Manish K. Pandey · Alison Bentley
Haile Desmae · Manish Roorkiwal
Rajeev K. Varshney   *Editors*

Frontier Technologies for Crop Improvement

Springer

Chapter 9
Genomic Selection in Crop Improvement

**H. V. Veerendrakumar, Rutwik Barmukh, Priya Shah,
Deekshitha Bomireddy, Harsha Vardhan Rayudu Jamedar,
Manish Roorkiwal, Raguru Pandu Vasanthi, Rajeev K. Varshney,
and Manish K. Pandey**

Abstract A boost in the crop improvement rate is essential for accomplishing a sustainable food supply and other demands of rapid population growth. Genomic selection (GS), a very promising breeding strategy used effectively in animal breeding, is now used in crop improvement. GS offers a reduced duration of breeding cycles by rapidly selecting better genotypes. Several empirical and simulated research on GS and their implications on agricultural production enhancement

H. V. Veerendrakumar · D. Bomireddy
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India

S. V. Agricultural College, Acharya N.G Ranga Agricultural University (ANGRAU), Tirupati, India

R. Barmukh · P. Shah
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India

Department of Genetics, Osmania University, Hyderabad, India

H. V. R. Jamedar
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India

Agricultural College, Acharya N.G Ranga Agricultural University (ANGRAU), Bapatla, India

M. Roorkiwal
Khalifa Center for Genetic Engineering and Biotechnology, United Arab Emirates University, Al-Ain, United Arab Emirates
e-mail: mroorkiwal@uaeu.ac.ae

R. P. Vasanthi
S. V. Agricultural College, Acharya N.G Ranga Agricultural University (ANGRAU), Tirupati, India

R. K. Varshney
Murdoch's Centre for Crop and Food Innovation, State Agricultural Biotechnology Centre, Food Futures Institute, Murdoch University, Murdoch, WA, Australia
e-mail: rajeev.varshney@murdoch.edu.au

M. K. Pandey (✉)
Center of Excellence in Genomics and Systems Biology (CEGSB), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India
e-mail: Manish.Pandey@icrisat.org

have lately been published. We briefly discuss the GS methodology, its present position, the GS advantages over alternative methods of breeding, commonly used prediction models of GS, and factors interfering with the prediction accuracy of GS to provide a comprehensive grasp of the technology. In addition, the integration of speed breeding and other modern techniques for increasing the effectiveness and speed of GS are discussed.

9.1   Introduction

Plant breeding programme aims to develop genotypes improved for desirable traits to fulfil the requirements of key stakeholders. The breeder needs to explore a large genetic landscape to identify the superior genotypes and the material required to surpass the capacity of breeding programs (Chenu 2015). Plant breeding may be viewed as a number game in which breeding plans are meant to enhance the likelihood of identifying genotypes with acceptable combinations of traits with minimal resources (Brown et al. 2008). The assessment, which includes multiple phenotyping processes aimed to quantify the heritable genetic value of selection genotypes, is a critical component of the breeding scheme (Lynch and Walsh 1998).

For a characteristic such as yield, a set of individuals identified for high heritable traits is often assessed in multi-environmental trials meant to resemble the target population of environments (TPE), where the product is predicted to perform (Chenu 2015). Throughout most breeding programmes, such assessment processes need large amount of resources and span several years (Brown et al. 2008). Several approaches and technologies have arisen over the last three decades to overcome these constraints and boost the efficiency of breeding programmes, due largely to developments in the characterisation of DNA polymorphisms and computational capacity (Xu et al. 2020). Among them, methods that can identify the phenotypic performance by using molecular information (marker-assisted selection) (Cobb et al. 2019) and GS (Meuwissen et al. 2001) are suitable tools that help modern breeding programmes to get maximum outcome from limited resources. In contrast to traditional marker-assisted selection (MAS), current genomic prediction algorithms account for both minor and major QTLs, capturing the large amount of genetic diversity in a trait. The GS original idea was initially proposed in the field of animal breeding by estimating the marker effects for two generations (Meuwissen et al. 2001). Genotyping based on NGS has increased the genomic estimated breeding values (GEBVs) accuracy of prediction over other platforms developed in cereals and also in other crops. In this way, the dream of GS in crop plants has come true. To get the most benefits out of GS, these marker techniques should be used along with high-throughput phenotyping to get higher genetic gain from complex traits.

GS is rapidly becoming the favourable strategy for accelerating breeding by utilising genetic markers. Prediction models for GS are built by regressing observed phenotypes in a training population (TP) on markers that are genotyped on same population (Meuwissen et al. 2001). Best individuals are found in future generations

by utilising these models, which are based purely on the genetic profile. For GS to be successful, genetic markers should be densely and widely dispersed across the genome, with a high probability that every quantitative trait locus (QTL) is really in linkage disequilibrium (LD) with at least one marker (de Resende and de Assis 2010). As the marker number increases, so does the prediction ability of model (Meuwissen and Goddard 2010). As a result, a low-cost, high-density, adaptable, and precise genotyping platform is required for the implementation of GS effectively. Many genome sequencing approaches in the earlier days have recorded a vast number of useful SNPs (single-nucleotide polymorphisms), and to reduce the subsequent challenging problem of high cost incurred characterising those, numerous DNA chip-based SNP genotyping systems are advanced and have become an extremely renowned platform of genetic variation analysis across the genome (Maresso and Broeckel 2008). All of these systems rely largely upon annealing single or several oligonucleotides near a point close to the variable site known before, followed by detection of extension reaction of the attached nucleotide on a chip. DNA chip techniques, including Axiom Affymetrix and Illumina Infinium, provide large-scale, simpler SNP analysis for a large number of loci (Crossa et al. 2013). The emergence of such platforms has laid the foundation for the development of GS as a suitable molecular breeding strategy in species having genotyping tool accessibility.

9.2 Basics of GS

GS is applicable to minor gene governed traits, and its accuracy of prediction is more time efficient than the usual phenotypic selection (Spindel and Iwata 2018). The main advantage of genomic prediction is that it may be used in making decisions for the effective selection of breeding material at different levels (generations) of the breeding program. As a result, efficient genomic prediction integration necessitates more grip over the breeding strategy and its various factors. The breeding strategy is often in the breeder's head, and converting this information into a systematic structure is an important aspect in thoroughly designing various methods (Cobb et al. 2019). Genomic prediction is indeed a continuous long-term commitment requiring breeding programme, and switching to an optimal GS breeding strategy is not always feasible. As a result, the breeding team and specialists must design a transition strategy that outlines precise ways to achieve the objectives (Bartholomé et al. 2021).

Optimal GS strategies are seldom simply evolutions of existing breeding methods. In general, pedigree breeding is used in the majority of conventional breeding systems in self-pollinated plants (Guimarães 2009), but GS is very much suitable to recurrent selection methods relying on superior by superior crosses to improve complex traits. Due to the greater LD among QTL and markers, undetectable or lower population structure, and greater similarity among genotypes, a well-structured breeding programme with clearly defined elite germplasm and a

low population size (Ne 40) seems to be more likely to benefit from prediction using GS (Bartholomé et al. 2021).

Just like with traditional marker-assisted selection (MAS), the GS will be using genome-wide markers for estimating the breeding values of individual genotypes rather than individual markers. The genomically determined breeding values could then be used for effective trait selection. A plethora of algorithms have been developed to model GEBV. In consideration of both dimensions (columns) and instances, modelling GEBVs generates a massive genotype (rows). The right feature combinations can let you figure out which phenotype is being represented. Preparing a solid TP sample is seen to be a practical strategy for dealing with such complicated genetic data (Purbarani et al. 2017). GEBV is measured on a set of beneficial loci present in each BP's genome, and it gives a direct assessment of each individual's likelihood of having a better phenotype (i.e. higher breeding value). Newer parent choices are made using the GEBVs. This shortens the breeding cycle since there is no need to phenotype quantitative parameters like yield and its components in successive generations. The validation population (VP) is a third set of individuals that undergoes genotyping and phenotyping. The GEBVs are calculated for VP, and the correlation between it and the phenotype is used to assess the precision of GS model (Bassi et al. 2015).

The genetic gain expected per unit time from GS is

$$\Delta G = ir\sigma A/T$$

i–intensity of selection,
r–accuracy of selection,
σ_A–the square root value of the additive variance,
T–time required to complete one cycle of breeding (Falconer and Mackay 1996).

Following the first study of Meuwissen et al., the researchers Bernardo and Yu first established the effect of GS on crop breeding (Bernardo and Yu 2007). The researchers utilised a computer-based simulation to show that the use of an entire set of genotyping markers produced better breeding value prediction accuracy than using only a couple of markers that were significantly associated with QTL. Years later, the first genomic-enabled predictions in the actual crop breeding conditions were demonstrated, indicating that strong genomic predictions may be obtained in a variety of corn and wheat data sets (De Los Campos et al. 2009). This study was the first to use pedigree and genetic relationship information to make wheat predictions, which researchers applied in different non-parametric and parametric statistical models. Following these initial discoveries, a significant number of scientific research into forecast accuracy in a different crop species have been conducted and published; some of them are mentioned in Table 9.1.

Higher genetic advance per generation can be attained if the reduction in breeding cycle duration by GS compensates for the drop in selection accuracy, assuming equal selection intensities and genetic variation for both GS and phenotypic selection (PS). Considering assumptions about breeding cycle lengths, selection accuracies,

Table 9.1 GS in different crops

Sl. No.	Crop	Traits	Model	Population	Markers	Reference
1.	Wheat	Plant height (PH), grain yield (GY), heading date (HD), test weight (TW)	RR-BLUP	816 breeding lines	21,643 SNPs	(Xu 2013)
2.	Wheat	Quality traits and agro-morphological traits	G-BLUP	329 genotypes	7748 SNPs	(Ward et al. 2019)
3.	Wheat	Fusarium head blight, Septoria tritici blotch, plant height, heading date	RR-BLUP	1120 winter	5628 SNPs	(Herter et al. 2019)
4.	Wheat	Grain yield, kernel number per spike, kernel weight per spike, thousand kernel weight	RR-BLUP	156 RILs, 239 lines,100 DHs	1188, 5661, and 2780 SNPs	(Lozada et al. 2019)
5.	Wheat	Grain yield	G-BLUP	1100 lines	27,000 SNPs	(Belamkar et al. 2018)
6.	Wheat	Gluten index, protein content, and alveograph	G-BLUP, RR-BLUP, RKHS, BayesA, B, BL	324 lines	9752 SNPs	(Haile et al. 2018)
7.	Buck wheat	Nodes, stem length, flower clusters, primary branches, selection index	G-BLUP	92 lines	14,598–50,000 markers	(Yabe et al. 2018)
8.	Wheat	Fusarium head blight resistance	RR-BLUP	273	19,992 SNPs	(Arruda et al. 2016)
9.	Wheat	Grain yield, flour yield, softness equivalence, fusarium head blight index	RR-BLUP, BL, RF	470 lines	4858 SNPs	(Hoffstetter et al. 2016)
10.	Wheat	Drought and heat stress	G-BLUP	8416 and 2403	40,000 DArTs	(Crossa et al. 2016)
11.	Wheat	Stem rust resistance	BRR and G-BLUP	374 lines	18,653 GBS	(Rutkoski et al. 2013)
12.	Wheat	Quality of grain	BayesCπ, RR-BLUP	174, 209	574 DArTs and 399 various markers	(Heffner et al. 2011)
13.	Maize	Starch, oil content, and protein	G-BLUP	257 inbreds	48,814 SNPs	(Guo et al. 2014)

(continued)

Table 9.1 (continued)

