GreenariA
Go Green, Learn Green
An Agricultural Monthly e-Magazine
Volume 04, Issue 02, 2026
ISSN: 2584-153X
Article ID: G-26-0216

PR GENOMIC SELECTION – A MODERN TOOL FOR CROP IMPROVEMENT

Kommula Uday[1], Kommula Srija[2], Azmeera Swetha Sahithi[1] and Parigi Kanthi Kiran[3]

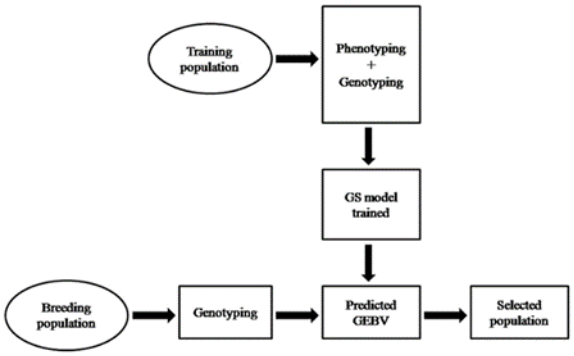
[1]PhD Scholar, Department of Genetics and Plant Breeding, ICAR-IARI, New Delhi - off campus
ICAR-Central Research Institute for Dryland Agriculture, (CRIDA), Hyderabad
[2]Ph.D Scholar, Department of Vegetable Science, SKLTGHU, Mulugu, Siddipet, Telangana
[3]M. Sc in Genetics and Plant Breeding, PJTAU, College of Agriculture, Rajendra nagar, Hyderabad
**Corresponding Author Mail ID: kommulauday364@gmail.com*

1. Introduction:

Commercial plant breeding primarily targets complex quantitative traits such as yield, drought tolerance, and nutritional quality, which are controlled by a few major QTLs and many minor ones. Traditional Marker-Assisted Selection (MAS) focuses only on statistically significant markers, capturing mainly large-effect QTLs and ignoring the cumulative contribution of numerous small-effect genes, making it inadequate for improving such traits. To overcome this limitation, Meuwissen *et al.* (2001) introduced Genomic Selection (GS), which uses genome-wide marker information to estimate an individual's genetic value in terms of Genomic Estimated Breeding Value (GEBV) without requiring prior QTL detection or significance testing.

2. Theoretical Framework and Operational Workflow

Genomic Selection is predicated on the existence of extensive Linkage Disequilibrium (LD) across the genome. It assumes that with sufficient marker density, every QTL affecting a trait will be in LD with at least one marker. The workflow is bifurcated into two distinct populations:



2.1 The Training Population

The Training Population serves as the foundation for the GS model. It consists of individuals that are both genotyped for genome-wide markers and phenotyped for the target traits. The data from this population is used to "train" the statistical model, estimating the effect of every marker allele simultaneously.

2.2 The Breeding Population

The Breeding Population consists of the selected candidates. These individuals are genotyped but not phenotyped. Using the marker effects derived from the Training Population, the model calculates a GEBV for each candidate. Selection is then applied based solely on these predicted values, allowing for the identification of superior genotypes at the seedling stage.

3. Determinants of Prediction Accuracy

The efficacy of GS is measured by the accuracy of the GEBV, defined as the correlation between the predicted value and the true

breeding value. Several factors critically influence this accuracy:

- **Training Population Composition:** The training population must be genetically related to the breeding population to ensure that LD phases (associations between markers and genes) are consistent. Ideally, the training population should include parents or recent ancestors of the breeding candidates. If the populations are unrelated, prediction accuracy diminishes significantly.
- **Population Size:** There is a linear relationship between the size of the training population and prediction accuracy. Simulation studies have demonstrated that reducing the training population size from 2,200 to 500 individuals can cause prediction accuracy to plummet from 0.848 to 0.708.
- **Marker Density:** Adequate marker coverage is essential to capture QTL effects. Cross-pollinated species, which generally exhibit faster LD decay, require a much higher marker density than self-pollinated species to achieve comparable accuracy.
- **Trait Heritability:** GS is particularly advantageous for low-heritability traits like yield, where phenotypic selection is inefficient. While the absolute accuracy of GEBV estimates declines with lower heritability, increasing the size of the training population can overcome this problem.

4. Statistical Models for GEBV Estimation

The estimation of marker effects in GS presents a "large *p*, small *n*" statistical problem, where the number of markers (*p*) far exceeds the number of phenotypic observations (*n*). Standard least-squares regression cannot solve this; therefore, specialized models are employed:

4.1 Ridge Regression - Best linear unbiased prediction (RR-BLUP)

Proposed for GS by Meuwissen et al. (2001), Ridge Regression treats marker effects as random and assumes a normal distribution with equal variance for all markers. This method shrinks all marker effects toward zero. While the assumption of equal variance is biologically unrealistic (as genes have varying effect sizes), RR-BLUP is computationally efficient and performs well for traits controlled by many genes with small effects.

