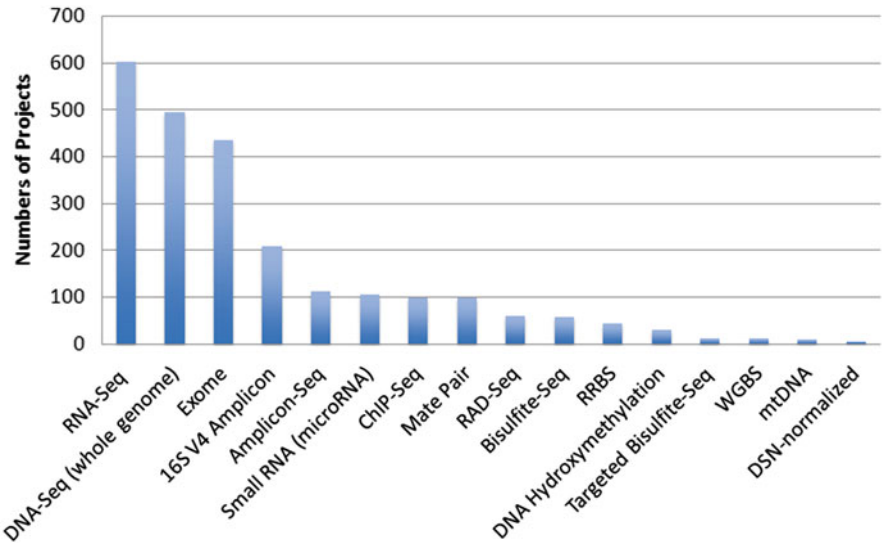

Fig. 19.2 Sequencing projects initiated in the period October 2013–December 2013 (Source: http://genohub.com)

of reads in gigabases throughput has qualitatively revolutionized development of bioinformatic tools. In this way, high coverage will provide more sufficient data to correct the error rates for short and long reads. In this respect, several fragment-assembly algorithms, such as PCAP (Huang et al. 2003), Phrap (http://www.phrap.org) and Velvet (Zerbino and Birney 2008) have been developed to identify the contiguous reads to enable clear nonbranching paths.

Then again, analysis of repeat sequences is another formidable challenge to overcome in the assembling process. Even with the amount of coverage increased to high levels, it is difficult to resolve the mispairing errors in repeat sequences. At the whole genome scale, several algorithms supported with a graphical view and running under short scripts have been developed to manipulate the huge amount of sequence data. In every sequencing cycle, each error which occurs in a read creates a *bulge* in the de Bruijn graph. To remove these bulges from the graphs, short assembler algorithms, like EULER-SR (Pevzner et al. 2001), Velvet and SOAP (Li et al. 2009) de novo, have been developed to extract the error reads from the original sequence data. To adjust genome reconstruction, basically, reads are mapped as k-mers (short words) paths to construct the de Bruijn graph under the EULER-SR algorithm. During sequence data analysis, capturing error-free k-mers and counting their frequency can be prompted to follow a nonbranched path and provide perfect nucleotide strings. Unfortunately, these misassembly errors and the polyploid genome structure can cause a mass complexity during allel mining in plants. To reduce the clutter in sequence data, reads are typically generated in pairs by sequencing at the end of long DNA fragments (Deschamps and Campbell 2009; Iqbal et al. 2012).

The current era of marker-assisted breeding began with the Restriction Fragment Length Polymorphisms (RFLPs), followed by other molecular marker types such as Random Amplified Polymorphic DNA (RAPD), Simple Sequence Repeats (SSR) and Single Nucleotide Polymorphism (SNP) (Rafalski 2010). Molecular markers permit the characterization of modified genes, germplasm diversity and agronomically-important traits. As of 2012, some of the features of plant genome sequences and their structural data had been released as sequencing project outputs. As a known member of Brassicaceae, thale cress (*Arabidopsis thaliana*) was the first sequenced plant genome (Kaul et al. 2000). Subsequently, high-quality draft genomes were released for other plant groups; publicly available completed plant genome sequence numbers have now reached 55 (Michael and Jackson 2013). Plant genome sequencing studies directly impact future plant breeding perspectives. Post-domestication of wild plant relatives has been focused on screening diverse gene homologues between wild and cultivated species. Therefore, designing species-specific primer regions can facilitate covering universal target gene regions and increase genome coverage (Malory et al. 2011). Understanding genetic variations helps overcome problems in rapid genotype selection. Sequence-based genotyping methods are being incorporated into plant breeding (Deschamps et al. 2012). Moreover, sequencing is a supportive approach for defining plant evolutionary history. Identification of rare mutants with targeted sequencing can facilitate selection of precise genotypes (Henry 2012). At the genotype level, providing stress-tolerant plants and clarifying plant-pathogen interactions are major challenges of plant breeding.

Next-generation sequencing technologies can furnish evidence to test and select the focused gene region, expressed as a genome part or whole. In particular, SNP arrays are one of the significant methods of genotyping by sequencing. In crops, Restriction Site Associated DNA (RAD) sequencing is a novel approach used to filter the SNPs without a reference genome (Chutimanitsakun et al. 2011). High-throughput genotyping requires a sufficient and effective number of markers which can generate more accurate data for assembling the plant material into different clusters. In this respect, due to direct contributions to the phenotypic variation, SNPs are potential DNA markers related to trait of interest (Nielsen et al. 2011).

The recent emergence of NGS technologies such as 454 Life Sciences (Roche Applied Science, Indianapolis, IN), HiSeq (Illumina, San Diego, CA), SOLiD and Ion Torrent (Life Technologies Corporation, Carlsbad, CA) has solved problems associated with low throughput and the high cost of gene discovery. Transcriptome sequencing quickly identifies upregulated marker regions and incorporates them into breeding programs and the resequencing of plant genomes (Varshney et al. 2009). Many of the new, informative breeding markers have been identified using RNA sequencing. RNA-seq appears to be replacing older expression microarray technologies for transcriptome profiling in agricultural research applications.

Sequencing technology is evolving at a rapid rate and has been modified and strengthened by scientific and technological collaboration. Steadily, novel sequencing technology platforms are becoming available and researchers have begun to use sequencing data obtained from different sources. Therefore, an extended comparison for the selected genome part or whole genome can be incorporated into the

presented output data. For example, a combined Illumina Solexa and Roche 454 sequencing approach has been used to characterize the cotton genome (Wilkins et al. 2009). A potato draft genome released with the help of three different sequencers (Illumina, Sanger and Roche 454) allowed raw data comparisons to be made in 2011 (Potato Genome Sequencing Consortium 2011). A hybrid sequencing approach can effectively reduce cost and save time, and allow the pooling of more data before the read assembly. Due to their significant roles in global food production, plant pathogen genomes have been attracting attention. Currently, the number of fully sequenced genomes for bacteria has reached 50, while there are 15 for fungal plant pathogens (Sanmiguel 2011). Taken together, studies of both plants and plant pathogens are providing a direct contribution to modern plant breeding research.

19.1.1 Sequencing: A Revolution for Plant Breeding

Plants complete their life cycle under varying local environment conditions. As a result of uncertain biotic stress events and unexpected shifts in climatic conditions, there is great concern about farming practices which are closely related to food and feed production. In addition to the environmental stress factors, breeders give priority to enriching the range of high yielding, highly nutritious plants and bacterial toxin-free seeds for human consumption. Genetic improvement of oil-rich crops has also been reported as another promising area in breeding.

To select feasible individuals with stable traits, mapping populations and pure lines, both genotyping and phenotyping approaches can be jointly used in plant science (Varshney et al. 2014). During the current molecular era, genotyping has been performed with molecular markers carrying dominant and co-dominant characteristics. For example, SSRs in plants have been used to illuminate the population structure and genomic differences of what are called *polymorphic sites*. As an example, Jennings et al. (2011) have provided a multiplexed massively parallel sequencing method that lowers the genotyping costs and offers unique microreads, which is important for finding a distinctive genotype. Today, surveying different genotypes with fragment-based marker approaches have begun to mine with the use of single nucleotide polymorphic markers (SNPs). As compared to previous types of markers, from RAPD to the SSRs, SNPs can help to generate more dense data at lower cost, using hundreds of genotypes resulting in a large quantity of raw data. The SNP-based approach has been used in a number of species such as chickpea (Roorkiwal et al. 2014), wheat (Cavanagh et al. 2013; Oliver et al. 2011), barley (Guo et al. 2014), rice (Yu et al. 2014), maize (Wu et al. 2014) and oats (Tinker et al. 2014). During SNP conjugated genotype selection, data handling and analyzing can be carried out on different platforms (Ray and Satya 2014).

Generally, missing markers and *ghost* QTL regions are other obstacles in plant breeding. QTLs lying in neighboring parts of chromosomes usually carry numerous candidate markers. Anchoring one or more marker to a heritable trait requires detailed analysis. Recently, QTL sequencing was performed to identify the candi-

date plant architecture of individuals from opposite populations. Hence, the effectiveness of the QTL was increased and resequencing of highly allelic variations analyzed (Takagi et al. 2013).

In the past decade, sequencing technologies have been playing an important role in visualizing the details of plant genomes and have provided comparative data for point mutations (Austin et al. 2014; Henry et al. 2014). Leach et al. (2014) applied a transcriptome sequencing approach in wheats (*Triticum aestivum*, *T. urartu*) and goatgrasses (*Aegilops speltoides*, *A. tauschii*) and chromosomes 1 and 5 nullisomic-tetrasomic lines by using Illumina GAII and compared the homologous gene expression to understand the individual roles of each wheat genome. One of the milestones in wheat genome research is the complete sequence of 1 Gb chromosome 3B (Choulet et al. 2014). Thus, it is primarily aimed at providing a chromosome-specific gene catalog for unraveling the regions related to drought stress tolerance, pathogen resistance and high yield performance in wheat.

Some sequencing platforms generate millions of short reads and pooled short fragments, called *barcodes*, making it possible to dissect the dissimilar DNA sequences. On the basis of genotyping by sequencing, these barcodes serve as specific primer adaptors, including a complementary region for the restriction enzyme digested site (Poland and Rife 2012). Also, these regions can be amplified with improved sequencing methodologies and integrated into plant evolutionary studies as species specific taqs (Li et al. 2014).

Future plant research projections indicate the necessity for quick and effective experimental design to develop stress-tolerant commercial varieties, including several major crops. From this perspective, the rapid evolution in next-generation tools can help to provide different plant models for clarifying traits related to important agronomical characters and pave the way for more precise next-generation plant breeding. In addition, unravelling the codes of plant-pathogen genomes will unlock the door to understanding the host genetic resistance mechanism. In particular, ongoing sequencing efforts will allow the evolution of plant genomes, identify the critical genes and transcription factors and present a powerful bias for the isolation of new gene targets from economically important species.

19.2 Basic Principles of Next-Generation Sequencing (NGS)

Complex genome structures with high repetitive elements need to perform high sequencing coverage. The term *next-generation* has become a popular designation after improvement of the first next-generation DNA sequencing machine, the GS20 by 454 Life Sciences (Basel, Switzerland) in 2005. The next-generation sequencing (NGS) approach provides an opportunity for the sequencing of millions of single stranded DNA molecules captured on a bead or a glass slide. NGS read lengths vary depending to the system (Table 19.1). For example, Roche 454 output reads can be obtained up to 1 Kb. Read lengths from Illumina and Ion Torrent are under 600 bp, while PacBio covers longer reads at less than 9 Kb in length (Quail et al. 2012).

Table 19.1 A list of sequencing platforms and their technical features

Instrument class	Instrument name	Sequencing principle	Sample preparation	Detection type	Read length (bp)	Number of bp per run	Sensitivity
1st generation sequencing	Automated Sanger sequencing (ABI 3730xl)	Enzymatic chain termination during in vivo amplification	Cloning-PCR	Optical	900	96 Kb	Low
2nd generation sequencing	454 FLX System	Pyrosequencing on solid support	Emulsion PCR	Optical	200–300	80–120 Mb	Medium
	454 GS FLX/ Titanium	Paired end read sequencing	Emulsion PCR	Optical	500	400 Mb	Medium
	Illumina/Solexa	Sequencing by synthesis with reversible dye terminators	Bridge amplification	Optical	30–40 (2 × 100)	1 Gb	High
	Illumina MiSequencing	Sequencing by synthesis with reversible dye terminators	Bridge amplification	Optical	36–150	1 Gb	High
	Illumina HiSequencing 2000	Sequencing by synthesis with reversible dye terminators	Bridge amplification	Optical	100	600 Gb	High
	SOLID	Sequencing by ligation	Emulsion PCR	Optical	35	1–3 Gb	High
3rd generation sequencing	Ion Torrent	Ion sensitive Sequencing by synthesis	Emulsion PCR	Electronic	200	1 Gb	Medium
	Helicos	Sequencing by synthesis with reversible single dye terminators	Amplification free	Optical	25–55	35 Gb	Medium
	Pacific BioScience	ZMW sequencing	Amplication free	Optical	<9000	200 Mb–1Gb	Medium
	Nanopore Sequencing	Ion current shift		Electronic/optical	Up to 50 Kbp	Tens of Gb	Medium

Resequencing and de novo sequencing approaches include the vast majority of sequencing applications. The most significant differences between the two methods can be expressed as the application of resequencing which covers the known genome structures; there are assemblies originated from unknown genomes in de novo sequencing. Therefore, reference genome exploitation is important for resequenced fragment analysis. Resequencing is one of the comparative methods in plant breeding for traits of interest and the extent of their ability to contribute to resistant plant selection. In addition, two important model plants, *Arabidopsis* and the draft maize genome have been finalized with a BAC-by-BAC sequencing approach which represent the first representatives of Sanger-based sequencing of bacterially-cloned long DNA fragments. However, there are still physical gaps present on these two genomes. NGS tools have made an obvious improvement in whole genome sequencing of other commercially important plants (Jackson et al. 2011).

19.3 Second Generation Sequencing Platforms

With the development of advanced hardware technologies after 2005, sequencing methodologies improved significantly. The change from manual methods to automated platforms has decreased sequencing costs almost 100 K-fold since 2000. In addition to the traditional Sanger sequencing, second generation technologies have been used to frame several objectives in plant science, such as whole genome sequencing, transcriptome sequencing, indel characterization, de novo and resequencing (Feuillet et al. 2011; Metzker 2010).

The second-generation sequencing (SGS) approach is independent of bacterial cloning but still needs initial fragment enrichment (Fig. 19.3). Data generated by SGS instruments have the capability of mining and comparing first- and third-generation sequencing results. For example, the banana genome (D'Hont et al. 2012) has been drafted with generated data obtained from reciprocal sequencers (Illumina, Sanger and Rıche 454). In SGS, they have been used in three major sequencing methodologies; a) pyrosequencing, b) sequencing by ligation and c) sequencing-by-synthesis approach, as illustrated in simplified form in Fig. 19.4.

19.3.1 Roche 454GS FLX

In the 454 sequencing system, adapter-ligated DNA fragments are fixed to beads to perform a large-scale parallel pyrosequencing amplification in a water-in-oil emulsion. Each bead in the emulsion acts as an independent PCR. The 454GSFLX with titanium chemistry can enable read lengths up to 1 Kbp. However, indel specific errors are characteristic of 454 sequencing; SNPs can be screened from these blocks. With this system, de novo whole-genome sequencing, resequencing of whole

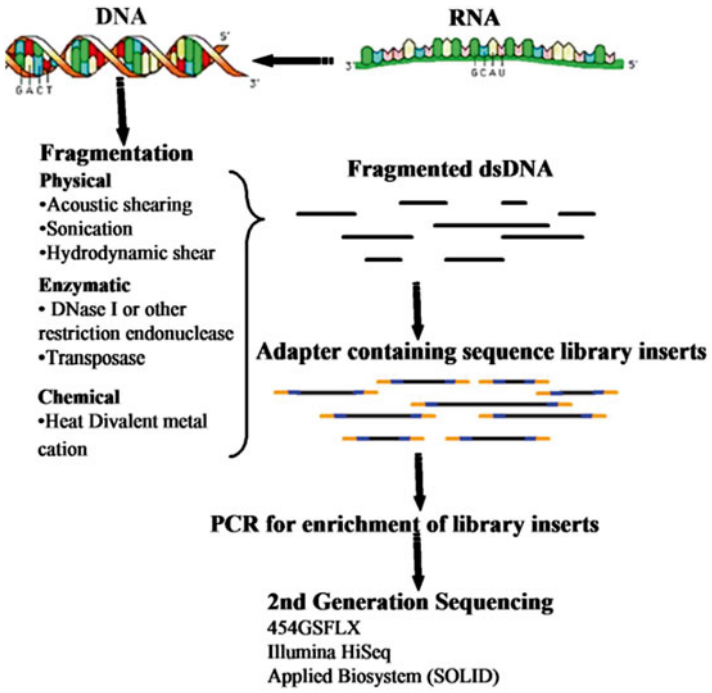

Fig. 19.3 Simple library preparation pipeline for some NGS platforms

genomes and target DNA regions, metagenomics and RNA analysis can efficiently be applied. Several studies in plant genome research have been performed with the 454 sequencing system (Table 19.2). The amount of the generated sequence data ranks second after the Sanger sequencing approach. According to the 454 sequencing data, recent studies in plants have been focused on SNPs and full transcriptome sequences.

19.3.2 Illumina

Since 2007, the Illumina Company (San Diego, CA) has been developing sequencing chemistry platforms for full automation, from sample preparation to computer-based analysis. At short time intervals, an intense focus on chemistry improvements and computational methods have supported the introduction of new systems with core updates. Thus, more complex genomes can be deciphered by presenting cost-effective and user-friendly tools. The Illumina Genome Analyzer II (GAII) and HiSeq2000 systems are the most commonly used systems run according to the sequencing-by-synthesis approach. Especially for the Illumina GAII system, the same flow cell can be used more than once without reducing the optical detection intensity.

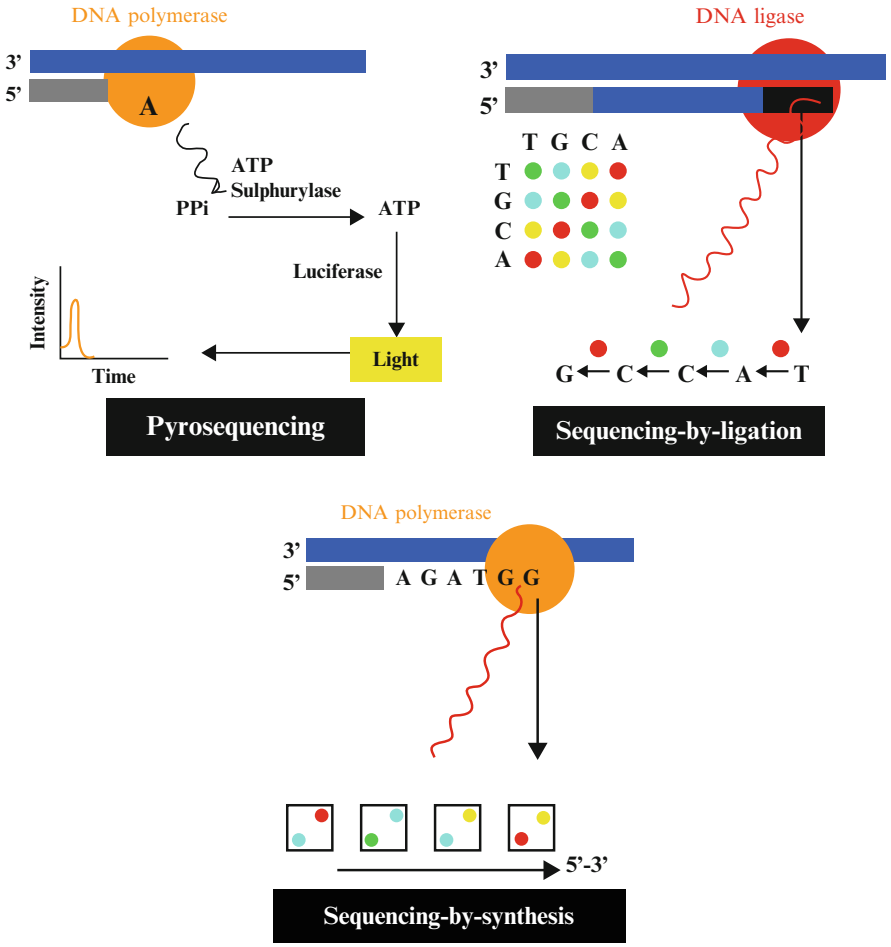

Fig. 19.4 Schematic representation of three major methodologies used in SGS platforms

Generally, with the Illumina sequencing approach, fragmented DNA pieces are hybridized to primers that are stocked on a solid substrate and bridge amplification of the bound DNA template is performed in an isothermal reaction to create a lawn for the flow cell. The sequence order of each molecule is achieved by cleavage of a fluorescently-labeled reversible terminator nucleotide that is labeled with four different fluorescent colors to distinguish nucleotides. During detection, complementary nucleotides, carrying different fluorescent dye signals, are detected by a laser. Single molecule incorporation occurs in each cycle as in Sanger sequencing. Short read lengths (35 bp in average) with 1 Gb throughput can be generated using Illumina HiSeq2000 with a 0.1 % error rate in raw sequencing (Turcatti et al. 2008). Illumina HiSeq2000 is capable of paired-end reads (80–150 bp in length) that permits de novo assemblies at a high rate of efficiency. This second generation platform has

Table 19.2 Applications performed with Roche 454 system in plants

Plant species	Generated data	Sequencing instrument	Reference
Ammopiptanthus mongolicus (desert shrub)	672,000 reads from root transcriptome	454 GS XLR70 Titanium pyrosequencer	Zhou et al. (2012)
Hordeum vulgare (barley)	complex BAC sequencing	Roche 454	Wicker et al. (2008)
Manihot esculenta (cassava)	61 million reads covering 416 Mb of total genome	Roche 454 GS-FLX Titanium	http://www.phytozome.net/cassava
Phaseolus vulgaris (common bean)	3.487 SNPs	Roche 454 FLX System	Hyten et al. (2010)
Podophyllum hexandrum Royle (Indian mayapple)	1.5 million high quality reads for de novo assembly of transcriptome	Roche 454	Bhattacharyya et al. (2013)
Vitis vinifera (grape)	504 Mbp full genome sequencing	Roche 454	Velasco et al. (2007)
Zea mays (maize)	261,000 EST sequences	Roche 454	Emrich et al. (2007)

been utilized in gene prediction studies, organelles genome sequencing, SNP calls for genotype based sequencing and expression profiling studies in plants (Table 19.3).

19.3.3 Applied Biosystems SOLiD

The SOLiD biosystem generally produces shorter reads than its competitors, but with a higher amount of reads. The transcribed region of genomes and noncoding RNA profiling are the two most commonly used analysis types for the SOLiD system. Because short reads are compatible with the sequence census approach, they massively increase high-throughput (Torres et al. 2008). There are microreactors containing DNA templates, beads, primers and these PCR components are the key players of library preparation. Before sequencing, next-generation library enrichment is performed on these beads on bounded templates by emulsion PCR. Basically, 3' end modified DNA templates attach to the glass slides. Then, the primer and a mixture of four different-colored fluorescent oligos are pumped into the cell flow. The result is a complex of templates bounded with fluorescent oligo ligated onto the primer. The remaining nonligated oligos are removed from the flow cell by washing and a CCD device captures the signals excited by the different colors attached to the

Table 19.3 Applications performed with Illumina system in plants

Plant species	Generated data	Sequencing instrument	Reference
Ammopiptanthus mongolicus (desert shrub)	82,795 cDNA unigenes from 132.35 million high quality reads	Illumina/Solexa	Pang et al. (2013)
Brassica napus (rapeseed)	120 Kb chloroplast genome	Illumina	Wu et al. (2012)
Brassica napus (rapeseed)	23,330–41,593 putative SNPs from transcriptome sequences	Illumina/Solexa	Trick et al. (2009)
Cajanus cajan (pigeonpea)	237.2 Gb predicted 48,680 genes	Illumina	Varshney et al. (2012)
Cajanus cajan (pigeonpea)	128.9 million short single end reads	Illumina GA IIx	Kudapa et al. (2012)
Glycine max (soybean)	67.64 Gb sequence reads from multiplexed panel	Illumina	Xu et al. (2013)
Hordeum vulgare (barley)	Sequences related to plant height from three different GBS libraries in barley RILs	Illumina GAII and Hiseq2000	Liu et al. (2014)
Oryza sativa (rice)	1464 SNPs mapped by genotyping by sequencing	Illumina	Spindel et al. (2013)
Pinus (pine)	120 Kb plastome genome	Illumina GA	Cronn et al. (2008)
Triticum aestivum (wheat) and *Hordeum vulgare* (barley)	34,000 SNPs in barley and 20,000 in wheat	Illumina GAII and Hiseq2000	Poland et al. (2012)
Zea mays (maize)	17.1–29.9 million purified reads from RNA-seq.	Illumina	Davidson et al. (2011)

primer. In Table 19.4, studies performed with the SOLiD sequencing platform from different plant groups have been listed.

19.4 Third-Generation Sequencing Platforms

Third-generation sequencers are massive parallel systems that directly sequence individual DNA molecules rather than relying on amplification prior to sequencing. One of the most obvious features that third-generation technologies use is the single-molecule DNA template to provide real-time detection with a large number of high-quality longer reads from a single molecule. Third-generation sequencing can be divided into three different categories: a) sequencing-by-synthesis (PacBio single molecule real time SMART), Heliscope true single molecule sequencing (tSMS) and Ion Torrent, b) sequencing where single nucleotides pass through a nanopore and c)

Table 19.4 Applications performed with SOLiD system in plants

Plant species	Generated data	Sequencing instrument	Reference
Arabidopsis thaliana	114,797 SNPs	SOLiD	Ashelford et al. (2011)
Fragaria vesca (wild strawberry)	220 Mb full genome de novo sequence	SOLiD	Shulaev et al. (2011)
Hevea brasiliensis (rubber tree)	Full genome sequence (2150 Mb read data)	SOLİD	Rahman et al. (2013)
Solanum lycopersicum (tomato)	760 Mb assembled sequence reads	SOLİD	Tomato Genome Consortium (2012)

direct imaging of individual molecules (IBM). In the near future, it is estimated that current sequencing costs will be reduced by 85 % and sequencing speed and capacity will be increased to complete genome sequence data in about 15 min.

19.4.1 Ion Torrent Semiconductor Sequencing

This sequencing system works by the detection of hydrogen ions released during polymerization of DNA and provides faster and simpler sequencing at low cost. Average read lengths are 100–200 bp and exhibit differences according to the chip set used per run. An ion chip has a million wells covering pixels which capture chemical information from the DNA and translate it into digital information for base calls. The first step of Ion Torrent sequencing is the shearing the DNA to provide millions of fragments. Each fragment is attached to a bead and each bead deposited into a chip well. Then the chip is flooded with four different DNA nucleotides in turn. The hydrogen ion changes the pH of the solution in the well. An ion sensitive layer beneath the well measures the pH changes and converts it into the electrical voltage during sequencing. Voltage changes are recorded indicating the nucleotide type that was incorporated into the DNA fragment. In essence, each well works as a very small pH meter detecting voltage changes during the reaction. Every 15 s the same process is repeated for the different nucleotide binding after washing of the chip. If there is a complementary nucleotide present, a hydrogen ion is released and the pH of the solution is changed, and base calling is achieved. The voltage signal doubles when the same two subsequent nucleotides are presentenced on the DNA strand. This voltage detection simultaneously occurs in the millions of chip wells and helps to sequence massively-scaled DNA (Fig. 19.5). Recently, this type of sequencing has been adopted in genotyping in plants by sequencing (Table 19.5).

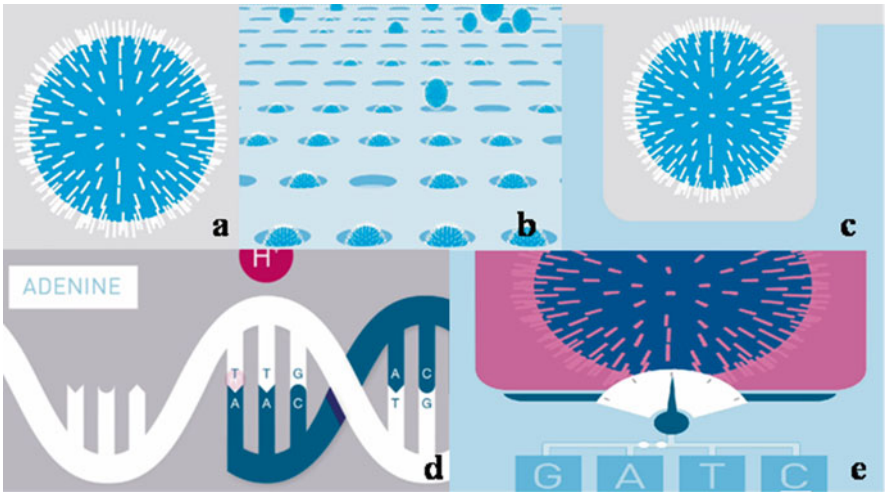

Fig. 19.5 Ion Torrent semiconductor sequencing components. (**a**) Bead attached fragmented DNA, (**b**) beads loaded into chips, (**c**) a single bead in chip well, (**d**) nucleotide incorporation, (**e**) a simple view of pH dependent voltage change in each chip well (Source: http://www.lifetechnologies.com)

Table 19.5 Applications performed with the Ion Torrent system in plants

Plant species	Generated data	Sequencing instrument	Reference
Hordeum vulgare (barley)	30 million and 17 million reads generated for SNP calling	Ion Torrent PGM and Ion Torrent Proton	Mascher et al. (2013)
Sedum album (white stonecrop)	Full genome sequence	Ion Torrent PGM	Michael (2012)

19.4.2 Pacific Biosciences (PacBio)

Pacific Biosciences first released their PacBio RS system in 2011; a new version called PacBio RS II was released in April 2013 which produces longer sequence reads than the first version and offers higher throughput than the original system. Typically, this long-read sequencing platform enables single molecule real time (SMRT) sequencing. This system includes some practical advantages beyond the current second next-generation platforms and provides significantly longer reads, single molecule sequencing and low composition bias. The system has also been used to identify methylation, DNA damage and other epigenetic information. The significant advantage of the PacBio system is that there is no need for library preparation (Coupland et al. 2012; Eid et al. 2009).

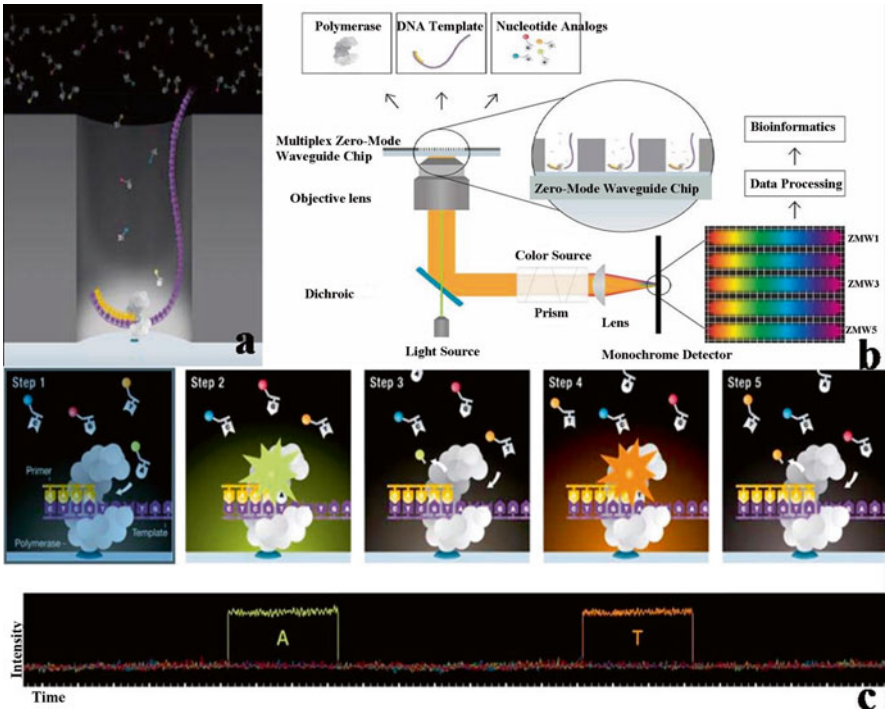

Fig. 19.6 A schematic overview of single molecule real time sequencing. (**a**) Zero-mode waveguide with Phi29 DNA polymerase, DNA template and free nucleotides, (**b**) PacBio detection pipeline, (**c**) single polymerase DNA sequencing steps:
Step 1: Fluorescent phospholinked labeled nucleotides are introduced into the ZMW
Step 2: The base being incorporated is held in the detection volume for tens of milliseconds, producing a bright flash of light
Step 3: The phosphate chain is cleaved, releasing the attached dye molecule
Step 4–5: The process repeats
(Source: http://www.pacificbiosciences.com/products/smrt-technology/)

Technically, SMRT cells as structures of the PacBio RS system are slightly less than one centimeter square and contain thousands of zero-mode waveguides (ZMWs), which are a type of array slide. There are 50 nm-wide nanophotonic structures on these arrays for attachment of DNA polymerase. The volume of the nanometer-sized aluminum layer wells is in the zeptoliter scale. Single-fluorophore detection occurs simultaneously in each nanophotonic structure via an excited signal being incorporated into the growing DNA strand (Fig. 19.6).