Sl. No.	Crop	Traits	Model	Population	Markers	Reference
14.	Maize	Resistance ear rot	RR-BLUP	238 lines	23,154 DArTs	(Dos Santos et al. 2016)
15.	Maize	Drought stress	G-BLUP	3273 lines	58,731 SNPs	(Zhang et al. 2015)
16.	Maize	Grain yield, silking-anthesis interval, and anthesis date	RKHS and GBLUP	504 DH and 296 inbreds	158,281 and 235,265 SNPs	(Crossa et al. 2013)
17.	Maize	Kernel rows, kernels per row, and ear length	G-BLUP	635, 5 DH populations	16,741 SNPs	(Riedelsheimer et al. 2013)
18.	Maize	10 agro-morphological traits	RR-BLUP, BLUP	441 hybrids and 294 RILs	261 SSRs	(Guo et al. 2013)
19.	Maize	Lignin content, biomass, female flowering, plant height, starch, and sugar	RR-BLUP	285 diverse inbred lines	56,110 SNPs &130 metabolites	(Riedelsheimer et al. 2013)
20.	Rice	Flowering, florets per panicle, protein, height	G-BLUP	405 lines	36,901 SNPs	(Isidro et al. 2015)
21.	Rice	Panicle weight, days to flowering; grain yield, height	BL, G-BLUP, LASSO, RR-BLUP, BRR	343 lines	8336 SNPs	(Grenier et al. 2015)
22.	Rice	Plant height, grain yield, head rice percentage, milling yield, and grain chalkiness percentage	GK and G-BLUP	327 indica and 309 japonica	Japonica: 44598 and indica: 92430 SNP	(Monteverde et al. 2018)
23.	Barley	12 malting traits	RR-BLUP	Winter: 148 Spring: 116–294	Winter:4359 SNPs Spring:4095	(Schmidt et al. 2016)
24.	Barley	Grain protein content, plant height, amylase activity, malt extract, grain yield	BayesA and B, G-BLUP	2 datasets, 140 and150 DH lines	223 RFLP/107 RFLP/AFLP	(Lorenzana and Bernardo 2009)
25.	Barley	Resistance to fusarium head blight	BayesCπ, BL, RR-BLUP	691 lines	3072 SNPs	(Lorenz et al. 2012)
26.	Rye	Protein content and grain yield		219 and 201 lines		

		Traits	Model	Population	Markers	Reference
			Multi-trait RR-BLUP		584 and 394 DArTs	(Schulthess et al. 2016)
27.	Rye	Total pentosan content, grain yield, starch content, and plant height	RR-BLUP	220 (2 sets)	1048 DArTs	(Wang et al. 2014)
28.	Sorghum	Condensed tannins, polyphenols, total antioxidant capacity, flavonoids	BayesB, GBLUP, BL, Bayesian RR	114 genotypes	61,976 SNPs	(Habyarimana and Lopez-Cruz 2019)
29.	Pearl millet	Plant height, grain weight, grain yield, days to flowering	RR-BLUP	37 inbreds, 320 hybrids	306, 15, 463, 32 SNPs	(Liang et al. 2018)
30.	Soybean	Protein, yield, oil	RR-BLUP	483 elite lines	5403 SNPs	(Stewart-Brown et al. 2019)
31.	Soybean	15 amino acids content	RR-BLUP	249 accessions	23,279 SNPs	(Qin et al. 2019)
32.	Groundnut	Total yield/plant, days to 50% flowering, hundred seed weight, days to maturity, rust, late leaf spot, oleic acid, pods/plant, and shelling %	BGLR	340 lines	13,355 SNPs	(Pandey et al. 2020)
33.	Chickpea	Early vigour score, grain yield, seed number, and hundred seed weight	RR-BLUP, BRR, and BL	119 and 13, respectively, Indian and Australian	144,777 SNPs	(Li et al. 2018)
34.	Cowpea	Seed size, flowering time (FT), maturity	SVR, RRBLUP, RKHS	305 RILs	51,128 SNPs	(Olatoye et al. 2019)
35.	n	Seed weight, grain yield, lodging susceptibility, and onset of flowering	BL, G-BLUP, RR-BLUP	306 genotypes	6058 SNPs	(Annicchiarico et al. 2019)

and selection intensities, GS can outperform PS in terms of genetic advance per year (Schaeffer 2006; Shengqiang et al. 2009; Wong and Bernardo 2008). Furthermore, when evaluating traits with a long generation time, GS becomes easier or cheaper than PS, enabling more number of candidates to be characterised for a unit cost, boosting higher selection intensity. As a result, GS is currently utilised in all crops; some of the crops in which GS is used is mentioned in Table 9.1.

9.3   Methodology of GS

The main principle of GS is to estimate breeding values of genotypes under the test purely depending on genotypic data with the usage of statistical models designed on TP (Meuwissen et al. 2001). TP is a group of individuals having a lot of genotypic and the phenotypic data that can be used to get GS model parameters. Following that, using genotypic data, these GS methods are used to calculate breeding values called as GEBV of breeding population (BP) genotypes. A TP is a group of individuals who are linked and have known ancestors, such as half-sibs and closely related groups. The BP is composed of TP descendants or aristocratic lineages with strong ties to the TP. To forecast genetic values of a BP for various traits, allelic similarities with loci connected to the phenotype in the TP is employed. As a result, the amount of genetic resemblance between BP and TP in LD of markers with trait loci determines GS (Edwards et al. 2019).

9.3.1   Designing Training Population (TP)

The overall design of TP is crucial for the success of GS because it adds to higher accuracy of prediction in BP, allowing actual individuals to be selected in present breeding activities (Zhang et al. 2017). The TP should be made up of individuals from single biparental family or germplasm collection accessions. The BP makeup is the most important issue in TP creation; therefore, the BP must always be described first; then, the TP design should be done, which is focused on lowering phenotyping/genotyping costs and improving the accuracy of candidate prediction. First of all, the size of TP, its composition, and its relation with BP will act as key elements to estimate the GS's accuracy of prediction. One of the most challenging aspects to optimise is the selection of individuals to be included in the TP, yet it is critical to achieve excellent accuracy of prediction. Generally, the most accurate GS methods with a TP of high population size are all closely related to BP, and no population structure (Isidro et al. 2015).

Crop breeders must make it a priority to design ad hoc TP for every BP, and the ideal TP will be made up of half-sibs or full-sibs of BP. Maintaining the similarity is crucial around the GS cycles. BP individuals may accrue some genetic variation, and gene frequencies may alter to the point that the TP deviates from BP with each round

of recombination and selection. As a result, the plant breeder must be well prepared to continuously update its TP every cycle (Heffner et al. 2010) or use closed recurrent selection strategies, such as crossing only half-sib or full-sibs. The first option has been extensively researched and is now in use due to the relative ease of execution; nevertheless, as stated previously, it has substantial downsides in selection accuracy. A hybrid strategy integrating these two alternatives was utilised to create the breeding schemes in the second and third rounds of recurrent GS to give better accuracies. Inter-mating full-sibs with half-sibs further ensures a quick progression towards inbreeding as well as the fixing of advantageous genes. Moreover, because the TP stays linked to the BP during closed recurrent selection cycles, multiple morphological scorings for such BP may be accumulated through time and space, boosting the accuracy of prediction (Heffner et al. 2010). The level of LD in TP and BP must be always similar, and it has been shown that more LD results in better predictions (Bassi et al. 2016). Insufficient marker density may lead to an unnatural exaggeration of LD, which, when paired with the near homozygosity of later filial generations (i.e. F6 and later), leads to a considerable reduction in the accuracy of prediction (Shengqiang et al. 2009). As a consequence, if BP and TP are not at all from same parents or have different levels of inbreeding, the density of marker should be raised to account for the larger effective size of the population and rate of recombination (Heffner et al. 2010). However, several TP designs can be feasible for particular breeding conditions, like the following:

1. TP and BP are segregants developed from same cross: genotypes from a biparental cross are chosen for both BP and TP. All genotypes (progenies) in a cross are genotyped, but only a portion of population is phenotyped to act as TP required for deciding a model for genomic prediction; this model is then used to forecast the genetic value of genotypes (BP), which are not phenotyped. Furthermore, the trained model may be utilised to forecast future rounds of selection in populations formed by the intercrossing the family's chosen genotypes (Combs and Bernardo 2013). This type of TP design was widely investigated in several breeding programmes, including huge biparental families and doubled haploids. Numerous investigations on rice (Cui et al. 2020), wheat (Miedaner et al. 2013; Thavamanikumar et al. 2015), rye (Wang et al. 2014), and maize (Zhang et al. 2015) have been conducted. The benefit of within-family genomic predictions would be that a high accuracy of prediction may be obtained with a limited marker number as well as a small population. Because of the large LD found in segregating lines of early hybridisation cycle, improved precision is feasible here (Zhao et al. 2012). This is comparable to the utility of biparental populations QTL mapping. The demerits of this type of TP design include the high cost associated with genotyping a larger number of genotypes in highly segregating generations as well as phenotyping data from replicates and trials at multi-locations, as well as the non-fixation of trait-related alleles in populations, which may impact the identification of an effective GS model.

2. Population lines for TP and BP that include both related as well as unrelated individuals: Prediction models derived from mono biparental populations offer

minimal practical use, except in a few breeding systems. Plant breeders might benefit more from TP schemes that integrate data both from related and unrelated groups. TP must be generated by combining progenies from multiple pedigrees and genetic lineages with varying degrees of relatedness, such as half-sibs, full-sibs, and other people with related heritage. From a large number of studies, we came to know that, when TP is not related to BP, the precision of GS decreases significantly. In hybrid wheat, genomic accuracy of prediction for resistance to disease was substantially higher for sets that are related (0.65–0.92, compared to unrelated sets (0.06–0.43)) (Gowda et al. 2014). The findings also reveal that aggregating a large number of population to forecast a specific target family yields improved prediction accuracies if the TP families have one parent in common with the target population (BP). Families with one parent included, in addition to the family-specific TP, may aid to improve prediction accuracy when solely compared to the family-specific TP, especially for small-sized target families (Schulz-Streeck et al. 2012). Furthermore, it has been proposed that using high-density markers may increase the predictive performance in unrelated families via exchanging marker data among families (Hickey et al. 2014). The essential benefit of TP design is that it would be ideally suited to implementing GS in current breeding strategies, as it comprises both closely related and distantly related genotypes. In general, TPs made up of exclusively unrelated individuals to the BP have very poor to nil accuracy of prediction.

3. Breeding and training genotypes are drawn out from a wide germplasm's: GS predicts the variability of germplasm in addition to measuring the breeding value of individuals from the successful breeding approach. There are large collections of numerous accessions, making it difficult to choose few best genotypes through phenotyping all collections of gene bank. High-throughput genotyping methods have allowed the genotyping of large number of germplasm, enabling the breeding value of germplasm accessions to be predicted (Jarquín et al. 2014). GS has been shown to be effective in harnessing germplasm potential in a variety of agricultural species, including sorghum (Fernandes et al. 2018), wheat (Daetwyler et al. 2014), soybean (Qin et al. 2019), sugar beet (Würschum et al. 2013), and lentil (Haile et al. 2018). Even with low-density genotyping techniques and a well-represented selection population in TP, it was revealed that GS could be used to efficiently unleash the potential of larger germplasms. In spring wheat, 1163 genotypes were phenotyped for resistance against *Puccinia striiformis* that causes stripe rust and then genotyped with a 9 K SNP array, and multiple genomic prediction techniques were tested. The results showed that increasing the marker density and TP size improved the accuracy of prediction. Whenever the size of TP was increased from 210 to 959, the accuracy of prediction increased from 0.50 to 0.63, by an average of 1% improvement for each 50 individuals added towards the size of TP. Beyond an SNP marker density of 1 per 3.2 cM, there was no further improvement in prediction accuracy observed. However, when subpopulations were created based on kinship and structure analysis, prediction accuracy increased. In one subpopulation, it ranged from 0.75 to 0.79, while in another, it was between 0.51 and 0.58. These

variations in prediction accuracy were linked to the genetic relatedness among accessions within each subpopulation, highlighting the importance of genetic relationships between the TP and selection population for germplasm collection decisions (Muleta et al. 2017). Utilising GS, that TP structure may be used to discover promising accessions with higher GEBVs out of huge number of accessions. It is one of several alternative methods for utilising essential gene bank data, as phenotyping the entire collection is a time-consuming process due to a number of practical challenges.

9.4 Statistical Tools and Models Adopted in GS

Many approaches in GS have progressed in recent decades, such as general approaches and their extensions. General GS approaches rely on additive models, but their accuracies might differ due to differences in assumptions and algorithms in regard to complicated trait variations. These general approaches can be enhanced by integrating multiple variates or non-additive effects.

9.4.1 Prediction Methods for Additive Genetic Effects

To generate reliable GEBVs of individuals for selection, GS utilises correlations of a huge number of markers across the entire genome with phenotype. For whole genome regression, though, the marker number (k) is typically much greater than the number of observations (n). Degrees of freedom are inadequate to simultaneously assess the effects of all markers, which is enhanced by multicollinearity (Neves et al. 2012). This ordinary least squares approach will be invalid if k marker effects are evaluated simultaneously. For GS, several approaches, such as GBLUP, machine learning, and Bayes, are being used to overcome these difficulties. The GBLUP and Bayesian techniques consider marker effects to be random effects, and the fundamental model is (Meuwissen et al. 2001)

$$y = X\beta + Z\alpha + \varepsilon$$

where

y–a vector of phenotypes,
β–a vector of non-genetic fixed effects, for which a flat prior is often used,
X–an incidence matrix for the fixed effects β.
α–a vector of random regression coefficients of all the marker effects,
Z–an n × k genotypic matrix for markers.
ε–a vector of residuals.