4.2 Bayesian Methods (BayesA and BayesB)

Bayesian approaches relax the assumption of equal variance.

- **BayesA:** assigns a separate variance for each marker using an inverted chi-square distribution.
- **BayesB**: assumes that many markers have no effect (zero variance) while others have large effects.

Simulation studies suggest that BayesB outperforms Ridge Regression when the trait is influenced by large-effect QTLs, as it better captures the underlying genetic architecture.

4.3 Semi-Parametric and Machine Learning Models

To account for non-additive effects such as epistasis (gene-gene interactions), models like Reproducing Kernel Hilbert Spaces (RKHS) and Neural Networks are employed. Machine learning methods, including Random Forest and Support Vector Machines, are particularly useful when epistasis contributes significantly to genetic variance, as they can model complex, non-linear relationships that linear models, such as RR-BLUP, might miss.

5. Applications

5.1 Accelerating the Breeding Cycle

Genomic Selection (GS) allows DNA-based selection at the seedling stage, enabling rapid cycling with multiple breeding rounds per year, and in perennials like oil palm it can reduce the breeding cycle from about 19 years to around 6 years.

5.2 Management of Exotic Germplasm

The use of exotic or wild germplasm is hindered by linkage drag and lengthy pre-breeding timelines (10–20 years). Genomic Selection enables a faster two-step introgression approach, selecting F_2 and later generations to quickly accumulate favorable alleles and purge unwanted background, achieving in about 3 years what once took decades.

6. Future Directions

- **New Marker Systems:** GS now uses SNPs along with CNVs and epigenetic markers, improving the prediction of complex traits.
- **AI-Driven Models:** Machine learning and semi-parametric models better capture non-linear effects and enhance GEBV accuracy.
- **Breeder-Friendly Software:** GS tools are shifting to GUI-based platforms, enabling easier routine use. Robust databases now manage large genotype–phenotype–environment datasets for real-time decisions.
- **Optimized Training Populations:** Dynamically updated training sets maintain accuracy while reducing phenotyping costs.
- **Managing Inbreeding:** Marker-based diversity control and within-family selection help sustain long-term gains.
- **Non-Additive Effects:** Inclusion of dominance and epistasis improves predictions for complex traits.
- **G×E Consideration:** Modern GS incorporates genetic background and environment interactions for stable performance.
- **Biological Integration:** Linking GS with QTLs, gene annotation, and functional genomics strengthens interpretation.
- **Lower Genotyping Costs:** Falling sequencing costs make GS economically viable and widely adoptable.

7. Conclusion

Genomic Selection represents a decisive shift from the observational to the predictive. By leveraging the entire genome, it addresses the limitations of Marker-Assisted Selection regarding quantitative traits. While implementation requires significant infrastructure and careful management of training populations to maintain accuracy, the potential of GS to increase genetic gain per unit of time makes it an indispensable tool for modern crop improvement. As genotyping costs decrease and statistical models evolve to capture environmental and epistatic interactions better, GS will likely become the standard operating procedure for plant breeding globally.

References

1. Cui Y, Li R, Li G, Zhang F, Zhu T, Zhang Q, Ali J, Li Z, Xu S. Hybrid breeding of rice via genomic selection. *Plant Biotechnol J.* 2020 Jan;18(1):57-67. doi: 10.1111/pbi.13170. Epub 2019 Jun 26. PMID: 31124256; PMCID: PMC6920338.
2. Huang, M., Balimponya, E.G., Mgonja, E.M. et al. Use of genomic selection in breeding rice (*Oryza sativa* L.) for resistance to rice

blast (*Magnaporthe oryzae*). *Mol Breeding* 39, 114 (2019). https://doi.org/10.1007/s11032-019-1023-2

3. Meuwissen, T. H., Hayes, B. J., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *genetics*, *157*(4), 1819-1829.
4. Michel, S., Löschenberger, F., Ametz, C. et al. Combining grain yield, protein content and protein quality by multi-trait genomic selection in bread wheat. *Theor. Appl. Genet.* 132, 2767–2780 (2019). https://doi.org/10.1007/s00122-019-03386-1
5. Shu YJ, Yu DS, Wang D, Bai X, Zhu YM, Guo CH. Genomic selection of seed weight based on low-density SCAR markers in soybean. *Genet Mol Res.* 2013 Jul 3;12(3):2178-88. doi: 10.4238/2013.July. 3.2. PMID: 23884761.