This new technology is being used successfully in clinically-based applications such as the resequencing of human medical amplicons (Carneiro et al. 2012) and human transcriptomes (Sharon et al. 2013). An evolutionary study of the pink barren strawberry's (*Potentilla micrantha*) chloroplast genome resequencing was carried out with a PacBio RS platform (Ferrarini et al. 2013). Sequencing with XL chemistry doubled the maximum read length to 20 kilobases and made second-generation

sequencing data much more practicable for genome annotation. Publicly available assembly algorithms such as PBJelly (https://sourceforge.net/projects/pb-jelly/) have been adopted to fill gaps in the four different draft genomes, with PacBio RS long reads used as templates during de novo assembly. The lack of high accuracy (up to about a 15 % error rate) is preventing the overlapping of long reads. For sequencing errors, short interval time between binding and releasing of nucleotides has been stated as a main reason (Metzker 2010). This is one of the handicaps of the PacBio system, which was also noted by English et al. (2012). As of October 2013, an updated chemistry had been released with a mean read length of 8500 bases exceeding 30 Kbp. The PacBio sequencing system is used for the following tasks in genome biology studies: characterization of genetic variation, methylation SNP detection and validation, haplotypes and phasing, base modification detection to understand gene expression, host-pathogen interactions, DNA damage and DNA repair.

19.4.3 HeliScope™ Single Molecule Sequencer by Helicos Bioscience Cooperation

The Helicos™ Single Molecule Sequencing (SMS) provides direct sequencing of cellular nucleic acids in an unbiased manner. DNA is simply fragmented into 100–200 bp pieces with restriction enzyme digestion, tailed with poly(A) and hybridized to a Helicos flow cell plate containing oligo(dT) for sequencing-by-synthesis of billions of molecules. The fluorescent tag is chemically cleaved to permit subsequent elongation of the DNA (Thompson and Steinmann 2010). The Helicos genetic analyzer applications include DNA sequencing and RNA sequencing.

19.4.4 Nanopore Sequencing by Oxford Nanopore Technologies

The Oxford Nanopore GridION system is a scalable network device in which a single nanopore is inserted into a polymer membrane across the top of the microwell (Fig. 19.7). During nanopore sequencing, an applied electrical field controls the DNA movement and the template DNA strand is inserted into the top of pore with Phi29DNA polymerase. A complementary strand is directed into the bottom of the pore and an ionic current change obtained from DNA through the nanopore is detected. The size of the nanopore is also effective on negatively-charged DNA movement through the tiny 1.4 nm wide space. In this way, DNA mutations can be detected by nanopore sequencing. Damaged DNA molecules pass through the membrane in one-thousandth of a second, whereas normal-shaped DNA complete their passage in one-millionth of a second.

Each microwell has its own electrodes for individual sensing. Multiple microwells are fabricated on an array chip using standard materials of the semiconducting

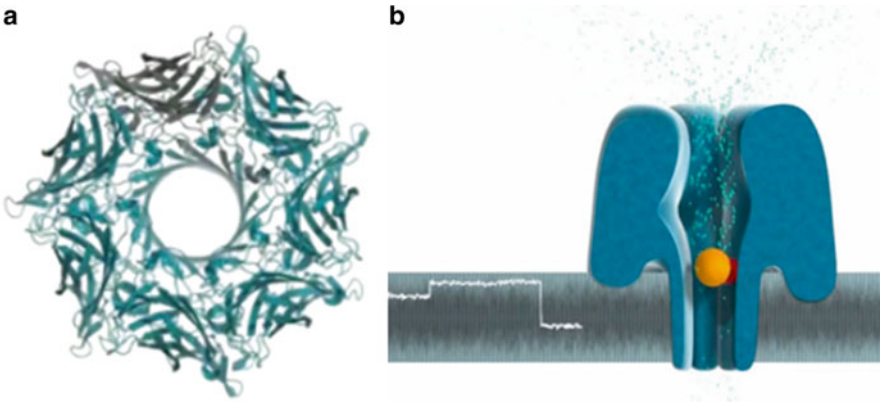

Fig. 19.7 Single molecular nanopore. (**a**) Three dimensional overview of nanopore protein from the upper part, (**b**) nanopore embedded in polymer membrane (Source: https://www.nanoporetech.com/technology)

industry. This array chip may be scaled depending on the application use, many tens or hundreds of thousands of nanopores. The array chip is housed within a single cartridge used for the specific type of molecule. The user adds the sample analyte to the single node for the GriION system. Also, the company designed a MiniION device which can be directly plugged into a computer. Each node contains high performance electronics for local real-time processing. Primary analysis is performed by the system in real time. Nodes communicate with each other and the network storage systems. This is supported by an embedded computer system. A single node may be used as a network benchtop instrument for a single molecule analysis. In addition, the system can be transferred to a multiple node instrument. Individual nodes can cross-communicate and each node can be run under computer management (Fig. 19.8).

Up to now, there have been no real sequence data released for plant genomes. This new generation system only demonstrated two bacterial genome analyses with average read lengths of 4.9 Kbp and 5.4 Kbp (http://ngs-expert.com/2014/03/17/will-oxford-nanopores-minion-hold-its-promises/).

19.5 Conclusions and Prospects

Almost 50 years ago, researchers began to clarify the sequence information about genomes of different organisms. Halley et al. (1965) sequenced an alanin tRNA molecule. Previously, one of the most critical points was whether reaching the full sequence data of an organism would be possible within a short time period. Subsequently, with the development of new techniques in molecular biology and chemistry, increased understanding and knowledge of genomics have brought genomics up to a new level. Today, progress in plant genomic research is made via next-generation sequencing

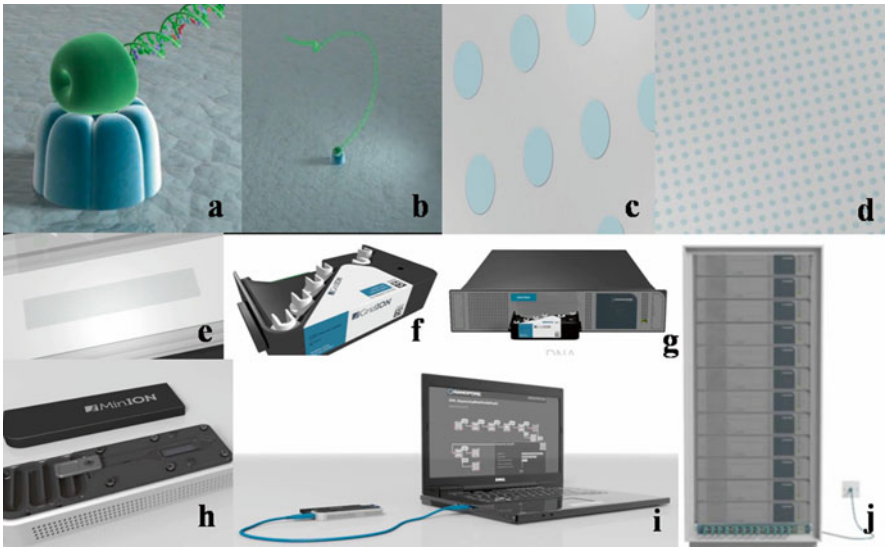

Fig. 19.8 Main items of Nanopore sequencing system. (**a**) Enzyme (*green*)-DNA complex on the *top* of the nanopore (*blue*), (**b**) general overview of single nanopore, (**c**) microwells, (**d**) multiple microwells, (**e**) microwell containing chip, (**f**) single use cartridge containing an individual specific chip, (**g**) GriION single node ready for cartridge loading, (**h**) MiniION benchtop node, (**i**) computer plugged with a MiniION node, (**j**) multiple nodes parallel to each other (Source: https://www.nanoporetech.com/technology)

technologies and this stimulates the agrigenomics revolution into a new era. Hence, new requirements emerged to re-assess plant genomes with updated molecular marker information by using new generation technologies. To accelerate the release of new cultivars with corresponding new traits such as disease resistance, it is necessary to pursue modern systems that have the ability to generate data and automatically mine information in the computational pipelines. Importantly, sequencing instruments can be created that reduce data producing time and cost to a lower level.

Established high-throughput method combinations for selection is expected to become available. To achieve this, work should continue toward a new technology platform by developing knowledge of plant genomics. Conventional plant breeding which considers various stress conditions is a laborious and time-consuming process. Using molecular-marker technology and its future applications for selection of interesting alleles on a target locus is considered an ideal system. In this way, functional allelic differences among genotypes will be selected more accurately (Masojc 2002). Because allel frequency estimations for polyploid genomes are useful for analysis of a very large number of individuals in the germplasm pool, an effective marker should allow the selection of heterozygotes from homozygote genotypes. In addition, selected genotypes should consist of genes/traits of interest in a homozygote state that can be easily used as potential starting material. Consequently, the marker strategy and the type of system play an important role in breeding for particular agronomic character(s). Moreover, recent advances in DNA sequencing, bioinformatics and comparative genetics afford new opportunities to identify the genes

of interest for almost every plant species. Improved breeding technologies must include advanced molecular markers. Single nucleotide polymorphism (SNP) markers may help to distinguish phenotypic traits. In particular, DNA sequencing technology has the potential to contribute new steps in the selection of better varieties under different growth regimes. It is necessary to manage genetic diversity at the whole genome level with the technology to accurately analyze the results (Egan et al. 2012). Apart from the discovery of regular sequence data, future objectives in plant genome research have been extended from second-generation sequencing to a third-generation system. To exploit and analyze the sequenced data, there are now bioinformatic tools with high throughput hardware infrastructure. It will be possible to narrow the connections between genes and traits which hold the promise of completing mass sequence data.

This chapter has attempted to summarize sequencing technologies and their infrastructures with relation to sequencing principles. Each next-generation system is highlighted with the current state of the art and their utilization methods for plant genome development, classified according to the sequencing type and a system-based approach based upon significant research publications. Research on plant genome sequencing continues to expand rapidly in step with the development of new systems. It is obvious that genome research will help to *rediscover* DNA and its role in plant breeding through advanced sequencing facilities.

References

Ashelford K, Eriksson ME, Allen CM et al (2011) Full genome re-sequencing reveals a novel circadian clock mutation in *Arabidopsis*. Genome Biol 12:R28

Austin RS, Chatfield SP, Desveaux D et al (2014) Next-generation mapping of genetic mutations using bulk population sequencing. Methods Mol Biol. doi:10.1007/978-1-62703-580-4-17

Bhattacharyya D, Sinha R, Hazra S et al (2013) *De novo* transcriptome analysis using 454 pyrosequencing of the Himalayan may apple, *Podophyllum hexandrum*. BMC Genomics 14:748. doi:10.1186/1471-2164 14-748

Carneiro MO, Russ C, Ross MG et al (2012) Pacific biosciences sequencing technology for genotyping and variation discovery in human data. BMC Genomics 13:375. doi:10.1186/1471-2164-13-375

Cavanagh CR, Chao S, Wang S et al (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. PNAS. doi:10.1073/pnas.1217133110

Choulet F, Alberti A, Theil S et al (2014) Structural and functional partitioning of bread wheat chromosome 3B. Science. doi:10.1126/science.1249721

Chutimanitsakun YR, Nipper W, Cuesta-Marcos A et al (2011) Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. BMC Genomics 12:4. doi:10.1186/1471-2164-12-4

Coupland P, Chandra T, Quail M et al (2012) Direct sequencing of small genomes on the pacific biosciences RS without library preparation. Biotechniques 53:365–372

Cronn R, Liston A, Parks M et al (2008) Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Res 36. doi:10.1093/nar/gkn502

Davidson RM, Hansey CN, Gowda M et al (2011) Utility of RNA-seq for analysis of maize reproductive transcriptomes. Plant Genome 4:191–203

Deschamps S, Campbell MA (2009) Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. Mol Breed 25:553–570
Deschamps S, Liaca V, May GD (2012) Genotyping-by-sequencing in plants. Biology 1:460–483
D'Hont A, Denoeud F, Aury J-M et al (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature 488:213–217. doi:10.1038/nature11241
Egan AN, Schlueter J, Spooner DM (2012) Applications of next-generation sequencing in plant biology. Am J Bot 99:175–185
Eid JA, Fehr J, Gray K et al (2009) Real-time DNA sequencing from single polymerase molecules. Science 323:133–138
Emrich SJ, Barbazuk WB, Li L, Schnable PS (2007) Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Res 17:69–73
English AC, Richards S, Han Y et al (2012) Mind the Gap: upgrading genomes with pacific biosciences RS long-read sequencing technology. PLoS One 7:e47768. doi:10.1371/journal.pone.0047768
Ferrarini M, Moretto M, Ward JA et al (2013) An evaluation of the PacBio RS platform for sequencing and *de novo* assembly of a chloroplast genome. BMC Genomics 14:670. doi:10.1186/1471-2164-14-670
Feuillet C, Leach JE, Rogers J et al (2011) Crop genome sequencing: lessons and rationales. Trends Plant Sci 16:77–88
Guo G, Dondup D, Zhang L et al (2014) Identification of SNPs in barley (*Hordeum vulgare* L.) by deep sequencing of six reduced representation libraries. Crop J. doi:10.1016/j.cj.2014.06.008
Henry RJ (2012) Next-generation sequencing for understanding and accelerating crop domestication. Brief Funct Genomics 11:51–60
Henry IM, Nagalakshmi U, Lieberman MC et al (2014) Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. Plant Cell. dx.doi.org/10.1105/tpc.113
Holley RW, Apgar J, Everett GA et al (1965) Structure of a ribonucleic acid. Science 147:1462–1465
Huang X, Wang J, Aluru S et al (2003) PCAP: a whole-genome assembly program. Genome Res 13:2164–2170
Hyten DL, Song Q, Fickus EW et al (2010) High-throughput SNP discovery and assay development in common bean. BMC Genomics 11:475. doi:10.1186/1471-2164-11-475
Iqbal Z, Caccamo M, Turner I et al (2012) *De novo* assembly and genotyping of variants using colored de Bruijn graphs. Nat Genet 44:226–232
Jackson SA, Iwata A, Lee SH et al (2011) Sequencing crop genomes: approaches and applications. New Phytol 191:915–925
Jennıngs TN, Knaus BJ, Mullıns TD et al (2011) Multiplexed microsatellite recovery using massively parallel sequencing. Mol Ecol Resour. doi:10.1111/j.1755-0998.2011.03033.x
Kaul SH, Koo L, Jenkins J et al (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815
Kudapa H, Bharti AK, Cannon SB et al (2012) A comprehensive transcriptome assembly of Pigeonpea (*Cajanus cajan* L.) using Sanger and second-generation sequencing platforms. Mol Plant 5:1020–1028
Laurie MT, Bertout JA, Taylor SD et al (2013) Simultaneous digital quantification and fluorescence-based size characterization of massively parallel sequencing libraries. Biotechniques 55:61–67
Leach LJ, Belfield EJ, Jiang C et al (2014) Patterns of homoeologous gene expression shown by RNA sequencing in hexaploid bread wheat. BMC Genomics 15:276. doi:10.1186/1471-2164-15-276
Li R, Li Y, Fangi X et al (2009) SNP detection for massively parallel whole-genome resequencing. Genome Res 19:1124–1132
Li X, Yang Y, Henry RJ et al (2014) Plant DNA barcoding: from gene to genome. Biol Rev. doi:10.1111/brv.12104

Liu H, Bayer M, Druka A et al (2014) An evaluation of genotyping by sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. BMC Genomics 15:104. doi:10.1186/1471-2164-15-104
Malory S, Shapter FM, Elphinstone MS et al (2011) Characterizing homologues of crop domestication genes in poorly described wild relatives by high-throughput sequencing of whole genomes. Plant Biotechnol J 9:1131–1140
Mardis ER (2013) Next-generation sequencing platforms. Annu Rev Anal Chem 6:287–303
Mascher M, Wu S, Amand PS et al (2013) Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. PLoS One 8(10):e76925. doi:10.1371/journal.pone.0076925
Masojc P (2002) The application of molecular markers in the process of selection. Cell Mol Biol Lett 17:499–509
Metzker ML (2010) Sequencing technologies – the next generation. Nat Rev Genet 11:31–46
Michael T (2012) Rapid interrogation of a small plant genome *Sedum album* using the Ion Torrent (PGM). Plant Anim Genom XX. 14–18 Jan 2012, San Diego
Michael TP, Jackson S (2013) The first 50 plant genomes. Plant Genome 6:1–7
Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet 12:443–451
Oliver RE, Lazo GR, Lutz JD et al (2011) Model SNP development for complex genomes based on hexaploid oat using high-throughput 454 sequencing technology. BMC Genomics 12:77
Pang T, Ye Cy, Xia X, Yin W (2013) De novo sequencing and transcriptome analysis of the desert shrub, *Ammopiptanthus mongolicus*, during cold acclimation using Illumina/Solexa. BMC Genomics 14. doi:10.1186/1471-2164-14-488
Pevzner PA, Tang H, Waterman MS (2001) An eulerian path approach to DNA fragment assembly. Proc Natl Acad Sci U S A 98:9748–9753
Poland J, Rife T (2012) Genotyping-by-sequencing for plant breeding and genetics. Plant Genome 5:92–102
Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7(2):e32253. doi:10.1371/journal.pone.0032253
Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195
Quail MA, Smith M, Coupland P et al (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13:341
Rafalski A (2010) Association genetics in crop improvement. Curr Opin Plant Biol 13:174–180
Rahman AYA, Usharraj AO, Misra BB et al (2013) Draft genome sequence of the rubber tree *Hevea brasiliensis*. BMC Genomics 14:75. doi:10.1186/1471-2164-14-75
Ray S, Satya P (2014) Next generation sequencing technologies for next generation plant breeding. Front Plant Sci 5:367. doi:10.3389/fpls.2014.00367
Roorkiwal M, Nayak SN, Thudi M et al (2014) Allele diversity for abiotic stress responsive candidate genes in chickpea reference set using gene based SNP markers. Front Plant Sci. doi:10.3389/fpls.2014.00248
Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94:441–448
Sanmiguel P (2011) Next-generation sequencing and potential applications in fungal genomics. Methods Mol Biol 722:51–60. doi:10.1007/978-1-61779-040-4
Sharon D, Tilgner H, Grubert F, Snyder M (2013) A single-molecule long-read survey of the human transcriptome. Nat Biotechnol 31:1009–1014. doi:10.1038/nbt.2705
Shulaev V, Sargen DJ, Crowhurst RN et al (2011) The genome of woodland strawberry (*Fragaria vesca*). Nat Genet 43:109–116
Spindel J, Wright M, Chen C et al (2013) Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental

mapping and breeding populations. Theor Appl Genet 126:2699–2716. doi:10.1007/s00122-013-2166-x
Stranneheim H, Lundeberg J (2012) Stepping stones in DNA sequencing. Biotechnol J 7:1063–1073
Takagi H, Abe A, Yoshida K et al (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J 74:174–183
Thompson JF, Steinmann KE (2010) Single molecule sequencing with a HeliScope genetic analysis system. Curr Protoc Mol Biol. doi:10.1002/0471142727.mb0710s92
Tinker NA, Chao S, Lazo GR et al (2014) A SNP genotyping array for hexaploid oat (*Avena sativa* L.). Plant Genome. doi:10.3835/plantgenome2014.03.0010
Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641
Torres TT, Metta M, Ottenwalder B, Schlotterer C (2008) Gene expression profiling by massively parallel sequencing. Genome Res 18:172–177
Trick M, Long Y, Meng J, Bancroft I (2009) Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. Plant Biotechnol J 7:334–346
Turcatti G, Romieu A, Fedurcom M, Tairi AP (2008) A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. Nucleic Acids Res 36(4):e25. doi:10.1093/nar/gkn021
Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522–530
Varshney RK, Chen W, Li Y et al (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30:83–89
Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoS Biol 12(6):e1001883. doi:10.1371/journal.pbio.1001883
Velasco R, Zharkikh A, Troggio M et al (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS One 2:e1326
Wicker T, Narechania A, Sabot F et al (2008) Low-pass shotgun sequencing of the barley genome facilitates rapid identification of genes, conserved non-coding sequences and novel repeats. BMC Genomics 9:518
Wilkins TA, Mudge J, Abidi N et al (2009) The sequencing and resequencing of cotton. In: Plant and animal genomes XVII. San Diego
Wu J, Liu B, Cheng F et al (2012) Sequencing of chloroplast genome using whole cellular DNA and Solexa sequencing technology. Front Plant Sci 3:243. doi:10.3389/fpls.2012.00243
Wu X, Li Y, Shi Y et al (2014) Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping. Theor Appl Genet 127:621–631
Xu X, Zeng L, Tao Y et al (2013) Pinpointing genes underlying the quantitative trait loci for root-knot nematode resistance in palaeopolyploid soybean by whole genome resequencing. Proc Natl Acad Sci U S A 110:13469–13474
You FM, Huo N, Deal KR et al (2011) Annotation-based genome-wide SNP discovery in the large and complex *Aegilops tauschii* genome using next-generation sequencing without a reference genome sequence. BMC Genomics 12:1–19
Yu H, Xie W, Li J et al (2014) A whole-genome SNP array (RICE6K) for genomic breeding in rice. Plant Biotechnol J 12:28–37
Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829
Zhou Y, Gao F, Liu R et al (2012) *De novo* sequencing and analysis of root transcriptome using 454 pyrosequencing to discover putative genes associated with drought tolerance in *Ammopiptanthus mongolicus*. BMC Genomics 13:266. doi:10.1186/1471-2164-13-266

Chapter 20
Next-Generation Sequencing (NGS) Tools and Impact in Plant Breeding

Vinay K. Singh, Anil K. Singh, Sakshi Singh, and Brahma D. Singh

Abstract Next-generation sequencing (NGS) of plant genomes provides opportunities to understand the genetic makeup and behavior of plant genomes. Plant whole genome sequencing, transcriptome sequencing and exome sequencing will allow discovering agronomically-important genes that regulate yield and tolerance to biotic and abiotic stresses. The availability of the genome and transcriptome sequence data helps in the development of genetic markers. At present, available NGS tools are powerful enough to provide high-resolution analysis of plant genomes. NGS generates huge amounts of sequenced data in a cost-effective manner and allows profiling for nucleotide variation and large-scale discovery of functional markers. These markers will help in selection of economically-important traits in plant breeding. Plant breeding has been beneficial in developing improved varieties using conventional tools, techniques and methodologies. The availability of NGS tools and online resources is leading to a gene revolution of plant breeding, as they facilitate the study of the genome and its relationship with the phenome for complex traits. The analyses of NGS data allow plant breeders to discover regulatory sequences and their relative positions and subsequent development of molecular markers for marker-assisted selection (MAS).

V.K. Singh (✉)
Centre for Bioinformatics, School of Biotechnology, Faculty of Science, Banaras Hindu University, Varanasi 221005, India
e-mail: vinaysingh@bhu.ac.in

A.K. Singh
Department of Genetics and Plant Breeding, Institute of Agriculture Science, Banaras Hindu University, Varanasi 221005, India

College of Agriculture and Research Station, Korea 497335, Chhattisgarh, India
e-mail: aksingh25@gmail.com

S. Singh
Department of Molecular and Human Genetics, Faculty of Science, Banaras Hindu University, Varanasi 221005, India
e-mail: sakshi.rajput55@gmail.com

B.D. Singh
School of Biotechnology, Banaras Hindu University, Varanasi 221005, India
e-mail: brahmadsingh@gmail.com

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_20

Keywords Abiotic stresses • Biotic stresses • Functional markers • Gene identification • NGS • Plant breeding • Sequencing tools

20.1 Introduction

DNA sequencing is a method to determine the precise order of the four nitrogenous bases, viz. adenine, guanine, cytosine and thymine, within a DNA molecule. The advent of DNA sequencing technologies has greatly accelerated biological, agricultural and medical research, discovery and development. The Maxam-Gilbert sequencing and the chain-termination methods are basic methods of sequencing, first reported in 1977 (Maxam and Gilbert 1977; Sanger et al. 1977). Subsequently, the shotgun sequencing method was developed for analysis of >1000 base pairs (Staden 1979). By this method the target DNA is broken into random fragments and individual fragments which can be used for sequencing, and their sequences can be reassembled on the basis of their overlapping regions. Then, several different DNA sequencing techniques were developed; these sequencing methods are faster, cheaper and have very high throughput. This chapter is devoted to a brief description of some of these technologies that have been commercialized, and examines their application for plant genome and transcriptome sequencing. It also provides the salient features of many of the bioinformatic tools that are used for acquisition, processing and analysis of the huge amounts of sequence data generated by these technologies.

The genome sequences of several plant species, viz. thale grass, rice, grapevine, poplar, papaya, sorghum and many more, are either completed or in the pipeline. The sequences of most crop genomes will probably be in hand by the next decade. This information will seed the in silico birth of plant biology for functional plant genomics, and is likely to reveal relationships between DNA sequence variation and genetic diversity (Hamilton and Buell 2012; Pennisi 2007). The Omics-based approaches e.g. genomics, trascriptomics, metabolomics etc., are very important for disease diagnostics and treatment, and equally important for improvement of crops for food and fuel production (Edwards et al. 2013; Egan et al. 2012; He and Jiao 2014; Henry 2011; Jain 2012; Kelly and Leitch 2011; Schatz et al. 2012; Sucher et al. 2012; Varshney and May 2012). Genomics and transcriptomics have radically altered the scope of genetics by providing a landscape of genes and their epigenetic states, analysis of enormous range of genetic variation, and the potential to measure gene expression with high quality and accuracy. Systems breeding approaches can be used to study the diverse genomic information, based on which phenotypic information can be predicted from genotypic information and thereby accelerate crop-improvement programs to address food security issues (Bevan and Uauy 2013; Edwards and Batley 2010).

Genomics resources are used to derive molecular markers that enable indirect selection for traits that are not very amenable to phenotypic selection. Marker-assisted selection may offer one or more of the following advantages over pheno-

typic selection: ease of implementation, lower costs, faster assays that are independent of the environmental and developmental factors, and increased efficacy and reliability. Markers allow selection for traits like yield, in the greenhouse and off-season nurseries early in the growing season so that the selected plants can be ordered in crosses in the same growing season (Babu et al. 2004; Collard et al. 2005). These factors accelerate the crop development process and the private seed industry is extensively using markers in their breeding programs (Eathington et al. 2007). One of the major limitations of markers has been the cost of genotyping, which is being effectively addressed by NGS technologies that have allowed the development of strategies for detection and genotyping for SNPs in a single step. In addition, reduced representation genotyping approaches are being developed to genotype the individuals of a population at a fraction of the cost that would be incurred if whole genome resequencing were done. Linkage mapping has facilitated positional cloning of several genes and quantitative trait loci (QTLs), are providing the base material for use of these genes in creation of the desired transgenic plants.

20.1.1 Definition of Next-Generation Sequencing

Low-cost sequencing technologies, commonly referred to as of next-generation sequencing (NGS) technologies, produce millions of sequencing reads concurrently (Church 2006). NGS rapidly generates huge amounts of sequence data in a very cost-effective way, and allows profiling for nucleotide variation and large-scale discovery of genetic markers. These markers aid in the indirect selection for economically-important traits based on gene/quantitative trait locus (QTL) mapping and/or genome-wide association studies (GWAS). Furthermore, comparative genomics using NGS data provides better chances of identifying loci under selection. High-throughput sequencing technologies cut down the cost of sequencing, while ultra-high-throughput sequencing (UHTS) technologies are very fast and help to reduce the time required for sequencing (Schuster 2008; Tucker et al. 2009).

20.1.1.1 Massively Parallel Signature Sequencing (MPSS)

Massively parallel signature sequencing (MPSS) was one of the first NGS technologies developed in the early 1990s at Lynx Therapeutics (a company founded by Sydney Brenner and Sam Eletr in 1992). It is a sequence-based approach used to identify and quantify mRNA transcripts present in a given sample. MPSS is similar to serial analysis of gene expression (SAGE), but their biochemical manipulation and sequencing approaches differ quite significantly. MPSS captures data by counting virtually all mRNA molecules in a given sample of tissue/cell. All genes are analyzed simultaneously using bioinformatics tools to sort out the number of mRNA molecules obtained from each gene to the total number of RNA molecules present in the given sample. Counting mRNAs with MPSS is based on the capability

to clearly identify every mRNA in the sample. This is done by generating a 17-base sequence for each mRNA at a specific site upstream of its poly(A) tail (the first DpnII site in double stranded cDNA). The 17-base sequence is then used as an mRNA identification *signature*. To measure the level of expression of any given gene, the total number of signatures for that gene's mRNA is counted. MPSS was generally used for sequencing cDNA for measuring the gene expression levels in plants, e.g. rice MPSS contains 46,971,553 mRNA transcripts from 22 libraries, and 2,953,855 small RNAs from 3 libraries (Nakano et al. 2006; Nobuta et al. 2007), *Arabidopsis* MPSS database currently contains more than 36,991,173, 17-base signature sequences and more than 31,404,553 20-base signatures derived from more than 14 *Arabidopsis* libraries (Meyers et al. 2004).

20.1.1.2 Polony Sequencing

The Polony sequencing method developed by Church in his laboratory at Harvard is performed on the Polonator G.007 system. Polony sequencing was among the first NGS platforms used to sequence the full *Escherichia* coli genome in 2005. It combined an in vitro paired-tag library with emulsion PCR, an automated microscope, and ligation-based sequencing chemistry to sequence *E. coli* genome at of >99.99 % accuracy and a cost approximating 1/9th that of Sanger sequencing. Library preparation for sequencing on the Polonator system is accomplished using emulsion PCR for amplification of template DNA, loading of the beads onto the flow cells and a fully-automated polymerase colony (referred to as a *polony*) sequencing (Shendure et al. 2005) by ligation.

20.1.1.3 454 Pyrosequencing

Margulies et al. (2005) reported a high-throughput pyrosequencing method involving sequencing by synthesis. This platform was created by 454 Life Sciences Corporation and, later acquired by Roche (http://www.roche.com). This method amplifies random DNA fragments inside water droplets in an oil solution (emulsion PCR). Each droplet contains a single DNA template attached to a single primer-coated bead that forms a clone. Each bead in the emulsion serves as an independent PCR reaction, where millions of copies of the original template are produced and bound to the capture bead; these copies serve as templates for the subsequent sequencing reaction. The beads are individually deposited into the wells of a picotiter plate along with DNA polymerase, primers, and the enzymes necessary to create fluorescence by using the inorganic phosphate produced during sequencing. The instrument washes the picotiter plate free from the dNTPs before each new dNTP is added. Sequencing occurs on a flow cell within the picoliter wells, in which addition of a nucleotide to the growing strand by DNA polymerase results in release of a pyrophosphate (PPi). This PPi is detected by an enzymatic luminometric inorganic pyrophosphate detection assay (ELIDA) through the generation of a light signal

following the conversion of PPi into ATP. Pyrosequencing uses luciferase to generate light for detection of the individual nucleotides added to the nascent DNA, and the combined data are used to generate sequence read-outs. This technology provides intermediate read length and price per base compared to Sanger sequencing on one end and Solexa and SOLiD methods on the other. Since 2012, plant genome sequencing has gained momentum using 454 prosequencing due to low cost and high efficiency. A number of plant genomes, including apple, cacao, strawberry, potato, cassava, tomato, melon, barley, bread wheat etc. have been sequenced using this technique (Bevan and Uauy 2013).

20.1.1.4 Illumina (Solexa) Sequencing

In 2008, a high-throughput sequencing-by-synthesis method that utilized reversible dye terminators was developed by Solexa, which was later acquired by Illumina (Bentley et al. 2008). This method of sequencing is based on reversible dye-terminator technology and engineered polymerases developed by the company. In this method, DNA molecules and primers are first attached onto a slide and amplified with polymerase so that local clonal DNA colonies (=DNA clusters) are formed. To determine the sequence, four types of reversible terminator bases (RT-bases) are added, and the un-incorporated nucleotides are washed away before addition of the new dNTP. A camera takes images of the fluorescently-labeled nucleotides, then the dye, along with the terminal 3" blocker, is chemically removed from the DNA, allowing for the next cycle to begin. Unlike 454 sequencing, all four bases are present for the polymerization step and only a single nucleotide is incorporated per cycle. Illumina sequencing allows up to 3 billion reads per run with a read length of 50–300 bp. The Illumina platform has been widely used in plant genomics for expression profiling (Filichkin et al. 2010), de novo sequencing (Dassanayake et al. 2011) and re-sequencing (Cao et al. 2011).

20.1.1.5 SOLiD Sequencing

Sequencing by the Oligo Ligation and Detection (SOLiD) platform was created by Applied Biosystems (now a Life Technologies brand). It employs sequencing by ligation to determine DNA sequence composition. In this instance, pools of all possible oligonucleotides of a fixed length are labeled according to the sequenced position. Oligonucleotides are annealed and ligated; the preferential ligation by DNA ligase for matching sequences results in a signal informative of the nucleotide at that position. Before sequencing, the DNA is amplified by emulsion PCR. The resulting beads, each containing copies of the same single DNA molecule, are deposited on a glass slide. The library preparation in SOLiD is very similar to that of Roche/454 sequencing. This method is often used in resequencing studies (Ashelford et al. 2011), transcriptomics or in genomic sequencing alongside other technologies (Shulaev et al. 2011).

20.1.1.6 Ion Torrent Semiconductor Sequencing

Ion Torrent Systems Incorporated (now owned by Life Technologies) developed a NGS platform based on standard sequencing chemistry, but with a novel, semiconductor-based detection system. The Ion Torrent system is unique among NGS technologies in that the detection for sequencing is not based upon fluorescent dyes, but rather on measuring of the pH change as a result of the release of an H^+ ion upon nucleotide incorporation (Rothberg et al. 2011). A microwell containing a template DNA strand to be sequenced is flooded with a single nucleotide. If the introduced nucleotide is complementary to the leading template nucleotide it is incorporated into the growing complementary strand. This causes the release of one hydrogen ion that triggers a hypersensitive ion sensor, which indicates that a reaction has occurred. If homopolymer repeats are present in the template sequence, multiple nucleotides will be incorporated in a single cycle. This leads to a corresponding number of released H^+ ions and a proportionally higher electronic signal. This method generates sequence of up to 80 million nucleotides per run, the read length is of up to 400 bp and accuracy of 98 %.

