

Empowering Crop Improvement Through Genomic Selection

Somsol Somsol^{1*}, Uttaran Dasgupta² and Shridhar Ragi³

^{1,2,3}Ph.D. Research Scholar, Division of Genetics, ICAR – IARI, New Delhi -110012

*Corresponding Author: bharathsomsol@gmail.com

Introduction

Economically valuable traits, like yield, quality, and stress resilience, are predominantly governed by genome-wide minor effect alleles, contributing to the complexity of quantitative traits. Two primary breeding approaches reliant on markers exist—Marker-Assisted Selection (MAS) and Genomic Selection (GS)—aiming to identify plants with superior traits. MAS involves selecting crops based on Quantitative Trait Loci (QTLs) pinpointed through either linkage mapping (LM) or Genome-Wide Association Studies (GWAS). MAS struggles to detect genes with minor impacts related to intricate traits. When the linked markers represent only a small part of genetic diversity, MAS outcomes fall behind those of phenotypic selection. Additionally, MAS relies on a limited number of statistically significant markers, leaving out others from the analysis. Consequently, traditional MAS usually involves few specific markers per trait, making it less effective when multiple genes with subtle effects contribute to one trait. Therefore, MAS is best suited for categorical traits like specific metabolic traits and disease resilience, rather than quantitative traits.

A novel approach called Genomic Selection (GS) addresses these challenges. Enhancing complex traits necessitates evaluating phenotypes across multiple locations and years to establish the relationship between genotype and environment. However, this was hindered by the prohibitive costs and labor involved. With the advent of Next-Generation Sequencing (NGS) technology, sequencing expenses have decreased, enabling easy access to high-resolution genome data. Progress in sequencing techniques has facilitated the implementation of GS. All the high-performing markers across the genome are available for selecting GS crops. When provided with a marker set spanning the entire genome, GS models consider every marker's influence on a trait without specific cutoffs. In contrast, traditional MAS concentrates on a limited number of key genes or

QTLs, while GS integrates all accessible markers in the genome for predictions. By accounting for all genetic effects, GS prevents the loss of genetic diversity attributed to minor genes or QTLs. Thus, GS proves more efficient than MAS for traits governed by numerous markers. The primary advantage of GS lies in its ability to forecast phenotype information of mature individuals using genotype data obtained from seeds or seedlings. This method negates the necessity for comprehensive phenotypic assessments across years or environments, thereby accelerating crop varietal development. Previous research demonstrated that genomic prediction (GP) achieves significantly greater accuracy compared to MAS. Specifically, for maize (*Zea mays* L.), GP is three times more accurate, and for wheat (*Triticum aestivum* L.), it is twice as accurate. Hence, it is anticipated that this genome-based predictive approach holds the potential to supplant phenotypic selection or marker-assisted breeding.

Steps in GS

Genomic selection (GS) is a breeding method with the potential to hasten the rate of genetic improvement in crops. To initiate GS, a training group is assembled, comprising individuals with observable traits for the desired characteristic(s) and genome-wide DNA marker profiles. This data is then utilized to formulate a predictive model, estimating the impact of each marker on the trait simultaneously. If the markers are closely linked to the genetic mutations influencing the trait, they can capture a significant portion of its genetic variability. Subsequently, potential selections (such as seedlings) undergo genotyping to calculate genomic estimated breeding values (GEBVs) using the established genetic model and prediction equation. Superior genotypes are then chosen based on their GEBVs, predicting their performance. Over the past two decades, various statistical models and machine learning methods have been proposed for genomic selection (GS). These include approaches assuming different distributions

of SNP effects, such as normal distributions (e.g., RR-BLUP, Genomic BLUP) or prior distributions favoring moderate to large effects (e.g., BayesA, wBSR). Some methods assume zero effects for certain SNPs (e.g., BayesB, BayesC π), while non-parametric methods like random forest and RKHS have also been explored. When selecting a method, its suitability for the intended application should be carefully considered. For example, RKHS may excel in predicting future phenotypes but may require explicit extraction of additive breeding values for selection programs.

Key points to be considered while implementing GS

From the implementation of genomic selection in crop and livestock breeding programs, several practical insights have emerged.

- ✓ Firstly, the training set should encompass individuals (lines/varieties) closely related to the selection candidates.
- ✓ Secondly, the training populations need to be substantial due to the numerous loci and their minimal effect on typical quantitative traits like yield. It is estimated that such traits are influenced by 2000 to 4000 loci.
- ✓ To maintain the accuracy of Genomic Estimated Breeding Values (GEBVs) over time, the reference population should be regularly updated with new individuals genotyped and phenotyped.