9.4.1.1 GS Based on a Single Trait

In the literature, many single-trait GS (STGS) approaches are being suggested. We'll go through some of the most common methods in this section.

Linear Regression Model

Our major goal in GS is designing a model to connect between genotypic and phenotypic variables in order to predict GEBVs and choose desirable individuals. Simple linear regression is the simplest model for analysing this relationship

$$Y_i = \mu + \sum_{j=1}^{P} X_{ij}\beta_j + e_i$$

where μ is an intercept,

X_{ij} is the genotype of the ith individual $i = 1$ to n of the jth marker $j = 1$ to p,
β_j is effect of marker,
e_i is i^{th} individual's associated random residuals.
This may be expressed as follows:

$$Y = X\beta + e'$$

where

Y is phenotypic character observation vector,
X is a design matrix,
β is a vector of marker effects that are unknown, and,
e is a random residual.

The advantage of employing a linear model is that it is a relatively basic model with straightforward inference because it is statistically sound. However, when the fundamental criteria of normality, linearity, explanatory variable independence, and $p < n$ are met, this model performs effectively.

Ridge Regression

Multi-collinearity has been discovered among markers. As a result, the individual's GEBV that is estimated becomes inaccurate (Hoerl and Kennard 1970). RR was determined to be a preferable option in this case. RR obtains a smaller variance estimation of b, but as a price to pay estimator, it becomes biased. One more benefit of RR is that it could be utilised in cases where $p > n$ problem. Instead of reducing the sum of the squared residuals, just like in linear regression models, RR minimises $(\mu, \beta) = (Y-X\beta)' (Y-X\beta) + \lambda\beta' \beta$, where $k = 0$ is a regularisation parameter that

controls the intensity of the penalty. The higher the value of k, the more shrinkage variables there are. In this scenario, the estimate of regression coefficients, or marker effects, could be provided by

$$\beta = \left(X^{'} X + \lambda I\right) X^{'} Y$$

Best Linear Unbiased Prediction

Henderson for the first time presented the theory of mixed random effect model, and that has been widely utilised in conventional animal breeding programmes earlier (Henderson et al. 1959). In the situation of imbalanced data, BLUP is commonly utilised. Because of its versatility and adaptability, BLUP is used not just in animal breeding programs but also in crop improvement (Bernardo 1994). BLUP has the added benefit of being able to accommodate any family information. Hayes and Goddard were the first to use BLUP in GS (Meuwissen et al. 2001). The model is as follows:

$$Y = X\beta + Zm + e^{'}$$

where β is a $p \times 1$ fixed effect vector that needs to be calculated and m is a random effect that is also a parameter of interest, that is, $m \sim N(0, G)$ and $e \sim N(0, R)$. BLUE is the fixed effect β estimator, whereas BLUP is the random effect m estimator. The main downside of BLUP is that it necessitates complicated statistical calculations, which demands enough computational resources. However, as computer power improves, this problem becomes limited.

Least Absolute Shrinkage and Selection Operator (LASSO)

RR may utilise LASSO for GS to overcome the restrictions of linear regression. This is an RR variation that is generated by changing the penalty function in RR, i.e., assigning a linear penalty rather than quadratic penalty. As a result of this, the influence of certain least important markers is reduced, and the effects of the less important markers are set to zero, thereby solving the $p > n$ problem. Tibshirani was the first to develop LASSO. It is written as

$$\beta \text{ (lasso)} = \text{argmin} (Y - X\beta)^{'} (Y - X\beta) + \lambda\beta$$

Bayesian Methods

In a Bayesian method, we must assume a prior distribution of such model's parameters based on previous knowledge and experience. The previous distributions of such model variables are combined well with likelihood function to get the posterior distribution. The posterior distribution is used to get model parameter inferences. Hayes and Goddard (2001) applied the Bayesian technique for the first time in GS, utilising the previously mentioned linear regression model. The parameters of the model are computed using the prior and posterior distributions of the variables. Bayesian approaches include BayesA, BayesB, BayesCp, and BayesDp. The assumptions of such prior distribution of model parameters, model variance, and other factors differ across different approaches.

BayesA implies that the variance of every marker site has the same prior, whereas BayesB considers that the location of all markers doesn't contribute to overall genetic variation. In comparison to BayesA, the BayesB technique is more realistic for GS. Other Bayes variants have been created to address the shortcomings of BayesA and BayesB. BayesC assumes and utilises the same variance for all markers, whereas BayesD, a scaled parameter, is computed rather than given by the user. BayesCp and BayesDp are derived from BayesC and BayesD, respectively, with the probability p determined for low-effect SNPs.

Support Vector Machine (SVM)

The methods given in this chapter are parametric in nature. These approaches always need the data to be subjected to a number of assumptions. However, parametric model assumptions do not always hold. Parametric approaches perform poorly in such situations. In this scenario, nonparametric approaches may perform better. They believe that the response and predictors have an unknown connection. In GS, nonparametric approaches, such as neural network (NN), SVM, and RKHS (reproducing Kernel Hilbert space), have been employed (Budhlakoti et al. 2019). SVM is a machine learning approach. Supervised learning was the principle on which it was based. It creates a separating hyperplane with the goal of classifying data into distinct categories. It is based on the maximum separation hyperplane idea. The SVM approach is used extensively in support vector regression.

9.4.1.2   Multi-Trait-Based GS

The models outlined before are based on single attribute's genetic information. However, we now have access to data on various traits. We lose lot of details related to association among many traits if we employ approaches based on single factors, since they fit the method by evaluating each attribute separately. Multivariate-based approaches have been developed to use this kind of info in the model. A number of multivariate regression-based methods have been developed. Multivariate methods

are indeed extensions of basic regression models, in which users regress two or more responses (q > 1) upon p predictors instead of one response onto p predictors. Consider a straightforward multivariate regression model.

$$Y = X\beta + e$$

Here,

X is the marker's n × p matrix,
Y is the n × q matrix,
β is the dimensional vector p of coefficients of regression, and,
e is the random residual.

Multivariate Regression with Covariance Estimation

By applying a LASSO-like penalty to β and Ω while accounting for correlated errors, this method minimises the parameter number that are to be estimated (Rothman et al. 2010). Based on $\Omega = [\omega_{j\cdot j}]$ and $\sum_{jk} |b_{jkl}|$, the LASSO constraint (Tibshirani 1996) on the entries of b, two penalties are employed to construct a sparse estimator of b. The following is the form in which the function can be written

$$(\beta^\wedge, \Omega^\wedge) = \min_{\beta, \Omega} \left\{ f(\beta, \Omega) + \lambda_1 \sum_{i \neq j} |\omega_{j\cdot j}| \; \lambda_2 \sum_{j=1}^{p} \sum_{k=1}^{q} |b_{jk}| \right\}$$

where λ1 ≥ 0 and λ2 ≥ 0 are tuning parameters.

We impose a related penalty for off-diagonal entries of the inverse error covariance matrix O like in the case of LASSO because (1) it makes sure a solution to the issue of (q > n), i.e. more response than that of the total count of sample population, and (2) it reduces the amount of parameters in O and has been found to be effective whenever the number of response parameters is greater.

Multivariate Mixed-Model-Based Approach

This is simply a multivariate version of the univariate mixed model technique. Multivariate BLUP is another name for it. Because a mixed model incorporates random along with fixed variables, it employs the covariance structure for random effects in the multivariate situation. It employs the very same model as that of the mixed-model-based method, namely, BLUP, with the exception that Y is a response matrix rather than a vector.

Conditional Gaussian Graphical Models

cGGMs are multivariate linear regression models that have been reparametrised (Chiquet et al. 2016). It makes use of the predictor-response variable covariance structure. Using multivariate regression parameterisation, many statistical formulations may be generated. Because partial covariance reflects crucial relationships among variables, these models are valuable. Multivariate penalised techniques, like univariate penalised methods, resemble regularisation. By applying penalty-based approaches, the multivariate frame takes use of the sparsity within the predictors. Partial Gaussian graphical model is another name for it.

9.5 Factors Influencing GS Predictions

The use of GS in ordinary crop breeding programmes is dependent on the precision of prediction; hence, it is crucial to use cross-validation to check a trained model for high accuracy of prediction. A second genotyped and also phenotyped population, termed as the validation population (VP) or BP, is employed for this. Based on TP phenotypic data, marker effects throughout the particular genome are computed. To measure the model's accuracy of prediction, comparison of GEBVs is calculated from VP with its true breeding values (TBV). A trained model of TP and VP genotypic data is used to construct the GEBV for VP. Correlation between TBV and calculated GEVBs based on VP's phenotypic data is being used to assess the validity of the GS model. Cross-validation is commonly used in TP to train and design the optimal model of prediction that can be used to calculate BP's GEBV (Pérez-Cabal et al. 2012). The Pearson product moment correlation among calculated TBVs and GEBVs is being used to determine predictive performance. GS accuracy is influenced by below mentioned variables:

- Genetic relatedness (Duangjit et al. 2016; Endelman et al. 2014).
- Efficient population size (Poland and Rife 2012; Zhao et al. 2012).
- Structure of population (Isidro et al. 2015).
- Genetic inheritance of the trait (Heffner et al. 2009; Ornella et al. 2012).
- DNA marker characteristic and density (Zhao et al. 2012).
- The distribution and level of LD between markers and genomic sequences correlated with the desired trait (Rajsic et al. 2016).
- Statistical models that are frequently used to calibrate any best-fitted model (Heslot et al. 2012).
- Gene effects (Akdemir 2013).
- Genotype environment interactions (Rajsic et al. 2016).

An investigation of corn doubled haploid lines found that when full-sibs were replaced with half-sibs, accuracy of prediction dropped by 42% (Riedelsheimer et al. 2013). To increase accuracy of genomic prediction, BP and TP must be close enough to exchange long-range haplotypes (Lorenz and Smith 2015). The size of TP is also a

significant component in achieving improved prediction accuracy. More genotypic and phenotypic data might be accessible for precise assessment of the genetic components impacting the characteristics' expression, which would enhance the accuracy of genetic effect estimation (T. Guo et al. 2013), especially for low inherited characteristics (Lian et al. 2014). Since GS intends to eliminate the necessity of phenotyping as well as the costs associated with it, determining the ideal TP size to produce relevant accuracy of prediction is critical when using it in a breeding programme. Raising TP size, boosted accuracy of prediction for many characters, these prediction accuracies plateaued around 700 lines (Cericola et al. 2017). Likewise, many studies with respect to the effect of size of TP on genomic predictions have been published (Liu et al. 2018).

Based on one consensus finding, the optimal TP size required for improved prediction accuracy ranged from less than one hundred full-sibs or thousands of half-sibs of a BP to hundreds of unrelated individuals. As a consequence, enhancing relatedness of TP by incorporating populations that are strongly related could be preferred to increasing size of TP through including distantly associated groups (Brandariz and Bernardo 2019). However, in long-term GS, the usage of strongly related individuals may cause the response of selection to be delayed (Moeinizade et al. 2019). As a result, the TP-BP interaction must be carefully considered in order to achieve better genetic gains and effective GS implementation in breeding operations. Another major element controlling prediction accuracy in GS investigations is trait heritability. The typical GS model's prediction accuracy has a positive relationship with heritability (Hayes et al. 2009). Previous research has shown that as trait heritability rises, accuracy of prediction improves. For example, under wet and water-stressed situations in maize, heritability of grain yield, a complex trait, was significantly less than less complex characteristics like days to anthesis and height of the plant.