20.1.1.7 Heliscope Single Molecule Sequencing

The Heliscope method of single-molecule sequencing was developed by Helicos Biosciences (http://www.helicosbio.com). It was the first single-molecule sequencing system (SMS) to become commercially available (Harris et al. 2008). Heliscope sequencing involves DNA library preparation based on DNA shearing followed by addition of a poly-A tail to the sheared DNA fragments. These poly-A tailed DNA fragments are attached to flow cells through poly-T anchors. Sheared DNA templates are either hybridized to immobilized oligonucleotide primers adhered to a solid surface or directly immobilized by covalent bonding to the solid surface. These DNA templates are then primed with a universal primer (Thompson and Steinmann 2010). The next step is extension-based sequencing with cyclic washes of the flow cell with fluorescently-labeled nucleotides (one type of nucleotide added at a time, as applied in the Sanger method). The reads are performed by the Heliscope sequencer, which generates short reads of up to 28 Gb in a single sequencing run.

20.1.1.8 Single Molecule Real Time (SMRT) Sequencing

SMRT (single-molecule real-time) sequencing is based on the sequencing- by-synthesis approach. According to Pacific Biosciences, the SMRT technology developer, the PacBio platform measures the enzymatic activity of a single DNA polymerase enzyme in real-time using zero-mode waveguides. The DNA is synthesized in zero-mode wave-guides, (ZMWs), small well-like containers with the capturing tools located at the bottom of the well. The sequencing is performed by unmodified DNA polymerase (attached to the bottom of ZMW) and fluorescently-labeled nucleotides flow freely in the solution. The wells are constructed in such a way that only the

fluorescence occurring at the bottom of the well is detected. The fluorescent label is detached from the nucleotide as it is incorporated into the DNA strand. This SMRT sequencing permits tens of thousands of zero-mode waveguides to be captured in a 30 min run. This approach allows the longest read length (reads of 20,000 nucleotides or more) with average read length of 5 kb in a microbial genome resequencing project (Rasko et al. 2011). But its use in de novo sequencing and assembly of large genomes has yet to be demonstrated.

20.1.2 Importance of Plant Genome and Transcriptome Sequencing Studies

One of the foremost objectives of plant genome sequencing is to make it easier and faster for plant breeders to develop new crop varieties that better meet the growing needs for food, fiber and fuel. In 2010, the genome of *Brachypodium distachyon*, a wild, annual grass, native to the Mediterranean and Middle East, with little agricultural importance, was published. So, why sequence it? The reason is that *Brachypodium* is a close relative of grasses, such as wheat, which are critical to world nutrition, but whose massive and complex genomes make them extremely hard to work with. *Brachypodium*, on the other hand, has one of the smallest known genomes among grasses, it is easy to grow in the lab and manipulate genetically, and has a short life-cycle. Thus, by working with *Brachypodium* instead of wheat, scientists can more quickly make advances that can be used to improve vital cereal crops like wheat and oats.

Plant genomics will allow discovery and isolation of important genes and to analyze their functions to allow insight into the genes that regulate yields and tolerance to biotic and abiotic stresses. Genomics-based interactions can aid and accelerate crop improvement, including the improvement of staple and *orphan* crops, and to facilitate the utilization of untapped allelic variation. Thale cress (*Arabidopsis thaliana*) is the plant species whose genome sequence was first to be completed and now serves as the model species in plant molecular biology research due to its small plant and genome size, short generation time and high efficiency transformation (AGI 2000). Later in 2002, the draft genome sequence of both *japonica* and *indica* varieties of rice, an important staple food as well as a model monocotyledon, was completed (Goff et al. 2002; Yu et al. 2002). Whole-genome sequencing provides information about important genomic features, e.g. identification of protein-coding and non-coding genes, gene families, regulatory elements, repetitive sequences, simple sequence repeats (SSRs) and guanine-cytosine (GC) content. High levels of homology and gene synteny based on whole genome sequencing allows us to understand important genomic functions to be applied across plant species by comparison to high-quality reference genome sequences (Henry 2011; Mochida and Shinozaki 2010).

Whole-genome comparisons for identifying chromosomal duplication and conserved synteny among related species provide evidence for hypotheses on comparative evolutionary histories with regard to the diversification of species in a related lineage (Paterson et al. 2009; Schnable et al. 2009). NGS technologies generate vast

amounts of sequence data, which presents many problems to computational biologists, bioinformaticians, and other end-users endeavoring to assemble and analyze NGS data in novel ways (Egan et al. 2012). These data sets have become primary sequence material for the design of genome sequence-based platforms such as microarrays, tiling arrays and molecular markers. Similarly, transcriptome analysis is a significant approach to identify candidate genes, predict gene functions and discover cis-regulatory motifs. The hybridization-based methods, such as microarrays and *gene chips*, are now well established for acquiring large-scale gene expression profiles for various plant species. The recent rapid accumulation of data sets containing large-scale gene expression profiles and the availability of such large repositories of data in the public domain are an efficient and valuable resource for many secondary uses, such as co-expression and comparative analyses. Furthermore, as a NGS application, deep sequencing of short fragments of expressed RNAs, including sRNAs, is quickly becoming an efficient tool for use with genome-sequenced species (De Hoon and Hayashizaki 2008).

20.1.3 Size of Plant Genomes

The genomes of plants vary greatly in size. The smallest known plant genome of merely 63 Mb belongs to the carnivorous corkscrew plant, *Genlisea aurea*; the largest plant genome is 148 Gb (gigabase pairs) of the rare Japanese plant, *Paris japonica*. Of those plant genomes that have been sequenced and published, the smallest so far belongs to a close relative of the corkscrew plant, the bladderwort (*Utricularia gibba*). A carnivorous denizen of nutrient-poor bogs that gets nutrition from feeding on insects, the bladderwort has a genome of 77 Mb. In contrast to this, the largest plant genome sequenced to date is that of the Norway spruce at 20 Gb. For comparison, the genome of the bacterium *Escherichia coli* is about 4.6 Mb, while the human genome is 3.2 Gb, about 1/6th the size of the spruce genome. Despite this great variation in genome size, plants tend to have roughly the same number of genes at about 32,000. Bladderwort, for example, retains a standard number of 28,500 genes, even though its overall genome is very small.

What scientists know from analyzing the sequenced plant genomes is that this broad range in plant genome size appears to be driven by the proliferation of the so called *copy-and-paste long terminal repeat (LTR) retrotransposons*. Retrotransposons are DNA sequences that can copy themselves to RNA and then back to DNA. Their DNA copy may integrate back into the genome, increasing the retroposon size. Retrotransposons are found in humans and other animals, but they are especially abundant in plants. The maize genome, for example, is bloated with 75 % LTRs. Bladderwort's genome, on the other hand, comprises only 3 % LTRs. This is the main reason for the great size range of plant genomes. While people and most other animals are diploid, i.e. contain two sets of chromosomes, one of which is contributed by the mother and the other by the father, many plants species are tetraploid (4 sets of chromosomes) or even hexaploid (6 sets of chromosomes). For example, the

bread wheat (*Triticum aestivum*) has 6 sets of chromosome, which makes its analysis complicated. Thus polyploidy is the other major reason for the large variation in plant genome size.

20.2 The NGS Approaches for Genome Sequencing

At present, the available NGS technologies are powerful enough to provide high-resolution analyses of an entire eukaryotic genome. The popularly-applied strategies for plant genome sequencing include whole genome sequencing (WGS), transcriptome sequencing (RNA-seq) and exome sequencing (Fig. 20.1).

20.2.1 Whole Genome Sequencing (WGS)

Whole genome sequencing, also known as full genome sequencing, complete genome sequencing, or entire genome sequencing, determines the complete DNA sequence of an organism's genome at a single time. In the case of plants, WGS comprises sequencing of chromosomal DNA as well as the DNA contained in the

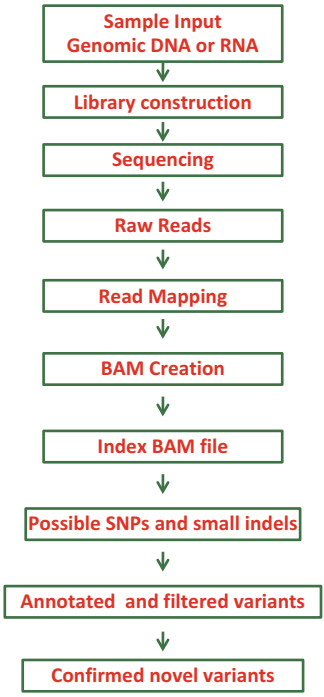

Fig. 20.1 Flow chart for whole genome sequencing (WGS), whole transcriptome shotgun sequencing (WTSS), exome sequencing assembly and annotation. *BAM* Binary (sequence) Alignment Map

mitochondria and chloroplasts. The genome sequences of organisms are fundamentally important for understanding the functions of individual genes and their networks, for defining evolutionary relationships and processes, and for revealing previously unknown regulatory mechanisms that coordinate the activities of genes (Bevan and Uauy 2013). Among plants, *Arabidopsis thaliana* was the first plant species to be targeted for whole genome sequencing; it was successfully completed in 2000 (AGI 2000) using a BAC-by-BAC approach with Sanger sequencing technique. The sequenced genome of *A. thaliana* (119 Mb) is considered as the gold standard for plant genomes due to the high quality and finished nature of the sequence (Hamilton and Buell 2012).

Although, NGS technologies are ideal for resequencing, de novo sequencing can also be undertaken using these technologies. NGS technologies have emerged as highly-efficient methods for generation of whole genome sequences in a much faster, less cumbersome (as compared to BAC by BAC approach) and cost-effective manner. They generate a sequence resource for molecular marker and gene discovery, comparative genomics and genome assembly. NGS enables the discovery of millions of DNA polymorphisms such as SNPs and Indels (insertions/deletions) by comparing the whole genome sequences of several individuals within a species. Presently, SNPs and InDels are preferred marker systems for molecular breeding due to their high frequency, stability, high-throughput capability and cost-effectiveness over other DNA markers (Henry and Edwards 2009). For instance, large-scale SNP markers have been developed in a wide array of crops such as rice (Arai-Kichise et al. 2011; Gopala Krishnan et al. 2011; Yamamoto et al. 2010), wheat (Akhunov et al. 2009), maize (Yan et al. 2010), soybean (Kim et al. 2010) and grapes (Myles et al. 2010).

WGS of the sorghum genome (Mace et al. 2013) revealed untapped genetic potential in this Africa's only indigenous cereal crop; the study involved analysis of high coverage (16–45 x) resequenced genomes of 44 sorghum lines comprising the primary gene pool. It enabled the identification of a relatively large number of DNA markers (8 M high-quality SNPs, 1.9 M Indels) and, specific gene loss and gain events in sorghum. Recently, Wang et al. (2014) reported a high-quality genome sequence of sesame (*Sesamum indicum*) assembled de novo with a contig N50 of 52.2 kb and a scaffold N50 of 2.1 Mb, containing an estimated 27,148 genes. The results reveal novel, independent whole-genome duplication and the absence of the Toll/interleukin-1 receptor domain in resistance genes. Studies on candidate genes and oil biosynthetic pathways contributing to high oil content revealed the expansion of type 1 lipid transfer genes by tandem duplication, the contraction of lipid degradation genes, and the differential expression of essential genes in the triacylglycerol biosynthesis pathway, particularly in the early stage of seed development. Resequencing data of 29 sesame accessions collected across 12 countries suggested that the high genetic diversity of lipid-related genes might be associated with the wide variation in oil content. Furthermore, the results shed light on the pivotal roles of the stage of seed development, oil accumulation and potential key genes for *sesamin* production, an important pharmacological constituent of sesame. Similarly, in capsicum the whole genome sequencing of cultivated and wild peppers along with

resequencing of 20 accessions was reported by Qin et al. (2014). The genome sequence of the cultivated pepper Zunla-1 (*Capsicum annuum*) and its wild progenitor chiltepin (*C. annuum* var. *glabriusculum*) provides approximately 2.75 Gb (79 % of reference) scaffolds containing 34,476 protein-coding genes anchored to 12 linkage groups by 7657 SNP markers in a high-density genetic map. A comparison of the cultivated and wild pepper genomes with 20 resequenced accessions revealed molecular footprints of artificial selection and a number of candidate domestication genes. They also found that the dosage compensation effect of tandem duplication of genes probably contributed to the pungent diversification in pepper.

To the present, genome sequencing projects of a number of plant species have been completed or are underway. Draft genome sequences have been assembled based on NGS technology in several crop species such as *Arabidopsis thaliana* (AGI 2000), *A. lyrata* (Hu et al. 2011), rice (*Oryza sativa*) (Yu et al. 2002), *Zea mays* (Schnable et al. 2009), *Brassica rapa* (Wang et al. 2011b), *Cajanus cajan* (Singh et al. 2011a; Varshney et al. 2011), *Cannabis sativa* (Van et al. 2011), castor bean (Chan et al. 2010; Rivarola et al. 2011), cocoa (Argout et al. 2010), capsicum (Qin et al. 2014), sesame (Wang et al. 2014), *Cicer arietinum* (Jain et al. 2013; Varshney et al. 2013), foxtail millet (Doust et al. 2009), grapevine (Jaillon et al. 2007), soybean (Schmutz et al. 2010), *Populus trichocarpa* (Tuskan et al. 2006), potato (Xu et al. 2011b), *Pyrus bretschneideri* (Wu et al. 2013b), *Sorghum bicolor* (Mace et al. 2013; Paterson et al. 2009), wheat (Brenchley et al. 2012) and tomato (TGI 2012). These genome sequences are now available in public domain; details of some of them are listed in Table 20.1.

20.2.2 Transcriptome Sequencing (RNA-Seq)

The term *transcriptome*, proposed by Charles Auffray in 1996 (Pietu et al. 1999), describes *the complete complement of mRNA molecules generated by a cell or group of cells*. Formerly, mRNA expression was measured by microarray or real-time PCR techniques. The traditional microarray analysis requires all the genes of interest to be known and represented on the microarray. Further it does not generate sequence/genotype data for a particular individual. On the other hand, real-time PCR analysis is quite expensive and cannot be used for a genome-wide survey of gene expression (Mardis 2008a, b). But the rapid and inexpensive NGS methods offer high throughput gene expression profiling, genome annotation and discovery of non-coding RNA. Transcriptome sequencing not only defines the sequences of all transcripts de novo, revealing previously undiscovered genes and splice variants, but also indicates the relative abundance of the various transcripts. The transcriptomics variant of pyrosequencing technology is called short-read massively parallel sequencing or RNA-seq (Denoeud et al. 2008). In recent years, RNA-seq has rapidly emerged as the major quantitative transcriptome profiling system (Wang et al. 2011a). Different companies have developed different RNA sequencing platforms all based on variations of pyrosequencing.

Table 20.1 Details of some complete crop genomes available at NCBI database

Plant species	Bioproject	Genomes size (Mb)	Chromosome complement ($2n=2x$)	Approx no. protein- coding genes
Arabidopsis thaliana	PRJNA116	119.668	10	35,378
Glycine max	PRJNA48389	973.779	40	44,551
Medicago truncatula	PRJNA30099	314.478	16	46,092
Solanum lycopersicum	PRJNA66163	781.509	24	26,365
Populus trichocarpa	PRJNA17973	480.871	38	67,023
Vitis vinifera	PRJNA33471	486.261	38	23,418
Cicer arietinum	PRJNA190909	530.894	16	31,988
Citrus sinensis	PRJNA225998	327.83	18	26,599
Cucumis sativus	PRJNA185270	323.986	14	26,587
Oryza sativa var. *indica*	PRJNA361	426.337	24	37,358
Oryza sativa var. *japonica*	PRJNA13139	391.148	24	35,394
Zea mays	PRJNA10769	2066.91	20	62,681
Sorghum bicolor	PRJNA38691	739.15	20	33,490
Phaseolus vulgaris	PRJNA41439	521.077	22	32,720
Solanum tuberosum	PRJNA13297	844.00	$2n=4x=48$	39,031
Triticum aestivum	PRJEB217	3800.33	$2n=6x=42$	94,000–96,000

Once mRNA is extracted and purified from cells, it is sent to a high-throughput sequencing facility. Here, the RNA is reverse transcribed to create a cDNA library. This cDNA can then be suitably fragmented depending on the platform used for sequencing. Sequencing of the fragments yields raw images, which are converted into short read sequences, and then aligned to the reference genome or transcriptome. The amount of mapped reads is counted and the gene expression level is calculated by peak calling algorithms. Finally, differential gene expression is determined using statistical tests. The most popular technology for RNA-seq has been the Illumina Genome Analyzer and Hi-Seq. Illumina sequencing technology has steadily increased the read length and overall number of reads generated per run. RNA-seq (RNA sequencing), also called *whole transcriptome shotgun sequencing* (WTSS), is a technology that utilizes the capabilities of NGS to reveal a snapshot of RNA presence and quantity from a tissue at a given moment in time. In the last few years, several algorithms have been developed for transtriptome analysis and adapted to specific applications. Now researchers are able to combine a variety of bioinformatics tools to obtain an appropriate analysis system optimized to their requirements (Pagani et al. 2011), though principles of operation as well as advantages and limitations vary (Garber et al. 2011; Pepke et al. 2009; Treangen and Salzberg 2011).

Use of reverse transcriptase to convert RNA into cDNA has been shown to introduce biases and artifacts that may interfere with proper characterization and quantification of transcripts. To overcome this problem, single molecule direct RNA sequencing (DRSTM) technology was developed by Helicos. DRSTM sequences RNA molecules directly in a massively-parallel manner without their conversion to

cDNA or other biasing sample manipulations such as ligation and amplification (Ozsolak et al. 2009).

20.2.3 Exome Sequencing

An exome is the protein coding fraction of the genetic material, i.e. all the exons in the genome, and comprises about 1–2 % of the genome of a species. Exons are short, functionally-important sequences of DNA representing the regions in genes that are translated into protein. Exons are flanked by untranslated regions (UTR) that are usually not included in exome studies. In the human genome there are about 180,000 exons constituting about 1 % of the human genome, i.e. 30 Mb. Exome sequencing (also known as targeted exome capture) is an efficient strategy to selectively sequence the coding regions of the genome as a cheaper but still effective alternative to WGS. Because sequencers can read only a limited number of bases per run, exome sequencing can produce more reads quickly, at greater resolution and lower cost in comparison to WGS. The recent advent of NGS technology permits the examination of the whole genome in a comprehensive and unbiased manner. Genetic variations in the coding exons are of significant interest as a majority of Mendelian disorders are caused by mutations in the protein coding sequences (Stenson 2009). A significant decrease in the cost and time associated with exome sequencing compared to WGS has made exome capture a preferred method for the identification of genes involved in disease/stress tolerance.

In allohexaploid (AABBDD) wheat, the development of molecular markers has, until recently, been problematic due to the presence of homoeologous and paralogous copies of the various genes (Kaur et al. 2012). Many of the recently developed genotyping platforms rely on the identification of single nucleotide polymorphisms (SNPs) (Paux et al. 2011). To overcome the various problems associated with SNP generation, characterization and most importantly validation in wheat, many researchers have previously used NGS technology to identify and map relatively large numbers of gene-based SNP loci (Akhunov et al. 2009; Allen et al. 2013; Chao et al. 2010). However, these studies used cDNA and EST sequences and were, therefore, subject to variation in expression of homoeologous and paralogous genes. Genomic DNA is likely to be a more reliable source of putative SNPs; however, the size of the wheat genome makes WGS of multiple varieties to the depths required for successful SNP identification impractical, time consuming and costly (Biesecker et al. 2011). In view of the above, Allen et al. (2013) used a recently-developed sequence capture targeted resequencing approach to characterize a significant proportion of the wheat exome (Winfield et al. 2012). In this study, the exomes of the UK wheat varieties Alchemy, Avalon, Cadenza, Hereward, Rialto, Robigus, Savannah and Xi19 were captured using the NimbleGen capture array (NimbleGen array reference 100819_Wheat_Hall_cap_HX1) as described in Winfield et al. (2012). This generated between 9.8 and 48.7 million reads on the Illumina GAIIx platform. Sequence data were analyzed to discover 95,266 putative SNPs in 26,551

distinct reference sequences (Winfield et al. 2012). Examination of these SNPs suggested that 10–20 % of them were co-dominant.

NGS analyzes millions of short DNA fragments during one sequencing run. The read length depends on the type of NGS platform and can be in the range of 25–450 bp. While these reads are much shorter than those created by Sanger sequencing, NGS has a higher throughput and creates data sets of up to 50 Gb per run (Nowrousian 2010). This demands improved algorithms that are capable of processing the huge amounts of raw sequence data. Comparison of the transcriptome sequences of different tissues can be used to follow changes in gene expression through development, in different tissues or in different genotypes to understand genetic differences that are due to differential gene expression (Gillies et al. 2012). The genetic variation within commercial crop varieties is usually not well characterized or quantified, and the effect of intra-varietal genetic variation on crop performance under stress is poorly understood. Transcriptome sequencing allows genome-wide analysis of large, complex plant genomes and identifies biologically significant SNPs. Genome-wide marker sets, such as those developed from transcriptome sequencing may become a valuable and complementary addition to the plant breeding toolbox and aid both selection and management of diversity within species. The transcriptome sequences of the following crop species are available in the public domain: buckwheat (*Fagopyrum esculentum*) (Logacheva et al. 2011), barbados nut (*Jatropha curcas*) (Natarajan and Parani 2011), pigeon pea (*Cajanus cajan*) (Dubey et al. 2011), chickpea (*Cicer arietinum*) (Agarwal et al. 2012; Garg et al. 2011; Hiremath et al. 2011; Jhanwar et al. 2012; Kudapa et al. 2014), lentil (*Lens culinaris*) (Kaur et al. 2011; Verma et al. 2013), rubber tree (*Hevea brasiliensis*) (Xia et al. 2011b), watermelon (*Citrullus lanatus*) (Guo et al. 2011), Hall's panicgrass (*Panicum hallii*) (Meyer et al. 2012), groundnut (*Arachis hypogaea*) (Zhang et al. 2012a), switchgrass (*Panicum virgatum*) (Wang et al. 2012b), mustard greens (*Brassica juncea*) (Sun et al. 2012a), garlic (*Allium sativum*) (Sun et al. 2012b), safflower (*Carthamus tinctorius*) (Lulin et al. 2012), wheat (*Triticum aestivum*) (Duan et al. 2012; Oono et al. 2013), European yellow lupine (*Lupinus luteus*) (Parra-González et al. 2012), pepper (*Capsicum annuum*) (Ahn et al. 2014; Ashrafi et al. 2012), bananito (*Musa acuminate*) (Passos et al. 2013), coconut palm (*Cocos nucifera*) (Fan et al. 2013), upland cotton (*Gossypium hirsutum*) (Dubey et al. 2013), winter rocket (*Barbarea vulgaris*) (Wei et al. 2013), field penny-cress (*Thlaspi arvense*) (Dorn et al. 2013) and gold-of-pleasure (*Camelina sativa*) (Mudalkar et al. 2014).

20.3 Next Generation File Formats for Sequence Data

20.3.1 SAM/ BAM Format

SAM (Sequence Alignment Map) is a generic format used for storing large nucleotide sequence alignments. A SAM file (.sam) is a tab-delimited text file that contains sequence alignment data. The SAM format is a text format for storing sequence data

in a series of tab-delimited ASCII (American Standard Code for Information Interchange) columns. This format is flexible enough to store all the alignment information generated by various alignment programs. It can be generated simply by alignment programs or converted from existing alignment formats. While BAM file (.bam) is the compressed binary version of the SAM format, i.e. a compact and index-able representation of nucleotide sequence alignments. Most of the next-generation sequencing and analysis tools use SAM/BAM format (Li et al. 2009).

20.3.2 FASTQ Format

FASTQ format is a common text-based format for nucleotide sequence with quality scores. The sequence letter and quality score are encoded with a single ASCII character. FASTQ format was originally developed at the Wellcome Trust Sanger Institute, United Kingdom (Cock et al. 2010).

20.4 Next Generation Sequencing Tools

A large number of bioinformatics tools have been developed for various operations related to processing and analysis of sequence data sets (Table 20.2). Some of the important and widely used tools are briefly described here, and certain others are listed in Table 20.2.

20.4.1 Tools for Verification, Quality Control and Filtering of Data

ABySS (Assembly by Short Sequencing) An assembler for short read sequence data, ABySS is a parallel, paired-end sequence assembler, useful for assembling genomes of >100 Mb (Simpson et al. 2009).

AMOS (A Modular Open-Source Assembler) AMOS is designed for genome assembly using modular approach. It includes different tools for sequence assembly and analysis. Some of the de novo assemblers Bambus2, Minimus and Minimo are available with AMOS to handle polymorphic and metagenomic data (Treangen et al. 2011b).

Clean reads Clean reads can trim poor quality regions and adaptors. Clean reads also filters out the reads that do not meet a minimum quality criteria based on the sequence length and the mean quality. It uses several algorithms and third party tools (lucy, blast, mdust and trimpoly) to carry out the cleaning. The functionality

Table 20.2 Various NGS tools along with their references

Tool's name	Function classes and tool details	References
	Verification, quality control and filtering data	
Arapan-S	A whole-genome assembly software for smaller genomes, based on the de Bruijn graph and uses short reads; tested on raw data retrieved from the NCBI trace archive	Sahli and Shibuya 2012
ALLPATHS	An assembly tool predecessor of ALLPATHS-LG works on ~30 base read lengths	Butler et al. 2008; Maccallum et al. 2009
ARACHNE	An assembly tool designed for long Sanger-chemistry reads	Batzoglou et al. 2002;; Gnerre et al. 2009; Jaffe et al. 2003
ALLPATHS-LG	A short-read assembler used for both small and large (mammalian size) genomes	Gnerre et al. 2011
Cutadapt	A tool for removal of adapter sequences. It is basically written in Python	Martin 2011
Edena	A tool for short-reads assembly	Hernandez et al. 2008
Euler	A powerful tool for fragment assembly	Pevzner et al. 2001
Euler-sr	A tool used for analysis of the read length and assembly of mate-paired short reads	Chitsaz et al. 2011
Flexbar	A tool used for removal of adapter sequences, filtering features and trimming	Dodt et al. 2012
FreClu	A tool for improvement of alignment file; clustering methodology can be used for trimming of short reads and error correction	Qu et al. 2009
Geneious	An extendable desktop and integrated software platform for the analysis and organization of sequence data	Kearse et al. 2012
HTSeq	A python-based framework for analysis of high-throughput sequencing data	Anders et al. 2014
MaSuRCA	MaSuRCA Maryland super read – celera assembler is a genome assembler	Zimin et al. 2013
PASHA	PASHA is a parallel short-read assembler for large genomes using de Bruijn graphs	Liu et al. 2011
QRQC	Quick Read Quality Control (QRQC) is an R package used for read quality-control analysis	Perkins et al. 2013
Ray	A tool to assemble reads obtained from combinations of NGS platforms	Boisvert et al. 2010

(continued)

Table 20.2 (continued)

Tool's name	Function classes and tool details	References
RNA-SeQC	A Java-based tool for quality control and optimization; provides three types of quality control: read counts, coverage and expression correlation. It can take one or more BAM files as input, and generate output as HTML reports	DeLuca et al. 2012
RSeQC	RSeQC is a tool for strand specificity, sequence quality and read distribution analysis. It accepts different input files like chromosome size file in plain text file or two-column or SAM, BAM, FASTA and BED files	Wang et al. 2012a
SAMStat	A program to generate quality reports at different phases of the data process. SAMStat evaluates unmapped, poorly mapped and accurately-mapped sequences independently to infer possible reasons for bad mapping	Lassmann et al. 2011
SEECER	An error correction tool for RNA-seq data sets; uses the raw read sequences produced by NGS platforms. It removes mismatch and indel errors to produce the best quality of the de novo transcriptome assembly	Le et al. 2013
Sequencher	A program developed by Gene Codes Corporation, Ann Arbor, Michigan; used for contigs assembly and DNA sequence analysis	Tippmann 2004
ShortRead	A package written in the R programming language for manipulation of NGS data, exploration and quality assessment	Morgan et al. 2009
SGA	A grammar-based alignment algorithm based on statistical estimation called super-pairwise alignment (SPA)	Hu et al. 2007
SOPRA	A Scaffolding algorithm for paired reads via statistical optimization	Dayarian et al. 2010
SparseAssembler	A tool utilizing a new sparse k-mer graph structure evolved from the de Bruijn graph	Ye et al. 2012
SPAdes	A new assembler for both standard multicell and single-cell assembly	Bankevich et al. 2012
Taipan	An algorithm for fast hybrid short read fragment assembly	Schmidt et al. 2009
VCAKE	Verified Consensus Assembly by K-mer Extension (VCAKE) is a modification of simple k-mer extension that overcomes error by using high depth coverage	Jeck et al. 2007
	Pre-processing data analysis tools (performed before alignment)	
DeconRNASeq	A package written in R language for deconvolution of heterogeneous tissue based on mRNA-seq data	Gong and Szustakowski 2013

(continued)

Table 20.2 (continued)

Tool's name	Function classes and tool details	References
FLASH	A tool for read pre-processing; combines paired-end reads which overlap and converts them to single-long reads	Magoč and Salzberg 2011
IDCheck	This tool can be used for sample identity check of RNA sequencing data	Huang et al. 2013
	Alignment tools	
GNUMAP	A tool based on Needleman-Wunsch method for alignment of sequence data. It can be used for alignment in repetitive regions of a genome	Clement et al. 2010
Maq	A versatile and user-friendly aligner, which aligns reads to reference sequences and performs a consensus	Li et al. 2008a
Mosaik	Mosaik based on Smith-Waterman algorithm, aligns reads containing short gaps, deletions and insertions	Lee et al. 2014
NovoAlign	Based on Needleman-Wunsch algorithm, this short aligner is used for NGS data derived from Illumina platform. It is able to deal with bisulphite data and give output in SAM format	Yu et al. 2012
RazerS	A program that allows users to align sequencing reads of arbitrary length. RazerS can use either the edit distance or Hamming distance method	Weese et al. 2009, 2012
SEAL	System for Easy Analysis of Lots (SEAL) tool uses BWA to perform alignment and Picard Mark duplicates to detection and removal of duplicate read	Walker and Koonin 1997
ZOOM	Zillions of oligos mapped (ZOOM) is a commercial short read aligner	Lin et al. 2008
	Spliced aligners	
Erange	A tool for data quantification to transcriptomes and alignment	Mortazavi et al. 2008
OSA	Omicsoft sequence aligner (OSA) is a fast alignment tool for RNA-seq data with accuracy. Benchmarked with existing methods, Omicsoft sequence aligner improves mapping speed 4 to 10-fold with less false positives and better sensitivity	Hu et al. 2012
RNA-MATE	RNA-MATE is applied biosystems SOLID system pipeline used for alignment of NGS data; useful in quality control management and trimming of reads. RNA-MATE tool allows tag visualization of alignments	Cloonan et al. 2009
RNASEQR	RNASEQR analyzes raw sequence data obtained from RNA-seq experiments effectively and outputs results in a manner that is compatible with a wide variety of specialized downstream analyses on desktop computers	Chen et al. 2012

(continued)

Table 20.2 (continued)

Tool's name	Function classes and tool details	References
SpliceSeq	A tool for identification of potential functional change that results from splice variation and provides a clear view of alternative splicing. SpliceSeq aligns reads to gene splice graphs and complex transcript variants	Ryan et al. 2012
X-Mate	An advanced version of RNA-MATE; capable of mapping both DNA and RNA-seq data sets; improved performance, configuration files, output file formats and flexibility in core mapping software	Wood et al. 2011
	De novo splice aligners	
ABMapper	Basically developed for exploring all probable locations of reads that are mapped to splice junctions or repetitive in nature	Lou et al. 2011
ContextMap	A stand-alone program and context-based approach for identification of most likely mapping for RNA-seq experiments. The main function of this tool is to consider reads in gene expression context and alignment improvement	Bonfert et al. 2012
CRAC	An integrated approach for analysis of RNA-seq reads. It can be used for analysis of reads that integrates genomic locations and local coverage. This tool is able to detect candidate mutations, indels, fusion or splice junctions in single reads	Philippe et al. 2013
GSNAP	Genomic Short-read Nucleotide Alignment Program (GSNAP) is a program for single- and paired-end reads alignment for both short (14 nt) and long reads. GSNAP is able to detect interchromosomal splicing in individual reads	Wu and Nacu 2010
HMMSplicer	A tool for non-canonical and canonical detection of splice junctions in RNA-seq data	Dimon et al. 2010
OLego	A program to score exon junctions by intron size and splice-site strength; implemented in C++ and allows fast processing of large-scale data	Wu et al. 2013a
Pass	A tool for aligning short sequences basically aligns gapped and ungapped reads based on Smith-Waterman and Needleman-Wunsch algorithms; performs alignment in 3 stages. (1) scanning positions of seed sequences in the genome, (2) testing the contiguous regions and (3) refining the alignment	Campagna et al. 2009
PASSion	It can investigate junctions and permit discovery of differential and shared splicing patterns among multiple samples	Zhang et al. 2012b

(continued)

Table 20.2 (continued)

