

## Genomic Selection (GS) for improvement of crops

Shalu Kumari<sup>1</sup>, Namata kumara<sup>2</sup>, Bishawajit Kumar<sup>2\*</sup>

<sup>1</sup>Department of Genetics and Plant Breeding, IGKV, Raipur, Chhattisgarh

<sup>2</sup>Department of Genetics and Plant Breeding, RPCAU, Pusa, Samastipur, Bihar

\*Corresponding author email: [bishawajitpbg@gmail.com](mailto:bishawajitpbg@gmail.com)

### Introduction

Genomic Selection (GS) has emerged as a potent selection strategy in breeding programs, thanks to breakthroughs in crop genome sequencing and the availability of genome-wide density marker systems. Genotype selection in GS is based on genome estimated breeding values (GEBVs), which have a lot of potentials to improve selection efficiency. So, *GS is a specialized form of MAS, in which information from genotype data on marker alleles covering the entire genome forms the basis of selection.* Meuwissen *et al.* (2001) proposed GS in dairy cattle breeding for the first time 2001. Since then, GS has been effectively implemented in numerous cattle breeding programs, with genetic gains per generation doubled when compared to traditional cattle breeding approaches. The application of GS in crop plants and tree species have been recognized over time due to its significant potential for improved/ accelerated breeding.

### Advantages of GS over MAS

GS facilitates rapid selection of genotypes with superior performance and accelerates the breeding cycle by increasing the intensity and accuracy of selections thereby providing a reliable selection needed for faster genetic progress in breeding for complex traits. So, GS is a form of MAS with extended scope and advantages and these are briefly mentioned below in table 1.

Table 1: Difference between GS and MAS in terms of advantages over MAS.



MAS	GS
Uses only a subset of markers used for selection	Uses several genome-wide markers
Prior identification and mapping of genes or QTLs related to the traits of interest, estimations of marker-trait associations, and their validation in different populations are required	GS eliminates the establishment of the marker-trait association and its rigorous validation in the breeding population.
Markers in MAS explain only a limited part of the genetic variations for a trait, i.e., QTLs with small effects would not be detected.	The use of genome-wide markers makes it capable of identifying all the QTLs, including even those with small effects.
MAS follows the “breeding by design” approach	Uses an exhaustive approach for trait improvement and can minimize the costs and time

### *Genomic Estimated Breeding Values (GEBV) and Breeding Values (BV)*

The GEBV of an individual is the sum total of effects associated with all the marker alleles present in the individual and included in the GS model applied to the population under selection. The breeding value (BV) of an individual/line represents the expected phenotype of its progeny. The BV is thus determined through progeny testing and is solely based on additive genetic influences. The genotypic value of an individual/line, on the other hand, is the phenotype predicted by its genotype. As a result, the genotypic value is determined by both additive and non-additive genetic effects.

### *A Generalized Procedure for Genomic Selection*

The GS method is based on two separate but related populations i.e., the Training population & Breeding population. The training population is used for training the GS model and for obtaining estimates of the marker-associated effects needed for the estimation of GEBVs of individuals/lines in the breeding population. The breeding population, on the other hand, is the population subjected to GS for achieving the desired improvement and isolation of superior lines for use as new varieties/parents

of new improved hybrids. A simple schematic representation of the genomic selection (GS) scheme is given below in figure 1. (Singh B.D. & Singh A.K., 2016)

### Stepwise procedure for GS to apply in breeding programs are given below

1. The first step in a GS program is to create a training population suitable for the concerned breeding population
2. The individuals/lines in the training population are genotyped for a large number of markers evenly distributed over the entire genome at an adequate density
3. The individuals/lines in the training population are subjected to extensive phenotypic evaluation for the target trait(s) in replicated trials.
4. The phenotype and marker genotype data are used for computing the GS model parameters, this is called model training.
5. The breeding population is evaluated for the same set of markers that were used for the estimation of the model parameters in the training population
6. The GEBVs of individuals/lines of the breeding population are calculated from their marker genotype data and the marker-associated effects are estimated from the training population
7. The superior individuals/lines are selected from the breeding population on the basis of their GEBV estimates

