**Genomic Selection for Plant Improvement: An Introduction**

*Rajeev K Varshney*1,2,\**, Manish Roorkiwal1, Mark E Sorrells3*

1International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India

2University of Western Australia (UWA), Crawley, Australia

3Cornell University, Ithaca, New York, USA

**ABSTRACT**

Marker assisted selection (MAS) exploits the markers associated with trait of interest for selecting lines with superior allele for developing improved superior lines. However use of MAS is restricted to simple traits due to its inability to handle complex traits as well as higher genotyping cost. Advancement in genomic technologies has been able to reduce the cost of genotyping drastically enabling the use of genome-wide marker data for selecting lines with higher breeding value. Genomic selection (GS), a modern breeding approach that uses using genome-wide maker data to estimate the breeding value, has the potential to address the complex traits. GS exploits the genotyping and phenotyping data on training population and establishes the prediction models to calculate the genomic estimated breeding value (GEBV) using genotyping data. GS has the capability to reduce selection duration, increase selection accuracy, intensity, efficiency and gains per unit of time, and thereby enhancing the rate of genetic gains. Availability of cost effective genotyping platforms has enabled the generation of large scale genotyping data in much less cost and faster manner, enabling the deployment of GS in several plant species. This chapter provides an introduction about the book, highlighting the basic and advance principles of GS breeding and its applications for plant improvement.

**Introduction**

Producing sufficient food to meet the demand of vastly growing population and eradication of rural poverty is one of the critically important issues that world is facing. With current growing pace, the World population is expected to cross mark of 9 billion by 2050 adding further pressure to already exhausted food production systems. Considering the volatile climate condition, it would be difficult to maintain the crop production in conjugation with the demand, resulting increase in the food price affecting people who spend the highest percentage of their disposable income on food. In addition to climate condition, limited water resource availability and poor soil health have the potential to restrict the food crop production. Furthermore with increase in the world population, the availability of agricultural land is decreasing. Under these several constraints, to meet the rising demand for food, agricultural production must increase by an estimated 50% without greatly increasing water usage or expanding the total land area dedicated to agriculture. Smallholder farmers especially from under developed and developing part of world with limited access to agricultural inputs or agricultural markets are also likely to be affected by rising production costs and climate volatility. As per the United Nations estimates, more than 790 million people globally do not have access to sufficient nutritious food (<https://sustainabledevelopment.un.org/sdg2>), placing a threat to achieve the Sustainable Development Goal (SDG) target of zero hunger (universal access to safe, nutritious and sufficient food at all times of the year).

In the event of these huge challenges, there is a need to look for new ways of breeding for food crops and other plant species by using modern technologies. Modern breeding approaches that have the capability to reduce breeding cycle time, provide more precision in selection and more efficient use of genetic variation can be exploited to increase the rate of genetic gains in breeding programs. The rapid decline in the cost of sequencing and genotyping has led to the development of new tools and strategies that can transform the way we breed plant species. Higher genotyping cost, in past, restricted the regular use of markers in breeding. Only in select cases, limited number of markers for the target regions were used for selecting the lines based on presence or absence of agriculturally important alleles. Development of crop varieties using conventional breeding approaches has been effective but time consuming and labor intensive. Recent advances in the next generation sequencing (NGS) technologies have been able to reduce the cost of genotyping and sequencing. This has enabled use of the high-throughput and cost effective high density genotyping. These low cost genotyping platforms have accelerated the use of markers in the breeding programs using genome-wide approaches (Varshney et al. 2014).

Integrating genomics tools with conventional breeding can make path breaking impact for current and future environmental challenges efficiently in less time and more accurately. In such conditions, three major resources (germplasm, genetics and genomics) are mandatory in all the plant species for speedy genetic gains of these species using decision support tools. First generation molecular breeding approaches (marker-assisted backcrossing, marker-assisted recurrent selection) follow a lengthy process for developing genetic populations for identification of linked markers for few simply traits. Majority of the economically important traits such as drought tolerance and yield are polygenic in nature and controlled by multiple genes with small effects. In order to improve complex traits, such as drought tolerance and yield, the modern breeding approach such as genomic selection (GS) (Meuwissen et al. 2001) can be deployed which aim specifically at improving quantitative traits by using genome-wide marker profiled data without need of identification markers associated with traits of interest. GS uses a ‘training population’ of individuals that have been both phenotyped and genotyped to develop a model for calculating genomic estimated breeding values (GEBVs). Subsequently by using this model, GEBVs can be calculated for untested individuals from a ‘candidate population’ and selection candidates (SCs) for making the crosses can be identified. Although GEBVs are not the function of the underlying QTLs/genes for the trait, they are the ideal selection criterion (Jannink et al. 2010). GS basically attempts to capture the total additive genetic variance with genome-wide marker coverage and effect estimates (Rutkoski et al. 2011). Therefore, selection of an individual without phenotypic data can be performed based on this model (predicting the individual’s breeding value). The models can be used to calculate GEBVs that help breeder to identify offspring that will be good parents in the next generation, based solely on genotypic information about an existing line. Use of GEBVs in the context of genome wide prediction promises to help accelerate the rate of genetic gain in breeding.

