Next-generation BREEDING

Deeper research needed to quickly and inexpensively identify plant genomes

By C.J. SCHWARTZ

Traditional breeding in plants and animals depends on the identification of individual subjects with superior performance or traits. These plant qualities are referred to as the organism's phenotype. Individuals with desired phenotypes are selected for additional cycles of breeding, thereby stabilizing the genetic architecture (inbreeding) and producing consistent offspring with the properties of interest. While traditional breeding relies on phenotype alone, marker-assisted breeding also uses genetic information — the genotype of the plant — to identify the genes that confer the specific phenotype, allowing for a vast acceleration in the development of desirable cultivars.

To perform successful marker-assisted breeding, two strains with differing phenotypes are crossed for the purpose of comparing genetic differences or similarities that allow identification of a correlation between phenotypes and genotype (genes). The progeny of this cross is the F1 generation. F1 plants often display hybrid vigor, which means the progeny's phenotype is improved compared to either parent.

For some crops, F1 plants are the goal and are propagated in the field. When F1 plants undergo sexual reproduction, the resulting progeny (F2 generation) will display a wide range of phenotypes between those of the original parents. For example, if the mother plant flowered at 45 days and the father plant flowered at 65 days, the F2 plants might flower at 45, 47, 53, 60 or 65 days — a continuous distribution. To identify the gene or genes underlying the flowering time differences, the two extreme classes are genotyped to identify a correlation between the genes and the phenotype. Different genes will likely be identified when using different parents. Additional crosses may be required to assist in gene identification, including a backcross, which is crossing F1 plants back to one of

the parents. The progeny from this cross (BC1) will have 75% of its genes from one parent and 25% from the other parent.

"DEVELOPMENTS IN RECENT YEARS HAVE LED TO A MASSIVE INCREASE IN THE USE OF DNA SEQUENCING

This substantially decreases the genetic variability of the resulting progeny, thus increasing the probability of identifying the genes of interest.

Once a gene is identified as being important for a trait, the specific difference identified in the DNA sequence can be used to track the maintenance of that version of the gene during additional generations. This is required to stabilize the strain and also allows further selection for

