JUST AGRICULTURE
multidisciplinary e-Newsletter

Genomic Selection: A Key to Rapid Genetic Gain in Crop Plants

Narayana Bhat Devate[1*], Karthik Kumar M[1], Manoj Gowda M[1] and Revanasiddayya[2]

[1]ICAR-Indian Agricultural Research institute, New Delhi, India
[2]University of Agricultural Sciences, Raichur, Karnataka, India

ARTICLE ID: 33

Introduction

Food is a basic requirement for all living things. Humans rely on a variety of crops, mostly cereals for dependable energy sources. Since the beginning of agriculture, man has made a constant attempt to choose appropriate plant forms and high yielding cultivars. The previous century saw a rise in food yields due to systematic crop breeding and the work of thousands of scientists, especially after the green revolution. To achieve the food and nutritional security and the other needs of the growing population, an increase in the yield potential and reduction in the yield gaps of crop plants is essential. The current rate of genetic improvement is about 0.8-1.2% for the major food crops and in recent years these values have plateaued out. This genetic improvement rate falls short of the 2.4% required to fulfil food demand of the projected world population of 9.5 billion people by 2050's.

It is challenging to attain this pace of genetic gain using conventional breeding techniques, particularly in view of the depletion of land and water resources under the influence of climate change. Additionally, the majority of yield and agronomic traits are genetically complex and greatly affected by environmental fluctuations, making it challenging to increase them using traditional breeding techniques. As a result, using markers in breeding programmes gave the advantage of breeding with less environmental effect. However, oligo genes or quantitative trait loci (QTLs) are the main targets of marker-assisted selection and backcrossing. Hence, genome wide distributed markers to cover all available minor genes can be achieved by one of the marker assisted prediction based selection method called genomic selection. Meuwissen et al. (2001) proposed the GS for the first time in cattle breeding. Since then, several breeding programmes have effectively used GS to increase genetic gains.

Genetic gain through genomic selection

In breeding programmes, the genetic improvement or response to selection is often measured by the genetic gains made yearly or per unit of time, i.e., the amount of improvement in breeding population/line performance brought about by selections in a year or unit of time. The expected value of genetic gains is estimated using the popular breeder's equation:

$$GA = \frac{i\sigma_a h}{L}$$

where,

i is the selection intensity,

h is the narrow sense heritability,

σ_a is the additive genetic variance,

L is the length of breeding cycle interval or generation

Genomic selection increases the genetic gain by rapid increase in selection intensity, heritability of traits by lower environmental influence and reduction in the length of the breeding cycles.

Basic principle of genomic selection

Trait-determining genes are dispersed across the genome. Some may be having a significant impact, while others may be having a less significant impact. when we use genome wide dense markers all the genes responsible for the trait will be in linkage disequilibrium with one or the other marker. As a result, when the effects of each marker are calculated across the genome, all the genes involved in the trait of interest may be indirectly captured due to linkage disequilibrium with marker. Sum of the marker effects present in an individual is noted as genomic estimated breeding value (GEBV) based on which potentiality of individuals can be predicted. In a population desirable genes/QTLs were distributed randomly among the individual during crossing over and segregation. Some of them are likely to fix more favourable allele than others able to produce superior phenotypic expression. Such individuals can be screened based on GEBV without any phenotypic observation.

Training and breeding population

Genomic selection can be applied with trained GS model for the selection of lines based on GEBV from breeding population with only genotypic data without any

phenotyping. However, model training requires a population called training population with both genotyping and phenotyping data. The broad scheme of genomic selection is given in the figure below.
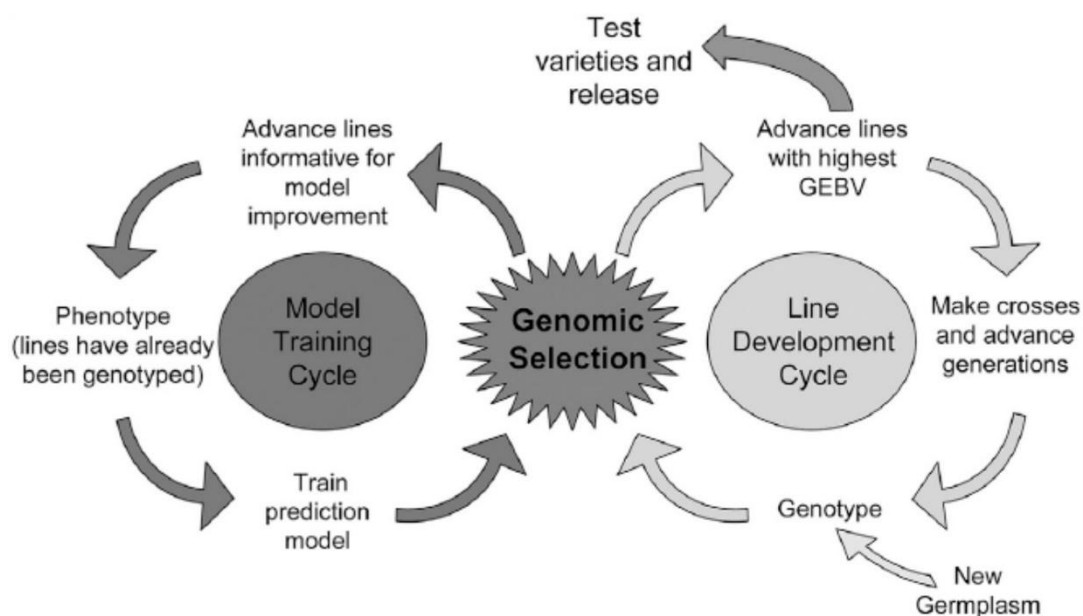

Figure 1: Flowchart of GS by Heffner et al., (2009)

As a general procedure of genomic selection training population should be phenotyped preferably at multi location over the years to nullify environmental effect. Each lines of the training population should be genotyped for genome wide markers preferably SNP or SSR. With this data suitable GS model should be trained. With the trained GS models GEBV of breeding population can be calculated with only genotypic data for same set of marker. Selection of lines from the breeding population can be done based on predicted GEBV.