Tool's name	Function classes and tool details	References
PASTA	A highly sensitive and efficient tool for identifying splicing junctions; provides highly annotated output files about characteristics and their locations. PASTA is useful for large-scale investigation of alternative splicing and transcription	Tang and Riva 2013
QPALMA	A tool for optimal spliced alignments of short-sequence reads; can be used for prediction of splice junctions based on machine-learning algorithms	De Bona et al. 2008
SuperSplat	An application software for discovering potential splice junctions in high throughput sequencing data. The Supersplat splits each read in all possible combinations for alignment	Bryant et al. 2010
Subread	Subread is superfast and powerful read aligner tool used to determine the mapping location on the basis of largest mappable region. It automatically decides whether the read should be locally or globally mapped	Liao et al. 2013b
MapNext	An software tool for unspliced and spliced alignments, and SNP detection of short sequence reads	Bao et al. 2009b
STAR	An ultrafast universal aligner implemented in C++; detects canonical, non-canonical splice and fusion (chimeric) transcripts; mainly involved in alignment of long reads derived from NGS technologies	Dobin et al. 2013
TrueSight	A tool for splice-junction detection. Both real data evaluations and simulations showed that TrueSight achieved higher specificity and sensitivity than other methods	Li et al. 2013b
TopHat	A software for detection of de novo junctions. TopHat aligns reads using Bowtie and Maq tool. The Bowtie tool aligns unspliced reads. After the alignment, reads are assembled using Maq tool	Trapnell et al. 2009
	Quantitative analysis and differential expression	
ALDex	An ANOVA-like differential expression method for identification of genes with between- to within-condition differences	Fernandes et al. 2013
Alexa-Seq	A program to study transcript specific expression analysis and quantitative alternative analysis. It can be used for gene expression analysis and exon junction expression studies	Griffith et al. 2010
BaySeq	A package for differential expression analysis based on Bayesian statistics.	Hardcastle and Kelly 2010

(continued)

Table 20.2 (continued)

Tool's name	Function classes and tool details	References
CEDER	A tool for detection of differentially-expressed genes	Wan and Sun 2012
DEB	A web-interface for identification of significantly expressed transcripts and comparisons of the reads or genes	Yao and Yu 2011
DESeq	A bioconductor package to perform differential gene expression analysis based on negative binomial distribution	Anders and Huber 2010
DEGSeq	A package for identification of differentially-expressed genes or isoforms from RNA-seq data from different samples	Wang et al. 2010b
DiffSplice	A genome-wide analysis tool for visualization and detection of differential expression and study of differential splicing modules; also used for identification of alternative-splicing events that diverge in the different isoforms	Hu et al. 2013
DEXSeq	A statistical package to test differential exon usage in RNA-seq data. It facilitates the discovery on a genome-wide scale of regulation and function of alternative-exon usage	Anders et al. 2012
EBSeq	EBSeq is useful in identification of differentially-expressed isoforms between two or more biological conditions	Leng et al. 2013
EdgeR	A package for differential expression analysis of digital gene expression data derived from different DNA sequencing methods such as ChIP-Seq or SAGE, RNA-seq	Robinson et al. 2010
ERANGE	A useful tool for alignment, quantification and normalization of expressed genes	Mortazavi et al. 2008
FeatureCounts	An efficient read quantifier and part of the SourceForge Subread package and Rsubread package	Liao et al. 2013a
FDM	The flow difference metric is a statistical method for identification of differential transcription using RNA-seq data and identifies regions between pairs of splice graphs, without requirement of an underlying gene model or transcripts catalog	Singh et al. 2011b
MATS	A flexible and effective approach for investigation of differential alternative splicing from RNA-seq data	Shen et al. 2012
MMSEQ	A tool for estimating allelic imbalance and isoform expression in diploid organisms. The pipeline contains different tools like Bowtie, ArrayExpressHTS, TopHat and SAMtools	Turro et al. 2011
Myrna	A pipeline tool that runs in elastic map reduce environment (cloud platform) for estimating differential gene expression in RNA-seq datasets	Langmead 2010

(continued)

Table 20.2 (continued)

Tool's name	Function classes and tool details	References
NEUMA	A tool to estimate RNA abundances using length normalization based on mRNA isoform models and uniquely aligned reads. NEUMA uses known transcriptome data available in databases like NCBI Reference Sequence Database RefSeq	Lee et al. 2011
NOISeq	A novel data-adaptive approach for the identification of differentially expressed genes from count data that aims to be robust against the number of available reads	Tarazona et al. 2011
NPEBseq	A tool for nonparametric empirical differential expression analysis using Bayesian statistics	Bi and Davuluri 2013
RNAeXpress	A Java based graphical user interface for read counting and used for GTF comparison and feature detection	Forster et al. 2013
RseqFlow	A virtual machine with all the necessary software, which eliminates complex configuration and installation steps	Wang et al. 2011c
RSEM	A user-friendly software package for isoform abundances and quantifying gene from paired-end or single-end data is requires	Li and Dewey 2011
Scotty	A web server to estimate the number of replicates, to measure differential gene expression and depth of sequencing required to call differential expression	Busby et al. 2013
SpliceTrap	A program to quantify exon inclusion levels and estimation of expression-level of each exon. SpliceTrap is useful in identification of major classes of alternative-splicing events	Wu et al. 2011
Splicing compass	A method to predict genes, which are differentially spliced	Aschoff et al. 2013
	Open source solutions	
ArrayExpressHTS	A package that permits preprocessing, estimation of expression and quality assessment. ArrayExpressHTS may be run locally or through cloud computing	Goncalves et al. 2011
EasyRNASeq	An R package that simplifies the processing of RNA sequencing data, hiding the complex interplay of the required packages behind a single functionality	Delhomme et al. 2012
Galaxy	A workbench for computational biologists. There are several publicly accessible Galaxy servers that contain several RNA-seq tools	Blankenberg et al. 2010
GENE-Counter	A perl pipeline for RNA-seq differential gene expression analysis. It performs alignments using Bowtie, BWA or CASHX and other SAM output aligners	Cumbie et al. 2011

(continued)

Table 20.2 (continued)

Tool's name	Function classes and tool details	References
GenePattern	It offers integrated software for RNA-seq analysis developed by the Broad Institute	Kuehn et al. 2008
GeneProf	A freely-accessible software, easy to use, analysis pipelines for ChIP-seq experiments and RNA-seq data	Halbritter et al. 2014
MultiExperiment Viewer (MeV)	A Java-based microarray data analysis tool for clustering, visualization and classification. It includes a variety of tools to execute tasks like gene set enrichment analysis and *t*-test or significance analysis	Chu et al. 2008
NGSUtils	A suite of software tools for manipulating data derived from NGS experiments; uses such as BED, FASTQ and BAM format files. These tools provide a modular and stable platform for data analysis and management	Breese and Liu 2013
RobiNA	A user friendly graphical interface to work with bioconductor packages. It provides a diversity of quality control techniques and possibility to produce many plots and tables detailing results for differential expression	Lohse et al. 2012
Taverna	A graphical user interface based tool, which helps in the composition and enactment workflow	Oinn et al. 2004
	Alternative splicing analysis	
Asprofile	A suite of programs for quantifying, comparing and extracting alternative splicing events from RNA-seq data	Florea et al. 2013
	Bias correction	
SysCall	A tool for identification, classification and correction of systematic error in high-throughput sequence data	Meacham et al. 2011
	Fusion genes/chimeras/translocation finders/ structural variations	
ChimeraScan	ChimeraScan is used for the discovery of chimeric transcription between two independent transcripts in high-throughput transcriptome sequencing data	Iyer et al. 2011
EBARDenovo	Extension, Bridging And Repeat-sensing Denovo (EBARDenovo) is for chimera-detection tool and resolves the complications of RNA-seq assembly due to mainly repetitive sequences, sequencing errors and aberrant chimeric amplicons	Chu et al. 2013
FusionHunter	A tool for identification of fusion transcripts without depending on known annotations. FusionHunter uses Bowtie tool for alignment	Li et al. 2011a
TopHat-Fusion	A tool based on TopHat program. It was developed to handle reads from fusion genes. It does not require information about known genes; uses the Bowtie tool for alignment of continuous reads	Kim and Salzberg 2011

(continued)

Table 20.2 (continued)

Tool's name	Function classes and tool details	References
ViralFusionSeq	A high-throughput sequencing tool for reconstruction of single-base resolution transcripts	Li et al. 2013a
DeFuse	A software package for gene fusion discovery using RNA-seq data detect fusion sequences with better sensitivity than previously reported methods	McPherson et al. 2011
	RNA-seq simulators	
BEERS	A Benchmarker for Evaluating the Effectiveness of RNA-seq Software (BEERS); a simulation tool for generating RNA-seq data	Grant et al. 2011
Dwgsim	A tool to generate simulated reads according to their needs	Torri et al. 2012
Flux simulator	A computer pipeline simulation technique to mimic a RNA-seq experiment	Griebel et al. 2012
RSEM	A user-friendly software for quantification of transcript abundances from RNA-seq data	Li and Dewey 2011
	Transcriptome assemblers	
iReckon	A tool for estimating the abundance of novel and known isoforms; has a superior ability to discover novel isoforms with a significantly reduced number of false-positive predictions	Mezlini et al. 2013
IsoLasso	A new transcriptome assembly tool for RNA-seq data; based on the well-known LASSO algorithm; a multivariate regression method to maximize; uses prediction accuracy. It is able to make the set of assembled transcripts	Li et al. 2011b
Scripture	A method for transcriptome reconstruction that relies solely on an assembled genome and RNA-seq reads to build a transcriptome ab initio. Scripture contains modules for ChIP-Seq peak calling	Guttman et al. 2010
SLIDE	A tool for incorporating transcriptomic data such as CAGE, RACE, and EST into its model to increase isoform discovery accuracy. SLIDE can also work downstream of RNA-seq assembly algorithms to integrate newly discovered genes and exons	Li et al. 2011c
Rnnotator	An automated software pipeline that generates transcript models without the need for a reference genome by de novo assembly of RNA-seq data	Martin et al. 2010
STM	Scaffolding using Translation Mapping (STM) is a method that uses mapping against the closest available reference proteome	Surget-Groba and Montoya-Burgos 2010

(continued)

Table 20.2 (continued)

Tool's name	Function classes and tool details	References
	Visualization tools	
GBrowse	GBrowse supports NGS data by providing for the direct display of SAM and BAM sequence alignment files. BAM/SAM tracks provide semantic zooming and support both remote and local data sources	Stein 2013
Tablet	An analysis software and high-performance graphical viewer; provides quality based visualizations data in stacked views. It is memory efficient and supports a number of formats	Milne et al. 2010
Samscope	A fast OpenGL interface of SAMSCOPE provides instantaneous browsing of complex data at all levels and can be used for multiple experiments	Popendorf and Sakakibara 2012
SeqMonk	SeqMonk is capable of importing mapping information in various mapping formats or as a tab-delimited format. SeqMonk is a program to enable the analysis and visualization of mapped sequence data. It was written for mapping of NGS data	Chatterjee et al. 2012
Vespa	Visual Exploration and Statistics to Promote Annotation (VESPA) is an interactive visual analysis software to assist scientists in the annotation of prokaryotic genomes through the integration of transcriptomics and proteomics data with genome location coordinates	Peterson et al. 2012
	Commercial solutions	
DNASTAR	A bioinformatics software for DNA, RNA, and protein sequence analysis, including next-generation sequence assembly	Burland 2000

offered by clean reads is similar to the cleaning capabilities of NGS backbone pipeline (Garcia-Mas et al. 2012).

Condetri ConDeTri is content dependent read trimming. It can process paired-end and single-end sequencing data of arbitrary length. To get better results from assemblies, ConDeTri is able to remove reads having low quality scores. Basically, large genomes or low coverage sequencing projects gain quality data from trimming reads (Smeds and Künstner 2011).

FastQC FastQC is a Java-based tool for high-throughput sequence data analysis developed by the Babraham Institute. It helps in quality control checks for raw sequence data and imports the data in BAM, SAM or FastQC formats and exports the result reports in HTML format. This tool also provides an overview for summary graphs, problematic areas and tables for rapid assessment of next-generation sequence data (Andrews 2011).

FASTX Toolkit is a set of tools for manipulating reads in FASTA or FASTQ formats. FASTX can be used for conversion of FASTQ format into FASTA format. It is a command-line tool to process the files before read mapping. Furthermore, Bowtie tool can be used for mapping. FASTX can be used for conversion of DNA/RNA sequences, quality analysis, removal of sequencing adapters, and for filtering and cutting sequences.

HTSeqTools HTSeqTools is a bioconductor package, which can be used for data processing, visualization and quality control. HTSeqTools helps in visualization, correction of strand bias, removal of over-amplification artifacts and visualization of hits (Planet et al. 2012).

IDBA (Iterative De Bruijn graph short read Assembler) IDBA is able to assemble both low-expressed and high-expressed transcripts. It outperforms existing assemblers in terms of specificity and sensitivity for both real and simulated data (Peng et al. 2013).

Newbler Newbler is a software package for de novo DNA sequence assembly, designed specifically for assembling sequence data generated by the 454 GS-series of pyrosequencing platforms (Mundry et al. 2012).

Phrap Phrap is widely used software for DNA sequence assembly; it is part of the Phred-Phrap-Consed package. It was originally developed by Phil Green for the assembly of cosmids in large-scale cosmid shotgun sequencing within the Human Genome Project (De la Bastide and McCombie 2007).

PRINSEQ PRINSEQ can be used for rapid and easy quality control and data preprocessing of metagenomic and genomic data sets. Detailed statistics of FASTQ or FASTA files are generated in graphical and tabular forms (Schmieder and Edwards 2011).

SHARCGS SHARCGS (SHort-read Assembler based on Robust Contig extension for Genome Sequencing) is capable of assembling millions of very short reads; it virtually never generates misassemblies and copes with sequencing errors (Dohm et al. 2007).

SSAKE SSAKE is a tool for assembling millions of short nucleotide sequences by progressively searching through a prefix tree for the longest possible overlap between any two sequences (Warren et al. 2007).

TIGR Assembler The TIGR Assembler from The Institute for Genomic Research (TIGR) is the sequence assembly program used in genome sequencing projects. Development of the TIGR Assembler was based on the experience obtained in more than 20 sequencing projects completed at TIGR (Pop and Kosack 2004).

Trimmomatic Trimmomatic is a flexible trimmer for Illumina sequence data and works with FASTQ reads. It is used to cut adapters and bases in optional positions based on quality thresholds and cut reads.

Phusion assembler The Phusion assembler takes WGS reads, mostly paired with known insert sizes, as input along with quality score assigned for each base and produces a set of supercontigs or scaffolds (Mullikin and Ning 2003).

20.4.2 Alignment Tools

After control assessment of sequence data, analysis involves alignment of the sequenced reads to reference genomes, for which different tools can be used. Some of the sequence alignment software and high-throughput sequencing mappers are listed below.

20.4.2.1 Unspliced Aligners

Short read aligners are able to align continuous reads to a reference genome. Generally, there are two types of short read aligners: (1) Burrows-Wheeler transform method such as BWA and Bowtie, and (2) Needleman-Wunsch/Seed-extend methods or Smith-Waterman algorithm aligners. The first group (BWA and Bowtie) is many times faster, more sensitive than the second group and generates more aligned reads correctly.

BFAST BFAST is an alignment tool for large scale genome resequencing. It is sensitive to insertions, deletions and errors. BFAST is based on the Smith-Waterman algorithm (Homer et al. 2009).

Bowtie Bowtie is an ultrafast, memory-efficient short read aligner. It aligns short DNA reads to the human genome at a rate of over 25 million 35-bp reads per hour. Bowtie indexes the genome with a Burrows-Wheeler index to keep its memory footprint small, which is typically about 2.2 Gb for the human genome (2.9 GB for paired-end reads) (Lagmead 2010).

Burrows-Wheeler Aligner (BWA) BWA is a tool for aligning sequencing reads against reference genomes, based on the algorithms Burrows-Wheeler transform and Smith-Waterman method. BWA allows detection of deletions and insertions. In the case of BWA, the output file is in SAM format (Li and Durbin 2009).

Short Oligonucleotide Analysis Package (SOAP) SOAP is a command-driven package compatible with numerous applications, including single-read or pair-end resequencing. It can be used in mRNA tag sequence mapping and small RNA discovery. SOAP contains a batch module for multiple query sets and supports multi-threaded parallel computing (Li et al. 2008b).

SeqMap SeqMap is a tool used for mapping large numbers of short sequences. SeqMap supports different formats including FASTA. A complex mapping can be done in a few hours using SeqMap (Jiang and Wong 2008).

SHRiMP SHRiMP contains two main techniques for alignment of short reads. Q-gram filtering technique is based on multiple seeds, and is used for candidate regions identification. Smith-Waterman method is the second technique; it can be used for investigation of potential regions (Rumble et al. 2009).

Stampy Stampy is a statistical algorithm for mapping of Illumina sequence reads. Stampy was developed for alignment of reads and detection of the sequence variation like deletions and insertions. Stampy gives the output file in SAM format (Lunter and Goodson 2011).

20.4.2.2 Spliced Aligners

IsoformEx IsoformEx estimates gene expression levels and transcript expression levels from mRNA-Seq data. IsoformEx parses Bowtie alignment files and generates two files (1) expression levels of all transcripts, (2) expression levels of all genes (Kim et al. 2011).

MapAL MapAL is a tool that builds on the established programs Bowtie and Cufflinks for RNA-seq expression profiling. In the post-processing of RNA-seq reads, MapAL incorporates gene models already at the stage of read alignment, consistently increasing by 50 % the number of reliably measured known transcripts (Labaj et al. 2012).

RUM RNA-Seq unified mapper (RUM) tool is used for alignment using Bowtie and BLAT pipeline. It is able to manipulate reads with splice junctions. It performs alignment of unmapped sequences to the reference genome using the BLAT tool. The input files for RUM are in FASTQ or FASTA format, but the output files are in SAM and RUM formats (Grant et al. 2011).

SAMMate SAMMate tool allows quick processing of SAM/BAM files. The SAMMate tool is compatible with both single-end and paired-end sequencing data. It also automates some standard procedures in RNA-seq and DNA-seq data analysis. Using either customized or standard annotation files, SAMMate allows users to accurately calculate the short read coverage of genomic intervals. In particular, SAMMate can accurately calculate the gene expression abundance scores for customized genomic intervals using short RNA-seq reads originating from both exons and exon-exon junctions. Furthermore, to solve an array of bioinformatics problems, SAMMate can quickly calculate a whole-genome signal map at base-wise resolution. SAMMate can export both a wiggle file for alignment visualization in the UCSC genome browser and an alignment statistics report (Xu et al. 2011a).

20.4.2.3 De Novo Splice Aligners

De novo Splice aligners allow for detection of new splice junctions without the requirement of previously analyzed information.

MapSplice MapSplice can be applied to both long reads (≥75 bp) and short reads (<75 bp). MapSplice is not dependent on splice-site features or intron length. Consequently, it can detect non-canonical as well as novel canonical splices. MapSplice leverages the diversity and quality of read alignments of a given splice to increase accuracy (Wang et al. 2010a).

PALMapper PALMapper is a fast and easy tool designed to compute both spliced and unspliced alignments for millions of RNA-seq reads accurately. PALMapper combines the efficient read mapper GenomeMapper with the splice aligner QPALMA, which exploits read-quality information and predictions of splice sites to improve the align-

ment accuracy. The PALMapper package is a command-line Unix or Mac OS X systems tool or it can be used through a web interface based on Galaxy tools (Jean et al. 2010).

SeqSaw SeqSaw is useful for the detection of splice junctions with or without the canonical GT-AG splicing signal, applied to two ENCODE RNA-seq datasets and compared with two existing methods. The SeqSaw method was found to produce better results in finding novel splice junctions (Wang et al. 2011a).

SoapSplice SoapSplice is able to detect splice junctions with low false positive rates. It is a genome-wide ab initio tool for the detection of splice junction sites from RNA-seq. SOAPsplice performs better by predicting more true junctions when the coverage is low with low false positive rates, which is very useful for mRNAs with relatively lower expression levels (Huang et al. 2011).

SpliceMap SpliceMap is useful in the detection of splice junctions from RNA-seq data. This method is capable of finding novel splice junctions with high specificity and sensitivity and does not depend on any existing annotation of gene structures. SpliceMap can handle long reads (50–100 nucleotides) and can exploit paired-end read information to improve mapping accuracy. Several parameters are included in the output to indicate the reliability of the predicted junctions and help filter out false predictions (Au et al. 2010).

SplitSeek It is a program for de novo prediction of splice junctions from RNA-seq data. SplitSeek is based on split read alignment where the reads are split into two parts that are mapped separately. The alignments are then processed by the SplitSeek through a number of analysis steps. The SplitSeek results can be directly uploaded to the UCSC genome browser and used as input to the BEDTools software suite, which enables the user to analyze and visualize the predicted events at the genomic level (Ameur et al. 2010).

Subjunc Subjunc is an advanced version of Subread tool. It uses all mappable regions to investigate exon-exon junctions. Subjunc uses the receptor or donor signals to find the exact splicing locations. It can be used for genomic variation and junction detection (Liao et al. 2013a).

GEM The GEM (GEnome Multi-tool) Library is a set of optimized tools for querying/indexing huge genomes/files. GEM split mapper is an unconstrained split mapper, and very fast program to compute genome mappability, simultaneously delivering precision and speed (Marco-Sola et al. 2012).

20.5 Quantitative Analysis and Differential Expression

Differential Expression and Quantitative analysis tools calculate the abundance of each gene expressed in a RNA-seq sample. Different programs have also been developed to study the differential expression based on the variability of genetic expression between the samples.

BBSeq BBSeq is a package that incorporates two approaches (i) a simple beta-binomial generalized linear model, which has not been extensively tested for RNA-seq data and (ii) an extension of an expression mean-variance modeling approach, involving modeling of the overdispersion as a function of the mean (Zhou et al. 2011).

BitSeq Bayesian inference of transcripts from sequencing data (BitSeq) is a Bayesian approach for estimation of transcript expression level from data derived from RNA-seq experiments. Inferred relative expression is represented by Markov chain Monte Carlo samples from the posterior probability distribution of a generative model of the read data (Glaus et al. 2012).

CPTRA Cross-Platform TRanscriptome Analysis (CPTRA) is a powerful tool for the analysis of transcriptome profiling data generated by different methods. CPTRA is a powerful package for NGS-based transcriptome profiling in species with limited reference genome information (Zhou et al. 2009).

DEXUS DEXUS is a package used for the identification of differentially-expressed genes in RNA-seq data. This package does not require replicates for detection of differentially-expressed transcripts, and for each transcript the replicates are estimated by the Expectation-Maximization method (Klambauer et al. 2013).

eXpress eXpress is a tool for transcript-level RNA-seq quantification, and haplotype and allele-specific analysis. It can estimate transcript abundances of the multiple isoforms present in a gene. eXpress is directly associated with aligners like Bowtie, and it does not require a reference genome for alignment (Roberts et al. 2011).

NSMAP NSMAP is a program for spliced isoforms quantification and identification from RNA-seq datasets. NSMAP is used for estimation of expression levels without annotation information. Furthermore, TopHat can be used for alignment of exons and identification of splice junctions (Xia et al. 2011a).

rQuant rQuant is based on quadratic programming and web service of Galaxy that determines abundances of transcripts per gene locus. In this server, a combination of tools is employed, viz. PALMapper for read alignment, and mTiM and mGene for inferences of new transcripts (Bohnert and Ratsch 2010).

20.6 Open Source Solutions

Chipster Chipster is user-friendly analysis software for high-throughput sequencing data. Chipster has an intuitive graphical user interface that enables biologists to access a powerful collection of integration tools and data analysis, and to visualize data interactively. Users can collaborate by sharing workflows and analysis sessions (Kallio et al. 2011).

Expression Plot Expression Plot is a software package consisting of a default back end, which prepares Affymetrix microarray or raw sequencing data, and a web-based front end, which offers a biologically centered interface to browse, compare and visualize different data sets (Friedman and Maniatis 2011).

FX The FX tool runs in parallel on cloud computing infrastructure. It is used for genomic variant calling and for the estimation of gene expression levels. In the mapping of short RNA-seq reads, FX uses a transcriptome-based reference generated from Ensembl, UCSC and Ref-seq databases (Hong et al. 2012).

NGS-Trex NGS-Trex is a simple tool for the analysis of RNA-seq data. It is designed for researchers with limited bioinformatics skills. NGS-Trex offers simple data mining tools to explore transcriptome profiles of samples investigated using NGS technologies (Boria et al. 2013).

S-MART S-MART is used for data manipulation, differential expression analysis, selection/exclusion of reads, clustering and visualization of read information, and comparison with epigenomic ChIP-Seq data. S-MART has a user-friendly graphical user interface (Zytnicki and Quesneville 2011).

TCW Transcriptome Computational Workbench (TCW) is a Java graphical interface used for analysis of both comparative and single transcriptome data without the use of any reference genome. TCW allows in-depth data mining, which can help achieve a better understanding of the transcriptome (Soderlund et al. 2013).

WapRNA WapRNA contains four different modules for processing of miRNA-seq and mRNA-seq data from a Solexa or SOLiD platform. WapRNA accepts raw sequence data with an optional reads filter, followed by miRNA prediction, gene annotation and mapping. WapRNA also integrates downstream functional analyses such as KEGG pathway, gene ontology, comparison of genes, miRNA target prediction and miRNA's differential expression in different samples (Zhao et al. 2011).

20.7 Alternative Splicing Analysis

Alt Event Finder Alt Event Finder is a tool for identifying novel splicing events by using transcript annotation derived from genome-guided construction tools, such as Scripture and Cufflinks. With a proper combination of transcript reconstruction and alignment tools, it can identify novel splicing events in the genome (Zhou et al. 2012).

AStalavista ASTALAVISTA is an alternative splicing transcriptional landscape visualization tool. It employs complete notation system to identify alternative splicing events unequivocally. This method extracts AStalavista events dynamically from custom gene annotations, visualizes a comprehensive picture of the resulting AStalavista landscape and classifies them into groups of common types (Foissac and Sammeth 2007).

20.8 Bias Correction

GeneScissors GeneScissors is a comprehensive approach for detecting and correcting spurious transcriptome inferences. GeneScissors can detect spurious transcriptome calls owing to misalignment with close to 90 % accuracy. It provides substantial improvement over the widely used Cufflinks, TopHat, MapSplice pipelines in F-measurement (Zhang et al. 2013).

20.9 Fusion Genes, Chimeras, Translocation Finders and Structural Variations

Genome arrangements can produce aberrant genetic modifications like translocations or fusions. Identification of these modifications plays an important role in disease diagnostics. Different tools have been developed for translocation detection, fusion identification and finding structural variations.

BreakDancer The BreakDancer algorithm is used for analysis of structural variants, indels, inversions and translocations. BreakDancer is used to detect indels ranging from 10 bp to 1 Mb sensitively and accurately, which are difficult to detect via a single conventional approach (Chen et al. 2009).

FusionAnalyser FusionAnalyser is a tool used for the identification of driver fusion rearrangements through the analysis of paired-end high-throughput transcriptome sequencing data. FusionAnalyser has a fully event-driven graphical interface. A flexible filtering system allows complex analyses to be run in the absence of any a priori scripting knowledge or programming (Piazza et al. 2012).

FusionMap FusionMap is useful in the characterization of fusion genes. It detects fusions from both genomic DNA (gDNA)-Seq and single-end RNA-Seq data sets. The FusionMap provides systematic and accurate solutions for detecting fusion events through junction-spanning reads (Ge et al. 2011).

FusionSeq FusionSeq is a useful tool for identification of fusion transcripts and for removal of spurious candidate fusions resulting from random pairing of transcript fragments or misalignment. It has a module to identify exact sequences, and detect novel and known fusions at breakpoint junctions (Sboner et al. 2010).

SOAPFuse SOAPfuse is used to identify fusion transcripts from paired-end RNA-seq data. To construct a library of fusion junction sequences, SOAPfuse can apply partial exhaustion algorithm, and employs a series of filters to nominate high-confidence fusion transcripts (Jia et al. 2013).

SOAPfusion SOAPfusion tool is used for fusion discovery with paired-end RNA-seq reads. It is efficient and accurate for fusion discovery when the coverage is as low as 10x. Its high sensitivity (≥93 %) and low false-positive rate (≤1.36 %) high-

lights its ability to detect fusions efficiently at low-sequencing cost. From real data of Universal Human Reference RNA (UHRR) samples, SOAPfusion detected 7 novel fusion genes (Wu et al. 2013c).

20.10 Copy Number Variation Identification

CNVseq CNVseq is a method for detecting copy number variations based on a statistical model. CNVseq is based on an array-comparative genomic hybridization technique (Xie and Tammi 2009).

20.11 RNA-Seq Simulators

The RNA-seq simulators generate in silico reads and are useful tools for testing and comparing the efficiency of algorithms developed to handle RNA-seq data. The RNASeqRead Simulator is a python-based script for the generation of random expression levels of transcripts; it simulates reads with a specific positional bias pattern and generates random errors (Grant et al. 2011).

20.12 Transcriptome Assemblers

There are two types of approaches for assembling transcriptomes. The genome-guided methods use a reference genome (finished high quality genome sequence) as a template to align and assemble reads into transcripts. Genome-independent methods, on the other hand, do not require a reference genome and are normally used when a genome is not available. In this case, reads are assembled directly in transcripts.

20.12.1 Genome-Guided Assemblers

Cufflinks Cufflinks assembles aligned reads into transcripts, estimates their abundances, and tests for transcriptome-wide differential expression and regulation. Cufflinks can illuminate the substantial regulatory flexibility and complexity and can improve transcriptome-based genome annotation (Trapnell et al. 2010).

IsoInfer IsoInfer is able to calculate the expression levels of RNA isoforms with accuracy comparable to the state-of-the-art statistical method. It first formulates the

relationship among isoforms, single-end reads and exons as a convex quadratic program, and then uses an efficient algorithm called IsoInfer to search for isoforms (Feng et al. 2011).

RNAeXpress RNA-eXpress enables identification of transcripts and other genomic and transcriptional features independently. It accepts mapped reads in the standard binary alignment format and produces a study-specific feature annotation in comparison statistics, sequence extraction, GTF format and feature counts (Forster et al. 2013).

20.12.2 Genome-Independent Assemblers

KISSPLICE KisSplice is a part of software that enables the analysis of RNA-seq data with or without a reference genome. It is a local transcriptome assembler for detection of SNPs, indels and alternative splicing events. KISSPLICE is more accurate in comparison to other transcriptome assemblers (Sacomoto et al. 2012).

Velvet/Oases The Velvet program is for de novo short read assembly or genomic sequence assembly using the de Bruijn graphs algorithm (Zerbino and Birney 2008). The preliminary data obtained from Velvet are further transferred to Oases, which uses long-read and paired-end read information to build transcript isoforms (Schulz et al. 2012).

SOAPdenovo-Trans SOAPdenovo-Trans is a de novo transcriptome assembler designed for transcriptome assembly with alternative splicing. The assembler provides a comprehensive way to construct the full-length transcripts (Xie et al. 2014).

Trans-ABySS Assembly By Short Sequences (ABySS) is a parallel and paired-end sequence assembler written in Python and Perl languages. This tool first reduces the dataset into smaller sets of non-redundant contigs, and identifies splicing events including novel exons, exon-skipping, novel introns, retained introns and alternative splicing. The Trans-ABySS algorithms are also able to estimate gene expression levels, identify potential polyadenylation sites and candidate gene-fusion events (Robertson et al. 2010).

Trinity Trinity was developed at the Broad Institute and the Hebrew University of Jerusalem. It is a novel method for a robust and efficient de novo reconstruction of transcriptomes from RNA-seq data. Trinity combines three independent software modules, Inchworm, Chrysalis and Butterfly, applied sequentially to process large volumes of RNA-seq reads. It can be used for identification of differentially expressed transcripts across samples and for identifying protein-coding genes (Haas et al. 2013).

20.13 Visualization Tools

Artemis Artemis is a DNA sequence annotation and visualization tool that allows sets of analyses to be viewed in the context of the sequence and its six-frame translation. It is especially useful for analyzing the compact genomes of archaea, bacteria and lower eukaryotes, and can cope with sequences of any size from small genes to whole genomes (Rutherford et al. 2000).

Apollo Apollo is a genomic annotation editing platform. It can visualize similarities based on BLAST searches and alignments, and enables biologists to utilize computational evidence to edit and create gene models and other genomic features, e.g. exon-intron structure prediction using gene prediction algorithms (Misra and Harris 2006).

EagleView EagleView is useful for large genome assembly of millions of NGS reads. EagleView supports multiple navigation modes, compact assembly view and a pinpoint view of technology-specific trace information. It supports viewing coassembly of mixed-type reads from different technologies and supports integration of genome feature annotations into genome assemblies (Huang and Marth 2008).

Integrated Genome Browser (IGB) IGB is on open source desktop graphical display tool implemented in Java that supports real-time panning and zooming through a genome. The layout of genomic features and data sets in moveable, incremental or genome-scale data loading from local files or remote web servers and dynamic manipulation of quantitative data via genome graphs (Nicol et al. 2009).

Integrative Genomics Viewer (IGV) Integrative Genomics Viewer is a visualization tool for large-scale genomic data sets on standard desktop computers that enables real-time exploration of diverse genomic features. IGV supports flexible integration of a wide range of genomic data types including aligned sequence reads, copy number, mutations, gene expression, methylation, RNA interference screens and genomic annotations (Robinson et al. 2011).

GenomeView GenomeView is a genome browser designed for manipulation and visualization of genomics data, useful for alignment of short-read data. It contains semantic zooming and dynamic navigation, from single nucleotide to whole genome level. GenomeView helps in visualization of whole genome alignments of different plant genomes. It is a unique tool for handling of huge data sets consisting of tens of aligned genomes, millions of mapped short reads, both as editor and viewer and thousands of annotation features (Abeel et al. 2012).

MapView MapView is useful for representation of large-scale genetic variation analysis, short reads alignment data and automated genetic variation detection. MapView can handle hundreds of millions of short reads on a computer with limited memory. MapView supports a compact alignment view for paired-end and single-end short reads, multi-thread processing, multiple navigation and zoom modes (Bao et al. 2009a).

Savant Genome Browser is stand-alone software for computational and visual analysis of high-throughput sequencing data. Savant is useful for visualizing several types of genomic data and introducing innovative navigation interfaces. It implements automated analyses and is easy for non-expert users (Fiume et al. 2012).