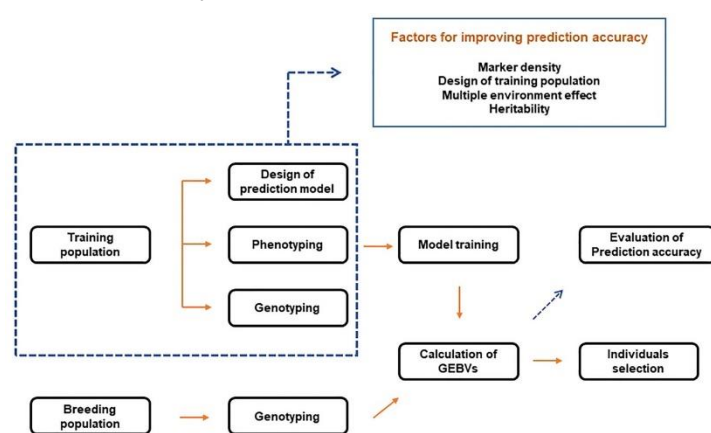


Fig. 1 A visual representation illustrating GS
(Source: Jeon *et al.*, 2023)

Approaches to enhance predictive precision in GS

Evaluation of genomic selection methods involves assessing prediction accuracy. This accuracy is determined by the correlation between the observed

Genomic Estimated Breeding Values (GEBVs) and the actual phenotype data. Hence, enhancing prediction accuracy holds significance for genomic selection applications.

✓ Marker density and selection

GS estimates the effects of markers using LD between quantitative loci and markers, a high density of markers is advantageous for GS and at the same time it adds more cost. Incorporating genomic selection with reduced marker density involves the elimination of highly correlated duplicate markers within the linkage disequilibrium (LD) block. This method diminishes multicollinearity without impacting the marker's effect.

✓ Design of training population

Genomic prediction accuracy hinges on the training population's design. Breeders optimize it to maximize accuracy and minimize costs. Traditional random sampling fails due to uneven genetic representation. Methods like coefficient of determination (CD) and prediction error variance (PEV) offer alternatives, with CD slightly outperforming PEV. Incorporating genetic information from the test population during training improves accuracy significantly, especially with genetic algorithms. This shared genetic background reduces recombination between markers and QTLs, enhancing accuracy further. Therefore, including unrelated individuals in the training population can notably boost genomic prediction accuracy.

✓ Multiple environments and heritability

Predicting GEBVs across environments poses challenges. Modified prediction models, like non-informed and informed models, enhance accuracy. Non-informed models include environment and genotype-environment interactions, improving accuracy. Informed models incorporate environmental covariates using kernel-based methods, further enhancing accuracy. Considering these models across environments is crucial for accurate GEBV prediction. Traits with high heritability show high prediction accuracy, while environmental considerations are vital for traits influenced by numerous minor-effect genes.

✓ High-throughput phenotyping (HTP)

The advent of Next-Generation Sequencing (NGS) and extensive marker data has boosted Genomic Selection (GS) in plant breeding. However, inaccurate phenotypic data poses a challenge. High-throughput phenotyping (HTP) addresses this, offering rapid and precise measurements through advanced sensors like RGB, NDVI, and LiDAR. Studies show significant improvements in prediction accuracy using HTP, ranging from 7% to 33% in wheat. Integration of HTP with ML and DL technologies promises enhanced breeding outcomes by enabling accurate selection.

Advantages of GS

- ✓ Marker effects inform genomic selection directly in breeding populations, eliminating the need for QTL mapping.
- ✓ Studies show GS outperforms phenotypic selection, particularly for low heritability traits.
- ✓ The selection index method combines multiple trait data, enhancing selection effectiveness, especially for low heritability traits.
- ✓ GS reduces inbreeding rates and preserves genetic variability without sacrificing gains. It shortens breeding cycles, particularly in perennial species like oil palm, and facilitates commercialization of improved genotypes.
- ✓ GS allows parent selection based on genomic estimated breeding values (GEBVs) for target environment adaptation, streamlining germplasm exchange.
- ✓ It estimates GEBVs for untested traits using diverse environment data. While theoretically applicable for hybridization programs, practical validation is needed.

Limitations of GS

- ✓ Despite potential benefits, genomic selection in plant breeding lacks practical evidence, primarily

relying on simulations rather than real-world validation.

- ✓ Fluctuations in gene frequencies and epistatic interactions necessitate constant updates to genomic selection models to maintain accuracy.
- ✓ Limited understanding of genetic traits hampers accurate genomic selection model development, impacting prediction accuracy.
- ✓ Genomic selection's cost-effectiveness is questionable without off-season facilities, and the high cost of genotyping poses financial challenges.
- ✓ Implementation of genomic selection requires substantial resources and a fundamental shift in breeding approaches, posing challenges for many breeding programs.

References

- Jeon, D., Kang, Y., Lee, S., Choi, S., Sung, Y., Lee, T. H., & Kim, C. (2023). Digitalizing breeding in plants: A new trend of next-generation breeding based on genomic prediction. *Frontiers in plant science*, 14, 1092584.
- Voss-Fels, K. P., Cooper, M., & Hayes, B. J. (2019). Accelerating crop genetic gains with genomic selection. *TAG. Theoretical and applied genetics*, 132(3), 669–686.
- Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de Los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., & Varshney, R. K. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in plant science*, 22(11), 961–975.

* * * * *