Chapter 1
Genomic Selection for Crop Improvement: An Introduction

Rajeev K. Varshney, Manish Roorkiwal, and Mark E. Sorrells

1.1 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 the world is facing. At the current pace, the world population is expected to cross the mark of nine billion people by 2050 adding further pressure to already exhausted food production systems. Considering the increasingly volatile climate, it will be difficult to maintain the crop production in conjugation with the demand, resulting in increased food prices affecting people who already spend the highest percentage of their disposable income on food. In addition to climate change, limited water resource availability and poor soil health have the potential to restrict food crop production. Furthermore, with increases in the world population, the availability of agricultural land is decreasing. Under these 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 underdeveloped and developing countries with limited access to agricultural inputs or agricultural markets, are likely to be affected by rising production costs and climate volatility. As per the United

---

R.K. Varshney (✉) • M. Roorkiwal
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Telangana, India
e-mail: R.K.Varshney@CGIAR.ORG

M.E. Sorrells
Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York 14853, USA

R.K. Varshney et al. (eds.), *Genomic Selection for Crop Improvement*,
DOI 10.1007/978-3-319-63170-7_1

Nations' estimates, more than 790 million people globally do not have access to sufficient nutritious food (https://sustainabledevelopment.un.org/sdg2), posing a threat to achieving 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 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. In the past, the cost of genotyping restricted the regular use of markers in breeding. In most cases a 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 the 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 genomic tools with conventional breeding can have a major impact for dealing with current and future environmental challenges more efficiently. In such conditions, germplasm, genetic, and genomic resources are mandatory in all plant species for rapid genetic gains in productivity of these species using decision support tools. First-generation molecular breeding approaches (marker-assisted backcrossing, marker-assisted recurrent selection) followed a lengthy process for developing mapping populations for identification of markers linked to quantitative trait loci (QTL) for a few simple traits. The majority of 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, genomic selection (GS) (Meuwissen et al. 2001), can be deployed which specifically aims at improving quantitative traits by using genome-wide marker data without requiring identification of markers associated with QTL for traits of interest. GS uses a "training population" of individuals that have been both phenotyped and genotyped to train a prediction 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 crosses or for advanced yield trials can be identified. Although GEBVs do not identify the function of the underlying QTLs/genes for the trait, they are an excellent selection criterion (Jannink et al. 2010). GS 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 the individual's predicted breeding value.

The models can be used to calculate GEBVs that help the breeder to identify offspring that will be good parents in the next generation, based solely on genotypic information about an existing line. The use of GEBVs in the context of genome-wide prediction promises to help accelerate the rate of genetic gain in breeding.

The purpose of this book is to bring up-to-date information on GS breeding and its application for crop species improvement. The editors believe that this book can serve as ready reference for geneticists and crop breeders. This chapter introduces the book and provides a summary of different chapters included in the book.

1.2 Methodologies and Models for GS

The first step toward deploying GS in crop breeding is to define a training set, which should be closely related to the selection candidate population. Chapter 2 entitled “Training population design and resource allocation for genomic selection in crop breeding” provides detailed information about composition and optimization of training population design related to population and trait architecture. In this chapter, Aaron Lorenz and Liana Nice highlight the importance of the training population design for predicting the breeding value of 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 new technologies, it has become possible to collect phenotyping data in a more precise manner with decreased error and increased efficiency and in larger quantities. NGS technologies are contributing to a continuing decrease in the genotyping cost and are enabling the prediction of breeding value using genome-wide marker profiling. This chapter also discusses the possible resource allocation in terms of the number of replications for calibration of GS models vs allocation of more plots for model training and allocation of plots within and across environment replication.

The Chap. 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. Model parameters are determined 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 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 environment types and genotypes. Prediction of the trait values of samples based on their genetic and environmental factors is explained using a prediction model that describes the relationship between the explanatory factors observed in the samples and the trait values. This 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. When the number of markers is greater than the number of samples, 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 Chap. 4 entitled "Bayesian genomic-enabled prediction models for ordinal and count data". Authors used these two models on simulated as well as a 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 crop breeding for the selection of candidate genotypes. In order to best use GS models, understanding the data type being analyzed is important before deciding on the modelling approach to be employed.

1.3 GS in Field Crop Breeding

GS has been used or is being used in several crop breeding programs. This book includes three chapters on applications offering both constraints and opportunities of GS in crop breeding. The Chap. 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 wheat, barley, oat, rye, durum wheat, perennial ryegrass, and intermediate wheatgrass. This chapter also discusses the factors affecting the GS prediction accuracies in small grains and 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 discussed the cost-benefit ratio for deploying GS in cereal crops.

In Chap. 6 entitled "Current status and prospects of genomic selection in legumes", Jain and colleagues from ICRISAT provide an update on molecular breeding in legumes and describe 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 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 GS in legume breeding were initiated. Authors have suggested that it is time for other legumes to start deployment of GS in those breeding programs to achieve a higher rate of genetic gain.

Hybrid breeding has been successful over varietal improvement in several crops. Schulthess and colleagues from the 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 Chap. 7, "Genomic selection in hybrid breeding". Authors have explained the basic concepts relevant to hybrid breeding including dominance, heterosis, combining abilities, and heterotic groups and patterns. The chapter also describes 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 many variables impact in hybrid breeding as compared to pureline breeding. Authors propose an integrated plan with multidisciplinary skills of breeders, scientists, and technicians before implementing GS in hybrid breeding.

1.4 GS for Improvement of Clonal Crops and Tree Species

Breeding in clonal crops and tree species is different from field crops. Therefore, Gemenet and Khan in Chap. 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 clonal crops. Authors highlight conventional breeding approaches for clonal crops that involve crossing and planting of true seed plants in different generations 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 the challenges including modelling of genetic effects and heritability, linkage disequilibrium between markers and QTLs, genetic architecture of traits, 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 the case for clonally propagated crops, as dominance and epistatic effects play an important role along with additive effects and need special consideration. Therefore, for clonal crops, GS models with the capability to include additive, dominance, and epistatic genetic effects need to be employed 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 Chap. 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 in a tree breeding program by reducing the selection cycle. In the case of a tree breeding program, the long generation time typically necessary to complete a full breeding cycle can be reduced by genotyping young seedlings and predicting their phenotype instead of waiting for long a breeding cycle of 4–20 years or more. The authors have 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.

1.5 Summary

As can be seen from the introduction of eight different chapters, GS, a modern breeding approach, is gaining popularity and becoming the choice for many breeders for improving complex traits. The book provides up-to-date information about models, methodologies, factors affecting prediction accuracy, and some examples of deployment of GS for crop improvement. This book will serve as reference 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 the germplasm to be used, phenotyping, 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.

References

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

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

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

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