Genomic Selection: A Tool for Accelerating the Efficiency of Molecular Breeding for Development of Climate-Resilient Crops

Neeraj Budhlakoti[1†], Amar Kant Kushwaha[2†], Anil Rai[1], K K Chaturvedi[1], Anuj Kumar[1], Anjan Kumar Pradhan[3], Uttam Kumar[4], Rajeev Ranjan Kumar[1], Philomin Juliana[4], D C Mishra[1*] and Sundeep Kumar[3*]

[1]ICAR- Indian Agricultural Statistics Research Institute, New Delhi, India, [2]ICAR- Central Institute for Subtropical Horticulture, Lucknow, India, [3]ICAR- National Bureau of Plant Genetic Resources, New Delhi, India, [4]Borlaug Institute for South Asia (BISA), Ludhiana, India

Since the inception of the theory and conceptual framework of genomic selection (GS), extensive research has been done on evaluating its efficiency for utilization in crop improvement. Though, the marker-assisted selection has proven its potential for improvement of qualitative traits controlled by one to few genes with large effects. Its role in improving quantitative traits controlled by several genes with small effects is limited. In this regard, GS that utilizes genomic-estimated breeding values of individuals obtained from genome-wide markers to choose candidates for the next breeding cycle is a powerful approach to improve quantitative traits. In the last two decades, GS has been widely adopted in animal breeding programs globally because of its potential to improve selection accuracy, minimize phenotyping, reduce cycle time, and increase genetic gains. In addition, given the promising initial evaluation outcomes of GS for the improvement of yield, biotic and abiotic stress tolerance, and quality in cereal crops like wheat, maize, and rice, prospects of integrating it in breeding crops are also being explored. Improved statistical models that leverage the genomic information to increase the prediction accuracies are critical for the effectiveness of GS-enabled breeding programs. Study on genetic architecture under drought and heat stress helps in developing production markers that can significantly accelerate the development of stress-resilient crop varieties through GS. This review focuses on the transition from traditional selection methods to GS, underlying statistical methods and tools used for this purpose, current status of GS studies in crop plants, and perspectives for its successful implementation in the development of climate-resilient crops.

Keywords: GS, climate change, STGS, MTGS, abiotic stress, biotic stress, GEBV, climate-resilient crops

INTRODUCTION

Sustainable food production is the utmost requirement for food and nutritional security. Based on reports, 821 million people are point below nourishment level; i.e., 151 million children under 5 years are stunted; in terms of micronutrients, two billion people are not able to meet the requirement for living a healthy life, globally. To meet these demands, the production and supply system has to be

sound. It has been projected that production has to be increased by 60% by 2050, amid different challenges related to the production system posed by climate change (WHO/FAO, 2015), which is further projected to worsen by an increase in the price of food to the extent of 1–29% by 2050. The development of climate-resilient varieties through conventional approaches of hybridization and selection is input-intensive (labor, land, and time), limiting the realized genetic gain. Improvement in the genetic gain as per the Lush equation (Lush, 1943) can be secured through i) better intensity of selection *via* accurate and high-throughput phenotyping and ii) having a broad genetic base representing diverse eco-geography in breeding program. The advancement in genomics approaches leads to the availability of huge resources like genome sequence information, transcriptome, and proteome that have paved the way to hasten the identification of target genes mitigating the effects of climate change (Varshney et al., 2018). This sequence of information also leads to the identification of several mutant loci at the nucleotide level which might be associated with characters of complex nature like yield in general and under different circumstances of stress, which are otherwise very difficult to decipher. Genomic selection emerged as an important tool which can utilize such information for modeling the crop yield for effective and rapid selection under different environmental conditions to meet the production challenges in a climate-changing world.