Furthermore, for each of the three attributes, heritability mean values during watered situations were greater than those under water-constrained situations in the investigated populations. In watered situations, heritability mean values for plant height, anthesis date, and grain yield were 0.59, 0.55, and 0.38, respectively, in comparison with 0.37, 0.47, and 0.27 in stressed conditions (Zhang et al. 2017). In peanut, though, it has recently been established that there's no correlation between heredity and trait accuracy of prediction in genomic investigations (Pandey et al. 2020). Among eleven agronomic traits included for the GS research, plant height had the maximum heritability (92.3%), while the main branches/plant had the lowest (78.7%). However, employing four GS models (Pandey et al. 2020), the accuracies of prediction of these two characteristics were 0.56 and 0.64, respectively, indicating that there is no link between the heritability of a characteristic and its accuracy of prediction. GS for sucrose solvent retention (heritability: 0.45) and flour amount of protein (heritability: 0.56) in wheat found that the accuracy of prediction for the two traits was 0.74 and 0.64, respectively. As a result, GS might be a useful technique for accelerating genetic gains, especially for traits with low heritability.

Another essential factor that influences the accuracy of prediction is density of markers that varies with the type of population, trait of interest, and plant species. A

few markers are needed for biparental populations and self-pollinated crops, and more markers are needed for crops that undergo cross-pollinated and also for natural populations (Juliana et al. 2019; Liu et al. 2018). In theory, genome-wide high-density markers assure nearly perfect LD among at least one marker with each QTL, leading to greater accuracy of prediction, but in fact, once the optimum density (markers) is reached, there is no further genetic gain in accuracy of prediction significantly (Wang et al. 2018). To obtain same levels of accuracy of prediction, GS in such a bi-parental population uses few markers (hundreds) in comparison with vast number of markers in a multiple family population (Crossa et al. 2014; Technow et al. 2012). In a research on quality of grain in bi-parental populations of wheat, GS accuracy of prediction plateaued at a lower density of markers (Hoffstetter et al. 2016), and a similar trend was reported in predicting the performance of rice hybrids (Wang et al. 2017). According to study, validated functional markers might potentially be utilised as fixed effects in the model to increase accuracy of prediction (Xu et al. 2020).

Furthermore, factual investigations have demonstrated that the statistical models used have an impact on the accuracy of GS prediction (Daetwyler et al. 2010; Resende et al. 2012). In comparison to the GS models, which examine the genetic relationship matrices among individuals of BP and TP, simpler models with the hypothesis that no link exists across genotypes result in lower estimates of genetic variation (Cericola et al. 2017). RR-BLUP, BayesA, GBLUP, RKHS, Bayesian LASSO, BayesC, and BayesB were used to examine the genomic prediction of six maize variables (yield per plant, 100-kernel weight, plant height, ear height, ear diameter, and ear length). For complex characteristics with lower heritability, additive-dominance and RKHS models demonstrated improved accuracy of prediction (Liu et al. 2018). Accurate phenotyping of TP for model construction could also improve the accuracy of prediction (Voss-Fels et al. 2019). As a result, precise phenotyping is critical for training the prediction model while implementing GS, since the risk of GEI is higher in plants than in cattle (Jonas and De Koning 2013).

The cost-benefit ratio of generating a new variety in crop improvement programmes is determined by factors like phenotyping and genotyping expenses, which are determined by the nature and heritability of traits, as well as the size of the TP and BP. The GS method was found to be advantageous for variables with less than 0.1% heritability when the size of TP was >400 and the effective chromosomal segments (Me) was >100 (Rajsic et al. 2016). GS also offers economic benefits for characteristics with heritability less than 0.25 and effective chromosomal segments less than 100, as far as phenotypic expenses per genotype remain less than the expenditures for genotyping. For example, the break-even cost ratio for resistance to common beans bacterial blight (heritability = 0.24) was significantly lower than that of maize tryptophan and lysine content (heritability = 0.96) that showed the heritability consistency. As a result, if phenotyping is less expensive or not complex, traditional selection may be highly cost-efficient for breeding tryptophan and lysine concentration in maize. Furthermore, when assessing total performance, GS was determined to be financially efficient (Rajsic et al. 2016). Same patterns in cost

efficiency of GS were seen in other investigations (Heffner et al. 2010; Wong and Bernardo 2008).

When utilising GS in breeding, the heritability of characteristics and the size of TP must be considered. One of the current obstacles to the effectiveness of GS is the massive price of genotyping. Breeders' interest is in using GS on fixed materials of later generations, such as preliminary yield trials, although the genetic gain is not as significant as with PS. As a result, breeders have been slow to include GS into ordinary varietal development efforts, particularly in research that is government-funded. When we include unrelated families in the TPs, the expense rises even more. A large-sized population of ~20,000 requires nearly 10,000 markers to achieve a higher prediction accuracy of 0.7 (Hickey et al. 2014). Yet, GS's financial efficiency can be improved if genotyping expenses continue to decline and predictions made in one generation are utilised to influence decisions in the following generations (Rajsic et al. 2016).

Moreover, having an open-source GS breeding network, wherein high-throughput systems and infrastructure—including high-throughput phenotyping (HTP), high-throughput genotyping (HTG), and effective models for prediction—are built and allowed to share among scientific organisations and private industries, will lead to significant savings and facilitate the most GS applications in crop improvement programmes.

9.6   Part Strategy of GS

9.6.1   Two-Part Strategy

The aims of developing inbred lines are (i) new inbred's identification and (ii) parent's identification for subsequent breeding cycles. Traditional breeding programmes would be reorganised into two different components under this two-part strategy: a product development (PD) component to produce as well as evaluate inbred lines and a population improvement (PI) component to boost the number of favourable alleles via rapid recurrent genomic selection. Most breeding strategies that produce inbred lines involve crossing to generate new germplasm and thereafter selfing to derive new inbred lines (Bernardo 2014). Instead, doubled haploid technique can be utilised to generate inbred lines quickly (Forster et al. 2007). These newly obtained inbred lines are then phenotyped for one or even more cycles prior to the final selection in order to achieve one or both of the previously mentioned goals of product development and germplasm enhancement. Within this context, genomic selection may be used to find promising lines earlier, lowering cycle time and boosting genetic gain per generation (Heffner et al. 2009). The practice of using inbred lines as parents might be avoided totally with the introduction of GS (Heffner et al. 2009). Strategies based on this concept have been detailed for crops that are simple to cross (Bernardo 2009; Bernardo and Yu 2007) as well as for those which are difficult to cross due to the self-pollination behaviour (Bernardo

2010). The two-part technique is an extension of previous methods and attempts to optimise the capacity of genomic selection over an entire breeding programme (Gaynor et al. 2017). The population improvement component of the two-part strategy employs fast recurrent selection via GS. The idea is to shorten the breeding cycle time in order to enhance genetic gains per year. Each phase of population improvement starts with a large number of genetically different plants. These plants are genotyped, and the finest ones are involved in intercrossing to generate a new generation. The procedure is repeated. Thus, in two-part strategy, population improvement is just a recurrent genomic selection scheme. To assure a consistent supply of enhanced germplasm, a part of the seed generated in few or all cycles is passed towards the product development component. The product development aspect of the two-part approach is primarily dedicated on producing inbreds to release as inbred varieties or hybrid parents. This component's structure is similar to existing breeding programmes and may thus be designed flexibly to fit current or newer breeding program. This design flexibility in the product development aspect also enables different ways to implement at GS. The fundamental distinction between the two-part strategy's product development aspect and conventional breeding programmes is that lines are not selected for subsequent breeding cycles, as this is managed by population improvement component. Furthermore, certain phenotyped plants must be genotyped as part of the product development component. This is required for revising the GS training population utilised in population improvement component aspect as well as for using GS in product development component. The product development component aspect of the two-part approach leads population improvement component by facilitating the development and revision of the training set. The two-part breeding method produced the largest genetic gain, and all genomic selection techniques generated genetic gain greater than conventional breeding.

9.6.2 Multi-Part Strategy

The multi-part approach is an expansion of the two-part strategy, in which exotic germplasm is brought into the PI component via pre-breeding bridges. The pre-breeding bridges take up the role of breaking the linkage between favourable and unfavourable alleles to lower the performance gap between elite germplasm and exPVP. Introgression into the PI component began a year after the last bridge was established. Individuals from the previous year's final breeding cycle were randomly separated into males and females. Individuals from the exPVP breeding programmes or a prior bridge were introduced into the female population when germplasm exchange occurred. Germplasm was pushed back from the PI component to improve pre-breeding performance, with the goal of decreasing the performance gap between exPVP and elite germplasm. Individuals that had been excluded from the PI component were returned into the male population. The number of crossings made between males and females for every bridging cycle was maximised. GS was used

to pick the best individuals for the following breeding cycle. When germplasm exchange occurred, outstanding individuals were advanced to the next stage of the breeding programme. After all bridges were constructed, the PI component was introduced. Each cycle of the PI component began with a large number of lines under the genotyping constraints. These individuals were evenly and randomly divided into males and females. When no germplasm exchange occurred, 50 males and 50 females were chosen. When germplasm exchange occurred, 50 females and 50 males were chosen, minus the number of males to be introgressed. The results of the multi-part strategy imply that it has the ability to improve quantitative (polygenic) traits while also giving a tool for avoiding suboptimal convergence of long-term genetic gain (Breider et al. 2022).

9.7   Advantage of GS over Other Breeding Methods Using MAS

In the last century, classical plant breeding together with agronomic practices has developed several high-yielding nutrient responsive varieties and achieved notable gains in terms of production and productivity. However, with diversified food consumption patterns and to address demands of the ever-increasing population, there is a need for increase in crop improvement (Krishnappa et al. 2021). Food and nutritional security in verge of the rapid population growth can be attained by increased yield potential and reduced crop yield gaps. For major food crops, 0.8–1.2% of the current crop improvement rate is short of 2.4% required for meeting food demands of the 10 billion population predicted over the next 30 years (Hickey et al. 2019; Ray et al. 2013). Globally, plant scientists are striving hard to achieve this required genetic improvement rate by developing high yielding, more nutritious, biotic, and abiotic resistant crops, peculiarly during the degrading land and water resources due to the scenario climate change. This indeed is a challenging task to the plant scientists to develop climate smart crops.

Most of the agriculturally important grain yield, stress adaptive and plant growth traits are controlled by quantitative and genes with minor effects leading to higher epistatic interactions (Mackay 2003). Conventional breeding methods are less precise and reliable for multi-genic, low heritable traits as they are highly influenced by environmental cues and G × E interactions, which make their improvement difficult. Besides, these methods require large land and maintenance of large breeding populations, which are often laborious, time-consuming, and cost-prohibitive. It also requires hybridisation of genetically distinct parents and continuous PS over successive generations, which indeed require 5–12 years for a variety development. This necessitated the development of efficient and rapid selection methods to address the restrictions of traditional breeding methods, as shown in Fig. 9.1.

Extensive use of DNA markers over the last two decades has enabled the usage of MAS in crop improvement programmes as it requires minimum phenotypic information for the indirect selection of traits of interest (Collard and Mackill 2008). In MAS, plants with desirable alleles are selected by using the markers related to the
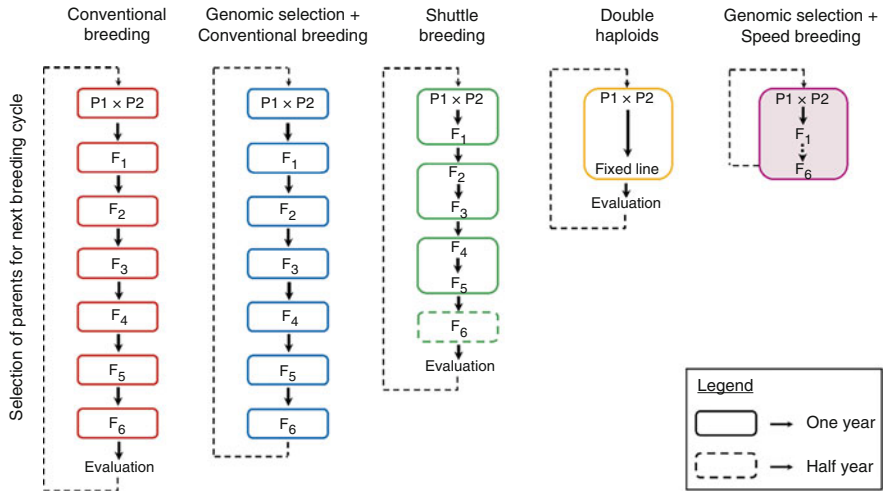
Fig. 9.1 Depicts comparison of breeding cycle length for conventional breeding versus modern strategies that exploit GS, shuttle breeding, double haploids, and speed breeding

desired trait, which makes it efficient only for the traits governed by few major effect QTLs. For the complex traits regulated by several minor QTLs, this approach is inferior to the traditional phenotypic selection as the effect of QTL estimation is biased through linkage or association mapping (Zhao et al. 2014). Initially, marker-assisted backcross breeding (MABB) was used for introgressing one or few genes (from donor) with large genetic effect into the background of adapted cultivars (recipient) with the main aim of recovering the genome content of the recipient. Later on, as emphasis was made only on foreground selection, marker-assisted recurrent selection (MARS) turned out as an alternative strategy for integrating multiple favourable alleles within the same population from different sources (Rai et al. 2018). MAS and MARS are being used for stacking multiple genes (gene pyramiding) into widely adapted cultivars to correct the drawbacks associated with it or for introgressing novel genes (Rana et al. 2019). As of today, the success of gene pyramiding has been shown mainly for major effect genes and is found inefficient for complex agronomic traits governed by multiple genes with minor (very few) effects. These strategies are also constrained mainly because of the usage of low-marker density systems. But, advancement in genome sequencing technologies has made GS a powerful selection tool to tackle these complex traits. In GS, an individual's genetic worth is estimated based on the information from a vast set of markers distributed across the genome instead of a few markers as in MAS. By integrating phenotypic and genotypic data of TP, GS develops a prediction model to derive GEBVs of all the base population individuals that substantially increases the selection efficiency (Poland and Rife 2012). Thus, GEBVs help in selecting the better performing individuals that can be used either as parents in hybridisation programmes or for generation advancement. Based on realistic assumptions of selection intensities, selection accuracies, and generation time, GS has a comparative advantage over conventional PS and MAS in enhancing the genetic gain per year