In view of above, Editors planned this book to bring up-to-date information on GS breeding and its application for crop/plant species improvement. Editors believe that this book can serve as ready reference for geneticist and plant breeders. This chapter introduces the book and provides a summary of different chapters included in the book.

**Methodologies and models for GS**

First step towards deploying GS in plant breeding is to define a training set, which should be closely related to the testing population. The Chapter 2 entitled “Training population design and resource allocation for genomic selection in plant breeding” provides detailed information about composition and optimization of training population design related to population and trait architecture. In this chapter, Liana Nice and Aaron Lorenz highlight the importance of training population for prediction the breeding value of a lines. The chapter focuses on the process to select a calibration set (training population) for model training and optimizes the resource allocation for field trials. With the advent of technologies, it has become possible to collect phenotyping data in more precise manner with decreased error, fewer inefficiencies and in larger quantities. NGS technologies are contributing to continuous decrease in the genotyping cost and are enabling the prediction of breeding value using genome wide marker profiling. The chapter also discusses about the possible resource allocation in view of number of replication for calibration of GS models vs allocation of more plots in breeding cycle and allocation of plots to within versus across environment replication.

The Chapter 3 entitled “Derivation of linear models for quantitative traits by bayesian estimation with Gibbs sampling” contributed by Akihiro Nakaya and Sachiko Isobe provides detailed information about construction of a prediction model using a linear model and determination of the model parameters using the Bayesian estimation with Gibbs sampling providing a theoretical background sufficient to implement practical software for the model construction. The chapter also provides a sample output by the implemented software. The chapter describes about the different prediction models including linear models, one-marker model, two-marker model without interactions and two-marker model with interactions to predict the trait value of a sample using the obtained environment types and genotypes. Prediction of the trait values of samples based on their genetic and environmental factors has been explained using a prediction model that describes the relationship between the explanatory factors observed in the samples and the trait values. The chapter suggests that defining a prediction model for the target trait enables the selection to be based on the predicted trait values, making it an essential part of genomic selection. In the case number of markers is relatively higher as compared to the number of the samples, or the number of the samples is not sufficient for the number of markers, the prediction model will be distorted. In order to address the issue related to model overfitting, detailed inspection of the prediction model is necessary and strategy based on the linear mixed models and the Bayesian estimation may be useful in the prediction of trait values of samples.

Montesinos-López and colleagues highlighted recent advances in models for genomic-enabled prediction developed for ordinal categorical and count data in Chapter 4 entitled “Bayesian genomic-enabled prediction models for ordinal and count data”. Authors used these two models on simulated as well as real dataset using Bayesian framework suggesting that models used are a good alternative for analyzing ordinal and count data in the context of genomic-enabled prediction. Tested models have an advantage to perform an exact logistic or probit ordinal regression without having to do approximations to perform a logistic ordinal regression. Genotype (G) and environment (E) interaction is expected to affect the prediction accuracies and therefore modelling G×E in the context of genomic-enabled prediction plays a central role in plant breeding for the selection of candidate genotypes. In order to best use of GS models, understanding the data type being analyzed is important before deciding on the modelling approach to be employed.

**GS in field crop breeding**

GS has been used or being used in several crop breeding programs. The books includes three chapters on applications offering both constraints and opportunities of GS in crop breeding. The Chapter 5 entitled “Genomic selection for small grains improvement” by Rutkoski and colleagues presents an overview of GS efforts being undertaken in the small grain cereals. Authors in the chapter have explained different approaches for implementation of GS in applied breeding programs. A total of 40 GS studies have been undertaken so far in small grains including bread wheat, barley, oat, rye, durum wheat, perennial ryegrass, and intermediate wheatgrass. This chapter also discusses about the factors affecting the GS prediction accuracies in small grains and also highlights the applicability of GS for analyzing and predicting G × E. They have discussed various scenarios affecting gain from selection and cost relative to conventional breeding. Authors have also discussed about the cost-benefit ratio for deploying GS in cereal crops.