20.14 Functional, Network and Pathway Analysis Tools

BLAT Tool Jim Kent developed a pairwise sequence alignment algorithm known as BLAT (BLAST-like alignment tool) at the University of California Santa Cruz in the early 2000s. It helped in the assembly and annotation of the human genome. It was designed primarily to decrease the time needed to align millions of mouse genomic reads and expressed sequence tags against the human genome sequence. The alignment tools of that time were not capable of aligning a large number of reads in a manner that would allow a regular update of the human genome assembly. Compared to pre-existing tools, BLAT is ~500 times faster in performing mRNA/DNA alignments and ~50 times faster with protein/protein alignments (Kent 2002).

GAGE GAGE is a tool for analysis of biological data for pathway details, gene ontology and gene set analysis. GAGE provides opportunities for RNA-seq pathway analysis and microarray data analysis (Salzberg et al. 2012).

Blast2GO (B2G) The Blast2GO (B2G) server is used for gene ontology study based on data mining. It is a tool used in functional genomics research focusing mainly on molecular functions, cellular components, and biological processes in which the putative proteins are involved.

AgriGO AgriGO is an integrated web-based gene ontology (GO) analysis toolkit for the agricultural community, using the advantages of GO enrichment tool (EasyGO). EasyGO is valuable for its proficiency, and has proved useful in uncovering biological knowledge in massive data sets derived from high-throughput experiments. AgriGO system architecture, performance and accessibility are good for in silico experimentation (Du et al. 2010).

20.15 Commercial Solutions

CLC Genomics Workbench CLC Genomics Workbench is a powerful tool for analysis and visualization of NGS data and is integrated with NGS workflow. It is available for Windows and Linux platforms. CLC Workbench contains a number of tools related to genomics, epigenomics and transcriptomics (Kim et al. 2013).

Partek Partek Genomics Suite (Partek GS) is a powerful statistical analysis and interactive visualization software for analyzing single-channel oligonucleotide

(Affymetrix) and two-color cDNA microarrays as well as data derived from genomic and proteomic technologies (Downey 2006).

20.16 Conclusions and Prospects

The genomics revolution, which began in the 1990s, has played a major role in enhancing an understanding of the genetic make-up of living organisms. Next-generation genome sequencing programs ultimately aim to determine the complete genome sequence features of plant and other organisms. The availability of complete plant genome sequence provides opportunities for annotation of protein-coding genes and other important functional features of the genome. Genome assembly is a very complex computational task, because many genomes contain large numbers of identical sequences, known as repeats. In the large plant genomes these repeats can be thousands of nucleotides long, and located in thousands of different locations. The draft plant genome sequence is produced by combining the information sequenced contigs and employing linking information to create scaffolds. Scaffolds are positioned along the physical map of the chromosomes creating a golden path. Short Oligonucleotide Analysis Package (SOAP) is a popular package developed by BGI for de novo assembly of larger genomes, alignment, SSR detection, SNP detection, indel finding and structural variation analysis. Genome annotation is the process of gene hunting, identification of non-coding features and attaching biological information to sequences. It consists of three main steps: (a) identification of those functional features of the genome which do not code for proteins; (b) identifying functional elements in the genome sequence, a process called gene hunting; and (c) elucidating the biological functions of these elements.

The basic level of genome annotation finds similarities using Basic Local Alignment Search Tool, and then analyzes genomes based on that information. However, nowadays more and more additional tools and modules are added to the annotation platform. Structural annotation involves the identification of genomic elements based on ORFs and their localization, determination of gene structure, coding regions and location of regulatory motifs. In functional annotation, biochemical and biological function, regulation, interaction and expression analyses can be used to obtain complete biological information about the genomic elements. These steps may involve in silico analysis. A variety of software tools have been developed to permit scientists to view and share genome assembly, annotations and analyses information. Genome annotation remains a major challenge for scientists investigating large genomes. Identifying the locations of genes and genetic control elements is often described as defining the biological *parts list* for the assembly of an organism. Scientists are still at an early stage in the process of delineating this parts list and in understanding how all the parts *fit together*. Genome annotation is an active area of investigation and involves a number of different organizations in the life science community, which publish the results of their efforts in publicly available biological databases accessible via the web and other electronic means.

The genome sequences are very important for understanding the functions of elements, individual genes and their networks, for defining phylogenetic relationships and evolutionary processes, and for revealing unknown regulatory mechanisms that coordinate the activities of genes. An integration of biological knowledge and computational system can be used for crop improvement program. Recent progress in the area of plant biotechnology and bioinformatics has the potential to initiate a new Green Revolution by aiding in the development of improved crop germplasm. Identification of functional markers and genes to traits will lead to more efficient plant breeding in the future.

References

Abeel T, Van Parys T, Saeys Y et al (2012) Genome view: a next-generation genome browser. Nucleic Acids Res 40(2), e12

Agarwal G, Jhanwar S, Priya P et al (2012) Comparative analysis of kabuli chickpea transcriptome with desi and wild chickpea provides a rich resource for development of functional markers. PLoS One 7(12), e52443

AGI (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Arabidopsis genome initiative. Nature 408(6814):796–815

Ahn YK, Tripathi S, Kim JH et al (2014) Transcriptome analysis of *Capsicum annuum* varieties Mandarin and Blackcluster: assembly, annotation and molecular marker discovery. Gene 533(2):494–499

Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploidy wheat with the Illumina Golden Gate assay. Theor Appl Genet 119:507–517

Allen AM, Barker GL, Wilkinson P et al (2013) Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). Plant Biotechnol J 11(3):279–295

Ameur A, Wetterbom A, Feuk L et al (2010) Global and unbiased detection of splice junctions from RNA-seq data. Genome Biol 11(3):R34

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11(10):R106

Anders S, Reyes A, Huber W (2012) Detecting differential usage of exons from RNA-seq data. Genome Res 22(10):2008–2017

Anders S, Pyl PT, Huber W (2014) HTSeq – a python framework to work with high-throughput sequencing data, bioRxiv posted online 20 Feb 2014. doi:10.1101/002824

Andrews S (2011) FastQ screen [online]. http://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/

Arai-Kichise Y, Shiwa Y, Nagasaki H et al (2011) Discovery of genome-wide DNA polymorphisms in a landrace cultivar of japonica rice by whole-genome sequencing. Plant Cell Physiol 52:274–282

Argout X, Salse J, Aury JM et al (2010) The genome of *Thebroma cacao*. Nat Genet 43:101–108

Aschoff M, Hotz-Wagenblatt A, Glatting KH et al (2013) SplicingCompass: differential splicing detection using RNA-seq data. Bioinformatics 29(9):1141–1148

Ashelford K, Eriksson ME, Allen CM et al (2011) Full genome re-sequencing reveals a novel circadian clock mutation in Arabidopsis. Genome Biol 12(3):R28

Ashrafi H, Hill T, Stoffel K et al (2012) De novo assembly of the pepper transcriptome (*Capsicum annuum*): a benchmark for *in silico* discovery of SNPs, SSRs and candidate genes. BMC Genomics 13:571

Au KF, Jiang H, Lin L et al (2010) Detection of splice junctions from paired-end RNA-seq data by SpliceMap. Nucleic Acids Res 38(14):4570–4578

Babu R, Nair SK, Prasanna BM, Gupta HS (2004) Integrating marker-assisted selection in crop breeding – prospects and challenges. Curr Sci 87:607–619

Bankevich A, Nurk S, Antipov D et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455–477

Bao H, Guo H, Wang J et al (2009a) MapView: visualization of short reads alignment on a desktop computer. Bioinformatics 25(12):1554–1555

Bao H, Xiong Y, Guo H et al (2009b) MapNext: a software tool for spliced and unspliced alignments and SNP detection of short sequence reads. BMC Genomics 10(3):S13

Batzoglou S, Jaffe DB, Stanley K et al (2002) ARACHNE: a whole-genome shotgun assembler. Genome Res 12(1):177–189

Bentley DR, Balasubramanian S, Swerdlow HP et al (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53–59

Bevan MW, Uauy C (2013) Genomics reveals new landscapes for crop improvement. Genome Biol 14(6):206

Bi Y, Davuluri RV (2013) NPEBseq: nonparametric empirical Bayesian-based procedure for differential expression analysis of RNA-seq data. BMC Bioinform 14:262

Biesecker LG, Shianna KV, Mullikin JC (2011) Exome sequencing: the expert view. Genome Biol 12:128

Blankenberg D, Von Kuster G, Coraor N et al (2010) Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol Chapter 19: Unit 19.10.1-21. doi:10.1002/0471142727.mb1910s89

Bohnert R, Ratsch G (2010) rQuant.web: a tool for RNA-Seq-based transcript quantitation. Nucleic Acids Res 38:348–351

Boisvert S, Laviolette F, Corbeil J (2010) Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J Comput Biol 17(11):1519–1533

Bonfert T, Csaba G, Zimmer R et al (2012) A context-based approach to identify the most likely mapping for RNA-seq experiments. BMC Bioinformatics 13(6):S9

Boria I, Boatti L, Pesole G et al (2013) NGS-Trex: Next Generation Sequencing Transcriptome profile explorer. BMC Bioinform Suppl 14(7):S10

Breese MR, Liu Y (2013) NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. Bioinformatics 29(4):494–496

Brenchley R, Spannagl M, Pfeifer M et al (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491(7426):705–710

Bryant DW Jr, Shen R, Priest HD et al (2010) Supersplat-spliced RNA-seq alignment. Bioinformatics 26(12):1500–1505

Burland TG (2000) DNASTAR's Lasergene sequence analysis software. Methods Mol Biol 132:71–91

Busby MA, Stewart C, Miller CA et al (2013) Scotty: a web tool for designing RNA-Seq experiments to measure differential gene expression. Bioinformatics 29(5):656–657

Butler J, MacCallum I, Kleber M et al (2008) ALLPATHS: de novo assembly of whole-genome shotgun microreads. Genome Res 18(5):810–820

Campagna D, Albiero A, Bilardi A et al (2009) PASS: a program to align short sequences. Bioinformatics 25(7):967–968

Cao J, Schneeberger K, Ossowski S et al (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nat Genet 43(10):956–963

Chan AP, Crabtree J, Zhao Q et al (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotech 28:951–956

Chao S, Dubcovsky J, Dvorak J et al (2010) Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). BMC Genomics 11:727

Chatterjee A, Stockwell PA, Rodger EJ et al (2012) Comparison of alignment software for genome-wide bisulphite sequence data. Nucleic Acids Res 40(10), e79

Chen K, Wallis JW, McLellan MD et al (2009) BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Meth 6(9):677–681

Chen LY, Wei KC, Huang AC et al (2012) RNASEQR--a streamlined and accurate RNA-seq sequence analysis program. Nucleic Acids Res 40(6), e42
Chitsaz H, Yee-Greenbaum JL, Tesler G (2011) Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. Nat Biotechnol 29(10):915–921
Chu VT, Gottardo R, Raftery AE et al (2008) MeV + R: using MeV as a graphical user interface for bioconductor applications in microarray analysis. Genome Biol 9(7):R118
Chu HT, Hsiao WW, Chen JC et al (2013) EBARDenovo: highly accurate de novo assembly of RNA-Seq with efficient chimera-detection. Bioinformatics 29(8):1004–1010
Church GM (2006) Genomes for all. Sci Am 294(1):46–54
Clement NL, Snell Q, Clement MJ et al (2010) The GNUMAP algorithm: unbiased probabilistic mapping of oligonucleotides from next-generation sequencing. Bioinformatics 26(1):38–45
Cloonan N, Xu Q, Faulkner GJ et al (2009) RNA-MATE: a recursive mapping strategy for high-throughput RNA-sequencing data. Bioinformatics 25(19):2615–2616
Cock PJ, Fields CJ, Goto N et al (2010) The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res 38(6):1767–1771
Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169–196
Cumbie JS, Kimbrel JA, Di Y et al (2011) GENE-counter: a computational pipeline for the analysis of RNA-Seq data for gene expression differences. PLoS One 6(10), e25279
Dassanayake M, Oh DH, Haas JS et al (2011) The genome of the extremophile crucifer *Thellungiella parvula*. Nat Genet 43(9):913–918
Dayarian A, Michael TP, Sengupta AM (2010) SOPRA: Scaffolding algorithm for paired reads via statistical optimization. BMC Bioinform 11:345
De Bona F, Ossowski S, Schneeberger K et al (2008) Optimal spliced alignments of short sequence reads. Bioinformatics 24(16):i174–i180
De Hoon M, Hayashizaki Y (2008) Deep cap analysis gene expression (CAGE): genome-wide identification of promoters, quantification of their expression, and network inference. Biotechniques 44(5):627–628
De la Bastide M, McCombie WR (2007) Assembling genomic DNA sequences with PHRAP. Curr Protoc Bioinform. Chapter 11: Unit11.4. doi:10.1002/0471250953.bi1104s17
Delhomme N, Padioleau I, Furlong EE et al (2012) easyRNASeq: a bioconductor package for processing RNA-Seq data. Bioinformatics 28(19):2532–2533
DeLuca DS, Levin JZ, Sivachenko A et al (2012) RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics 28(11):1530–1532
Denoeud F, Aury JM, Da Silva C et al (2008) Annotating genomes with massive-scale RNA sequencing. Genome Biol 9:R175
Dimon MT, Sorber K, DeRisi JL (2010) HMMSplicer: a tool for efficient and sensitive discovery of known and novel splice junctions in RNA-Seq data. PLoS One 5(11), e13875
Dobin A, Davis CA, Schlesinger F et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29(1):15–21
Dodt M, Roehr JT, Ahmed R et al (2012) FLEXBAR – flexible barcode and adapter processing for next-generation sequencing platforms. Biol 1(3):895–905
Dohm JC, Lottaz C, Borodina T (2007) SHARCGS, a fast and highly accurate short-read assembly algorithm for de novo genomic sequencing. Genome Res 17(11):1697–1706
Dorn KM, Fankhauser JD, Wyse DL et al (2013) De novo assembly of the pennycress (*Thlaspi arvense*) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock. Plant J 75(6):1028–1038
Doust AN, Kellogg EA, Devos KM et al (2009) Foxtail millet: a sequence-driven grass model system. Plant Physiol 149(1):137–141
Downey T (2006) Analysis of a multifactor microarray study using Partek genomics solution. Methods Enzymol 411:256–270
Du Z, Zhou X, Ling Y et al (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res 38:64–70

Duan J, Xia C, Zhao G et al (2012) Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. BMC Genomics 13:392
Dubey A, Farmer A, Schlueter J et al (2011) Defining the transcriptome assembly and its use for genome dynamics and transcriptome profiling studies in pigeonpea (*Cajanus cajan* L.). DNA Res 18(3):153–164
Dubey NK, Goel R, Ranjan A et al (2013) Comparative transcriptome analysis of *Gossypium hirsutum* L in response to sap sucking insects: aphid and whitefly. BMC Genomics 14(1):241
Eathington SR, Crosbie TM, Edward MD et al (2007) Molecular markers in a commercial breeding programme. Crop Sci 47S:154–163
Edwards D, Batley J (2010) Plant genome sequencing: applications for crop improvement. Plant Biotechnol J 8(1):2–9
Edwards D, Batley J, Snowdon RJ (2013) Accessing complex crop genomes with next-generation sequencing. Theor Appl Genet 126(1):1–11
Egan AN, Schlueter J, Spooner DM (2012) Applications of next-generation sequencing in plant biology. Am J Bot 99(2):175–185
Fan H, Xiao Y, Yang Y et al (2013) RNA-Seq analysis of *Cocos nucifera*: transcriptome sequencing and de novo assembly for subsequent functional genomics approaches. PLoS One 8(3), e59997
Feng J, Li W, Jiang T (2011) Inference of isoforms from short sequence reads. J Comput Biol 18(3):305–321
Fernandes AD, Macklaim JM, Linn TG et al (2013) ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. PLoS One 8(7), e67019
Filichkin SA, Priest HD, Givan SA et al (2010) Genome-wide mapping of alternative splicing in Arabidopsis thaliana. Genome Res 20(1):45–58
Fiume M, Smith EJ, Brook A et al (2012) Savant Genome Browser 2: visualization and analysis for population-scale genomics. Nucleic Acids Res 40:615–621
Florea L, Song L, Salzberg SL (2013) Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. Version 2. F1000Res. 16 Sept 2013 [revised 21 Nov 2013]; 2:188. doi:10.12688/f1000research.2-188.v2
Foissac S, Sammeth M (2007) ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. Nucleic Acids Res 35:297–299
Forster SC, Finkel AM, Gould JA et al (2013) RNA-eXpress annotates novel transcript features in RNA-seq data. Bioinformatics 29(6):810–812
Friedman BA, Maniatis T (2011) ExpressionPlot: a web-based framework for analysis of RNA-Seq and microarray gene expression data. Genome Biol 12(7):R69
Garber M, Grabherr MG, Guttman M et al (2011) Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Methods 8:469–477
Garcia-Mas J, Benjak A, Sanseverino W et al (2012) The genome of melon (*Cucumis melo* L.). Proc Natl Acad Sci U S A 109(29):11872–11877
Garg R, Patel RK, Jhanwar S et al (2011) Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. Plant Physiol 156(4):1661–1678
Ge H, Liu K, Juan T et al (2011) FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. Bioinformatics 27(14):1922–1928
Gillies S, Furtado A, Henry RJ (2012) Gene expression in the developing aleurone and starchy endosperm of wheat. Plant Biotechnol J 10:668–679
Glaus P, Honkela A, Rattray M (2012) Identifying differentially expressed transcripts from RNA-seq data with biological variation. Bioinformatics 28(13):1721–1728
Gnerre S, Lander ES, Lindblad-Toh K et al (2009) Assisted assembly: how to improve a de novo genome assembly by using related species. Genome Biol 10(8):R88
Gnerre S, Maccallum I, Przybylski D et al (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A 108(4):1513–1518
Goff SA, Ricke D, Lan TH et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *Japonica*). Science 296(5565):92–100

Goncalves A, Tikhonov A, Brazma A et al (2011) A pipeline for RNA-seq data processing and quality assessment. Bioinformatics 27(6):867–869
Gong T, Szustakowski JD (2013) DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. Bioinformatics 29(8):1083–1085
Gopala Krishnan S, Waters DLE, Katiyar SK et al (2011) Genome-wide DNA polymorphisms in elite *indica* rice inbreds discovered by whole-genome sequencing. Plant Biotechnol J 10:623–634
Grant GR, Farkas MH, Pizarro AD et al (2011) Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). Bioinformatics 27(18):2518–2528
Griebel T, Zacher B, Ribeca P et al (2012) Modelling and simulating generic RNA-Seq experiments with the flux simulator. Nucleic Acids Res 40(20):10073–10083
Griffith M, Griffith OL, Mwenifumbo J et al (2010) Alternative expression analysis by RNA sequencing. Nat Methods 7(10):843–847
Guo S, Liu J, Zheng Y et al (2011) Characterization of transcriptome dynamics during watermelon fruit development: sequencing, assembly, annotation and gene expression profiles. BMC Genomics 12:454
Guttman M, Garber M, Levin JZ et al (2010) Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol 28(5):503–510
Haas BJ, Papanicolaou A, Yassour M et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8(8):1494–1512
Halbritter F, Kousa AI, Tomlinson SR (2014) GeneProf data: a resource of curated, integrated and reusable high-throughput genomics experiments. Nucleic Acids Res 42:851–858. doi:10.1093/nar/gkt966
Hamilton JP, Buell CR (2012) Advances in plant genome sequencing. Plant J 70(1):177–190
Hardcastle TJ, Kelly KA (2010) baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11:422
Harris TD, Buzby PR, Babcock H et al (2008) Single-molecule DNA sequencing of a viral genome. Science 320(5872):106–109
He J, Jiao Y (2014) Next-generation sequencing applied to flower development: RNA-seq. methods Mol Biol 1110:401–411
Henry RJ (2011) Next-generation sequencing for understanding and accelerating crop domestication. Brief Funct Genomics 11(1):51–56
Henry RJ, Edwards K (2009) New tools for single nucleotide polymorphism (SNP) discovery and analysis accelerating plant biotechnology. Plant Biotechnol J 7(4):311
Hernandez D, François P, Farinelli L et al (2008) De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. Genome Res 18(5):802–809
Hiremath PJ, Farmer A, Cannon SB et al (2011) Large-scale transcriptome analysis in chickpea (*Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. Plant Biotechnol J 9(8):922–931
Homer N, Merriman B, Nelson SF (2009) BFAST: an alignment tool for large scale genome resequencing. PLoS One 4(11), e7767
Hong D, Rhie A, Park SS et al (2012) FX: an RNA-Seq analysis tool on the cloud. Bioinformatics 28(5):721–723
Hu G, Shen S, Ruan J (2007) SGA: a grammar-based alignment algorithm. Comput Methods Prog Biomed 86(1):17–20
Hu TT, Pattyn P, Bakker EG et al (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet 43(5):476–481
Hu J, Ge H, Newman M et al (2012) OSA: a fast and accurate alignment tool for RNA-Seq. Bioinformatics 28(14):1933–1934
Hu Y, Huang Y, Du Y et al (2013) DiffSplice: the genome-wide detection of differential splicing events with RNA-seq. Nucleic Acids Res 41(2), e39
Huang W, Marth G (2008) EagleView: a genome assembly viewer for next-generation sequencing technologies. Genome Res 18(9):1538–1543

Huang S, Zhang J, Li R et al (2011) SOAPsplice: genome-wide ab initio detection of splice junctions from RNA-Seq data. Front Genet 2:46
Huang J, Chen J, Lathrop M et al (2013) A tool for RNA sequencing sample identity check. Bioinformatics 29(11):1463–1464
Iyer MK, Chinnaiyan AM, Maher CA (2011) ChimeraScan: a tool for identifying chimeric transcription in sequencing data. Bioinformatics 27(20):2903–2904
Jaffe DB, Butler J, Gnerre S et al (2003) Whole-genome sequence assembly for mammalian genomes: Arachne 2. Genome Res 13(1):91–96
Jaillon O, Aury JM, Noel B et al (2007) French-Italian public consortium for grapevine genome characterization. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449(7161):463–467
Jain M (2012) Next-generation sequencing technologies for gene expression profiling in plants. Brief Funct Genomics 11(1):63–70
Jain M, Misra G, Patel RK et al (2013) A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). Plant J 74(5):715–729
Jean G, Kahles A, Sreedharan VT et al (2010) RNA-Seq read alignments with PALMapper. Curr Protoc Bioinform Chapter 11: Unit 11.6. doi:10.1002/0471250953.bi1106s32
Jeck WR, Reinhardt JA, Baltrus DA et al (2007) Extending assembly of short DNA sequences to handle error. Bioinformatics 23(21):2942–2944
Jhanwar S, Priya P, Garg R (2012) Transcriptome sequencing of wild chickpea as a rich resource for marker development. Plant Biotechnol J 10(6):690–702
Jia W, Qiu K, He M et al (2013) SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. Genome Biol 14(2):R12
Jiang H, Wong WH (2008) SeqMap: mapping massive amount of oligonucleotides to the genome. Bioinformatics 24(20):2395–2396
Kallio MA, Tuimala JT, Hupponen T et al (2011) Chipster: user-friendly analysis software for microarray and other high-throughput data. BMC Genomics 12:507. doi:10.1186/1471-2164-12-507
Kaur S, Cogan NO, Pembleton LW et al (2011) Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. BMC Genomics 12:265
Kaur S, Francki MG, Forster JW (2012) Identification, characterization and interpretation of single-nucleotide sequence variation in allopolyploid crop species. Plant Biotechnol J 10:125–138
Kearse M, Moir R, Wilson A et al (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649
Kelly LJ, Leitch IJ (2011) Exploring giant plant genomes with next-generation sequencing technology. Chromosom Res 19(7):939–953
Kent WJ (2002) BLAT--the BLAST-like alignment tool. Genome Res 12(4):656–664
Kim D, Salzberg SL (2011) TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol 12(8):R72
Kim MY, Lee S, Van K et al (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. And Zucc.) genome. Proc Natl Acad Sci U S A 107:22032–22037
Kim H, Bi Y, Pal S, Gupta R (2011) IsoformEx: isoform level gene expression estimation using weighted non-negative least squares from mRNA-Seq data. BMC Bioinformatics 12:305
Kim KU, Park SK, Kang SA et al (2013) Comparison of functional gene annotation of *Toxascaris leonina* and *Toxocara canis* using CLC genomics workbench. Korean J Parasitol 51(5):525–530
Klambauer G, Unterthiner T, Hochreiter S (2013) DEXUS: identifying differential expression in RNA-Seq studies with unknown conditions. Nucleic Acids Res 41(21), e198
Kudapa H, Azam S, Sharpe AG et al (2014) Comprehensive transcriptome assembly of Chickpea (*Cicer arietinum* L.) using Sanger and next generation sequencing platforms: development and applications. PLoS One 9(1), e86039

Kuehn H, Liberzon A, Reich M et al (2008) Using GenePattern for gene expression analysis. Curr Protoc Bioinform Chapter 7: Unit 7.12. doi:10.1002/0471250953.bi0712s22. Review
Labaj PP, Linggi BE, Wiley HS (2012) Improving RNA-Seq precision with MapAl. Front Genet 3:28
Langmead B (2010) Aligning short sequencing reads with Bowtie. Curr Protoc Bioinform Chapter 11: Unit 11.7
Lassmann T, Hayashizaki Y, Daub CO (2011) SAMStat: monitoring biases in next generation sequencing data. Bioinformatics 27(1):130–131
Le HS, Schulz MH, McCauley BM et al (2013) Probabilistic error correction for RNA sequencing. Nucleic Acids Res 41(10), e109
Lee S, Seo CH, Lim B et al (2011) Accurate quantification of transcriptome from RNA-Seq data by effective length normalization. Nucleic Acids Res 39(2), e9
Lee WP, Stromberg MP, Ward A et al (2014) MOSAIK: A hash-based algorithm for accurate next-generation sequencing short-read mapping. PLoS One 9(3), e90581
Leng N, Dawson JA, Thomson JA et al (2013) EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics 29(8):1035–1043
Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform 12:323
Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760
Li H, Ruan J, Durbin R (2008a) Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 18(11):1851–1858
Li R, Li Y, Kristiansen K et al (2008b) SOAP: short oligonucleotide alignment program. Bioinformatics 24(5):713–714
Li H, Handsaker B, Wysoker A et al (2009) 1000 Genome project data processing subgroup. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079
Li JJ, Jiang CR, Brown JB et al (2011a) Sparse linear modeling of next-generation mRNA sequencing (RNA-Seq) data for isoform discovery and abundance estimation. Proc Natl Acad Sci U S A 108(50):19867–19872
Li W, Feng J, Jiang T (2011b) IsoLasso: a LASSO regression approach to RNA-Seq based transcriptome assembly. J Comput Biol 18(11):1693–1707
Li Y, Chien J, Smith DI et al (2011c) FusionHunter: identifying fusion transcripts in cancer using paired-end RNA-seq. Bioinformatics 27(12):1708–1710
Li JW, Wan R, Yu CS et al (2013a) ViralFusionSeq: accurately discover viral integration events and reconstruct fusion transcripts at single-base resolution. Bioinformatics 29(5):649–651
Li Y, Li-Byarlay H, Burns P (2013b) TrueSight: a new algorithm for splice junction detection using RNA-seq. Nucleic Acids Res 41(4), e51
Liao Y, Smyth GK, Shi W (2013a) FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30(7):923–930. doi:10.1093/bioinformatics/btt656
Liao Y, Smyth GK, Shi W (2013b) The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res 41(10), e108
Lin H, Zhang Z, Zhang MQ (2008) ZOOM! Zillions of oligos mapped. Bioinformatics 24(21):2431–2437
Liu Y, Schmidt B, Maskell DL (2011) Parallelized short read assembly of large genomes using de Bruijn graphs. BMC Bioinform 12:354
Logacheva MD, Kasianov AS, Vinogradov DV (2011) De novo sequencing and characterization of floral transcriptome in two species of buckwheat (*Fagopyrum*). BMC Genomics 12:30
Lohse M, Bolger AM, Nagel A et al (2012) RobiNA: a user-friendly, integrated software solution for RNA-seq-based transcriptomics. Nucleic Acids Res 40:W622–W627
Lou SK, Ni B, Lo LY et al (2011) ABMapper: a suffix array-based tool for multi-location searching and splice-junction mapping. Bioinformatics 27(3):421–422
Lulin H, Xiao Y, Pei S et al (2012) The first Illumina-based de novo transcriptome sequencing and analysis of safflower flowers. PLoS One 7(6), e38653

Lunter G, Goodson M (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res 21(6):936–939
Maccallum I, Przybylski D, Gnerre S et al (2009) ALLPATHS 2: small genomes assembled accurately and with high continuity from short paired reads. Genome Biol 10(10):R103
Mace ES, Tai S, Gilding EK et al (2013) Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nat Commun 4:2320
Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27(21):2957–2963
Marco-Sola S, Sammeth M, Guigó R et al (2012) The GEM mapper: fast, accurate and versatile alignment by filtration. Nat Methods 9(12):1185–1188
Mardis ER (2008a) Next-generation DNA sequencing methods. Annu Rev Genom Hum Genet 9:387–402
Mardis ER (2008b) The impact of next-generation sequencing technology on genetics. Trend Genet 24:133–141
Margulies M, Egholm M, Altman WE et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057):376–380
Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J North Am 17(1):10–12
Martin J, Bruno VM, Fang Z et al (2010) Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. BMC Genomics 11:663
Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci U S A 74(2):560–564
McPherson A, Hormozdiari F, Zayed A et al (2011) deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. PLoS Comput Biol 7(5), e1001138
Meacham F, Boffelli D, Dhahbi J et al (2011) Identification and correction of systematic error in high-throughput sequence data. BMC Bioinform 12:451
Meyer E, Logan TL, Juenger TE (2012) Transcriptome analysis and gene expression atlas for *Panicum hallii* var. *filipes*, a diploid model for biofuel research. Plant J 70(5):879–890
Meyers BC, Vu TH, Tej SS (2004) Analysis of the transcriptional complexity of *Arabidopsis thaliana* by massively parallel signature sequencing. Nat Biotechnol 22(8):1006–1011
Mezlini AM, Smith EJ, Fiume M et al (2013) iReckon: simultaneous isoform discovery and abundance estimation from RNA-seq data. Genome Res 23(3):519–529
Milne I, Bayer M, Cardle L et al (2010) Tablet--next generation sequence assembly visualization. Bioinformatics 26(3):401–402
Misra S, Harris N (2006) Using Apollo to browse and edit genome annotations. Curr Protoc Bioinform Chapter 9, Unit 9.5
Mochida K, Shinozaki K (2010) Genomics and bioinformatics resources for crop improvement. Plant Cell Phys 51(4):497–523
Morgan M, Anders S, Lawrence M (2009) ShortRead. http://bioconductor.org/packages/release/bioc/html/ShortRead.html
Mortazavi A, Williams BA, McCue K et al (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5(7):621–628
Mudalkar S, Golla R, Ghatty S et al (2014) De novo transcriptome analysis of an imminent biofuel crop, *Camelina sativa* L. using Illumina GAIIX sequencing platform and identification of SSR markers. Plant Mol Biol 84(1–2):159–171
Mullikin JC, Ning Z (2003) The phusion assembler. Genome Res 13(1):81–90
Mundry M, Bornberg-Bauer E, Sammeth M et al (2012) Evaluating characteristics of de novo assembly software on 454 transcriptome data: a simulation approach. PLoS One 7(2), e31410
Myles S, Chia JM, Hurwitz B et al (2010) Rapid genomic characterization of the genus Vitis. PLoS One 5, e8219
Nakano M, Nobuta K, Vemaraju K et al (2006) Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. Nucleic Acids Res 34:731–735
Natarajan P, Parani M (2011) De novo assembly and transcriptome analysis of five major tissues of *Jatropha curcas* L. using GS FLX titanium platform of 454 pyrosequencing. BMC Genomics 12:191