(Crossa et al. 2017; Heffner et al. 2010). Furthermore, GS is cheaper and easier to evaluate difficult traits like insect resistance, where more individuals can be assessed in a given time and cost (Bhat et al. 2016). Hence, GS enables faster development of crop varieties than PS with increased genetic gains, selection intensity, efficiency, and reduced duration, hence saving resources and time (Desta and Ortiz 2014).

Major consideration for applying GS in crop plants is its cost-benefit ratio. Next-generation sequencing (NGS) technologies have provided accessibility to cost-effective, high-throughput genome-wide markers for model and non-model crops, which made implementation of GS successful mainly in large breeding populations (Poland and Rife 2012). In many crop species, GEBV estimates through NGS-based genotyping are more accurate than other established genotyping platforms. GS is efficient in capturing many additive and small effect loci, which might not be covered by MAS (Heffner et al. 2009). Deploying genome-wide dense markers, GS increases the precision of detecting the markers in LD with desired QTLs of interest (Bekele et al. 2014). Upgradation in high-throughput phenotyping techniques and continuous downswing in sequencing cost will enable GS in accelerating the genetic gains from complex traits required for crop improvement in the near future.

9.8   Limitations of GS

Despite all the positive sides of GS, there are some constraints in situations where the GS is not so good at performance. Let us know the things where this GS performs poorly.

1. To assess the efficiency of genetic prediction, accuracy alone will not be sufficient. Almost the majority of the studies assessed genetic selection using prediction accuracy. Even though accuracy is a crucial factor in establishing prediction model efficacy, it never shows which are all the individuals selected in good faith by using various methodologies. The realised selection differential is arguably a superior criterion for comparing different genomic prediction systems, since breeders examine numerous traits together to advance material, making trait evaluation individually less relevant. Finally, it was correctly pointed out that the phenotype is the final predictor of the genuine breeding value and, like a GEBV, does have an error variance (Bassi et al. 2016).

2. The accuracy of within-family prediction is not fully considered. There has been no comprehensive investigation of prediction accuracy of within-family utilising numerous biparental families or parental information as the training set. Indeed, with the exception of research involving a single biparental household, findings on within-family accuracy are sparse. This is also evident in the hybrid literature, where most studies concentrate on predicting particular hybrid combinations instead of assessing general combining ability in a group of new males or females. This is an essential consideration when adopting genomic prediction

because improved within-family accuracy may help to accelerate genetic advancement while optimising the fraction of inbreeding within the population. Differences among crosses are well predicted since the model accounts for both within and between family variables (Edwards et al. 2019).

9.9 Speed GS High-Throughput Genotyping

In order to meet the growing demand for food, it is required to double the crop production by 2050. Continuous advancement in genotyping and phenotyping technology offers great potential for enhancing genetic gain in crop plants (Phillips 2010) and improves the efficiency of breeders. In GS, to train a prediction model, the phenotypes are utilised. Through high-throughput phenotyping, a large number of lines can be phenotypes more rapidly and accurately, and the precise selection of the best progeny can be made. Similarly, advancement in the high-throughput genotyping have led to the generation of valuable genomic information in a fasten and cost-effective way that has eased the development and study of the large mapping population (McMullen et al. 2009).

Speed breeding in completely contained, controlled environment like growth chambers can speed up the development of superior varieties, such as phenotyping mature plant features, mutant research, and transformation. Additional illumination in a greenhouse leads to faster generation advancement through single seed descent (SSD) and adaptability to plant growth activities at a larger scale. Speed breeding significantly reduces generation time and speeds up the crop breeding process. Durum wheat, Spring wheat, barley, pea, and chickpea could now produce nearly up to six generations per year, especially in comparison to 2–3 under regular glasshouse situations, and rapeseed (*Brassica napus*) could now produce nearly four generations per year (Watson et al. 2018) (Fig. 9.1).

Selection by breeding (SB) helps to resolve difficulties related to the double haploid (DH) technique, like poor germination percentage, reduced vigour, and deformed development (Ferrie 2006). Recombinant inbred lines (RILs) formed during numerous generations of autogamy may be preferred to DH for genetic mapping purposes because of the many meiotic events that take place all through repetitive fertilisation as well as the associated increased recombination frequency. Likewise, SSD can create and assess segregating generations quickly under SB circumstances (Sinha et al. 2021), saving a lot of time over the conventional pedigree breeding approach (Jähne et al. 2020). SB methodologies may be used in many crops and integrated with other current techniques of crop breeding like high-throughput genotyping, genome editing, and GS to accelerate crop development. In orphan crops, SB approaches might be utilised to cut down the breeding cycles and to hasten research. High-throughput phenotyping is one of the major advances in agricultural research in the twenty-first century that has the ability to overcome long-standing barriers to crop improvement advancement. Carrying out high-throughput phenotyping under SB circumstances opens new possibilities for discovering and

incorporating favourable features while conserving resources (Al-Tamimi et al. 2016). Because of their precision and convenience of use, high-throughput phenotyping platforms (HTPPs) have sparked a lot of interest (Furbank and Tester 2011). The HTPPs feature completely automated facilities in greenhouses with regulated environmental conditions, as well as remote sensing technology that allows for exact assessment of crop growth and performance. Recently, efforts have been made to develop low-cost HTPP approaches to widen its adoption in breeding programs. Under SB circumstances that include increased planting density, temperature control, and prolonged photoperiod, targeting proxy variables like seminal root number and angle of seedlings permitted quick selection for superior root architecture of mature plants (Richard et al. 2015). SB has been used to assess different stages of plant breeding operations. In spring wheat, GS was paired with SB to maximise genetic gain for complex traits (Voss-Fels et al. 2019). SB was employed for certain trait phenotyping of carried in TP of wheat, as well as the development and phenotyping of selection candidates. Indirect selection in SB settings with targeted populations was found to predict plant height and blooming time with accuracy comparable to direct field selection. In comparison to field phenotyping directly, SB allows for a genetic gain at a higher rate (Watson et al. 2018).

Similarly, imaging technology allowed for the acquisition of field plot photos at a rate of 7400 plots per hour based on wheat colour features (Walter et al. 2019). When compared to terrestrial sensing, the technology using automated aerial vehicles demonstrated a substantial correlation with enhanced grain output. For phenotyping plant characteristics, several sensing methods have been identified, including proximal (remote) sensing and imaging, laboratory studies of samples, and near-infrared reflectance spectroscopy (NIRS) analysis with the harvestable portion of the crop. The approach chosen is determined by the kind of trait as well as the period of evaluation. Remote sensing techniques could be used to conduct in situ screening for a broad range of breeding objectives, such as yield potential, tolerance to abiotic and biotic limiting circumstances, and even quality-related traits. Several different sorts of properties, ranging from green biomass to photosynthetic transpirative gas exchange, quality features, and even grain production prediction, may be measured using remote sensing techniques in diverse environmental conditions (Weber et al. 2012).

NIRS is generally utilised in breeding for a wide range of feed and food quality parameters. In fact, NIRS could be used to evaluate for drought tolerance, nutrient efficiency, and other breeding/gene discovery objectives. Because of the employment of proximal sensing with VIS-NIR and far-infrared light in imaging formats, the measurement process has been upscaled: for example, from analysing a single plot towards dissecting an entire trial made up of several plots, providing the picture has an adequate resolution (pixels). Additionally, the aerial HTPPs have enabled to take measurements of all plots simultaneously in the trial, and this has made the phenotyping process to overcome the largest limitation, time, and allows rapid characterisation of several plots within a short duration. The use of these approaches in the GS would provide an overall understanding of the role of genetic factors in
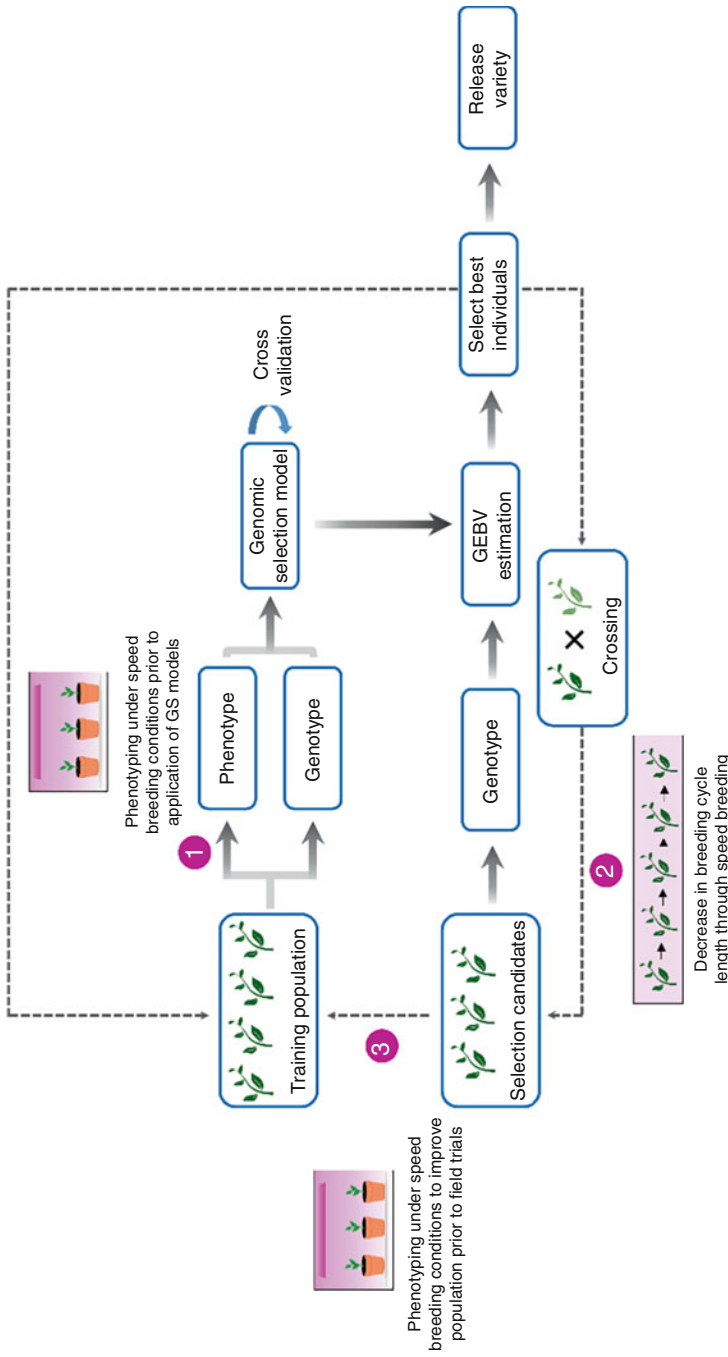
Fig. 9.2 Represents crop breeding cycle employing GS and points within the cycle wherein speed breeding can be applied to increase the rate of genetic gain

determining crop performance, the stages where SB can be efficiently applied in GS are mentioned in Fig. 9.2.