additional traits of interest, such as pest resistance, drought resistance or nutrient requirements.
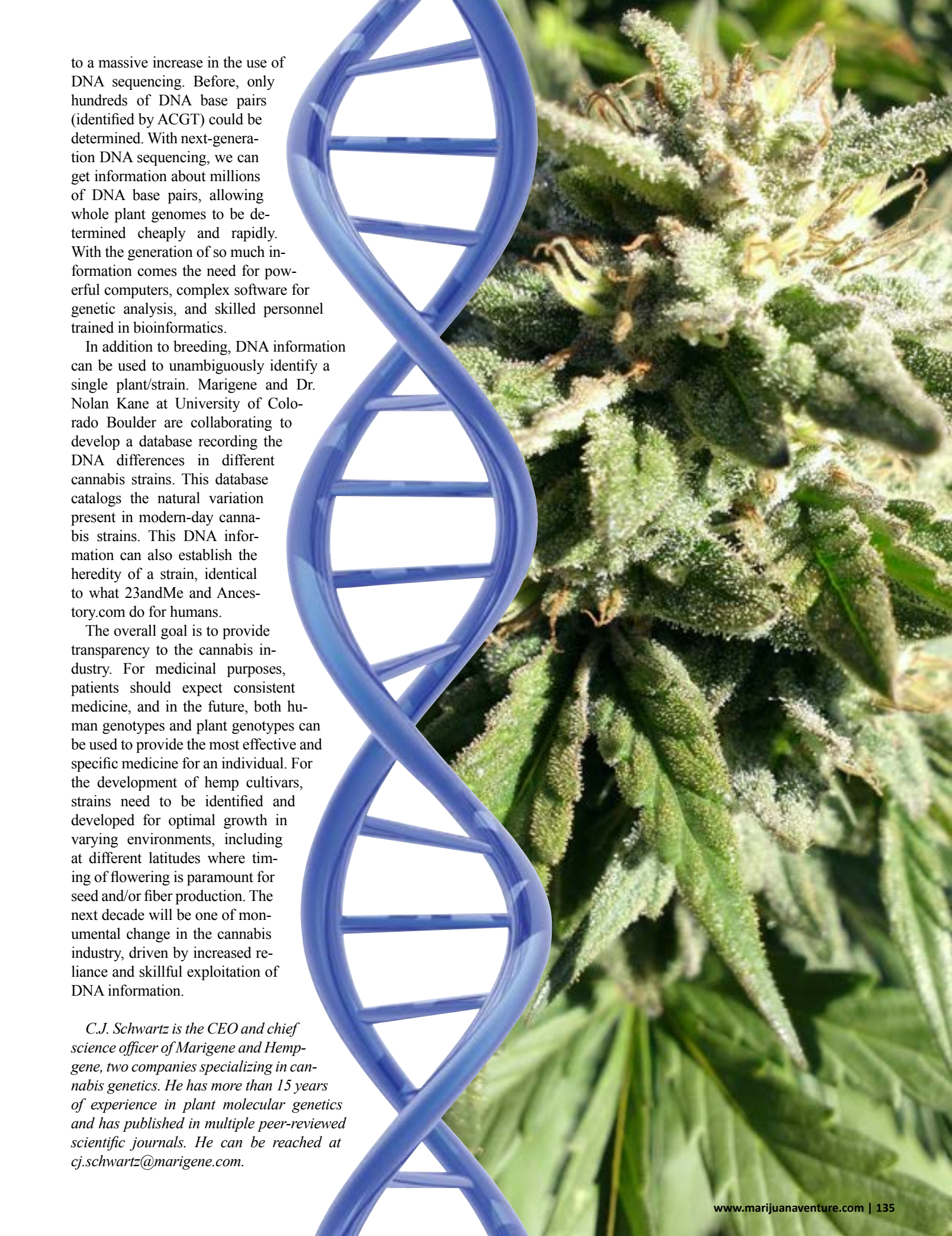
For marijuana, cannabinoid and terpene composition are the most important traits of interest. For example, a strain may have a very active THC synthase, but only produces 10% THC. Marker-assisted breeding can be used to identify, follow and retain the active THC synthase, while identifying additional genes that lead to an increase in THC, such as CBG synthase. CBG (cannabigerol) is the precursor to THC. Increasing the substrate concentration for THC synthase can result in more THC being produced.

In this scenario, we need to track the two genes using DNA sequence differences between the maternal genes and the paternal gene. If three genes are being tracked, only 1.5% of the plants will have the desired combination, underlying the reason why using DNA changes (genetic markers) can vastly accelerate breeding programs. Instead of growing 200 plants to maturity and phenotyping them, only three plants need to be retained, phenotypically verified and propagated.