Training population must have some relationship with breeding population. TP often consists of individuals with known pedigree and having close relationship. Natural populations having adequate genetic diversity and variation, such as germplasm accessions or breeding lines (elite inbred lines, bi parental, and multi-parental populations), might be used as TP. Relationship between training and breeding population is of three type in general.

1. Training and breeding population lines are segregating progenies from the same cross.
   (2) Training and breeding population lines comprising of both related and unrelated genotypes.

2. Training and breeding lines comprising of lines from a diverse germplasm collection. Increased genetic relation among the training and breeding population can enhance reliability of GS with higher prediction accuracy.

Basic model of genomic selection

The basic model of GS includes sum of the multiplication of effect associated with genotype at a marker locus (gi) and coded marker genotype in an individual (xi). The equation is:

$$P = \sum gixi + e$$

However, advance in the statistics and computational biology, there are many advanced models of GS are proposed. Genomic best linear unbiased prediction (G-BLUP) and ridge regression best linear unbiased prediction (RR-BLUP) are the two common prediction models used in various GS studies in crop plants. Apart from that model based on Bayesian approaches (BayesA, Bayes B, BayesCπ), machine learning approaches (Random forest (RF), Support vector machine (SVM)) and non-parametric approaches (RKHS-Reproducing kernel Hilbert space) were also proposed. Ultimate aim of any GS model is to identify marker effects precisely with high prediction accuracy of GS. High prediction accuracy from the GS model guarantees that the calculated GEBV is close to the real breeding value.

Prediction accuracy

The success of any genomic selection depends on the prediction accuracy. Higher the prediction accuracy higher will be the reliability of genomic selection. A crucial component of any effective genomic selection programme is validating the accuracy of prediction estimates for models trained using a training population. The Pearson's correlation between the estimated GEBVs and actual breeding values estimates the prediction accuracy of the GS models. Accuracy of prediction depends on the composition of the training population (size and genetic structure), markers (number, density and coverage), relationships between TP and BP, genetic architecture and heritability of the trait, precision of phenotyping of TP and the GS model used.

The estimation of prediction accuracy can be done by k fold cross validation in training population or by developing separate validation set. In the first method, leave-one-out cross-validation, a kind of k-fold cross validation is used in the training population

itself. If k-fold cross-validation is taken into account, the entire dataset is split into k groups and examined 'k' times, with one of the k groups being left out of the training model and used for validation in each analysis. Whereas in second method, a validation population (VP) that has been genotyped and phenotyped is utilized. Based on real TP phenotypic data, marker effects across the genome are computed. The estimated GEBV of VP is compared with the actual breeding values (TBV) of VP to evaluate the prediction accuracy of the model.

Integrated genomic selection

Although GS is known as a potential tool for genetic improvement of complex traits with a greater reduction in breeding cycles, enhanced genetic gain can be achieved by incorporation of various advanced breeding tool with genomic selection approaches. Genomic selection can be combined with modern breeding technologies such as speed breeding to reduce breeding cycle time combined with genome editing to introduce new variation for selection, along with high through put genotyping and phenotyping platforms, other marker technologies like marker assisted selection and transgenic technologies can be integrated in GS. There are research efforts made to achieve genetic gain by integrating genomic selection with other modern breeding approaches in major crops. The successful integration of genomic selection with other molecular and omics technologies enhances the potential of genomic selection for selecting elite genotypes for future breeding needs.

Conclusion

Genetic gain, which is essential for fulfilling future food demand, may be greatly increased by genomic selection. To attain higher genetic gain in the direction of boosting the productivity of major food crops to feed a growing population, it is necessary to develop an optimised genomic selection pipeline to use in existing breeding platforms. Modification and integration of GS with existing breeding programs increases genetic gain per unit time and cost. The cost-effective genotyping of large number of lines in plant populations has become possible because of advancements in genotyping technology. The development of an efficient breeding pipeline to achieve rapid success in increasing genetic gain requires the integration of GS with modern tools such as speed breeding. It has facilitated the broad scale deployment of GS for economically significant traits in several important crops and expanded its horizons to improve many more important traits in several crops.

References and Further reading

Anilkumar, C., Sunitha, N. C., Harikrishna, Devate, N. B., & Ramesh, S. (2022). Advances in integrated genomic selection for rapid genetic gain in crop improvement: a review. *Planta*, *256*(5), 87.

Heffner, E. L., Sorrells, M. E., & Jannink, J. L. (2009). Genomic selection for crop improvement. *Crop Science*, *49*(1), 1-12.

Krishnappa, G., Savadi, S., Tyagi, B. S., Singh, S. K., Mamrutha, H. M., Kumar, S., ... & Singh, G. P. (2021). Integrated genomic selection for rapid improvement of crops. *Genomics*, *113*(3), 1070-1086.

Singh, B. D., & Singh, A. K. (2015). Marker-assisted plant breeding: principles and practices.Springer New Delhi

Varshney, R. K., Roorkiwal, M., & Sorrells, M. E. (2017). *Genomic selection for crop improvement: new molecular breeding strategies for crop improvement.* Springer International Publishing.