The high-throughput genotyping platforms (HTGP) and high-throughput phenotyping platforms (HTPP) allow effective GS and enhance the success rate of breeding programs (Cobb et al. 2019). The ease in availability of high-throughput genotyping platforms, such as high-density, mid-density, and low-density platforms, has increased the precision and accelerated the genetic gain in crops with large and complex genetic makeup. To estimate GEBVs of (BP), phenotyping and genotyping are very critical for the identification of the appropriate gene and GS model to be utilised. Hence, the combination of these high-throughput technologies with appropriate genetic diversity and analytical tools, along with databases, would lead to the new variety development having better yield, quality, and resistance to stresses.

Over the last decade, the dependence on phenotypic selection has gradually shifted to a greater reliance upon genotypic-based approaches for plant selection, enabled feasible in part by NGS-based sequencing platforms (Bhat et al. 2020; Pandey et al. 2016). The advancements in the HTPP and HTGP have been successful in enhancing the accuracy of genomic prediction and mapping of genes. Platforms such as AgriSeq, DartTag, and RiCa are widely used in the GS. The HTGP technology has improved genome-wide genotyping throughput, cost-effectiveness, and speed (Getachew et al. 2020). Before the advent of NGS-based marker genotyping, the generation of markers was costly and time-consuming, especially for the GS, which was limited by the number of markers that could be tested economically (Bhat et al. 2016). As a result, to forecast the presence or absence of agriculturally beneficial characteristics, only markers in crucial genomic regions were used (Varshney et al. 2014). The HTGP has enabled cost and time effective genotyping with precise identification of the desirable genotypes. Hence, it enables the development of markers, which further caters the GS.

9.10 Conclusion

Plant breeders can use GS to forecast genomic-estimated breeding values of individuals by employing markers that span the whole genome. However, the best way to implement GS is still up for dispute. Predictions within the breeding cycle in the breeding programme can provide high selection accuracies, but selections across the breeding cycle might suffer from a poor association between the training and test populations, making predictions less accurate. More research on predicting distantly related individuals is required. Due to the lack of precision in prediction ahead of several breeding cycles, lower accuracies can be predicted when GS is paired with the usage of untested parents. The best answer for using GS in plant breeding initiatives can be a mix of diverse approaches. Pedigree information might be used in GS to improve forecast accuracy and provide breeding values for non-genotyped lines. The characteristics and relationships of individuals influence the size of the TP and marker set, which should be examined separately before implementing GS in a

breeding programme. GS is commonly used to estimate individual additive genetic value while ignoring non-additive genetic variation, which reflects the performing power of a line as a parent. Future research will look at the assessment of total genetic value, which would be ideal for variety marketing. We can conclude that within generation, GS is now an attractive and realistic alternative, with expenditures in genotyping being recovered through improved selection decisions, reduced phenotyping, and a reduction in the number of candidates retained in the breeding programme. Because forecast accuracies in such systems may be poor, cross-breeding cycle GS, and in particular the use of untested parents, has to be examined further. We also come to the conclusion that plant breeders may benefit more from employing pedigree data, as well as combined pedigree-genomic data than they really do now.

References

Akdemir D (2013) Locally epistatic genomic relationship matrices for genomic association, prediction and selection

Al-Tamimi N, Brien C, Oakey H, Berger B, Saade S, Ho YS, Schmöckel SM, Tester M, Negraõ S (2016) Salinity tolerance loci revealed in rice using high-throughput non-invasive phenotyping. Nat Commun 7(1):1–11. https://doi.org/10.1038/ncomms13342

Annicchiarico P, Nazzicari N, Pecetti L, Romani M, Russi L (2019) Pea genomic selection for Italian environments. BMC Genomics 20(1). https://doi.org/10.1186/S12864-019-5920-X

Arruda MP, Lipka AE, Brown PJ, Krill AM, Thurber C, Brown-Guedira G, Dong Y, Foresman BJ, Kolb FL (2016) Comparing genomic selection and marker-assisted selection for fusarium head blight resistance in wheat (Triticum aestivum L.). Mol Breed 36(7). https://doi.org/10.1007/S11032-016-0508-5

Bartholomé J, Parthiban TP, Cobb JN (2021) Genomic prediction : progress and perspectives for rice improvement. Methods Mol Biol 2467:569–617

Bassi FM, Bentley AR, Charmet G, Ortiz R, Crossa J (2015) Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). Plant Sci 242:23–36. https://doi.org/10.1016/J.PLANTSCI.2015.08.021

Bassi FM, Bentley AR, Charmet G, Ortiz R, Crossa J (2016) Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). Plant Sci 242:23–36. https://doi.org/10.1016/J.PLANTSCI.2015.08.021

Bekele WA, Fiedler K, Shiringani A, Schnaubelt D, Windpassinger S, Uptmoor R, Friedt W, Snowdon RJ (2014) Unravelling the genetic complexity of sorghum seedling development under low-temperature conditions. Plant, Cell and Environment 37(3):707–723. https://doi.org/10.1111/PCE.12189/SUPINFO

Belamkar V, Guttieri MJ, Hussain W, Jarquín D, El-basyoni I, Poland J, Lorenz AJ, Baenziger PS (2018) Genomic selection in preliminary yield trials in a winter wheat breeding program. G3: Genes, Genomes, Genetics 8(8):2735–2747. https://doi.org/10.1534/G3.118.200415

Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci 34(1):20–25. https://doi.org/10.2135/CROPSCI1994.0011183X003400010003X

Bernardo R (2009) Genomewide selection for rapid introgression of exotic germplasm in maize. Crop Sci 49(2):419–425. https://doi.org/10.2135/CROPSCI2008.08.0452

Bernardo R (2010) Genomewide selection with minimal crossing in self-pollinated crops. Crop Sci 50(2):624–627. https://doi.org/10.2135/CROPSCI2009.05.0250

Bernardo R (2014) Process of plant breeding. Essentials Plant Breed 57(2):9–13. https://www.nhbs.com/essentials-of-plant-breeding-book

Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative traits in maize. Crop Sci 47(3):1082–1090. https://doi.org/10.2135/CROPSCI2006.11.0690

Bhat JA, Ali S, Salgotra RK, Mir ZA, Dutta S, Jadon V, Tyagi A, Mushtaq M, Jain N, Singh PK, Singh GP, Prabhu KV (2016) Genomic selection in the era of next generation sequencing for complex traits in plant breeding. Front Genet 7(DEC):1–11. https://doi.org/10.3389/fgene.2016.00221

Bhat JA, Deshmukh R, Zhao T, Patil G, Deokar A, Shinde S, Chaudhary J (2020) Harnessing high-throughput phenotyping and genotyping for enhanced drought tolerance in crop plants. J Biotechnol 324:248–260. https://doi.org/10.1016/J.JBIOTEC.2020.11.010

Brandariz SP, Bernardo R (2019) Small ad hoc versus large general training populations for genomewide selection in maize biparental crosses. Theor Appl Genet 132(2):347–353. https://doi.org/10.1007/S00122-018-3222-3

Breider I, Gaynor R, Gorjanc G, Thorn S, Pandey MK, Varshney RK, Hickey J (2022) A multi-part strategy for introgression of exotic germplasm into elite plant breeding programs using genomic selection. Research Square. https://doi.org/10.21203/RS.3.RS-1246254/V1

Brown J, Caligari P, Peter DS (2008) An introduction to plant breeding. John Wiley & Sons, p 209

Budhlakoti N, Mishra DC, Rai A, Lal SB, Chaturvedi KK, Kumar RR (2019) A comparative study of single-trait and multi-trait genomic selection. J Comput Biol 26(10):1100–1112. https://doi.org/10.1089/CMB.2019.0032

Cericola F, Jahoor A, Orabi J, Andersen JR, Janss LL, Jensen J (2017) Optimizing training population size and genotyping strategy for genomic prediction using association study results and pedigree information. A case of study in advanced wheat breeding lines. PLoS One 12(1):e0169606. https://doi.org/10.1371/JOURNAL.PONE.0169606

Chenu K (2015) Characterizing the crop environment–nature, significance and applications. In: Crop physiology: applications for genetic improvement and agronomy: second edition, pp 321–348. https://doi.org/10.1016/B978-0-12-417104-6.00013-3

Chiquet J, Mary-Huard T, Robin S (2016) Structured regularization for conditional gaussian graphical models. Statistics and Computing 27(3):789–804. https://doi.org/10.1007/S11222-016-9654-1

Cobb JN, Biswas PS, Platten JD (2019) Back to the future: revisiting MAS as a tool for modern plant breeding. Theor Appl Genet 132(3):647–667. https://doi.org/10.1007/S00122-018-3266-4/FIGURES/12

Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc B Biol Sci 363(1491):557–572. https://doi.org/10.1098/RSTB.2007.2170

Combs E, Bernardo R (2013) Accuracy of Genomewide selection for different traits with constant population size, heritability, and number of markers. The Plant Genome 6(1). https://doi.org/10.3835/PLANTGENOME2012.11.0030

Crossa J, Beyene Y, Semagn K, Pérez P, Hickey JM, Chen C, de Campos GL, Burgueño J, Windhausen VS, Buckler E, Jannink JL, Cruz MAL, Babu R (2013) Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3: Genes, Genomes, Genetics 3(11):1903–1926. https://doi.org/10.1534/G3.113.008227

Crossa J, Jarquín D, Franco J, Pérez-Rodríguez P, Burgueño J, Saint-Pierre C, Vikram P, Sansaloni C, Petroli C, Akdemir D, Sneller C, Reynolds M, Tattaris M, Payne T, Guzman C, Peña RJ, Wenzl P, Singh S (2016) Genomic prediction of gene bank wheat landraces. G3: Genes, Genomes, Genetics 6(7):1819–1834. https://doi.org/10.1534/G3.116.029637

Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de los Campos G, Burgueño J, González-Camacho JM, Pérez-Elizalde S, Beyene Y, Dreisigacker S, Singh R, Zhang X, Gowda M, Roorkiwal M, Rutkoski J, Varshney RK (2017) Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci 22(11):961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Crossa J, Pérez P, Hickey J, Burgueño J, Ornella L, Cerón-Rojas J, Zhang X, Dreisigacker S, Babu R, Li Y, Bonnett D, Mathews K (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112(1):48–60. https://doi.org/10.1038/HDY.2013.16

Cui Y, Li R, Li G, Zhang F, Zhu T, Zhang Q, Ali J, Li Z, Xu S (2020) Hybrid breeding of rice via genomic selection. Plant Biotechnol J 18(1):57–67. https://doi.org/10.1111/PBI.13170

Daetwyler HD, Bansal UK, Bariana HS, Hayden MJ, Hayes BJ (2014) Genomic prediction for rust resistance in diverse wheat landraces. Theor Appl Genet 127(8):1795–1803. https://doi.org/10.1007/S00122-014-2341-8

Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. Genetics 185(3):1021–1031. https://doi.org/10.1534/GENETICS.110.116855

De Los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Weigel K, Cotes JM (2009) Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics 182(1):375–385. https://doi.org/10.1534/GENETICS.109.101501

Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci 19(9):592–601. https://doi.org/10.1016/j.tplants.2014.05.006

Dos Santos JPR, Pires LPM, de Castro Vasconcellos RC, Pereira GS, Von Pinho RG, Balestre M (2016) Genomic selection to resistance to Stenocarpella maydis in maize lines using DArTseq markers. BMC Genet 17(1). https://doi.org/10.1186/S12863-016-0392-3

Duangjit J, Causse M, Sauvage C (2016) Efficiency of genomic selection for tomato fruit quality. Mol Breed 36(3). https://doi.org/10.1007/S11032-016-0453-3

Edwards SMK, Buntjer JB, Jackson R, Bentley AR, Lage J, Byrne E, Burt C, Jack P, Berry S, Flatman E, Poupard B, Smith S, Hayes C, Gaynor RC, Gorjanc G, Howell P, Ober E, Mackay IJ, Hickey JM (2019) The effects of training population design on genomic prediction accuracy in wheat. Theor Appl Genet. https://doi.org/10.1007/S00122-019-03327-Y

Endelman JB, Atlin GN, Beyene Y, Semagn K, Zhang X, Sorrells ME, Jannink JL (2014) Optimal design of preliminary yield trials with genome-wide markers. Crop Sci 54(1):48–59. https://doi.org/10.2135/CROPSCI2013.03.0154

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. In: Trends in genetics, vol 12, 4th edn. Prentice Hall