In the Chapter 6 entitled “Current status and prospects of genomic selection in legumes”, Jain and colleagues from ICRISAT provide an update about molecular breeding in legumes and describes the ongoing GS efforts in some legume breeding programs including soybean, alfalfa, pea, chickpea and groundnut. Legumes have witnessed significant progress in the field of genomics and genetics in the past decade and efforts to deploy the MAS have yielded some success for developing superior legume varieties. However, as expected MAS has not been that successful for addressing complex traits such as drought and yield and therefore, efforts to deploy the GS in legumes breeding were initiated. Authors have also suggested that this is a high time for other legumes to start deployment of GS in those breeding programs to achieve the higher rate of genetic gain.

Hybrid breeding has been successful over varietal improvement in several crops. Schulthess and colleagues from Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany describe the basic concepts of hybrid breeding and deployment of GS methods to simplify the philosophy underlying GS with hybrid breeding in the chapter 7, “Genomic selection in hybrid breeding”. Authors have explained about the basic concepts relevant to hybrid breeding including dominance, heterosis, combining abilities and heterotic groups and patterns. The chapter also describes about the deployment of GS for hybrid genotypes using cross-validated prediction accuracy, accommodating dominance effects within the GS Model, and other GS approaches employed in hybrid breeding. Deployment of GS in hybrid breeding is very challenging as large variables are present in hybrid breeding as compared to regular breeding. Authors suggest to have an integrated plan with multidisciplinary skills of breeders, scientists, and technicians before implementing the GS in hybrid breeding.

**GS for improvement of clonal crops and tree species**

Breeding in clonal crops and tree species is different as compared to field crops. Therefore, Gemenet and Khan in the Chapter 8 entitled “Opportunities and challenges to implementing genomic selection in clonally propagated crops” discuss issues related to deployment of GS for improving the rate of genetic gain in these clonal crops. Authors highlight that conventional breeding approaches for clonal crops involves crossing, planting of true seed plants in different generation followed by evaluation of clones for several generations, making it a time and resource consuming process. Therefore, GS based selection of true seed plants can expedite the breeding process. The chapter also describes about the challenges including modeling of genetic effects and heritability, linkage disequilibrium between markers and QTLs, genetic architecture of traits and size of training population and number of generations following training model, to deploy GS in clonal crops. For instance, GS models generally handle additive effects and assume dominance and epistatic effects as part of the residual which is not case for clonally propagated crops, as dominance and epistatic effects play an important role beside additive effects and need special consideration. Therefore for clonal crops, GS models with capability to include additive, dominance, and epistatic genetic effects need to be developed for analysis.

For tree species, Dario Grattapaglia from EMBRAPA Genetic Resources and Biotechnology, Brazil provides perspectives of genomic selection and a comprehensive discussion on the factors relevant to GS in tree breeding in the Chapter 9, “Status and perspectives of genomic selection in forest tree breeding”. The chapter highlights the potential of GS in enhancing the rate of genetic gain of tree breeding program by reducing the selection cycle. In the case of tree breeding program, long generation time typically necessary to complete a full breeding cycle which can be reduced by genotyping young seedlings and predict their phenotype within few week instead of waiting for long breeding cycle of 4-20 years or more. Authors have also compiled and presented all GS experimental studies in forest trees along with their key attributes and performance of predictive abilities for different traits in the chapter.

**Summary**

As can be seen from the above mentioned introduction of 8 different chapters, GS, a modern breeding approach is gaining popularity and becoming the choice for many breeders for addressing the complex traits. The book provides an up-to-date information about models, methodologies, factors affecting prediction accuracy and some examples of deployment of GS for crop/tree improvement. This book will serve as handbook for users that provides basic as well as advanced understanding about GS. The book is expected to serve as a useful review for users that explains about the germplasm to be used, phenotyping evaluation, marker genotyping methods, and statistical models involved in GS. It also includes some examples of ongoing activities of genomic selection in cereal, legume, clonal crop and tree species. Editors believe that readers will be heavily benefitted from the book.

**References**

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819-1829.

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9:166-177.

Rutkoski JE, Heffner EL, Sorrells ME (2011) Genomic selection for durable stem rust resistance in wheat. Euphytica 179, 161–173

Rutkoski JE, et al. Genomic selection for durable stem rust resistance in wheat. BGRI Technical Workshop 2010; 30-31 May 2010, St Petersburg, Russia.

Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoS Biol 12(6):e1001883.