Developments in recent years have led

BC1

F1 Generation
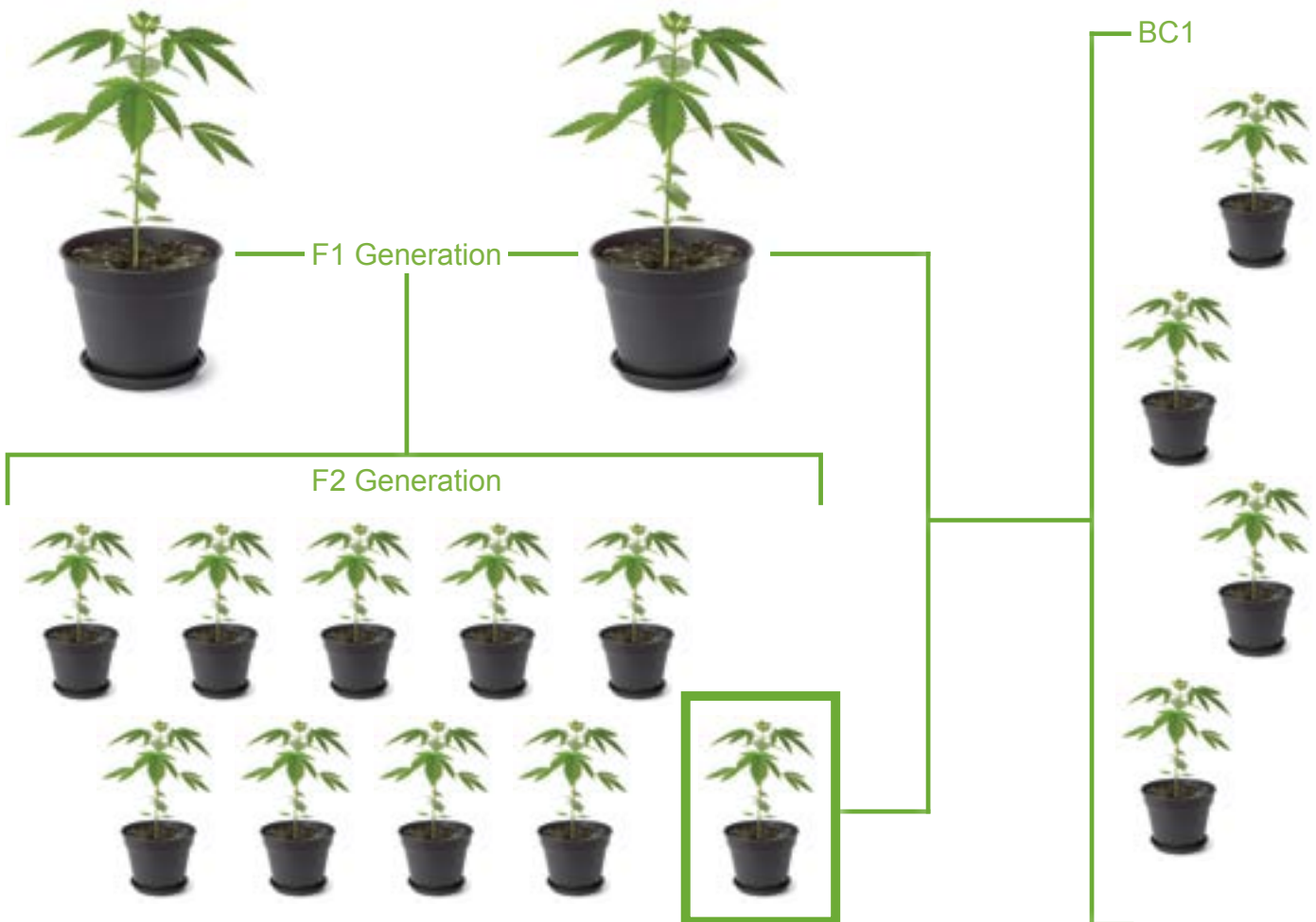
F2 Generation

to a massive increase in the use of DNA sequencing. Before, only hundreds of DNA base pairs (identified by ACGT) could be determined. With next-generation DNA sequencing, we can get information about millions of DNA base pairs, allowing whole plant genomes to be determined cheaply and rapidly. With the generation of so much information comes the need for powerful computers, complex software for genetic analysis, and skilled personnel trained in bioinformatics.

In addition to breeding, DNA information can be used to unambiguously identify a single plant/strain. Marigene and Dr. Nolan Kane at University of Colorado Boulder are collaborating to develop a database recording the DNA differences in different cannabis strains. This database catalogs the natural variation present in modern-day cannabis strains. This DNA information can also establish the heredity of a strain, identical to what 23andMe and Ancestory.com do for humans.

The overall goal is to provide transparency to the cannabis industry. For medicinal purposes, patients should expect consistent medicine, and in the future, both human genotypes and plant genotypes can be used to provide the most effective and specific medicine for an individual. For the development of hemp cultivars, strains need to be identified and developed for optimal growth in varying environments, including at different latitudes where timing of flowering is paramount for seed and/or fiber production. The next decade will be one of monumental change in the cannabis industry, driven by increased reliance and skillful exploitation of DNA information.

C.J. Schwartz is the CEO and chief science officer of Marigene and Hempgene, two companies specializing in cannabis genetics. He has more than 15 years of experience in plant molecular genetics and has published in multiple peer-reviewed scientific journals. He can be reached at cj.schwartz@marigene.com.