Nicol JW, Helt GA, Blanchard SG Jr et al (2009) The integrated genome browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics 25(20):2730–2731
Nobuta K, Venu RC, Lu C et al (2007) An expression atlas of rice mRNAs and small RNAs. Nat Biotechnol 25(4):473–477
Nowrousian M (2010) Next-generation sequencing techniques for eukaryotic microorganisms: sequencing-based solutions to biological problems. Eukaryot Cell 9:1300–1310
Oinn T, Addis M, Ferris J et al (2004) Taverna: a tool for the composition and enactment of bioinformatics workflows. Bioinformatics 20(17):3045–3054
Oono Y, Kobayashi F, Kawahara Y et al (2013) Characterisation of the wheat (*Triticum aestivum* L.) transcriptome by de novo assembly for the discovery of phosphate starvation-responsive genes: gene expression in Pi-stressed wheat. BMC Genomics 14:77
Ozsolak F, Platt AR, Jones DR et al (2009) Direct RNA sequencing. Nature 461(7265):814–818
Pagani I, Liolios K, Jansson J et al (2011) The Genomes OnLine Database (GOLD) v. 4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res 40:D571–D579
Parra-González LB, Aravena-Abarzúa GA, Navarro-Navarro CS et al (2012) Yellow lupin (*Lupinus luteus* L.) transcriptome sequencing: molecular marker development and comparative studies. BMC Genomics 13:425
Passos MA, de Cruz VO, Emediato FL et al (2013) Analysis of the leaf transcriptome of *Musa acuminata* during interaction with *Mycosphaerella musicola*: gene assembly, annotation and marker development. BMC Genomics 14:78
Paterson AH, Bowers JE, Bruggmann R et al (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457(7229):551–556
Paux E, Sourdille P, Mackay I, Feuillet C (2011) Sequence-based marker development in wheat: advances and applications to breeding. Biotechnol Adv 30(5):1071–1088
Peng Y, Leung HC, Yiu SM et al (2013) IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. Bioinformatics 29(13):i326–i334
Pennisi E (2007) Genome sequencing. The greening of plant genomics. Science 317(5836):317
Pepke S, Wold B, Mortazavi A (2009) Computation for ChIP-seq and RNA-seq studies. Nat Methods 6:S22–S32
Perkins TT, Tay CY, Thirriot F et al (2013) Choosing a benchtop sequencing machine to characterise *Helicobacter pylori* genomes. PLoS One 8(6), e67539
Peterson ES, McCue LA, Schrimpe-Rutledge AC et al (2012) VESPA: software to facilitate genomic annotation of prokaryotic organisms through integration of proteomic and transcriptomic data. BMC Genom 13:131
Pevzner PA, Tang H, Waterman MS (2001) An eulerian path approach to DNA fragment assembly. Proc Natl Acad Sci U S A 98(17):9748–9753
Philippe N, Salson M, Commes T et al (2013) CRAC: an integrated approach to the analysis of RNA-seq reads. Genome Biol 14(3):R30
Piazza R, Pirola A, Spinelli R et al (2012) FusionAnalyser: a new graphical, event-driven tool for fusion rearrangements discovery. Nucleic Acids Res 40(16), e123
Pietu G, Mariage-Samson R, Fayein NA et al (1999) The genexpress IMAGE knowledge base of the human brain transcriptome: a prototype integrated resource for functional and computational genomics. Genome Res 9:195–209
Planet E, Attolini CS, Reina O et al (2012) htSeqTools: high-throughput sequencing quality control, processing and visualization in R. Bioinformatics 28(4):589–590
Pop M, Kosack D (2004) Using the TIGR assembler in shotgun sequencing projects. Methods Mol Biol 255:279–294
Popendorf K, Sakakibara Y (2012) SAMSCOPE: an OpenGL-based real-time interactive scale-free SAM viewer. Bioinformatics 28(9):1276–1277

Qin C, Yu C, Shen Y et al (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. Proc Natl Acad Sci U S A 111(14):5135–5140

Qu W, Hashimoto S, Morishita S (2009) Efficient frequency-based de novo short-read clustering for error trimming in next-generation sequencing. Genome Res 19(7):1309–1315

Rasko DA, Webster DR, Sahl JW et al (2011) Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. N Engl J Med 365:709–717

Rivarola M, Foster JT, Chan AP et al (2011) Castor bean organelle genome sequencing and worldwide genetic diversity analysis. PLoS One 6(7), e21743

Roberts A, Trapnell C, Donaghey J et al (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol 12(3):R22

Robertson G, Schein J, Chiu R et al (2010) De novo assembly and analysis of RNA-seq data. Nat Methods 7(11):909–912

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139–140

Robinson JT, Thorvaldsdóttir H, Winckler W et al (2011) Integrative genomics viewer. Nat Biotechnol 29(1):24–26

Rothberg JM, Hinz W, Rearick TM et al (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature 475(7356):348–352

Rumble SM, Lacroute P, Dalca AV et al (2009) SHRiMP: accurate mapping of short color-space reads. PLoS Comput Biol 5(5), e1000386

Rutherford K, Parkhill J, Crook J et al (2000) Artemis: sequence visualization and annotation. Bioinformatics 16(10):944–945

Ryan MC, Cleland J, Kim R et al (2012) SpliceSeq: a resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts. Bioinformatics 28(18):2385–2387

Sacomoto GA, Kielbassa J, Chikhi R et al (2012) KISSPLICE: de-novo calling alternative splicing events from RNA-seq data. BMC Bioinform 13(6):S5

Sahli M, Shibuya T (2012) Arapan-S: a fast and highly accurate whole-genome assembly software for viruses and small genomes. BMC Res Notes 5:243

Salzberg SL, Phillippy AM, Zimin A et al (2012) GAGE: a critical evaluation of genome assemblies and assembly algorithms. Genome Res 22(3):557–567

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74(12):5463–5467

Sboner A, Habegger L, Pflueger D et al (2010) FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. Genome Biol 11(10):R104

Schatz MC, Witkowski J, McCombie WR (2012) Current challenges in de novo plant genome sequencing and assembly. Genome Biol 13(4):243

Schmidt B, Sinha R, Beresford-Smith B et al (2009) A fast hybrid short read fragment assembly algorithm. Bioinformatics 17:2279–2280

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27(6):863–864

Schmutz J, Cannon SB, Schlueter J et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463(7278):178–183

Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Schulz MH, Zerbino DR, Vingron M et al (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28(8):1086–1092

Schuster SC (2008) Next-generation sequencing transforms today's biology. Nat Methods 5(1):16–18

Shen S, Park JW, Huang J et al (2012) MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. Nucleic Acids Res 40(8), e61

Shendure J, Porreca GJ, Reppas NB (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309(5741):1728–1732
Shulaev V, Sargent DJ, Crowhurst RN (2011) The genome of woodland strawberry (*Fragaria vesca*). Nat Genet 43(2):109–116
Simpson JT, Wong K, Jackman SD et al (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19(6):1117–1123
Singh D, Orellana CF, Hu Y et al (2011a) FDM: a graph-based statistical method to detect differential transcription using RNA-seq data. Bioinformatics 27(19):2633–2640
Singh NK, Gupta DK, Jayaswal PK et al (2011b) The first draft of the pigeonpea genome sequence. J Plant Biochem Biotechnol 21:98–112
Smeds L, Künstner A (2011) ConDeTri – a content dependent read trimmer for Illumina data. PLoS One 6(10), e26314. doi:10.1371/journal.pone.0026314
Soderlund C, Nelson W, Willer M et al (2013) TCW: transcriptome computational workbench. PLoS One 8(7), e69401
Staden R (1979) A strategy of DNA sequencing employing computer programs. Nucleic Acids Res 6(7):2601–2610
Stein LD (2013) Using GBrowse 2.0 to visualize and share next-generation sequence data. Brief Bioinform 14(2):162–171
Stenson PD et al (2009) The human gene mutation database: 2008 update. Genome Med 1:13
Sucher NJ, Hennell JR, Carles MC (2012) DNA fingerprinting, DNA barcoding, and next generation sequencing technology in plants. Methods Mol Biol 862:13–22
Sun Q, Zhou G, Cai Y et al (2012a) Transcriptome analysis of stem development in the tumourous stem mustard *Brassica juncea* var. *tumida* Tsen et Lee by RNA sequencing. BMC Plant Biol 12:53
Sun X, Zhou S, Meng F et al (2012b) De novo assembly and characterization of the garlic (*Allium sativum*) bud transcriptome by Illumina sequencing. Plant Cell Rep 31(10):1823–1828
Surget-Groba Y, Montoya-Burgos JI (2010) Optimization of de novo transcriptome assembly from next-generation sequencing data. Genome Res 20(10):1432–1440
Tang S, Riva A (2013) PASTA: splice junction identification from RNA-sequencing data. BMC Bioinform 14:116
Tarazona S, García-Alcalde F, Dopazo J et al (2011) Differential expression in RNA-seq: a matter of depth. Genome Res 21(12):2213–2223
TGI (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Tomato genome consortium. Nature 485(7400):635–641
Thompson JF, Steinmann KE (2010). Single molecule sequencing with a HeliScope genetic analysis system. Curr Protoc Mol Biol Chapter: Unit7.10. doi:10.1002/0471142727. mb0710s92
Tippmann HF (2004) Analysis for free: comparing programs for sequence analysis. Brief Bioinform 5(1):82–87
Torri F, Dinov ID, Zamanyan A et al (2012) Next generation sequence analysis and computational genomics using graphical pipeline workflows. Genes (Basel) 3(3):545–575
Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25(9):1105–1111
Trapnell C, Williams BA, Pertea G et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28(5):511–515
Treangen TJ, Salzberg SL (2011) Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 13:36–46
Treangen TJ, Sommer DD, Angly FE et al (2011b) Next generation sequence assembly with AMOS. Curr Protoc Bioinform Chapter 11: Unit 11.8. doi:10.1002/0471250953.bi1108s33
Tucker T, Marra M, Friedman JM (2009) Massively parallel sequencing: the next big thing in genetic medicine. Am J Hum Genet 85(2):142–154
Turro E, Su SY, Gonçalves Â et al (2011) Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. Genome Biol 12(2):R13

Tuskan GA, Difazio S, Jansson S et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313(5793):1596–1604
Van BH, Stout JM, Cote AG et al (2011) The draft genome and transcriptome of *Cannabis sativa*. Genome Biol 12(10):R102
Varshney RK, May GD (2012) Next-generation sequencing technologies: opportunities and obligations in plant genomics. Brief Funct Genomics 11(1):1–2
Varshney RK, Chen W, Li Y et al (2011) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30(1):83–89
Varshney RK, Song C, Saxena RK et al (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nat Biotechnol 31(3):240–246
Verma P, Shah N, Bhatia S (2013) Development of an expressed gene catalogue and molecular markers from the de novo assembly of short sequence reads of the lentil (*Lens culinaris* Medik.) transcriptome. Plant Biotechnol J 11(7):894–905
Walker DR, Koonin EV (1997) SEALS: a system for easy analysis of lots of sequences. Proc Int Conf Intell Syst Mol Biol 5:333–339
Wan L, Sun F (2012) CEDER: accurate detection of differentially expressed genes by combining significance of exons using RNA-Seq. IEEE/ACM Trans Comput Biol Bioinform 9(5):1281–1292
Wang K, Singh D, Zeng Z et al (2010a) MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res 38(18), e178
Wang L, Feng Z, Wang X et al (2010b) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 26(1):136–138
Wang L, Wang X, Wang X et al (2011a) Observations on novel splice junctions from RNA sequencing data. Biochem Biophys Res Commun 409(2):299–303
Wang X, Wang H, Wang J et al (2011b) *Brassica rapa* genome sequencing project consortium. The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet 43(10):1035–1039
Wang Y, Mehta G, Mayani R et al (2011c) RseqFlow: workflows for RNA-Seq data analysis. Bioinformatics 27(18):2598–2600
Wang L, Wang S, Li W (2012a) RSeQC: quality control of RNA-seq experiments. Bioinformatics 28(16):2184–2185
Wang Y, Zeng X, Iyer NJ et al (2012b) Exploring the switchgrass transcriptome using second-generation sequencing technology. PLoS One 7(3), e34225
Wang L, Yu S, Tong C et al (2014) Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. Genome Biol 15(2):R39
Warren RL, Sutton GG, Jones SJ et al (2007) Assembling millions of short DNA sequences using SSAKE. Bioinformatics 23(4):500–501
Weese D, Emde AK, Rausch T et al (2009) RazerS-fast read mapping with sensitivity control. Genome Res 19(9):1646–1654
Weese D, Holtgrewe M, Reinert K (2012) RazerS 3: faster, fully sensitive read mapping. Bioinformatics 28(20):2592–2599
Wei X, Zhang X, Shen D et al (2013) Transcriptome analysis of *Barbarea vulgaris* infested with diamondback moth (*Plutella xylostella*) larvae. PLoS One 8(5), e64481
Winfield MO, Wilkinson PA, Allen AM (2012) Targeted re-sequencing of the allohexaploid wheat exome. Plant Biotechnol J 10:733–742
Wood DL, Xu Q, Pearson JV et al (2011) X-MATE: a flexible system for mapping short read data. Bioinformatics 27(4):580–581
Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26(7):873–881
Wu J, Akerman M, Sun S et al (2011) SpliceTrap: a method to quantify alternative splicing under single cellular conditions. Bioinformatics 27(21):3010–3016
Wu J, Anczuków O, Krainer AR et al (2013a) OLego: fast and sensitive mapping of spliced mRNA-Seq reads using small seeds. Nucleic Acids Res 41(10):5149–5163
Wu J, Wang Z, Shi Z et al (2013b) The genome of the pear (*Pyrus bretschneideri* Rehd.). Genome Res 23(2):396–408

Wu J, Zhang W, Huang S et al (2013c) SOAPfusion: a robust and effective computational fusion discovery tool for RNA-seq reads. Bioinformatics 29(23):2971–2978
Xia Z, Wen J, Chang CC (2011a) NSMAP: a method for spliced isoforms identification and quantification from RNA-Seq. BMC Bioinform 12:162
Xia Z, Xu H, Zhai J et al (2011b) RNA-Seq analysis and de novo transcriptome assembly of *Hevea brasiliensis*. Plant Mol Biol 77(3):299–308
Xie C, Tammi MT (2009) CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinform 10:80
Xie Y, Wu G, Tang J (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30(12):1660–1666. doi:10.1093/bioinformatics/btu077
Xu G, Deng N, Zhao Z et al (2011a) SAMMate: a GUI tool for processing short read alignments in SAM/BAM format. Source Code Biol Med 6(1):2
Xu X, Pan S, Cheng S, Zhang B et al (2011b) Genome sequence and analysis of the tuber crop potato. Nature 475:189–197
Yamamoto T, Nagasaki H, Yonemaru J et al (2010) Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. BMC Genomics 11:267
Yan JB, Yang XH, Shah T et al (2010) High-throughput SNP genotyping with the GoldenGate assay in maize. Mol Breed 25:441–451
Yao JQ, Yu F (2011) DEB: a web interface for RNA-seq digital gene expression analysis. Bioinformatics 7(1):44–45
Ye C, Ma ZS, Cannon CH et al (2012) Exploiting sparseness in de novo genome assembly. BMC Bioinform 13 Suppl 6:S1. doi:10.1186/1471-2105-13-S6-S1
Yu J, Hu S, Wang J et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296(5565):79–92
Yu X, Guda K, Willis J et al (2012) How do alignment programs perform on sequencing data with varying qualities and from repetitive regions? BioData Min 5(1):6
Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18(5):819–821
Zhang J, Liang S, Duan J et al (2012a) De novo assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.). BMC Genomics 13:90
Zhang Y, Lameijer EW, 't Hoen PA et al (2012b) PASSion: a pattern growth algorithm-based pipeline for splice junction detection in paired-end RNA-Seq data. Bioinformatics 28(4):479–486
Zhang Z, Huang S, Wang J et al (2013) GeneScissors: a comprehensive approach to detecting and correcting spurious transcriptome inference owing to RNA-seq reads misalignment. Bioinformatics 29(13):i291–i299
Zhao W, Liu W, Tian D et al (2011) wapRNA: a web-based application for the processing of RNA sequences. Bioinformatics 27(21):3076–3077
Zhou X, Su Z, Sammons RD et al (2009) Novel software package for cross-platform transcriptome analysis (CPTRA). BMC Bioinform 10 (Suppl 11):S16. doi:10.1186/1471-2105-10-S11-S16
Zhou YH, Xia K, Wright FA (2011) A powerful and flexible approach to the analysis of RNA sequence count data. Bioinformatics 27(19):2672–2678
Zhou A, Breese MR, Hao Y et al (2012) Alt event finder: a tool for extracting alternative splicing events from RNA-seq data. BMC Genomic Suppl 8:S10
Zimin AV, Marçais G, Puiu D et al (2013) The MaSuRCA genome assembler. Bioinformatics 29(21):2669–2677
Zytnicki M, Quesneville H (2011) S-MART, a software toolbox to aid RNA-Seq data analysis. PLoS One 6(10), e25988

Chapter 21
Molecular Breeding for Genetic Improvement of Cotton (*Gossypium* spp.)

N. Manikanda Boopathi, Selvam Sathish, Ponnaikoundar Kavitha, Ponnusami Dachinamoorthy, and Rajasekar Ravikesavan

Abstract Even after several economical transformations and technical advances in the synthetic fibers, cotton is still the most preferred fiber for its comfort and simplicity. Cotton has been an economic mainstay in both developed and developing countries and there is a huge demand for improved raw cotton in global textile industries due to their modernization. Besides its utility in cloth making, cotton lint is widely used in medicine, fire-extinguishing and more importantly in revealing the molecular mechanisms of cell elongation and polyploidization. Despite its importance and demand, the genetic improvement of cotton production through conventional breeding has shown slow progress due its complex genetic inheritance. To this end, recent advances in transcriptome profiling, functional genomics, proteomics and metabolomics approaches, coupled with molecular marker-assisted breeding and transgenic technology have made significant contributions in enhancing the efficiency of cotton breeding; these methods are collectively referred as *molecular breeding*. Efforts to link fiber quantitative trait loci, QTLs, and expression of genes involved in fiber development with molecular breeding tools provide novel targets for the development of desirable cotton fiber and economically and agronomically important traits. In this chapter, we describe progress made in these arenas, and discuss their limitations and perspectives relative to the genetic improvement of this economically-unique crop.

N.M. Boopathi (✉)
Department of Plant Molecular Biology and Bioinformatics, Tamil Nadu Agricultural University, Coimbatore 641003, India

Department of Fruit Crops, Horticultural College and Research Institute, Tamil Nadu Agricultural University, Periyakulam 625604, India
e-mail: nmboopathi@tnau.ac.in; biotechboopathi@yahoo.com

S. Sathish • P. Kavitha • P. Dachinamoorthy
Department of Plant Molecular Biology and Bioinformatics, Tamil Nadu Agricultural University, Coimbatore 641003, India
e-mail: ssksathishbt@gmail.com; kavi_ja@yahoo.com; dachinamoorthy@gmail.com

R. Ravikesavan
Department of Millets, Tamil Nadu Agricultural University (TNAU), Coimbatore 641003, India
e-mail: chithuragul@gmail.com

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0_21

Keywords Cotton • Functional genomics • Genomics assisted breeding • Marker assisted selection • Molecular breeding

21.1 Introduction

Considering the circumstances of an ever-growing human population, it is needless to emphasize the importance of increasing the productivity of cotton *(Gossypium* spp.) fiber, which is one among the three basic needs of mankind: food, cloth and shelter. Cotton is harvested as seed cotton which is then ginned to separate the seed and lint. The long lint fibers are further processed by spinning to produce yarn which is knitted or woven into fabrics. Though different kinds of fabrics (that are derived from synthetic or natural resources) are used worldwide, cotton is still the most preferred natural clothing material since it can be made into different weights and types of fabrics: batiste, jersey, flannel, terry, corduroy, twill, denim and duck. In other words, because of its versatile nature, it can be used for most types of clothing, from underwear to sweaters to household linens such as sheets and curtains. Cotton clothing is very agreeable to wear, especially in hot weather, since it is simple and provides great comfort to move around. Because of their hypoallergenic and dust-mite resistance, cottons are the best choice of material for people who suffer from asthma or allergies or those who have sensitive skin prone to irritation. Furthermore, they allow for better air circulation, which helps remove and absorb body moisture, drawing heat away from the skin and keeping the body cool and dry as this clothing breathes so well. Cotton clothing is the ideal wash-and-wear fabric and it has been used by humankind for time immemorial. For example, the antiquity of cotton in the Indian subcontinent has been traced to the fourth millennium BC (Santhanam and Sundaram 1997). The first reference to cotton is found in a Rig Veda hymn. Fabric dated from approximately 3000 BC recovered from the Mohenjodaro site in Sind, present Pakistan, were identified to have originated from cotton plants, and closely linked to *G. arboreum* (Gulati and Turner 1929), thereby confirming that cotton lint was spun and woven into cloth even before 3000 BC.

In general, cotton fiber is graded on three characteristics: color or degree of whiteness, amount of foreign or plant matter in the cotton, and preparation of the fiber done by the gin. There is yet another important property: staple length. Egyptian cotton is known for its long staple used for finer types of fabrics. Extra-long staple cotton is also used for thread. When it comes to production, cotton clothing is easy to dye and blend with other fibers. Furthermore, cotton is the only fiber that becomes stronger when wet and is the preferred choice in hospitals, since it can endure high temperatures and thus it can be easily sterilized. It is also the preferred choice for fire-fighter uniforms, since it can be easily coated with flame-retardant substances. In the light of environmental concerns, cotton clothing makes a lot of sense. Cotton is biodegradable and a renewable resource and during processing, less

than 10 % of it is lost or deemed non-usable. Moreover, it is inexpensive, durable, and easy to care for. Thus the production and utilization of the cotton fiber has positive implications in several aspects: economic, social and environmental benefits.

This chapter is intended to describe the potentials and problems of molecular breeding in cotton. The first part describes progress in genetic and quantitative trait loci (QTL) mapping and system quantitative genetics. Recent developments in functional genomics, transgenics and their application in cotton molecular breeding are highlighted later in the chapter.

21.2 Geographical Distribution

The genus *Gossypium* L. is indigenous to Africa, Central and South America, Asia, Australia, the Galapagos, and Hawaii. The word *cotton* is of Arabic origin and refers to four species in the genus *Gossypium* (family Malvaceae), namely *G. hirsutum* L., *G. barbadense* L., *G. arboreum* L., and *G. herbaceum* L., each domesticated independently as sources of textile fiber. Globally, the *Gossypium* genus comprises about 50 species (Fryxell 1992). Actually, the place of origin of the genus *Gossypium* is not known; however the primary centers of diversity are west-central and southern Mexico (18 species), northeast Africa and Arabia (14 species) and Australia (17 species). DNA sequence data from the existing *Gossypium* spp. suggests that the genus arose about 10–20 million years ago (Seelanan et al. 1997). Cotton plants grow well in warm climates that occur throughout the arid and semi-arid regions of Africa, Australia, Central and South America, the Indian subcontinent, Arabia, the Galapagos, and Hawaii (Fryxell 1992).

Worldwide, cotton is grown annually on an average of 33–35 million ha (about 2.5 % of the world's arable land) under a great diversity of agroclimatic conditions and widely varied farming practices (Townsend 2010). It is grown in more than 100 countries as a significant cash crop and has great importance in the economy of such countries (Gillham et al. 1995). More than 120 million family units are engaged directly in cotton production in the world and annually about 350 million people are estimated to work in the cotton industry. The most prominent cotton-growing countries (in descending of total cotton lint production; Fig. 21.1) include China, India, USA, Pakistan, Uzbekistan, Tajikistan, Mexico, Brazil, Turkey, Egypt, Sudan, Australia, some African nations and Israel (Gruère 2011). China, India and Pakistan continue to be the largest consumers of raw cotton, together accounting for around 65 % of the total world raw cotton consumption.

India continues to maintain the largest area under cotton and is the second largest producer after China with 34 % of world area and 21 % of world production. However, weather conditions in India during 2012–2013 were abnormal which led to reduction of 0.5–0.6 million ha of area in some cotton growing regions and lower production by 20–25 lakh (hundred thousand) bales (James 2013). Interestingly, more than 90 % of the area is under transgenic Bt cotton (Table 21.1).

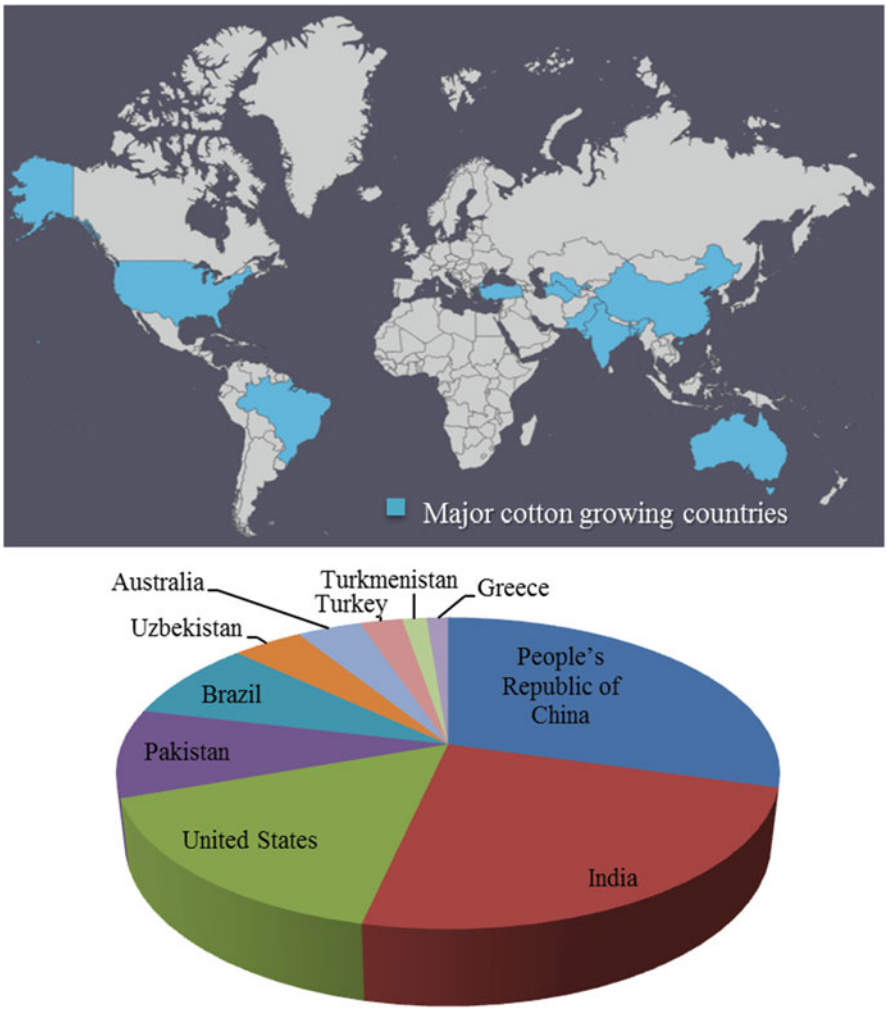

Fig. 21.1 Top ten cotton growing countries in the world (Source: USDA-Foreign Agricultural Service, Circular series May, 2012)

21.3 Biological Features of Cotton Crop and Cotton Fiber

The duration of annual cultivated cotton varieties/hybrids is 140–160 days. In most varieties, picking of the matured cotton boll (which contains the economic product of the plant, lint) begins 100–110 days after sowing (DAS). The length of each phenological phase (see below) differs depending upon the species, varieties and weather conditions as well as the cultivation techniques followed. Understanding the biological and genetic features of cotton indeed has a huge impact in designing a proficient cotton-breeding program.

Table 21.1 Comparison of total and transgenic cotton area in India

Year	Area (in million ha)	
	Total cotton area	Transgenic cotton area
2002	7.8	0
2003	7.8	0.2
2004	8.9	0.4
2005	8.7	1.4
2006	9.1	3.8
2007	9.4	6.1
2008	9.4	7.6
2009	9.5	8.2
2010	9.7	8.3
2011	10.1	9.09
2012	9.9	8.91
2013	10.75	9.675

Source: Ministry of Agriculture and Cotton Advisory Board, India

Germination of cotton seed depends on the genotype and availability of soil moisture, temperature and oxygen. Usually the radicle emerges within 2–3 DAS from the seed and newly germinated seedlings emerge above the soil 5–6 days after emergence of the radicle. The first cotton leaf appears 10–12 DAS and vegetative development reaches its peak during the fourth or fifth week. The first flower bud appears on the lowest fruiting branch at 45–50 DAS, depending upon prevailing temperatures. Fresh flower buds occur at regular intervals until shortly before flowering ceases. The time interval between the appearance of first flower bud and opening of the flower may be 25–30 days. Emergence of a large number of opened flowers is seen for a certain period and thereafter declines. During the peak period of flowering, vegetative growth is almost negligible and once the rate of flowering declines the vegetative growth restarts. The period of flowering is reduced by late sowing, strong plant competition and moisture stress. Furthermore, flowering in cotton is sensitive to both thermo- and photo-periods i.e. the day length alone or in combination with temperature determines the formation of flowering buds.

Cotton is predominately a self-pollinated crop; nevertheless, varying degrees of cross-pollination have been reported. Hence, breeding methods employed for cotton involve procedures for both self- and cross-pollinated crops. The hybridization and hybrid seed production is done according to the Doak (1934) method of hand emasculation and pollination. Genetic and cytoplasmic male sterility systems are also available (Kranthi 2013). Cotton flowers are extra-axillary, terminal and solitary and are borne on the sympodial branches. Cotton pollen is relatively large, heavy, sticky and watery and thus wind is not a factor in its pollination. Cross-pollination in cotton may vary from 0 to >20 %. Insects, more particularly, honey bees are attracted to the cotton flowers and they are active in cross-pollination. Pollination takes place usually in the morning during the opening of flower and anther dehis-

cence. Fertilization takes place 24–30 h after pollination (Govila 1969). Subsequently, the corolla along with anthers and filament drop from the fertilized ovary.

Cotton fibers are seed hairs and they originate from the epidermal cells of the ovular surface. As detailed by Basra and Malik (1984) fiber development is composed of four overlapping stages: (a) fiber cell initiation and enlargement (from −3 to 1 day post-anthesis, DPA), (b) fiber elongation (after anthesis until 25 DPA), (c) secondary cell wall synthesis especially cellulose deposition (15–50 DPA and (d) fiber cell dehydration and maturation (after 45 DPA). Thus, the near-synchronous growth of 500,000 terminally differentiated single-type fiber cells per ovule (Havov et al. 2008) is characterized by major discrete developmental stages – differentiation/initiation, expansion/elongation, primary cell wall synthesis, secondary cell wall synthesis and maturity (Wilkins and Jernstedt 1999). Among these the productivity and quality of cotton, that determine the higher economic profit, depend mainly on two developmental processes: (a) fiber initiation, which determines the number of fibers present on each ovule and (b) fiber elongation, which determines the final length and strength of each fiber.

21.4 Genetic Features of Cotton

21.4.1 Nuclear Genome

Gossypium includes 46 diploid ($2n = 2x = 26$) and 5 well-established and 1 purported tetraploid ($2n = 4x = 52$) species (Krapovickas and Seijo 2008; Wendel and Albert 1992). All diploid species of *Gossypium* have $n = 13$. Based on chromosome pairing relationships, those diploids were designated into 8 different genome types as A, B, C, D, E, F, G and K (Endrizjz et al. 1984). While retaining common chromosome number and largely-collinear gene order, the cotton genome types diverged into genome groups that vary in haploid genome size from 2500 Mb in the K genome, to less than 900 Mb in the D genome (Hendrix and Stewart 2005; Lin et al. 2010).

On the other hand, 5 totally tetraploid ($n = 2x = 26$) *Gossypium* species are recognized. All tetraploid species exhibit disomic chromosome pairing (Kimber 1961). Chromosome pairing in interspecific crosses between diploid and tetraploid cottons suggests that tetraploids contain two distinct genomes, which resemble the existing A genome of *G. herbaceum* ($n = 13$) and D genome of *G. raimondii* Ulbrich ($n = 13$), respectively (Wendel 1989).

It is believed that the two lineages of diploid ancestral A and D genomes may have diverged from a common ancestor 6–11 million years ago (Wendel 1989) as shown in Fig. 21.2.

Polyploidization was followed by radiation and divergence, with distinct $n = 26$, AD genome species is indigenous to Central America (*Gossypium hirsutum*), South America (G. *barbadense, G. mustelinum* Miers ex Watt), the Hawaiian Islands (*G. tomentosum* Nuttall ex Seemann), and the Galapagos Islands (*G. danuinii* Watt). The 5 polyploid *Gossypium* spp. recognized today, including cultivated cottons of

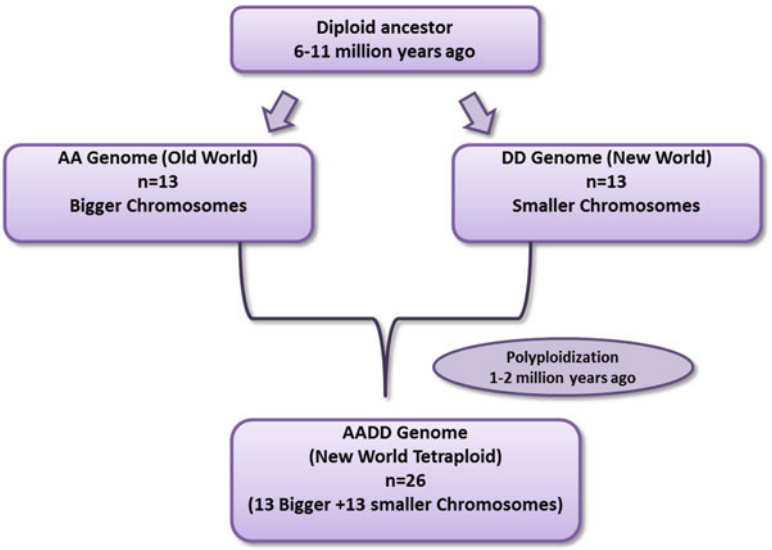

Fig. 21.2 Evolution of cultivated cotton from its phylogenetic relatives

G. hirsutum and *G. barbadense*, are thought to have originated 1–2 million years ago by transoceanic migration of Old World A genome progenitor, followed by hybridization with new world D genome progenitor (Endrizjz et al. 1984; Wendel 1989; Wendel and Cronn 2003). This hypothesis is a revised version of the simple origin of Old and New World diploids and their hybridization in the new world 1–2 million years ago (Fryxell 1992).

Most commercially-cultivated cotton is derived from two tetraploid species, *Gossypium hirsutum* (generally referred as upland or American cotton that occupies 90 % of world plantings) and *G. barbadense* (Pima, Egyptian or long-staple cotton). Two diploid species, *G. arboreum* and *G. herbaceum* (popularly referred in India as desi cottons) are also cultivated in Asia and Africa with uncertain areal extent.

All tetraploid cotton species came from interspecific hybridization between the A-genome species *Gossypium arboreum* and the D-genome species *G. raimondii* (Wendel and Cronn 2003). The A-genome species are cultivated, whereas the D-genome species do not produce spinnable fiber. Although *G. arboreum* and *G. raimondii* are the putative donor species for the A and D chromosome groups, respectively, tetraploid cotton species differ greatly with respect to plant morphology as well as economic characteristics such as fiber production, oil content, and pest and disease resistance.