Fernandes SB, Dias KOG, Ferreira DF, Brown PJ (2018) Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. Theor Appl Genet 131(3):747–755. https://doi.org/10.1007/S00122-017-3033-Y

Ferrie AMR (2006) Doubled haploid production in nutraceutical species: a review. Euphytica 158(3):347–357. https://doi.org/10.1007/S10681-006-9242-0

Forster BP, Heberle-Bors E, Kasha KJ, Touraev A (2007) The resurgence of haploids in higher plants. Trends Plant Sci 12(8):368–375. https://doi.org/10.1016/J.TPLANTS.2007.06.007

Furbank RT, Tester M (2011) Phenomics–technologies to relieve the phenotyping bottleneck. Trends Plant Sci 16(12):635–644. https://doi.org/10.1016/J.TPLANTS.2011.09.005

Gaynor RC, Gorjanc G, Bentley AR, Ober ES, Howell P, Jackson R, Mackay IJ, Hickey JM, Gaynor RC, Gorjanc G, Hickey JM, Institute R, Dick R, Bent-Ley AR, Ober ES, Howell P, Jackson R, Mackay IJ, Bingham J (2017) A two-part strategy for using genomic selection to develop inbred lines. Crop Sci 57(5):2372–2386. https://doi.org/10.2135/CROPSCI2016.09.0742

Getachew, T., Haile, A., Rekik, M., & Rischkowsky, B. (2020). A genetic database tool for data capture in small ruminant community-based breeding programs

Gowda M, Zhao Y, Würschum T, Longin CF, Miedaner T, Ebmeyer E, Schachschneider R, Kazman E, Schacht J, Martinant JP, Mette MF, Reif JC (2014) Relatedness severely impacts accuracy of marker-assisted selection for disease resistance in hybrid wheat. Heredity 112(5):552–561. https://doi.org/10.1038/HDY.2013.139

Grenier C, Cao TV, Ospina Y, Quintero C, Châtel MH, Tohme J, Courtois B, Ahmadi N (2015) Accuracy of genomic selection in a Rice synthetic population developed for recurrent selection breeding. PLoS One 10(8):e0136594. https://doi.org/10.1371/JOURNAL.PONE.0136594

Guimarães EP (2009) Rice breeding. Cereals 99–126. https://doi.org/10.1007/978-0-387-72297-9_2

Guo T, Li H, Yan J, Tang J, Li J, Zhang Z, Zhang L, Wang J (2013) Performance prediction of F1 hybrids between recombinant inbred lines derived from two elite maize inbred lines. Theor Appl Genet 126(1):189–201. https://doi.org/10.1007/S00122-012-1973-9

Guo Z, Tucker DM, Basten CJ, Gandhi H, Ersoz E, Guo B, Xu Z, Wang D, Gay G (2014) The impact of population structure on genomic prediction in stratified populations. Theor Appl Genet 127(3):749–762. https://doi.org/10.1007/S00122-013-2255-X

Habyarimana E, Lopez-Cruz M (2019) Genomic selection for antioxidant production in a panel of sorghum bicolor and s. bicolor × s. halepense lines. Genes 10(11). https://doi.org/10.3390/GENES10110841

Haile JK, N'Diaye A, Clarke F, Clarke J, Knox R, Rutkoski J, Bassi FM, Pozniak CJ (2018) Genomic selection for grain yield and quality traits in durum wheat. Mol Breed 38(6). https://doi.org/10.1007/S11032-018-0818-X

Hayes BEN, Goddard ME (2001) The distribution of the effects of genes affecting quantitative traits in livestock. Genet Sel Evol 33:1–21

Hayes BJ, Visscher PM, Goddard ME (2009) Increased accuracy of artificial selection by using the realized relationship matrix. Genet Res 91(1):47–60. https://doi.org/10.1017/S0016672308009981

Heffner EL, Jannink J-L, Sorrells ME (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome J 4(1):65. https://doi.org/10.3835/PLANTGENOME.2010.12.0029

Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. Crop Sci 50(5):1681–1690. https://doi.org/10.2135/CROPSCI2009.11.0662

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49(1):1–12. https://doi.org/10.2135/CROPSCI2008.08.0512

Henderson CR, Kempthorne O, Searle SR, von Krosigk CM (1959) The estimation of environmental and genetic trends from records subject to culling. Biometrics 15(2):192. https://doi.org/10.2307/2527669

Herter CP, Ebmeyer E, Kollers S, Korzun V, Miedaner T (2019) An experimental approach for estimating the genomic selection advantage for fusarium head blight and Septoria tritici blotch in winter wheat. Theor Appl Genet 132(8):2425–2437. https://doi.org/10.1007/s00122-019-03364-7

Heslot N, Yang HP, Sorrells ME, Jannink JL (2012) Genomic selection in plant breeding: a comparison of models. Crop Sci 52(1):146–160. https://doi.org/10.2135/CROPSCI2011.06.0297

Hickey JM, Dreisigacker S, Crossa J, Hearne S, Babu R, Prasanna BM, Grondona M, Zambelli A, Windhausen VS, Mathews K, Gorjanc G (2014) Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. Crop Sci 54(4):1476–1488. https://doi.org/10.2135/CROPSCI2013.03.0195

Hickey LT, Hafeez AN, Robinson H, Jackson SA, Leal-Bertioli SCM, Tester M, Gao C, Godwin ID, Hayes BJ, Wulff BBH (2019) Breeding crops to feed 10 billion. Nat Biotechnol 37(7):744–754. https://doi.org/10.1038/S41587-019-0152-9

Hoerl AE, Kennard RW (1970) Ridge regression: applications to nonorthogonal problems. Technometrics 12(1):69. https://doi.org/10.2307/1267352

Hoffstetter A, Cabrera A, Huang M, Sneller C (2016) Optimizing training population data and validation of genomic selection for economic traits in soft winter wheat. G3: Genes, Genomes, Genetics 6(9):2919–2928. https://doi.org/10.1534/G3.116.032532

Isidro J, Jannink JL, Akdemir D, Poland J, Heslot N, Sorrells ME (2015) Training set optimization under population structure in genomic selection. Theor Appl Genet 128(1):145–158. https://doi.org/10.1007/S00122-014-2418-4

Jähne F, Hahn V, Würschum T, Leiser WL (2020) Speed breeding short-day crops by LED-controlled light schemes. Theor Appl Genet 133(8):2335–2342. https://doi.org/10.1007/S00122-020-03601-4/FIGURES/3

Jarquín D, Crossa J, Lacaze X, Du Cheyron P, Daucourt J, Lorgeou J, Piraux F, Guerreiro L, Pérez P, Calus M, Burgueño J, de los Campos, G. (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 127(3):595–607. https://doi.org/10.1007/S00122-013-2243-1

Jonas E, De Koning DJ (2013) Does genomic selection have a future in plant breeding? Trends Biotechnol 31(9):497–504. https://doi.org/10.1016/J.TIBTECH.2013.06.003

Juliana P, Poland J, Huerta-Espino J, Shrestha S, Crossa J, Crespo-Herrera L, Toledo FH, Govindan V, Mondal S, Kumar U, Bhavani S, Singh PK, Randhawa MS, He X, Guzman C, Dreisigacker S, Rouse MN, Jin Y, Pérez-Rodríguez P et al (2019) Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. Nat Genet 51(10):1530–1539. https://doi.org/10.1038/S41588-019-0496-6

Krishnappa G, Savadi S, Tyagi BS, Singh SK, Mamrutha HM, Kumar S, Mishra CN, Khan H, Gangadhara K, Uday G, Singh G, Singh GP (2021) Integrated genomic selection for rapid improvement of crops. Genomics 113(3):1070–1086. https://doi.org/10.1016/J.YGENO.2021.02.007

Li Y, Ruperao P, Batley J, Edwards D, Khan T, Colmer TD, Pang J, Siddique KHM, Sutton T (2018) Investigating drought tolerance in chickpea using genome-wide association mapping and genomic selection based on whole-genome resequencing data. Front Plant Sci 9:190. https://doi.org/10.3389/FPLS.2018.00190/BIBTEX

Lian L, Jacobson A, Zhong S, Bernardo R (2014) Genomewide prediction accuracy within 969 maize biparental populations. Crop Sci 54(4):1514–1522. https://doi.org/10.2135/CROPSCI2013.12.0856

Liang Z, Gupta SK, Yeh CT, Zhang Y, Ngu DW, Kumar R, Patil HT, Mungra KD, Yadav DV, Rathore A, Srivastava RK, Gupta R, Yang J, Varshney RK, Schnable PS, Schnable JC (2018) Phenotypic data from inbred parents can improve genomic prediction in pearl millet hybrids. G3 (Bethesda, Md) 8(7):2513–2522. https://doi.org/10.1534/G3.118.200242

Liu X, Wang H, Wang H, Guo Z, Xu X, Liu J, Wang S, Li WX, Zou C, Prasanna BM, Olsen MS, Huang C, Xu Y (2018) Factors affecting genomic selection revealed by empirical evidence in maize. Crop J 6(4):341–352. https://doi.org/10.1016/J.CJ.2018.03.005

Lorenz AJ, Smith KP (2015) Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. Crop Sci 55(6):2657–2667. https://doi.org/10.2135/CROPSCI2014.12.0827

Lorenz AJ, Smith KP, Jannink JL (2012) Potential and optimization of genomic selection for fusarium head blight resistance in six-row barley. Crop Sci 52(4):1609–1621. https://doi.org/10.2135/CROPSCI2011.09.0503

Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120(1):151–161. https://doi.org/10.1007/S00122-009-1166-3

Lozada DN, Mason RE, Sarinelli JM, Brown-Guedira G (2019) Accuracy of genomic selection for grain yield and agronomic traits in soft red winter wheat. BMC Genet 20(1):1–12. https://doi.org/10.1186/S12863-019-0785-1/TABLES/2

Lynch M, Walsh B (1998) Chapter 27 BT–genetics and analysis of quantitative traits. *Genetics and analysis of quantitative traits*. papers2://publication/uuid/6E5E388C-7C1E-49F6-937F-13FB4B8E5A68

Mackay TFC (2003) The genetic architecture of quantitative traits. Annu Rev Genet 35:303–339. https://doi.org/10.1146/ANNUREV.GENET.35.102401.090633

Maresso K, Broeckel U (2008) Genotyping platforms for mass-throughput genotyping with SNPs, including human genome-wide scans. Adv Genet 60:107–139. https://doi.org/10.1016/S0065-2660(07)00405-1

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, Brown P, Browne C, Eller M, Guill K, Harjes C, Kroon D, Lepak N, Mitchell SE, Peterson B et al (2009) Genetic properties of the maize nested association mapping population. Science 325(5941):737–740. https://doi.org/10.1126/SCIENCE.1174320/SUPPL_FILE/MCMULLEN.SOM.PDF

Meuwissen T, Goddard M (2010) Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics 185(2):623–631. https://doi.org/10.1534/GENETICS.110.116590

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4):1819–1829. https://doi.org/10.1093/GENETICS/157.4.1819

Miedaner T, Zhao Y, Gowda M, Longin CFH, Korzun V, Ebmeyer E, Kazman E, Reif JC (2013) Genetic architecture of resistance to Septoria tritici blotch in European wheat. BMC Genomics 14(1):1–8. https://doi.org/10.1186/1471-2164-14-858

Moeinizade S, Hu G, Wang L, Schnable PS (2019) Optimizing selection and mating in genomic selection with a look-ahead approach: an operations research framework. G3 Genes|Genomes|Genetics 9(7):2123–2133. https://doi.org/10.1534/G3.118.200842

Monteverde E, Rosas JE, Blanco P, Pérez de Vida F, Bonnecarrère V, Quero G, Gutierrez L, McCouch S (2018) Multienvironment models increase prediction accuracy of complex traits in advanced breeding lines of Rice. Crop Sci 58(4):1519–1530. https://doi.org/10.2135/CROPSCI2017.09.0564

Muleta KT, Bulli P, Rynearson S, Chen X, Pumphrey M (2017) Loci associated with resistance to stripe rust (Puccinia striiformis f. sp. tritici) in a core collection of spring wheat (Triticum aestivum). PLoS One 12(6). https://doi.org/10.1371/JOURNAL.PONE.0179087

Neves HHR, Carvalheiro R, Queiroz SA (2012) A comparison of statistical methods for genomic selection in a mice population. BMC Genet 13. https://doi.org/10.1186/1471-2156-13-100