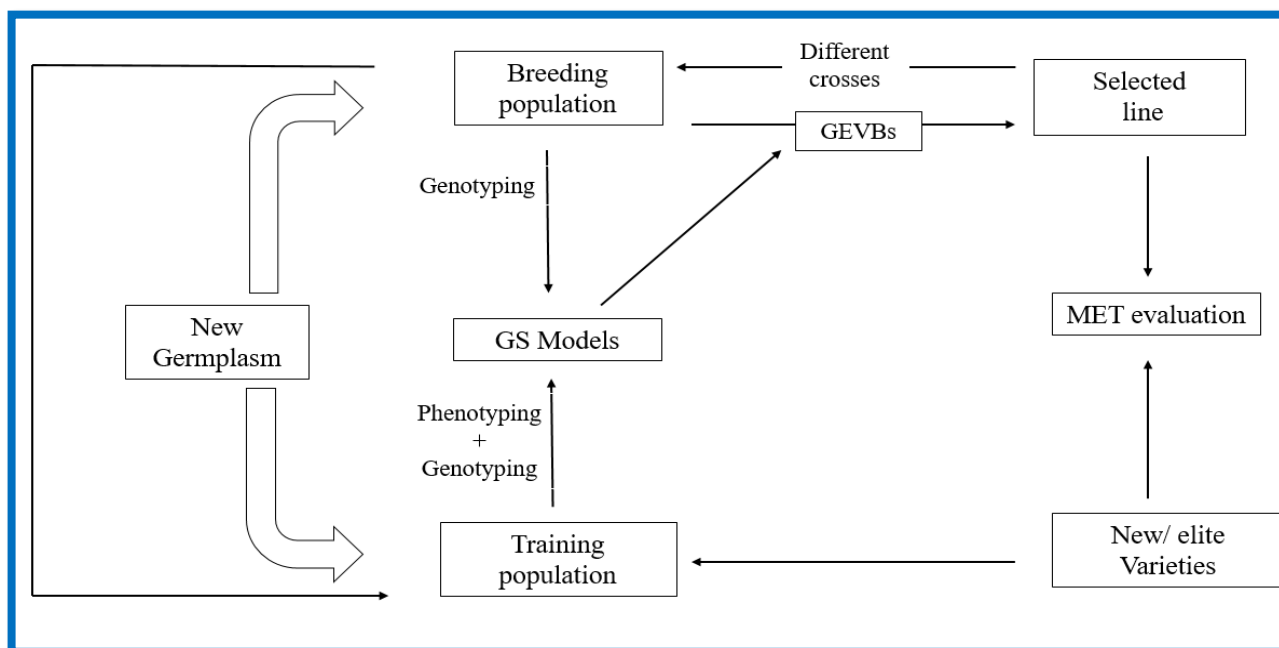


Fig. 1: Schematic representation of the basic genomic selection (GS) methodology. (Krishnappa G. et al., 2021)

## Statistical models for computation of Genomic Estimated Breeding Values

Many predictive models have been developed for GS studies in crop plants. The selection of an appropriate predictive model is critical to realizing high prediction accuracy and turning it into the success of GS. Two common prediction models used in various GS studies in crop plants:

- I. Genomic best linear unbiased prediction (G-BLUP) models
- II. Ridge regression best linear unbiased prediction (RR-BLUP) models

RR-BLUP results in high prediction accuracies when the trait is controlled by several loci with small effects. G-BLUP is also a widely used model which is equivalent to the RR-BLUP and it can estimate the additive genetic merits from a genomic relationship matrix. The common assumption of G-BLUP and RR-BLUP is that the effects of all loci have a common variance, making them more appropriate for complex traits that are controlled by numerous minor genes. However, assumptions of G-BLUP and RR-BLUP are rarely met as most of the markers have small or no effects and few markers have large effects. So, Bayesian methods (BayesA & BayesB) fit into the actual situations and allow different markers to have different effects and variances. (Krishnappa G. *et al.*, 2021)

## Applications of GS in crop breeding

GS can be applied to various groups of crops and even it is applied to the orphan crop. Here, GS is applied in different types of populations and with different numbers of markers for various types of traits in Table 2. By applying different models in these populations, a prediction accuracy of an average of 0.6 was obtained which helped in the selection of superior individuals.