Likewise, *Gossypium arboreum* (1746 Mb/1C) has a genome size that is almost twice that of *G. raimondii* (885 Mb/1C) and approximately 2500 Mb/1C for *G. hirsutum* (Hendrix and Stewart 2005). Despite intensive studies in generating high-resolution genetic and physical maps (see below), the identity of the parental diploids and the antiquity of polyploidization is still unclear.

21.4.2 Organelle Genome

Because of the low mutation and non-recombination properties of the chloroplast (ct)-genome, *Gossypium* phylogeny can be reconstructed independently of that of the nuclear genome (Li et al. 2013a). The ct-genome is quite conservative, while the nuclear genome has recombined continuously during the evolution and species formation of cotton. Restriction site mutations in chloroplast DNA confirmed that the ct-genome of *Gossypium* has descended through the female parent (Wendel and Cronn 2003). The complete cotton ct-genome is 160 Kb in length and contains a number of dispersed direct and inverted repeats that provide new insights into the evolution of the *Gossypium* genome (Lee et al. 2006). The inheritance of ct-genomes reflects patterns of seed flow and dispersal from progenitors to descendants (Li et al. 2013a). The *G. hirsutum* mitochondrial (mt)-genome is 621 Kb in length and possesses most of the common characters of higher plant mt-genomes such as the presence of conserved gene clusters, congruence of intergenic sequences and genic contents (Liu et al. 2013).

21.5 Potentials and Challenges in Cotton Production

Cotton is the premium world source of natural textile fiber, and has been used for thousands of years for clothing, fine paper, and other purposes such as an oilseed crop and animal feed. Its fiber, commonly known as cotton lint, is the principal source for the textile industry. Cotton lint has an annual textile global market value of approximately USD 630.6 billion in 2011 (http://marketpublishers.com/). In spite of the availability of synthetic fibers, demand for cotton fiber remains high and the international market showed susceptibility to cotton price shocks in 2010–2011, with high demand and low production in some major cotton producing countries.

Thus, cotton is the fifth most economically important crop in the world and it is cultivated on about 33 million ha across the world. India accounts for about 33 % of the global cotton area and contributes 21 % of the global cotton produced, currently ranking second after China. India's contribution to global cotton production increased from 14 % in 2002 to 20.5 % in 2007 (Kranthi 2013). Currently 14.5 million ha (i.e. 45 % of the total global cotton area) is under Bt cotton (cotton hybrid which contains the Cry gene isolated from *Bacillus thrungiensis* that provides resistance to lepidopteron pests; see below). Bt-cotton was first approved in 1996 for commercial cultivation in the USA. It was released for cultivation in China, Mexico and Australia in 1997, and later in Argentina (1998), South Africa (1998), Indonesia (2001), India (2002), Colombia (2003), Philippines (2003), South Korea (2003), Brazil (2005), Myanmar (2006), Burkina Faso (2008) and Egypt (2009). Among these countries, Brazil, China and India have made the most significant progress in the period 1999–2009. China made spectacular progress with an impressive increase to 47.5 million bales in 2007 from 22.9 million bales in 1999. India became a leading global exporter of raw cotton with exports of 0.6–1.5 million mt of raw cotton

each year from 2005 onwards, while concomitantly, imports declined from 0.43 to 0.09 million mt. Domestic consumption also increased from 15.8 million bales in 2002, to an estimated 25.8 million bales in 2010 (Kranthi 2013). James (2013) reported the benefits of Bt cotton in India and provided evidence of the advantages of transgenic cotton over non-transgenic cotton production in India.

Cotton production provides income for approximately 100 million families, and approximately 150 countries are involved in cotton import and export. Cotton fiber is an outstanding model for the study of plant cell elongation and cell wall and cellulose biosynthesis (Boopathi and Ravikesavan 2009). The fiber is composed of almost pure cellulose, the largest component of plant biomass. Compared to lignin, cellulose is easily convertible to biofuels. Translational genomics of cotton fiber and cellulose may lead to the improvement of diverse biomass crops. Therefore, it is apparent that cotton production has huge potential and determines the economic and social welfare of the several countries.

On the other hand, cotton production is limited by several factors. In India, despite the good progress made by public and private sector research and development, it is a matter of concern that productivity started to decline from 566 kg/ha in 2007 to 475 kg/ha in 2010 (Kranthi 2013). Several factors including erratic rainfall and emerging biotic and abiotic stresses were found to have contributed to the decline in yields (Boopathi et al. 2011). The cotton varieties and hybrids are usually more susceptible to insect pests such as jassids (leafhoppers), whiteflies, bollworms and diseases such as bacterial blight, verticillium wilt and leaf curl virus. The quality profile of Indian cotton fiber has also changed. Therefore, it is important to develop varieties suitable for dense plantings that are more efficient in utilizing water and nutrients, can resist pests and diseases and produce better fibers. For that reason, a priority area of research in Asian and African countries is to develop varieties through *ideotype breeding* of compact genotypes appropriate for narrow and ultra-narrow spacing, with specific fiber traits and insect resistant for specific locations. Such measures can enhance yields and provide sustainable options for optimal and efficient use of inputs.

Recently, it has been observed that high cost involved in cultivation of Bt cotton hybrids and low return from the product forced the Indian farmers to abandon cotton and cultivate other crops. Therefore, it is the appropriate time to genetically improve the cotton cultivars that suit the demands of both farmers and textile industries.

21.6 Achievements and Obstacles in Conventional Breeding

Success stories have been documented on several occasions in the development of varieties and hybrids in cotton with enhanced pest and disease resistance through conventional breeding procedures. Nevertheless, the cotton end-users demand more. The textile industry has been demanding quality cotton suitable for the recent spinning systems which were developed to achieve higher production rates, productivity and automation for cost reduction of yarn production. Open-end spinning systems such as rotor spinning, friction spinning and air-jet spinning which ensure

high rate of production and large size of yarn and package are widespread in modern textile mills (Kranthi 2013).

For such new systems, high fiber strength and fineness are currently more important. Especially, there is a need to (a) improve fiber strength (25–30 g/tex for 3 mm gauge) and percentage of mature fibers (75–80 %), ginning characteristics and optimum micronaire value without affecting maturity and (b) reduce stickiness and motes (neps and naps), short fiber content, trash content and seed-coat fragments. This demand has been met in the past to some extent by classical breeding. Strategic plant-breeding plans were employed to improve fiber quality traits and several outstanding varieties and hybrids with superior fiber length and strength were developed through traditional breeding methods and they were widely cultivated over the last three decades (Boopathi et al. 2014). For example, Indian breeders have been credited for their achievements particularly in the development of highly adaptable varieties (such as Bikaneri Narma, LRA 5166, Narasimha, SRT 1 and MCU 5), early maturing varieties, sucking pest tolerant varieties, disease resistant desi cotton varieties, high yielding and superior fiber quality varieties of *Gossypium arboreum* and *G. herbaceum*, development of inter-specific and intra-specific hybrid cotton, development of high yielding *G. hirsutum* varieties and the development of finest quality *G. barbadense* variety (e.g. Suvin).

One of the most spectacular achievements that stand out as a technology that had the greatest influence on cotton in India is the hybrid cotton (Kranthi 2013). In a revolutionary development in 1970, the world's first cotton hybrid H 4 (intra *hirsutum*) was developed in India and it was widely cultivated. In our laboratory, simple selections of accessions from the breeding population were used to identify a drought-tolerant cotton cultivar under water-stressed conditions (Fig. 21.3). The selected lines are being evaluated in multiple environments and seasons to release them as drought-tolerant varieties. On the other hand, genetic improvement of cotton for desirable traits (such as drought tolerance) through traditional plant-breeding techniques has several limitations including the longer time to develop a variety or hybrid with huge labor efforts and poor precision in genetic manipulation of desired traits.

21.7 Cotton Germplasm, Diversification and Enrichment of Primary Gene Pool

As stated above, genetic improvement of cotton with conventional breeding program is predominantly hindered by the lack of complete knowledge about and precise manipulation of agronomic traits. At the same time, naturally available cotton continues to be a resource for the upcoming breeding program. In order to safeguard the huge economic value of cotton and cotton by-products, it is essential to focus on coordinated efforts to collect and preserve cotton genetic resources and their effective utilization through widening the cotton gene pool (Boopathi et al. 2014). Major collections, including breeding materials and wild-species germplasm, genetic and

Fig. 21.3 Performance of cotton accessions derived from MCU5/TCH1218 under water stress in the field. Water stress was imposed during flowering phase to harvest. Performance of parents, MCU5 and TCH1218, under drought are shown in (**a**) and (**b**), respectively. Some of the progenies have shown excellent economic yield and fiber quality under severe water stress (**c**) whereas some accessions shown drought tolerance but poor yield (**d**)

cytogenetic stocks for cotton gene pools currently exist in Australia, Brazil, China, India, France, Pakistan, Turkey, Russia, Uzbekistan and the USA. Abdurakhmonov (2014) recently gathered detailed information on the status and contents of the major cotton germplasm collections of the world, targeting past and current progress, critical challenges and opportunities to preserve the cotton genetic resources within specific collections. The study also included characterization of the germplasm pools, the level of genetic and molecular diversity, and the on-going multinational communication and collaboration to enhance the protection, preservation, and evaluation of the global cotton germplasm resources.

The utilization of available germplasm in the cotton gene banks can be further enhanced by adding the most suitable genetic stocks that can be procured from various international sources and diversifying and strengthening the primary gene pool maintained at the given germplasm/breeding unit. Such activity will be useful in the development of breeding lines by introgressing beneficial alleles from the wild species, races and derivatives. Although elite x elite crosses are typical of traditional plant breeding, interspecific crosses are rarely used in cotton breeding because of numerous barriers. The finding that the *Gossypium hirsutum* allele is favorable at some loci and the *G. barbadense* allele at other loci shows that recombination of favorable alleles from each of these species may form novel genotypes distinct from either parent species. Similarly, the genomic exploration of other accessions of these species or other wild tetraploid cottons (*G. tomentosum, G. darwinii, and G.*

mustelinum), maintained at cotton germplasm bank could yield additional valuable alleles (Boopathi et al. 2014). However, the question is how to identify and characterize such valuable alleles and exploit them in breeding program.

Plant breeding has a long history of incorporating the knowledge obtained from the latest innovations in biology and genetics to enhance crop improvement. During the past four decades, the recent revolution in our understanding of the analysis and induction of genetic variation (through mutagenesis, marker-assisted selection and transgenics), cytogenetics, quantitative genetics, molecular biology, biotechnology and most recently, structural and functional genomics and bioinformatics tools, have established new breeding tools. These tools aid in the creation, analysis, and manipulation of genetic variation and the development of improved cultivars in several agricultural and horticultural crop plants.

Integration of the abovementioned knowledge into plant-breeding programs has created a new paradigm in genetic improvement of crop plants; such efforts are collectively referred as *molecular breeding*. Thanks to these developments, the diversification of the primary cotton gene pool can be achieved through breeding populations (either naturally occurring or synthetic), segregating progeny from a cross of selected parental lines, exotic materials that are not adapted to the target environment, wide interspecific crosses, naturally occurring or induced mutations, the introduction of transgenic events, or combinations of these sources. While molecular markers and genomic tools have been highly efficacious in widening existing genetic variation within species, transgenic technology generates new genetic diversity that often extends beyond species boundaries (covering all kingdoms of life or designed and assembled de novo in the laboratory) and creates an essentially infinite pool of novel genetic variation.

However, successful molecular breeding in cotton requires knowledge of the genomic organization and function of genes, a solid foundation in statistical approaches to estimate genetic effects and a strong background in plant biology. Required as well are experience with both the laboratory methods of molecular biology/functional genomics and field-based breeding practices, along with the ability to manage large datasets with diverse data types.

21.8 Genetic and QTL Mapping Resources

21.8.1 Bi-parental and Multi-parental Mapping Populations

Most of the desirable traits that plant breeders try to improve are quantitative, rather than qualitative, in nature. Quantitative traits vary continuously (e.g. yield, quality and stress tolerance), whereas qualitative ones are typically (but not always) binary (yes vs. no – e.g. resistance to a fungus or flower color). Quantitative traits are usually governed by a number of genes, while qualitative ones are often simply inherited (one or two genes; hence called simpler or major traits). The loci involved in the inheritance of quantitative traits are commonly called QTL (quantitative trait loci).

QTL mapping, in general, refers to identification of genomic region(s) that govern the expression of the given agronomic trait. Molecular markers (either DNA or RNA based) are employed to construct a genetic or linkage map which can be employed to understand the genetic basis and improvement of the complex polygenic traits (Boopathi 2013).

The identification of tightly-linked markers to the stable QTLs affecting the target agronomic traits across the generations would be useful in the process of marker-assisted selection (MAS) and thus increase the efficiency of a breeding program. Therefore, identification of molecular markers linked to the target trait QTLs would allow cotton breeders to trace such a trait in early plant growth stages or in early segregating generations (Fig. 21.4).

The main reasons supporting the utilization of molecular markers in cotton breeding programs are the percent of heritability of the markers and their lower cost (Winter and Kahl 1995). Hence, molecular markers are extensively being employed in selection of traits with low heritability, identification, and introgression of complex fiber productivity and quality traits from native or exotic germplasm into elite cultivar via MAS (Rong et al. 2007). In plant breeding, MAS is a relatively new concept, nevertheless the original selection concept per se has not changed; that is, the purpose of the selection is to search and preserve the best genotypes, but using molecular markers. At the same time, it is necessary to consider effectiveness and the cost of MAS (which is greatly influenced by the marker system used) besides polymorphism, technical feasibility and so forth. The value, ease and cost of measurement and the nature of genetic control of agronomic traits will also determine the way in which molecular markers may be effectively used in a breeding program (Winter and Kahl 1995). A highly saturated marker linkage map is necessary for effective MAS. Both intraspecific meiotic configuration analysis and interspecific linkage analysis have indicated that the cotton genome map is ~5000 cM or larger, considerably longer than genomes of bread wheat (3791 cM), soybean (3159 cM), maize (1807 cM), rice (1530 cM), tomato (1472 cM), and barley (1279 cM) (http://www.ncbi.nlm.nih.gov/mapview/). A highly saturated genetic map of cotton with a 5000 cM genome will require 3000 molecular markers to map at an average of 1 cM density (He et al. 2007).

The first and foremost requirement in QTL analysis is development of a mapping population to study the segregation of molecular markers and the trait of interest. To this end, different kinds of mapping populations are being employed including F_2 population, backcross population, recombinant inbred lines (RILs), near isogenic lines (NILs) and double haploid lines (DHLs) (Boopathi 2013). The major drawback to using F_2 or backcross populations is that the populations are not eternal. Therefore, the source of tissue to isolate DNA will be exhausted at some point in time and hence it would be necessary to begin mapping again in another population. Permanent mapping populations such as RILs and DHLs will not have this limitation. Furthermore, they permit multi-location experiments and thus they are expected to be superior over other population with regard to the power of QTLs detection. Compared to DHLs, it is relatively easy to develop RILs in cotton since it is recalcitrant to tissue culture protocols (Boopathi et al. 2011). Usually, RILs are devel-

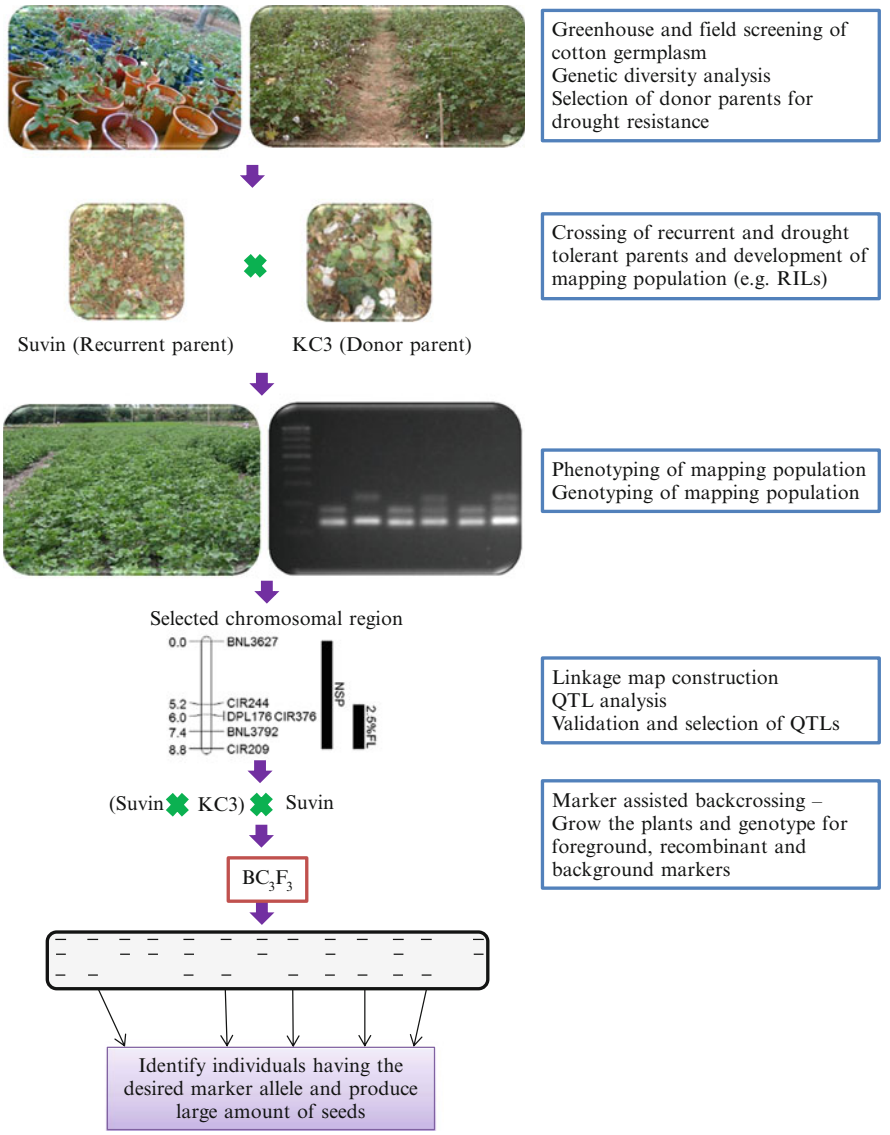

Fig. 21.4 Simplified illustration of marker-assisted selection in cotton. Screening of available germplasm helped to identify donor parent (in the above example, KC3 was selected as donor parent for drought-resistant improvement since it produced higher fiber yield under water-stressed conditions. Water stress was imposed during flowering phase (50 days after sowing) and the stress was continued until harvest). Recombinant inbred lines (RILs) were developed to genetically dissect out fiber quality traits under drought and tightly-linked markers were identified for each QTL. The QTLs were validated across the environments, seasons and genetic backgrounds and used for marker assisted selection as above

oped by single-seed selections from selfed individual plants of an F_2 population and single-seed descent is repeated for several generations (at least up to F_6).

The majority of the genetic maps in cotton were constructed using mapping populations derived from either interspecific or intraspecific single-cross hybridization. Due to a lower level of within-species and between-species polymorphism, most of the maps have included only a relatively small portion of the genome. For example, even a joint map from different mapping populations has shown 31 % coverage of the cotton genome. If such a poor coverage genetic map is used for QTL mapping, only a small portion of genome will be explored and large amounts of QTL information could not be revealed. Use of four parents of a double cross (otherwise referred as four-way cross) has been shown to increase the density of genetic maps (Qin et al. 2008). Thus, a four-way cross can increase the density of the linkage map and in some cases, it can counteract the lower levels of polymorphism found in certain crops such as cotton. Furthermore, use of a four-way cross can potentially reduce the type II error caused by a random sampling of parents and increase the probability of detecting QTL if they segregate in a single-line cross but not in the other single-line cross (Boopathi 2013). The features of the genetic structure of RILs can be studied using two-, four- and eight-way crosses following either selfing or sib mating. Although an eight-way cross RILs has been successfully shown in the mouse, it has yet to be demonstrated in major crops, particularly in cotton.

21.8.2 Molecular Markers: Development and Application

A molecular marker is defined as a particular segment of DNA that represents the difference at the genome level (Boopathi 2013). Molecular markers may or may not correlate with phenotypic expression of a trait. Molecular markers offer numerous advantages over conventional morphological markers and isoenzymes. They are stable and detectable in all tissues regardless of growth, differentiation, development and status of the cell. Furthermore, they are not confounded by the environment, pleiotropic and epistatic effects. Restriction fragment length polymorphism (RFLP) was the first reported molecular-marker technique in the detection of DNA polymorphism. After the invention of polymerase chain reaction (PCR), several PCR based markers were developed (Agarwal et al. 2008). Although several marker techniques are available at this point, an ideal molecular marker technique for cotton should have the following criteria: (a) be polymorphic and evenly distributed throughout the genome (b) provide adequate resolution of genetic differences (c) generate multiple, independent and reliable markers (d) be simple, quick and inexpensive (e) need only small amounts of tissue and DNA samples and (f) have linkage to distinct phenotypes. Unfortunately no single molecular-marker technique yet exists to fulfill the above conditions. Marker techniques differ from each other with respect to important features such as genomic abundance, level of polymorphism detected, locus specificity, reproducibility, technical requirements and cost.

With the significant reduction of genotyping cost and increased throughput, next-generation sequencing and detection of single nucleotide polymorphism (SNP), genotyping technologies are increasingly making MAS as a choice of cotton-breeding strategy. Now cotton genomic diversity can be characterized at the sequence level based on haplotype maps, and molecular breeding can be further accelerated through allele mining and haplotype-based design and selection (Xu et al. 2012). An obvious problem that usually arises is, how to choose the most appropriate marker among the myriad of different marker technologies. In general, the choice of a molecular-marker technique should represent a compromise between reliability and ease of analysis, statistical power and confidence of revealing polymorphisms.

21.9 Construction of Individual (or Genome Specific) and Integrated Genetic Maps

Since 1985, elucidating the structure of the cotton genome using morphological markers (and after 1994 with molecular markers) has progressed tremendously (see: Boopathi et al. 2011; Ulloa et al. 2007). The first diploid A-genome map comprised 18 morphological markers in 7 linkage groups, and the tetraploid AD genome map comprised approximately 80 markers distributed across 18 linkage groups (Endrizzi et al. 1985). Now almost fully resolved molecular genetic maps for the diploid A genome and D genomes and the tetraploid AD genome are available and all 26 AD linkage groups (and their A and D homoeologs) have been linked to the cytological map (Ulloa et al. 2007). With molecular markers distributed every 2–5 cM (on average) for the diploid A and D genomes and the tetraploid AD genome, genetic linkage maps can be used to describe and deploy the cotton genome in ways that were unbelievable two decades ago (Ulloa et al. 2007).

It was observed from physical-mapping studies that the D genome is approximately 980 Mbp long, the A genome contains about 1860 Mbp, and the length of the AD genome is additive, about 2835 Mbp (Grover et al. 2004). On the contrary, genetically the A_T and D_T sub-genomes are about 50–93 % longer than their extant diploid models (Rong et al. 2004). At the same time, the diploid A and D genomes, which differ in physical size by a factor of about 2, are recombinationally equivalent, as are the A_T and D_T tetraploid sub-genomes (Brubaker et al. 1999; Rong et al. 2004).

The conservation of synteny and collinearity across the cotton genomes is high, except for the well-characterized translocations that differentiate the two A-genome species (*Gossypium arboreum* and *G. herbaceum*) from each other and from the tetraploid A_T sub-genome. This level of conservation corresponds with the evidence that the cotton genomes are quite static structurally, particularly in the transition from diploidy to allopolyploidy, as was predicted by early cytological observations (Ulloa et al. 2007). This gross structural conservation will increase the transferability of information across genomes; however, the conservation of synteny and collinearity across the cotton genomes is not complete.

An individual genetic map has certain limitations, such as large gaps due to the lack of polymorphism in particular genomic regions. Moreover, marker order errors might be present and unnoticed in a map constructed using a single population. A consensus or integrated map constructed utilizing the information from multiple segregating populations provides a very important reference resource. It gives an opportunity to map a larger number of loci as compared to most single crosses, thus increasing the number of potentially-useful markers across divergent genetic backgrounds. In addition to providing opportunities to validate the marker order, a consensus map also provides greater genome coverage. Blenda et al. (2012) constructed a consensus genetic map with 8254 loci integrating 6 independent component maps and provided a very valuable resource for studies of the cotton genome structure.

Genetic or linkage map is primarily used as a tool for QTL or gene(s) discovery. As it is described in several papers (Blenda et al. 2012; Boopathi et al. 2011; Ulloa et al. 2007), QTL mapping is being used to dissect complex genetic traits such as fiber quality traits in cotton. The association of molecular markers (that were mapped in the genetic map) with agronomic traits (that are costly or laborious to score) will provide cheap, easily scored alternates for molecular-breeding programs. Simultaneously the saturated fully resolved genetic map is also useful in constructing the framework of physical map.

21.10 QTL Mapping

Genomic regions or QTLs governing the expression of important agronomic traits were reported in cotton (Boopathi et al. 2011; Ulloa et al. 2007) and selected recently published QTLs are listed in Table 21.2. Such studies were conducted with continuous efforts, including crossing parents that were differing for the given agronomic trait, studying the resulting progeny to establish linkage between phenotypes and DNA markers that were mapped genome-wide and finally identifying chromosomal regions controlling such traits (Boopathi 2013). Many QTL studies (listed in Boopathi et al. 2011; Ulloa et al. 2007) have indicated the surprising result that the D sub-genome, from an ancestor that did not produce spinnable fiber, has contributed at least as much as, or sometimes more than, the fiber-producing A genome, to the superior fiber quality and yield of allotetraploid cotton. These studies show that positive effects on all fiber characteristics are derived from alleles contributed by both inferior and superior parents.

Hundreds of QTLs were also detected recently for fiber quality, agronomic and yield related traits (for example, Gore et al. 2014; Mei et al. 2014; Zhiyuan et al. 2014). However, such QTLs cannot be directly deployed for molecular breeding since most of the detected QTLs are held up within family; the size of QTL effects that can be detected are limited, and inferences are restricted to a single population and set of conditions. Alternatively, combined QTL analysis by collecting information from several or many studies through meta-analysis was proposed (see below). Integration of QTL from different populations into a common map facilitates explo-

Table 21.2 Selected examples of QTLs associated with fiber quality traits that were recently reported in cotton

Trait	Chromosome[a]	Maximum[b] R^2	Reference
Seed-cotton yield	A12	8.4	Liu et al. (2012)
Lint yield	D2	13.2	
Bolls per plant	D8	8.5	
Boll weight	D2	20.8	
Lint percentage	A9	10.4	
Seed index	A5	11.7	
Lint index	D4	14.5	
Fruit branch number	D7	16.9	
Fiber strength	C7	19.8	Sun et al. (2012)
Fiber length	C7	13.7	
Micronaire	C7	18.2	
Fiber uniformity	C16	30.6	
Fiber elongation	C15	12.8	
Fiber strength	C7	27.8	
Fiber length	C25	20.6	
Micronaire	C16	48.0	
Fiber elongation	C25	13.4	
Elongation	15	47.1	Zhang et al. (2012)
Length	1	55.8	
Micronaire	24	34.7	
Strength	24	32.6	
Uniformity	8	50.5	
Uniformity index	9	15.4	
Boll weight	9	22.4	
Node of first fruiting branch	D3	19.1	Li et al. (2012)
Height of node of first fruiting branch	A9	13.9	
Fiber length	13	20.1	Liang et al. (2013)
Fiber uniformity ration	5	15.9	
Fiber strength	13	22.0	
Fiber elongation	24	10.7	
Fiber micronaire	26	16.0	
Lint percentage	A5	14.8	Zhiyuan et al. (2014)
Seed-cotton yield	A7	17.9	
Seed index	A7	15.5	
Boll weight	D11	19.5	
Fiber elongation	A7	10.6	
Fiber length	D11	25.3	
Fiber strength	D3	17.5	
FSCI	D3	9.4	
Fruit branch number	A9	19.3	

(continued)

Table 21.2 (continued)

Trait	Chromosome[a]	Maximum[b] R^2	Reference
Boll size	19	20.3	Gore et al. (2014)
Lint yield	17	17.8	
Fiber strength	11	18.1	
2.5 %-span length	25	19.3	
Length uniformity	15	22.3	
Length uniformity	25	14.3	

[a]Chromosome in which the given QTL was identified
[b]Maximum R^2 (percentage of phenotypic expression) reported for the particular QTL identified in the study

ration of their allelic and homeologous relationships, although the level of resolution is limited by comparative marker densities, variation in recombination rates in different crosses, variation in gene densities across the genome and other factors.

As an alternative, instead of using anonymous molecular markers, Li et al. (2013b) used fiber development specific genic markers and integrated them with the genetic map. QTL mapping and functional genomics (see below) results have clearly demonstrated that genomic distribution of these genes and their expression pattern were consistent with QTL analysis, quantitative real time polymerase chain reaction analysis and field experiments. Such analyses provide additional validated functional markers for cotton molecular breeding.

In summary, improved understanding of the cotton genome and compilations of QTLs linked with agronomic and economic traits has begun to contribute to molecular breeding in cotton, but with a long way yet to go. Because of the polyploid genome structure and its large genome size, cotton provides poor accessibility for genome mapping efforts. Furthermore, it is confronted with mapping populations and complex QTL x environment interactions which result in incongruence among QTL studies and greatly hamper progress in the molecular-breeding procedure. Hence, it is imperative to develop new tools, methods and statistical procedures to address these problems and such efforts should increase the efficiency of molecular breeding in cotton.

21.11 Mining Beneficial Alleles through Association Mapping and Nested Association Mapping

As stated above, identifying completely characterized QTLs that are useful in molecular breeding is an enormous task because of limitations in the scope of allelic diversity and resolution in available bi-parental and multi-parental mapping populations. Particularly, the poor resolution of the QTLs is mainly due to the limited number of recombination events that occur during population development. As an alternative, a large-scale genome wide association mapping procedure was

introduced to identify beneficial alleles from the natural plant population (Zhu et al. 2009). Association mapping uses genomic technologies for exploiting natural diversity, which had been utilized only on a small scale before the genomics era. It takes advantage of the remarkable recombinations over a long history as linkage disequilibrium generally decays within 2 kb. Association mapping was successfully used in cotton to identify markers linked to fiber quality traits (Abdurakhmonov et al. 2008; Kantartzi and Stewart 2008). However, because of the requirement of a large number of highly polymorphic molecular markers and the confounding effects of population structure, whole genome association analysis was found to be difficult in cotton. To circumvent these problems, nested association mapping (NAM) population can be constructed to enable high power and high resolution through joint linkage-association analysis by capturing the best features of both linkage and association mapping. The genetic structure of NAM population is a reference design of 25 families of 200 RILs per family. NAM has been successfully implemented in maize (McMullen et al. 2009) but it is yet to be demonstrated in cotton.

21.12 Genomics Assisted Breeding

In general, QTLs identified through either linkage mapping or association mapping-based approaches have low resolution and have been located at 10–20 cM intervals. Such intervals may span several hundreds of genes and identifying the right candidate gene(s) with causal effect on the given trait is like finding a *genic needle* in the *genomic haystack*. Therefore, to identify the causal gene(s), positional cloning of QTLs have been undertaken in several crop species. Although many reports are available on cloning of QTLs associated with different traits (see Salvi and Tuberosa 2007), there are very few reports addressing QTL cloning for fiber quality and other agronomic traits in cotton. Newer genomics approaches such as next-generation sequencing (NGS) and functional genomics hold great promise for accelerating identification of candidate genes for fiber quality and other economically-important traits. Recent advances in genomics make it possible to not only conduct large-scale and high-throughput marker genotyping, but also to sequence or resequence the genomes of germplasm collections, thus facilitating the identification of QTLs and candidate genes associated with target traits. These are briefly described below.

21.12.1 Cotton Genomic Resources and Genome Sequencing

In order to achieve the ultimate goal of characterizing the gamut of diversity existing among cotton genomes, the worldwide cotton community has prioritized the D genome progenitor, *Gossypium raimondii* for complete sequencing. As stated earlier, even though members of the D genome clade do not make spinnable fiber (see above), genetic mapping has shown that the majority of fiber QTLs mapped in tetraploid cotton fall on D genome (*G. raimondii*-derived) chromosomes, suggesting

that the D genome has been crucial to the evolution of the higher fiber quality and yield of cultivated tetraploid cottons. *G. raimondii* is the smallest genome in the *Gossypium* genus at ~60 % of the size of the diploid A genome and 40 % of the tetraploids.

A whole genome physical map of *Gossypium raimondii* was constructed by integrating genetically-anchored overgo hybridization probes, agarose-based fingerprints and high information content fingerprinting. In species where no whole-genome sequence is yet available, a physical map is a useful tool in a wide range of activities including comparative genomics and gene cloning. Physical mapping also provides a method of genome assembly independent of a sequence, and is useful in contributing to and/or validating whole-genome shotgun sequences. A total of 13,662 BAC-end sequences and 2828 DNA probes were used in genetically anchoring 1585 contigs to a cotton consensus genetic map (Lin et al. 2010). It was believed that such an integrated genetic-physical map is of value both in assembling and validating a planned reference sequence of *G. raimondii*. Wang et al. (2012) sequenced and assembled the draft genome of *G. raimondii* using a whole-genome shotgun strategy. Over 73 % of the assembled sequences were anchored on 13 *G. raimondii* chromosomes and consisted of 41,307 contigs and 4715 scaffolds that accounted for approximately 88.1 % of the estimated *G. raimondii* genome. The genome contains 40,976 protein-coding genes among which 92.2 % were further confirmed by transcriptome data.