Olatoye MO, Hu Z, Aikpokpodion PO (2019) Epistasis detection and modeling for genomic selection in cowpea (Vigna unguiculata. L. Walp.). Front Genet 10(JUN):677. https://doi.org/10.3389/FGENE.2019.00677/BIBTEX

Ornella L, Singh S, Perez P, Burgueño J, Singh R, Tapia E, Bhavani S, Dreisigacker S, Braun H-J, Mathews K, Crossa J (2012) Genomic prediction of genetic values for resistance to wheat rusts. The Plant Genome 5(3). https://doi.org/10.3835/PLANTGENOME2012.07.0017

Pandey MK, Chaudhari S, Jarquin D, Janila P, Crossa J, Patil SC, Sundravadana S, Khare D, Bhat RS, Radhakrishnan T, Hickey JM, Varshney RK (2020) Genome-based trait prediction in multi-environment breeding trials in groundnut. Theor Appl Genet 133(11):3101–3117. https://doi.org/10.1007/S00122-020-03658-1/TABLES/5

Pandey MK, Roorkiwal M, Singh VK, Ramalingam A, Kudapa H, Thudi M, Chitikineni A, Rathore A, Varshney RK (2016) Emerging genomic tools for legume breeding: current status and future prospects. Front Plant Sci 7(MAY2016):455. https://doi.org/10.3389/FPLS.2016.00455/BIBTEX

Pérez-Cabal MA, Vazquez AI, Gianola D, Rosa GJM, Weigel KA (2012) Accuracy of genome-enabled prediction in a dairy cattle population using different cross-validation layouts. Front Genet 3(FEB):27. https://doi.org/10.3389/FGENE.2012.00027/BIBTEX

Phillips RL (2010) Mobilizing science to break yield barriers. Crop Sci 50:S-99. https://doi.org/10.2135/CROPSCI2009.09.0525

Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. The plant. Genome 5(3). https://doi.org/10.3835/PLANTGENOME2012.05.0005

Purbarani SC, Wasito I, Kusuma I (2017) Adaptive genetic algorithm for reliable training population in plant breeding genomic selection. In: 2016 international conference on advanced computer science and information systems, ICACSIS 2016, pp 556–563. https://doi.org/10.1109/ICACSIS.2016.7872803

Qin J, Shi A, Song Q, Li S, Wang F, Cao Y, Ravelombola W, Song Q, Yang C, Zhang M (2019) Genome wide association study and genomic selection of amino acid concentrations in soybean seeds. Front Plant Sci 10:1445. https://doi.org/10.3389/FPLS.2019.01445/BIBTEX

Rai N, Bellundagi A, Kumar PKC, Kalasapura Thimmappa R, Rani S, Sinha N, Krishna H, Jain N, Singh GP, Singh PK, Chand S, Prabhu KV (2018) Marker-assisted backcross breeding for improvement of drought tolerance in bread wheat (Triticum aestivum L. em Thell). Plant Breed 137(4):514–526. https://doi.org/10.1111/PBR.12605

Rajsic P, Weersink A, Navabi A, Peter Pauls K (2016) Economics of genomic selection: the role of prediction accuracy and relative genotyping costs. Euphytica 210(2):259–276. https://doi.org/10.1007/S10681-016-1716-0/TABLES/4

Rana MM, Takamatsu T, Baslam M, Kaneko K, Itoh K, Harada N, Sugiyama T, Ohnishi T, Kinoshita T, Takagi H, Mitsui T (2019) Salt tolerance improvement in Rice through efficient SNP marker-assisted selection coupled with speed-breeding. Int J Mol Sci 20:2585. https://doi.org/10.3390/IJMS20102585

Ray DK, Mueller ND, West PC, Foley JA (2013) Yield trends are insufficient to double global crop production by 2050. PLoS One 8(6):e66428. https://doi.org/10.1371/JOURNAL.PONE.0066428

de Resende MDV, de Assis TF (2010) Seleção Recorrente Recíproca entre Populações Sintéticas Multi- Espécies (SRR-PSME) de Eucalipto. Pesquisa Florestal Brasileira 0(57):57. https://pfb.cnpf.embrapa.br/pfb/index.php/pfb/article/view/57

Resende MDV, Resende MFR, Sansaloni CP, Petroli CD, Missiaggia AA, Aguiar AM, Abad JM, Takahashi EK, Rosado AM, Faria DA, Pappas GJ, Kilian A, Grattapaglia D (2012) Genomic selection for growth and wood quality in eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytol 194(1):116–128. https://doi.org/10.1111/J.1469-8137.2011.04038.X

Richard CAI, Hickey LT, Fletcher S, Jennings R, Chenu K, Christopher JT (2015) High-throughput phenotyping of seminal root traits in wheat. Plant Methods 11(1):1–11. https://doi.org/10.1186/S13007-015-0055-9/TABLES/2

Riedelsheimer C, Endelman JB, Stange M, Sorrells ME, Jannink JL, Melchinger AE (2013) Genomic predictability of interconnected biparental maize populations. Genetics 194(2):493–503. https://doi.org/10.1534/GENETICS.113.150227

Rothman N, Garcia-Closas M, Chatterjee N, Malats N, Wu X, Figueroa JD, Real FX, Van Den Berg D, Matullo G, Baris D, Thun M, Kiemeney LA, Vineis P, De Vivo I, Albanes D, Purdue MP, Rafnar T, Hildebrandt MAT, Kiltie AE et al (2010) A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. Nature Genetics 42(11):978–984. https://doi.org/10.1038/ng.687

Rutkoski JE, Poland J, Jannink JL, Sorrells ME (2013) Imputation of unordered markers and the impact on genomic selection accuracy. G3: Genes, Genomes, Genetics 3(3):427–439. https://doi.org/10.1534/G3.112.005363

Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet 123(4):218–223. https://doi.org/10.1111/J.1439-0388.2006.00595.X

Schmidt M, Kollers S, Maasberg-Prelle A, Großer J, Schinkel B, Tomerius A, Graner A, Korzun V (2016) Prediction of malting quality traits in barley based on genome-wide marker data to assess the potential of genomic selection. Theor Appl Genet 129(2):203–213. https://doi.org/10.1007/S00122-015-2639-1

Schulthess AW, Wang Y, Miedaner T, Wilde P, Reif JC, Zhao Y (2016) Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. Theor Appl Genet 129(2):273–287. https://doi.org/10.1007/S00122-015-2626-6

Schulz-Streeck T, Ogutu JO, Karaman Z, Knaak C, Piepho HP (2012) Genomic selection using multiple populations. Crop Sci 52(6):2453–2461. https://doi.org/10.2135/CROPSCI2012.03.0160

Shengqiang Z, Dekkers JCM, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182(1):355–364. https://doi.org/10.1534/GENETICS.108.098277

Sinha R, Fritschi FB, Zandalinas SI, Mittler R (2021) The impact of stress combination on reproductive processes in crops. Plant Sci 311:111007. https://doi.org/10.1016/J.PLANTSCI.2021.111007

Spindel J, Iwata H (2018) Genomic selection in Rice breeding. In: Rice genomics, genetics and breeding, pp 473–496. https://doi.org/10.1007/978-981-10-7461-5_24

Stewart-Brown BB, Song Q, Vaughn JN, Li Z (2019) Genomic selection for yield and seed composition traits within an applied soybean breeding program. G3 (Bethesda, Md.) 9(7): 2253–2265. https://doi.org/10.1534/G3.118.200917

Technow F, Riedelsheimer C, Schrag TA, Melchinger AE (2012) Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. Theor Appl Genet 125(6):1181–1194. https://doi.org/10.1007/S00122-012-1905-8

Thavamanikumar S, Dolferus R, Thumma BR (2015) Comparison of genomic selection models to predict flowering time and spike grain number in two hexaploid wheat doubled haploid populations. G3: Genes, Genomes, Genetics 5(10):1991–1998. https://doi.org/10.1534/g3.115.019745

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J Royal Statistical Soc: Series B (Methodological) 58(1):267–288. https://doi.org/10.1111/J.2517-6161.1996.TB02080.X

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. https://doi.org/10.1371/JOURNAL.PBIO.1001883

Voss-Fels KP, Cooper M, Hayes BJ (2019) Accelerating crop genetic gains with genomic selection. Theor Appl Genet 132(3):669–686. https://doi.org/10.1007/S00122-018-3270-8

Walter JDC, Edwards J, McDonald G, Kuchel H (2019) Estimating biomass and canopy height with LiDAR for field crop breeding. Front Plant Sci 10:1145. https://doi.org/10.3389/FPLS.2019.01145/BIBTEX

Wang X, Li L, Yang Z, Zheng X, Yu S, Xu C, Hu Z (2017) Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. Heredity 118(3):302–310. https://doi.org/10.1038/HDY.2016.87

Wang X, Xu Y, Hu Z, Xu C (2018) Genomic selection methods for crop improvement: current status and prospects. Crop J 6(4):330–340. https://doi.org/10.1016/j.cj.2018.03.001

Wang Y, Mette MF, Miedaner T, Gottwald M, Wilde P, Reif JC, Zhao Y (2014) The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. BMC Genomics 15(1):1–12. https://doi.org/10.1186/1471-2164-15-556/FIGURES/4

Ward BP, Brown-Guedira G, Tyagi P, Kolb FL, van Sanford DA, Sneller CH, Griffey CA (2019) Multienvironment and multitrait genomic selection models in unbalanced early-generation wheat yield trials. Crop Sci 59(2):491–507. https://doi.org/10.2135/CROPSCI2018.03.0189

Watson A, Ghosh S, Williams MJ, Cuddy WS, Simmonds J, Rey MD, Asyraf Md Hatta M, Hinchliffe A, Steed A, Reynolds D, Adamski NM, Breakspear A, Korolev A, Rayner T, Dixon LE, Riaz A, Martin W, Ryan M, Edwards D, Hickey LT et al (2018) Speed breeding is a powerful tool to accelerate crop research and breeding. Nature Plants 4(1):23–29. https://doi.org/10.1038/s41477-017-0083-8

Weber N, Halpin C, Curtis Hannah L, Jez JM, Kough J, Parrott W (2012) Editor's choice: crop genome plasticity and its relevance to food and feed safety of genetically engineered breeding stacks. Plant Physiol 160(4):1842–1853. https://doi.org/10.1104/PP.112.204271

Wong CK, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theor Appl Genet 116(6):815–824. https://doi.org/10.1007/S00122-008-0715-5

Würschum T, Reif JC, Kraft T, Janssen G, Zhao Y (2013) Genomic selection in sugar beet breeding populations. BMC Genet 14. https://doi.org/10.1186/1471-2156-14-85

Xu S (2013) Genetic mapping and genomic selection using recombination breakpoint data. Genetics 195(3):1103–1115. https://doi.org/10.1534/GENETICS.113.155309/-/DC1

Xu Y, Liu X, Fu J, Wang H, Wang J, Huang C, Prasanna BM, Olsen MS, Wang G, Zhang A (2020) Enhancing genetic gain through genomic selection: from livestock to plants. Plant Communications 1(1):100005. https://doi.org/10.1016/j.xplc.2019.100005

Yabe S, Hara T, Ueno M, Enoki H, Kimura T, Nishimura S, Yasui Y, Ohsawa R, Iwata H (2018) Potential of genomic selection in mass selection breeding of an allogamous crop: an empirical study to increase yield of common buckwheat. Front Plant Sci 9:276. https://doi.org/10.3389/FPLS.2018.00276/BIBTEX

Zhang X, Pérez-Rodríguez P, Burgueño J, Olsen M, Buckler E, Atlin G, Prasanna BM, Vargas M, Vicente FS, Crossa J (2017) Rapid cycling genomic selection in a multiparental tropical maize population. G3 (Bethesda, Md.) 7(7):2315–2326. https://doi.org/10.1534/G3.117.043141

Zhang X, Pérez-Rodríguez P, Semagn K, Beyene Y, Babu R, López-Cruz MA, San Vicente F, Olsen M, Buckler E, Jannink JL, Prasanna BM, Crossa J (2015) Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. Heredity 114(3):291–299. https://doi.org/10.1038/HDY.2014.99

Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin FH, Ranc N, Reif JC (2012) Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet 124(4):769–776. https://doi.org/10.1007/S00122-011-1745-Y/TABLES/3

Zhao Y, Mette MF, Gowda M, Longin CFH, Reif JC (2014) Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. Heredity 112(6):638–645. https://doi.org/10.1038/HDY.2014.1