Table 2: Summary of genomic selection studies in cereals.

Crop	Population	Markers	Traits	Model	Prediction accuracy	Reference
Wheat	329 genotypes	7748 SNPs	Agro-morphological & quality traits	G-BLUP	0.29-0.67	Ward, B. P <i>et al.</i> (2019)

Rice	369 elite lines	73147 SNPs	Grain yield, days to 50% flowering, plant height	RR-BLUP	0.31-0.63	Spindel, J. <i>et al.</i> (2015)
Maize	479 hybrids	669 SNPs	Grain yield, moisture, root lodging, stalk lodging	BLUP, RR-BLUP	0.66-0.90	Massman, J. M. <i>et al.</i> (2013)

### Advantages of Genomic Selection

- i. The marker effects are estimated from the training population and used directly for GS in the concerned breeding population, and QTL discovery, mapping, etc. are not required.
- ii. Both simulation and empirical studies reveal that GS produces greater gains per unit time than phenotypic selection.
- iii. GS is able to predict the performance of breeding lines more accurately than that based on pedigree data
- iv. GS increases the effectiveness of selection, particularly for low heritability traits.
- v. GS would tend to reduce the rate of inbreeding and the loss of genetic variability in comparison to selection based on breeding values estimated from phenotype data without sacrificing selection gains
- vi. Phenotyping for every selection cycle in the breeding population is not required. So, reduces the length of the breeding cycle.
- vii. Allow breeders to select parents for hybridization programs from among those lines that have not been evaluated in the target environment

### Constraints in the adoption of GS

- i. GS has still not become popular with the plant breeding community primarily due to insufficient evidence for its practical usefulness.
- ii. The marker effects and GEBV estimates may change due to changes in gene frequencies and epistatic interactions. This would necessitate updating the GS model with every breeding cycle.
- iii. Most simulation models are based on additive genetic variance. These models ignore epistatic effects, which do not seem to be realistic.

- iv. Limited knowledge about the genetic architecture of quantitative traits limits our ability to develop appropriate models of GS to achieve maximum prediction accuracy.
- v. The need for genotyping of a large number of marker loci in every generation of selection adds considerably to the cost

## Conclusion

Genomic selection, which reduces the number of selection cycles and eliminates arduous, labor-intensive, and time-consuming phenotyping in BP, is a strong approach for accelerating genetic gain in crop plants for economically essential and complicated traits. Because it uses genome-wide dispersed markers in a trained model for selection, GS is a reliable method of selection. It eliminates the bottleneck in the usage of MAS and MARS by not requiring prior knowledge of genes/loci and marker-trait associations, as well as their rigorous validation. Integration of GS with other fast generation advancements interventions, such as HTP and HTG, will expand the scope of its utility and precision in achieving the expected genetic gain in many crop species to meet world food demand.

## References

- Krishnappa, G., Savadi, S., Tyagi, B. S., Singh, S. K., Masthigowda, M. H., Kumar, S., and Singh, G. P. (2021). Integrated genomic selection for rapid improvement of crops. *Genomics*, **113**: 1070-1086.
- Massman, J. M., Gordillo, A., Lorenzana, R. E., and Bernardo, R. (2013). Genomewide predictions from maize single-cross data. *Theoretical and applied genetics*, **126**(1): 13-22.
- Meuwissen, T. H., Hayes, B. J., and Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, **157**(4): 1819-1829.
- Singh, B. D., and Singh, A. K. (2016). Marker-assisted plant breeding: principles and practices (pp. 295-312). New Delhi: Springer.
- Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redona, E., and McCouch, S. R. (2015). Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS genetics*, **11**(2): e1004982.
- Ward, B. P., Brown-Guedira, G., Tyagi, P., Kolb, F. L., Van Sanford, D. A., Sneller, C. H., and Griffey, C. A. (2019). Multi environment and multitrait genomic selection models in unbalanced early-generation wheat yield trials. *Crop Sci*, **59**: 491–507.