A high-resolution genetic map that anchored and oriented 90.4 % of the *Gossypium arboreum* assembled scaffolds on 13 pseudo-chromosomes from paired-end sequencing of 193.6 Gb of clean sequence was obtained by Li et al. (2014). The final assembly of the *G. arboreum* genome showed that it was 1694 Mb in total length. This assembly was compared with that of *G. raimondii* to understand possible routes for genome evolution and species divergence. Collinearity was observed between *G. arboreum* assembly and the tetraploid genetic map that was reported previously. A total of 41,330 protein-coding genes were identified in the *G. arboreum* genome and over 96 % of predicted coding sequences were supported by transcriptome sequencing data, which indicated high accuracy of *G. arboreum* gene predictions from the genome sequence (Li et al. 2014).

Sequencing of these two genomes clearly indicates that although there is complexity in sequence assembly, using these sequences as a reference, it is now possible to sequence the remainder of the cultivated cotton tetraploid genomes, which would be a great resource for molecular breeding in cotton genetic improvement.

21.12.2 Transcriptomics

Cotton genes are highly conserved in the genomes of wild and cultivated species, as well as diploid and tetraploid species, despite millions of years of evolutionary history (Rong et al. 2004). The phenotypic variation in fiber properties therefore is more likely one of the quantitative differences in gene expression as opposed to differences in the genotype at the DNA level (Wilkins and Arpat 2005). Further

studies, hence, are required to understand the gene expression, their copy number and specific function in fiber development. Transcriptomics, a high-throughput method for investigating genome expression, is gaining importance as a new strategy for a better understanding of the molecular and genetic basis of complex trait expression. A large genetic variability can be revealed at the gene expression or transcriptome level, which could help to predict phenotypic performance on the basis of gene product variability. Techniques such as differential display, transcript imaging, DNA microarrays and RNA-seq allow large-scale transcriptome analyses. The resulting *Gossypium* transcripts or expressed sequence tags (ESTs) support to compute Digital Northerns that give clues: (a) on differential expression of numerous genes and (b) on transcript profiles in tissues at different developmental stages (such as fiber initiation, elongation and maturity phases) or exposed to various biotic and abiotic stresses (Boopathi and Ravikesavan 2009). At present, many ovule- and fiber-specific cDNA libraries have been constructed and sequenced and more than 270,000 ESTs from *Gossypium* are deposited in the NCBI database (http://www.ncbi.nlm.nih.gov). Several thousands of genes that were significantly up- or down-regulated during fiber initiation (which decide fiber yield), proper development (which is essential to ensure several fiber quality traits that are suitable for modern textile industry), secondary cell wall biosynthesis (which imparts fiber strength) were documented (Fang et al. 2014) and molecular, cellular and developmental changes related to fiber development have been identified. High-throughput, genome-wide transcriptomic analysis of cotton under drought stress revealed a significant down-regulation of genes and pathways involved in fiber elongation and an up-regulation of defense response genes (Padmalatha et al. 2012).

21.12.3 Proteomics

In addition to focusing on gene regulation at the transcriptional and post-transcriptional levels, it is also imperative to concentrate at the translational (i.e. protein synthesis and modifications) and metabolite (see below) levels. Proteins are considered the *workhorses* of cellular activities. However they are highly dynamic and require great attention to capture the entire proteome from the given tissue at a given point of time. Yao et al. (2006) and Wu et al. (2007) developed protocols for protein extraction from cotton fibers, which gave a higher yield, greater resolution and spot intensities during subsequent two-dimensional electrophoresis. Li et al. (2007) performed comparative proteomic analysis using 10 days post anthesis wild-type cotton ovules with fibers and (fiberless) fl mutant ovules that produced absolutely no fiber. A total of 1240 ± 20 protein spots were detected on both gels and among them, 61 were either highly up-regulated or down-regulated (Li et al. 2007). It is believed that phosphorylation is a critical post-translational modification involved in regulation of a wide range of cell activities. Ma et al. (2014) employed a mass spectrometry-based phosphoproteomics to conduct a global and site-specific phosphoproteome profiling between ovules of a fuzzless-lintless upland cotton mutant and its isogenic parental wild type. It was the first global

overview of phosphorylation during cotton fiber initiation that offered a dataset for clarification of signaling networks in fiber development of *Gossypium hirsutum.*

21.12.4 Metabolomics

Metabolome analysis provides a powerful tool to monitor the changes in primary and secondary metabolites during plant cell development. Documentation of such changes is an indispensable part in revealing the molecular mechanisms of plant growth and development under normal and adverse conditions. Identification of a complete set of metabolites is mainly based on application of mass spectrometric techniques coupled with chromatographic instrumentation and electrophoretic techniques. The proper choice of an ionization method and analyzer type used in mass spectrometer in metabolite analysis is the decisive component in metabolomics (Rodziewicz et al. 2014). Metabolite profiles of cotton fiber cells clearly indicated that cellulose biosynthesis, cell wall loosening and lipid biosynthesis were highly active during fiber elongation, which is consistent with transcriptome and proteomic data (Gou et al. 2007; Li et al. 2007; Shi et al. 2006).

The role of secondary metabolites in abiotic stress resistance was comprehensively documented in crop plants by Rodziewicz et al. (2014). For example, the polyphenol content in cotton leaves was found to increase in drought-stressed plants. The level of such a metabolite was significantly higher in the tolerant genotype than in the sensitive one. These results suggest that polyphenols are involved in the maintenance of the osmotic potential in cells and in scavenging free radicals under a water-stress condition (Parida et al. 2007).

Terpenoids constitute a broad class of lipophilic secondary metabolites synthesized in plants from isoprene units, which may be further assembled and modified in many different ways. These natural products exhibit a positive effect against both biotic and abiotic stress factors. Terpenoids show an antioxidant and antibiotic activity, take part in defense responses against herbivores, play an important role in the stabilization of the lipid membrane and improve environmental stress tolerance. Despite this fact, decreased content of these terpenoid compounds was reported in cotton under drought stress (Massacci et al. 2008; Parida et al. 2007).

The abovementioned two metabolite examples shown contradictory results and hamper the explanation of the role that metabolites play in stress response in cotton, at least in terms of drought resistance. Of course, this contradiction may be due to experimental error or interaction of other stresses. But the lesson to be learned from these metabolite studies is that the metabolite profile obtained with modern tools is still incomplete and there is an urgent need to find a high-throughput, efficient device that can provide the profile of comprehensive metabolome. This is owing to the point that the explanation of the role of particular metabolites during plant acclimation can contribute to the discovery of molecular markers associated with enhanced tolerance to the given adverse conditions and such markers would also be a great resource for molecular breeding in cotton.

21.12.5 Metagenomics

Although QTLs for several common agronomic and economic traits in cotton were reported, fruitful comparisons cannot be made because of non-availability of common markers in the published reports (Boopathi et al. 2011). In general, identified QTLs were held up within family, the size of detected QTL effects were limited, and inferences were restricted to a single population and set of conditions (Levi et al. 2009). To this end, combining QTL information from several studies by meta-analysis or meta-genomics is proposed. Such an integration of QTLs from different populations into a common map facilitates exploration of their allelic and homeologous relationships, although the level of resolution is limited by comparative marker densities, variation in recombination rates in different crosses, variation in gene densities across the genome and other factors (Boopathi 2013). Using a high-density reference genetic map which consists of 3475 loci in total, Rong et al. (2007) reported alignment of 432 QTLs mapped in 1 diploid and 10 tetraploid interspecific cotton populations and depicted in a CMap resource. Similarly, Lacape et al. (2010) conducted meta-analysis of more than 1000 QTLs obtained from the RIL and BC populations derived from the same parents and reported consistent meta-clusters for fiber color, fineness, and length. However, both of these studies pointed out that although their result on cotton fiber can hardly support the optimistic assumption that QTLs are accurate, they have shown that the reliability of QTL-calls and the estimated trait impact can be improved by integrating more replicates in the analysis. Hence, it is imperative to verify the regions of convergence with new maps which share common markers with the consensus map produced by Rong et al. (2007) and Lacape et al. (2010).

21.12.6 Phenomics

Further genetic improvement in cotton requires more information on the inheritance of the primary traits (agronomic traits) and associations with other traits (economic traits) that are needed in improved cultivars. Quantitative geneticists believe they could enhance breeding methods if the inheritance of quantitative traits was better understood. The data collection on a given trait is often hampered by the significant influence of the environmental factors and their variability. This is especially true for traits related to cotton yield under biotic and abiotic stresses. In addition to their sensitivity to environment and the phenomenon of genotype-by-environment interaction, such traits are often controlled by a large number of genes. The interactions of such a complex gene network make it difficult to analyze their genetic basis and, therefore, molecular breeding. The accuracy of phenotypic evaluation is of the utmost importance for the accuracy of QTL mapping and the success of molecular breeding (Boopathi 2013). A reliable QTL map can only be produced from reliable phenotypic data. Replicated phenotypic measurements can be used to improve the accuracy of QTL mapping by reducing experimental error or background noise.

High-throughput phenotyping for QTL mapping under highly controlled cotton development conditions provides the best basis for extracting a maximum of information from mapping populations. High-throughput and accurate analysis of fiber quality parameters in cotton, such as 2.5 % span length (mm), bundle strength (g/tex), fiber fineness (micronaire), uniformity ratio and elongation percentage, is now possible, thanks to high-volume instruments. However, accurate and large-scale phenotyping for biotic and abiotic stress resistance traits in cotton needs further major improvement. To this end, simple but efficient methods useful in biotic and abiotic stress resistance screening and realizing their importance in QTL analysis and molecular breeding have to be more focused.

21.12.7 Transgenics

Genetic improvement of cotton by classical breeding can achieve transfer of desirable alleles most frequently from within species and very rarely from distantly-related species. However, recent developments in transgenic or genetic engineering procedures, the desirable gene from any biological system can be efficiently transferred to cotton (and such cotton crop is called *transgenic* or *genetically modified* (GM) cotton). In today's transgenic era, introgression of one or few genes into a current elite cultivar via backcrossing is a common plant-breeding practice. Methods for marker-assisted backcrossing were developed rapidly for the incorporation of transgenic traits and reduction of linkage drag, where molecular markers were used in genome scans to select those individuals that contained both the transgene and the greatest proportion of desirable alleles from the recurrent (or elite) parent genome.

Although the first commercial GM crops were planted in 1994 (tomatoes), 1996 was the first year in which a significant area (1.66 million ha) of crops were planted containing GM traits. In terms of the share of the main crops in which GM traits have been commercialized (soybeans, maize, cotton and canola), GM traits accounted for >40 % of the world plantings to these four crops in 2011 (James 2013).

A comprehensive collection of methods for creating and monitoring transgenic cotton and its application in agricultural and basic research has been provided by Zhang (2013a, b). Although transgenic cotton plants with disease-resistance, abiotic stress tolerance and improved fiber quality have been developed in the past decades (see Table 21.3 for recent examples), insect-resistant and herbicide-tolerant cotton are the two dominant transgenic cottons in the transgenic cotton market (Chakravarthy et al. 2014; Zhang 2013a). Currently transgenic cotton hybrids occupy a larger area among cotton cultivating countries. For example, more than 90 % of the area in India during 2012 was occupied with Bt cotton hybrids (James 2013). Hybrids have contributed to wider adaptation, higher quality cotton production, and higher seed and seed oil output. However, there is a need for strong complementary research and development packages for both hybrids and varieties so that the full potential of both can be synchronized. Hybrids perform better under higher-input technology conditions and superior management. Hybrid seed production is labor intensive and

Table 21.3 Selected examples of recently published transgenic cotton researches

Trait	Transgene	Remarks	Reference
Resistance to *Fusarium oxysporum* and *Verticillium dahliae*	Plant defensin NaD_1	Provided substantial resistance in the field	Gaspar et al. (2014)
Drought resistance	AtRAV1/2 and AtABI5	Showed resistance to drought stress under field and greenhouse conditions	Mittal et al. (2014)
Fiber cell initiation	vacuolar invertase (VIN)	RNAi-mediated suppression resulted fiberless seed	Wang et al. (2014)
Abiotic stress resistance	SNAC1	Improved drought and salt tolerance by enhancing root development and reducing transpiration rate	Liu et al. (2014)
Fiber cell elongation	calcium sensor CaM7	by modulating reactive oxygen species production	Tang et al. (2014)

expensive. However, a focus on the development of useful male sterile systems and their deployment in hybrid-seed production make the process simple in cotton.

There is also another concern with respect to transgenic cotton in developing countries. For example, in India, the area under public research bred varieties and hybrids was reduced significantly to less than 8 % of the total cultivable area (Kranthi 2013). The area under hybrid cotton increased from 40 % in 2002 to 92 % in 2010. The area under *Gossypium hirsutum* varieties was 33 % in 2000, but reduced to less than 3 % in 2009. The area under *G. barbadense, G. arboreum* and *G. herbaceum*, 6.6 %, 25 % and 13 %, respectively, during 1995, declined to less than 7 % in 2010 for the three species together in India (Kranthi 2013). Furthermore, with severe selection pressure of Bt toxins used in over 90 % of the area of Bt crops, and less adoption of refugia and the declining area of inter crops in cotton cropping systems, development of bollworm resistance to the Bt toxins is an emerging concern. Therefore there is an urgent need to reorient molecular breeding research efforts to ensure productivity under these unfavorable conditions.

21.12.8 Mutagenesis

There is also a renewed interest in chemical mutagenesis and transposon mediated gene knock-outs as a means to obtain single-gene mutants affecting phenotypes of interest, since the prospects of gene identification are high and every gene affecting a trait is potentially a target (Boopathi and Ravikesavan 2009). Mutagenesis is considered as an additional strategy for creation of genetic variability using various mutagens (radiations and chemicals). In this regard, leading genotypes of *Gossypium hirsutum* and *G. arboreum* have been treated with ethyl methane sulfonate for developing targeted induced local lesions in genome (TILLING) populations that

would help in understanding the genes involved in conferring various traits of interest. The use of induced mutations for crop improvement has led to the development of several improved varieties of cotton (e.g. NIAB-78) at the Nuclear Institute for Agriculture and Biology, Pakistan, which clearly indicates the potential of this technique: a wealth of genetic variability has been developed by these improved cotton cultivars with significant positive impact on Pakistan's economy (Haq 2009). In vivo mutagenesis has also been used for genetic improvement of fiber traits (Ibragimov and Koval'chuk 1970; Lowery et al. 2007). Nevertheless, very few studies have employed in vitro mutagenesis to test its effectiveness in improving fiber quality (Herring et al. 2004) and other agronomically-important traits in cotton.

21.13 Bioinformatics

It is increasingly evident from reports (reviewed in Boopathi 2013) that development of user-friendly bioinformatics tools which facilitate efficient interaction of QTL x QTL and QTL x environment interactions during QTL analysis and appropriate modeling algorithms for gene network and metabolic pathway interactions require immediate attention. Bioinformatics tools available as of now need more technical skills; interpretation of results requires expert intervention in understanding the results. Progress has been made in the use of computational biology and comparative genomics approaches to explore genes/alleles involved in resistance of cotton cultivars to cotton leaf curl virus, cold tolerance, fiber initiation and development (USDA 2014).

21.13.1 Web Resources for Community Oriented Research

There is also an urgent prerequisite to expand bioinformatics infrastructure for managing, curating, and annotating the cotton genomic sequences that have been and will be generated in the near future. The cotton database of the future should be able to host and manage cotton sequence and functional genomics information resources using research community-accepted genome annotation, nomenclature and gene ontology. It is furthermore important to regularly update existing databases to effectively handle a large amount of data flow and community requests with additional resources that support key bioinformatics activities.

21.14 Challenges in Molecular Breeding in Cotton

In the current scenario, breeding for genetic improvement of cotton has several challenges to achieve objectives of benefit to cotton stakeholders. Due to modernization of ginning and garment machineries, breeding for improved fiber qualities as

per emerging needs is receiving primary attention. The possibility of utilizing donor genes from any biological system that favors fiber quality improvement should be explored, such as development of strategic plans for discovery and deployment of indigenous genes through molecular breeding to enhance yields with low input costs. Furthermore, current cotton production is challenged by quirks of the monsoon, damage by pests and diseases and increasing cost of production. Hence, appropriate reorientation and reinvention of resources should be considered while designing a molecular-breeding program.

Intensive plant-breeding programs of many crop plants have taken advantage of genetic resources and germplasm collections to develop improved genotypes with significant gains in yield. On the other hand such efforts have also unintentionally narrowed the genetic base and increased genetic vulnerability of many of the world's most important crops, including cotton. Hence, it is well-meaning that the molecular breeding program integrate the rich diversity present in the cotton germplasm through appropriate strategies such as automation or use of robotics in genomics laboratories, improved and efficient statistics for genomics studies starting from experimental design, data generation, transformation, normalization, and reducing error or false discovery rates. With the strength of the current germplasm pool and new molecular technologies, it would be possible to develop location-specific genotypes that have the potential to yield 3–4 times more than the current productivity of varieties elsewhere in the world. In this highly-developed information era, it is also imperative to discuss ethical, legal and social implications and issues in cotton molecular-breeding programs with the stakeholders for better and wider adaptation of the strategy.

21.15 Conclusions and Prospects

Molecular breeding in cotton includes traditional cotton breeding supplemented with marker-assisted breeding using advances in molecular-marker technology and QTL mapping (which includes marker-assisted backcrossing and marker-assisted recurrent selection), genomics (known as genomics-assisted breeding) and transgenics technology. Several lines of work have demonstrated the potential application of this new breeding strategy in cotton. Guo et al. (2003) were the first to screen molecular markers linked with important agronomical trait genes or QTLs and they studied the inheritance of high-quality fiber properties by segregation analysis using high fiber quality isogenic or near-isogenic cotton lines.

Similarly, near-isogenic lines were developed using drought related QTLs and the resulting lines manifested in many cases the expected drought-adaptive traits, but rarely exhibited an advantage in yield relative to the recipient parents (Levi et al. 2009). However, a molecular-breeding approach is not practiced routinely in cotton owing to several challenges that were outlined in the preceding sections. It is increasingly believed and deeply discussed in several forums that genes, proteins, metabolites and phenotypes should be considered simultaneously to unravel the

complex molecular integrated circuit that operates within the cell. A comprehensive revelation of the genotype-phenotype map does not seem to be achievable unless we can include all possible causal variables in the network-inference methodology. In order to achieve this, it is essential to have a global perspective on life processes instead of individual components of the system (or a reductionist approach). The network approach connecting all these sub-disciplines indicates the emergence of a system quantitative genetics (Narain 2010). This is considered as more challenging aspect of *omics* technologies since such refined analysis of quantitative dynamics in biological systems require highly skillful integrative biochemical network development. Several lines of developments in the integration of sub-disciplines such as quantitative trait nucleotide, expression QTLs or genetical genomics, mutagenesis and genomics, epistasis and epigenetics, phenomics were started to accumulate in biological systems (Long et al. 2008; Zhu et al. 2009) but their application has yet to be shown in cotton. Thus, the future and success of molecular breeding in cotton definitely depends on the development of system-quantitative genetics. Further, incorporation of cotton micro RNA and expression QTL research results into molecular breeding is still in its infancy. In a preliminary study, our laboratory has identified hundreds of drought responsive conserved and novel miRNAs in cotton. On the other hand, it is increasingly believed that the future of genetic improvement of cotton heavily depends on conventional breeding strategies coupled with molecular breeding approaches.

References

Abdurakhmonov IY (ed) (2014) World cotton germplasm resources. http://www.intechopen.com/books/world-cotton-germplasm-resources/cotton-germplasm-collection-of-uzbekistan. Accessed 3 Aug 2014

Abdurakhmonov IY, Kohel RJ, Yu JZ et al (2008) Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. Genome 92(6):478–487

Agarwal M, Neeta S, Harish P (2008) Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep 27(4):617–631

Basra AS, Malik CP (1984) Development of the cotton fiber. Int Rev Cytol 89:65–113

Blenda A, Fang DD, Rami JF et al (2012) A high density consensus genetic map of tetraploid cotton that integrates multiple component maps through molecular marker redundancy check. PLoS One 7(9):e45739

Boopathi NM (2013) Genetic mapping and marker assisted selection: basics, practice and benefits. Springer, India

Boopathi NM, Ravikesavan R (2009) Emerging trends in enhancement of cotton fiber productivity and quality using functional genomics tools. Biotechnol Mol Biol Rev 4(1):11–28

Boopathi NM, Thiyagu K, Urbi B et al (2011) Marker-assisted breeding as next-generation strategy for genetic improvement of productivity and quality: can it be realized in cotton? Int J Plant Genomics. doi:10.1155/2011/670104

Boopathi NM, Sathish S, Dachinamoorthy P et al (2014) Usefulness and utilization of Indian cotton germplasm. In: Abdurakhmonov IY (ed) World cotton germplasm resources. InTech, Rijeka, Croatia, p 119–136

Brubaker CL, Paterson AH, Wendel JF (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42:184–203

Chakravarthy VS, Reddy TP, Reddy VD, Rao KV (2014) Current status of genetic engineering in cotton (*Gossypium hirsutum* L): an assessment. Crit Rev Biotechnol 34(2):144–160

Doak CC (1934) A new technique in cotton hybridizing suggested changes in existing methods of emasculation and bagging of cotton flowers. J Hered 25:201–204

Endrizjz IE, Turcoite EL, Kohel RJ (1984) Qualitative genetics, cytology, and cytogenetics in cotton. In: Kohel RJ, Lewis CF (eds) Cotton, Agron Monograph 24. ASA-CSSA-SSSA, Madison, pp 81–129

Endrizzi JE, Turcotte EL, Kohel RJ (1985) Genetics, cytology, and evolution of *Gossypium*. Adv Genet 23:271–375

Fang L, Tian R, Chen J et al (2014) Transcriptomic analysis of fiber strength in upland cotton chromosome introgression lines carrying different *Gossypium barbadense* chromosomal segments. PLoS One 9(4):e94642

Fryxell P (1992) A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rheedea 2:108–165

Gaspar YM, McKenna JA, McGinnin BS et al (2014) Field resistance to *Fusarium oxysporum* and *Verticillium dahliae* in transgenic cotton expressing the plant defensin NaD1. J Exp Bot 65(6):1541–1550

Gillham FEM, Bell TM, Arin T et al (1995) Cotton production prospects for the next decade. The International Bank for Reconstruction and Development/The World Bank, Washington DC

Gore MA, Fang DD, Poland JA et al (2014) Linkage map construction and quantitative trait locus analysis of agronomic and fiber quality traits in cotton. Plant Genome 7(1):2014

Gou JY, Wang LJ, Chen SP et al (2007) Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. Cell Res 17:422–434

Govila OP (1969) Fertilization and seed development in crosses between *G. arboreum* and *G. hirsutum*. Indian J Gen 29:407–417

Grover CE, Kim HR, Wing RA et al (2004) Incongruent patterns of local and global genome size evolution in cotton. Genome Res 14:1474–1482

Gruère A (2011) ICAC Reprint from Cotton: review of the world situation, vol 64. No 5 May–June 2011

Gulati AN, Turner AJ (1929) A note on early history of cotton. J Text Inst Trans 20:1–9

Guo W, Sun J, Zhang T (2003) Gene cloning and molecular breeding to improve fiber qualities in cotton. Chin Sci Bull 48:709–717

Haq MA (2009) Development of mutant varieties of crop plants at NIAB and the impact on agricultural production in Pakistan. In: Shu QY (ed) Induced plant mutations in the genomics era. FAO, Rome

Havov R, Udall JA, Hovav E et al (2008) A majority of cotton genes are expressed in single-celled fiber. Planta 227:319–329

He DH, Lin ZX, Zhang XL et al (2007) QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of *Gossypium hirsutum* × *Gossypium barbadense*. Euphytica 153(1–2):181–197

Hendrix B, Stewart JM (2005) Estimation of the nuclear DNA content of *Gossypium* species. Ann Bot 95:789–797

Herring AD, Auld DL, Ethridge MD et al (2004) Inheritance of fiber quality and lint yield in a chemically mutated population of cotton. Euphytica 136:333–339

Ibragimov S, Koval'chuk R (1970) Mutagenesis in cotton. Tr. In-t selektsii i semenovodstva khlopschatnika (Tashkent) 3(3):237–245

James C (2013) Global status of commercialized biotech/GM crops: 2013. ISAAA Brief 35 No. 46. ISAAA, Ithaca, New York

Kantartzi SK, Stewart MJ (2008) Association analysis of fibre traits in *Gossypium arboreum* accessions. Plant Breed 127(2):173–179

Kimber G (1961) Basic of the diploid-like meiotic behavior of polyploid cotton. Nature 191:98–99

Kranthi KR (2013) Cotton production in India. In: Manickam S, Sankaranarayanan K, Prakash AH (eds) Training manual on relevance and techniques of organic cotton production. CICR, Nagpur, January 21–25, 2013

Krapovickas A, Seijo G (2008) *Gossypium ekmanianum* (Malvaceae), algodon silvestre de la Republica Dominicana. Bonplandia 17:55–63

Lacape J, Llewellyn MD, Jacobs J et al (2010) Meta-analysis of cotton fiber quality QTLs across diverse environments in a *Gossypium hirsutum*×*G. barbadense* RIL population. BMC Plant Biol 10(1):132

Lee SB, Kaittanis C, Jansen RK et al (2006) The complete chloroplast genome sequence of *Gossypium hirsutum*: organization and phylogenetic relationships to other angiosperms. BMC Genomics 7(1):61

Levi A, Ovnat L, Paterson AH, Saranga Y (2009) Photosynthesis of cotton near-isogenic lines introgressed with QTLs for productivity and drought related traits. Plant Sci 177:88–96

Li HB, Qin YM, Pang Y et al (2007) A cotton ascorbate peroxidase is involved in hydrogen peroxide homeostasis during fibre cell development. New Phytol 175:462–471

Li C, Wang C, Dong N et al (2012) QTL detection for node of first fruiting branch and its height in upland cotton (*Gossypium hirsutum* L.). Euphytica 188(3):441–451

Li P, Li Z, Liu H, Hua J (2013a) Cytoplasmic diversity of the cotton genus as revealed by chloroplast microsatellite markers. Genet Resour Crop Evol. doi:10.1007/s10722-013-0018-9

Li X, Yuan D, Zhang J et al (2013b) Genetic mapping and characteristics of genes specifically or preferentially expressed during fiber development in cotton. PLoS One 8(1):e54444

Li F, Fan G, Wang K et al (2014) Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat Genet 46(6). doi:10.1038/ng.2987

Liang Q, Hu C, Hua H et al (2013) Construction of a linkage map and QTL mapping for fiber quality traits in upland cotton (*Gossypium hirsutum* L.). Chin Sci Bull 58(26):3233–3243

Lin L, Pierce GJ, Bowers JE et al (2010) A draft physical map of a D-genome cotton species (*Gossypium raimondii*). BMC Genomics 11:395

Liu R, Wang B, Guo W et al (2012) Quantitative trait loci mapping for yield and its components by using two immortalized populations of a heterotic hybrid in *Gossypium hirsutum* L. Mol Breed 29(2):297–311

Liu G, Cao D, Li S et al (2013) The complete mitochondrial genome of *Gossypium hirsutum* and evolutionary analysis of higher plant mitochondrial genomes. PLoS One 8(8):e69476

Liu G, Li X, Jin S et al (2014) Overexpression of rice NAC Gene SNAC1 improves drought and salt tolerance by enhancing root development and reducing transpiration rate in transgenic cotton. PLoS One 9(1):e86895

Long TA, Brady SM, Benfey PN (2008) Systems approaches to identifying gene regulatory networks in plants. Ann Rev Cell Dev Biol 24:81–103

Lowery CC, Auld DL, Bechere E et al (2007) Use of chemical mutagenesis in improving upland cotton. In: World Cotton Research Conference-4, International Cotton Advisory Committee (ICAC), Lubbock, Texas, USA, 10–14 Sept 2007

Ma Q, Wu M, Pei W et al (2014) Quantitative phosphoproteomic profiling of fiber differentiation and initiation in a fiberless mutant of cotton. BMC Genomics 15(1):466

Massacci A, Nabiev SM, Pietrosanti L et al (2008) Response of the photosynthetic apparatus of cotton (*Gossypium hirsutum*) to the onset of drought stress under field conditions studied by gas-exchange analysis and chlorophyll fluorescence imaging. Plant Physiol Biochem 46:189–195

McMullen MD, Kresovich S, Villeda HS et al (2009) Genetic properties of maize nested association mapping population. Science 325(5941):737–740

Mei H, Ai N, Zhang X et al (2014) QTLs conferring FOV 7 resistance detected by linkage and association mapping in Upland cotton. Euphytica 197(2):237–249

Mittal A, Gampala SS, Ritchie GL et al (2014) Related to ABA-Insensitive3 (ABI3)/Viviparous1 and AtABI5 transcription factor coexpression in cotton enhances drought stress adaptation. Plant Biotechnol J 12(5):578–589

Narain P (2010) Quantitative genetics: past and present. Mol Breed 26(2):135–143

Padmalatha KV, Dhandapani G, Kanakachari M et al (2012) Genome-wide transcriptomic analysis of cotton under drought stress reveal significant down-regulation of genes and pathways involved in fibre elongation and up-regulation of defense responsive genes. Plant Mol Biol 78:223–246

Parida AK, Dagaonkar VS, Phalak MS et al (2007) Alterations in photosynthetic pigments, protein and osmotic components in cotton genotypes subjected to short-term drought stress followed by recovery. Plant Biotechnol Rep 1:37–48

Qin H, Guo W, Zhang YM, Zhang T (2008) QTL mapping of yield and fiber traits based on a four-way cross population in *Gossypium hirsutum* L. Theor Appl Genet 117:883–894

Rodziewicz P, Swarcewicz B, Chmielewska K et al (2014) Influence of abiotic stresses on plant proteome and metabolome changes. Acta Physiol Plant 36(1):1–19

Rong JK, Abbey C, Bowers JE et al (2004) A 3347- locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166:389–417

Rong JK, Feltus FA, Waghmare VN et al (2007) Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. Genetics 176(4):2577–2588

Salvi S, Tuberosa R (2007) Cloning QTLs in plants. In: Varshney RK, Tuberosa R (eds) Genomics-assisted crop improvement, vol 1. Springer, New York, pp 207–226

Santhanam V, Sundaram V (1997) Agri history of cotton India – an overview. Asian Agric-Hist 1(4):235–251

Seelanan T, Schnabel A, Wendel JF (1997) Congruence and consensus in the cotton tribe (Malvaceae). Syst Bot 22:259–290

Shi YH, Zhu SW, Mao XZ et al (2006) Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. Plant Cell 18:651–664

Sun FD, Zhang JH, Wang SF et al (2012) QTL mapping for fiber quality traits across multiple generations and environments in upland cotton. Mol Breed 30(1):569–582

Tang W et al (2014) The calcium sensor GhCaM7 promotes cotton fiber elongation by modulating reactive oxygen species (ROS) production. New Phytol 202(2):509–520

Townsend TP (2010) A balanced perspective on cotton: responding to valid problems, challenging irresponsible critics, Report of the Executive Director to the 69th Plenary Meeting of The International Cotton Advisory Committee, Lubbock, TX, USA, 20–25 Sept 2010

Ulloa M, Brubaker C, Chee P (2007) Cotton. In: Kole C (ed) Genome mapping and molecular breeding in plants, vol 6, Technical Crops. Springer, Berlin, pp 43–63

USDA (2014) http://search.usa.gov/search?utf8=%E2%9C%93&sc=0&query=cotton+bioinformatics&m=&affiliate=agriculturalresearchservicears&commit=Search, Accessed on 27 Aug 2014

Wang K, Wang Z, Li F et al (2012) The draft genome of a diploid cotton *Gossypium raimondii*. Nat Genet 44(10). doi:10.1038/ng.2371

Wang L et al (2014) Silencing the vacuolar invertase gene GhVIN1 blocks cotton fiber initiation from the ovule epidermis, probably by suppressing a cohort of regulatory genes via sugar signaling. Plant J 78(4):686–696

Wendel JF (1989) New world tetraploid cottons contain old world cytoplasm. Proc Natl Acad Sci U S A 86:4132–4136

Wendel JF, Albert VA (1992) Phylogenetics of the cotton genus (*Gossypium*): characterstate weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. Syst Bot 17:115–143

Wendel JF, Cronn RC (2003) Polyploidy and the evolutionary history of cotton. Adv Agron 78:139–186

Wilkins TA, Arpat AB (2005) The cotton fibre transcriptome. Physiol Plant 124:295–300

Wilkins TA, Jernstedt JA (1999) Molecular genetics of developing cotton fibers. Hawthorne Press, New York

Winter P, Kahl G (1995) Molecular marker technologies for plant improvement. World J Microbiol Biotechnol 11(4):438–448

Wu XP, Cheng YS, Liu JY (2007) Microwave-enhanced ink staining for fast and sensitive protein quantification in proteomic studies. J Protein Res 6:387–391

Xu Y, Li Z, Thomson MJ (2012) Molecular breeding in plants: moving into the mainstream. Mol Breed 29:831–832

Yao Y, Yang YW, Liu JY (2006) An efficient protein preparation for proteomic analysis of developing cotton fibers by 2-D. Electrophoresis 27:4559–4569

Zhang B (ed) (2013a) Transgenic cotton: methods and protocols. Humana Press, New Jersey, 277 p

Zhang B (2013b) Transgenic cotton: from biotransformation methods to agricultural application. Methods Mol Biol 958:3–15

Zhang K, Zhang J, Ma J et al (2012) Genetic mapping and quantitative trait locus analysis of fiber quality traits using a three-parent composite population in upland cotton (*Gossypium hirsutum* L.). Mol Breed 29(2):335–348

Zhiyuan N, Chen H, Mei H, Zhang T (2014) Molecular tagging of QTLs for fiber quality and yield in the upland cotton cultivar Acala-Prema. Euphytica 195(1):143–156

Zhu M, Yu M, Zhao S (2009) Understanding quantitative genetics in the systems biology era. Int J Biol Sci 5(2):161–170

Index

J.M. Al-Khayri et al. (eds.), *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, DOI 10.1007/978-3-319-22521-0