Changes brought about by climate change have affected the phenology of different crop species leading to a detrimental effect on production and productivity. Different stresses, viz., heat, cold, drought, and flood, are specific manifestations of climate change. Genetic improvement of crops based on phenotypic selection has been successfully achieved through traditional breeding. However, in recent past, genomics led to the identification of several underlying genes/QTLs providing tolerance to these specific conditions, which have been utilized in marker-assisted selection (MAS). MAS is an indirect selection process, where individuals for a particular trait of interest are selected based on the known markers linked to it (Fernando and Grossman, 1989). This method has been efficiently used in the past for selection of individuals in plant breeding to increase the selection accuracy compared to the traditional phenotype-based selection process (Mohan et al., 1997). In cereals, MAS resulted in a number of varieties, viz., Improved Pusa Basmati1 (Gopalakrishnan et al., 2008), Pusa Basmati 1728 (Singh et al., 2017a), Pusa Basmati 1637 (Singh et al., 2017b), Pusa Samba 1850 (Krishnan et al., 2019), Improved Samba Mahsuri (Madhavi et al., 2016), and Swarna-Sub1 (Neeraja et al., 2007) in rice, HUW510 in wheat (Vasistha et al., 2017), and HHB67-Improved in pearl millet (Rai et al., 2008). C214 in chickpea (Varshney et al., 2014a), JTN5503 and DS880 in soybean (Arelli et al., 2006, 2009), and JL24 and TAG24 in groundnut (Varshney et al., 2014b) have been derived using MAS. However, MAS is practically feasible only if the trait of interest is associated with one or very few major genes, and it is impractical or irrelevant for quantitative traits (i.e., polygenic traits that are governed by few hundreds of minor genes) (Bernardo, 2008), which most of the stress tolerance–related traits are based on. To overcome this issue, a
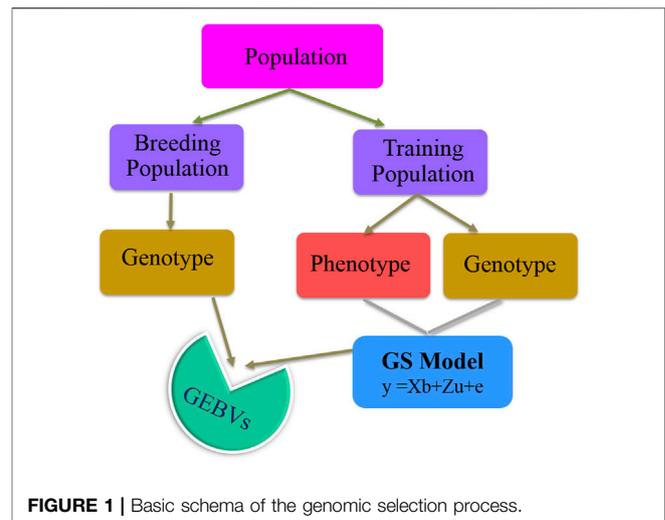
FIGURE 1 | Basic schema of the genomic selection process.

new selection tool called genomic selection (GS) was proposed that can facilitate selection for such traits, by means of net genetic merit of an individual obtained using the effects of dense markers distributed across the genome (Meuwissen et al., 2001). In this approach, the individual effect of each marker is estimated, and the additive sum of all the marker effects is used for calculation of the genomic-estimated breeding values (GEBV) of each individual. In the current scenario of climate change, GS is a promising tool for improving the genetic gain of individuals under the breeding program (Yuan et al., 2019). The basic process of any genomic selection process starts with the creation of training population, i.e., individuals having both genotypic and phenotypic information, and this information is used to build a model, where the phenotype is used as a response and genotype as a predictor. The information from the developed model is later used to estimate the GEBV of breeding population, i.e., individuals having only genotypic information. The basic process of GS is also explained in **Figure 1**.

The major advantage of using GS is that it allows for a drastic reduction in the duration of the breeding cycle as compared to traditional breeding and also minimizes the cost associated with extensive phenotyping, thereby subsequently accelerating genetic gains and ensuring food and nutritional security (Heffner et al., 2010). However, there are certain factors such as the size of training and breeding populations, genetic diversity of breeding population, heritability of the underlying trait, influence of genotype–environment (GxE) interaction, density of markers, and genetic relationship between training population and breeding population or selection candidates, which may influence the genomic prediction's accuracy (De Roos et al., 2009; Lorenzana and Bernardo, 2009; Luan et al., 2009; Daetwyler et al., 2010; Clark et al., 2011; Howard et al., 2014). Hence, successful implementation of GS in breeding programs requires careful consideration of all these factors. Apart from these factors, there are certain limitations of genomic selection. Changes in gene frequencies and epistatic interactions drastically affect the estimates of GEBV. Most of the models used to estimate GEBV ignore the effect of epistasis which plays a prime role

especially in cross pollinated plants (Heffner et al., 2009). The rate of declination of selection response is more in GS than pedigree based selection, which can be minimized through the addition of new markers to the model (Nakaya and Isobe, 2012). However, the cost of implementation of GS is more than that of the traditional breeding program.

The choice of models is an important factor in implementing GS, and several parametric and non-parametric genomic prediction models are available for this purpose. One of the most common and widely used parametric genomic selection model is the best linear unbiased prediction (BLUP). It is a mixed model–based whole-genome regression approach that is used to estimate the marker effects, and the same has been successfully applied to predict complex traits (Habier et al., 2009, 2013; de los Campos et al., 2013). In general, it was observed that the performance of parametric models found to be efficient only for traits with additive genetic architectures. For traits that are highly affected by epistatic or non-additive interactions, it becomes challenging to use parametric models (Moore and Williams, 2009). Epistatic interactions play a key role in explaining genetic variation for quantitative traits. Hence, ignoring such type of information in the prediction model might result in lower genomic prediction accuracies (Cooper et al., 2002). Due to these factors, it is not always advisable to practice simple linear or parametric models. Gianola et al. (2006) first used non-parametric and semiparametric methods for modeling the complex genetic architecture. Subsequently, several statistical methods were implemented to model both additive and epistatic effects for genomic selection (Xu, 2007; Cai et al., 2011). For a detailed comparison of various parametric, non-parametric and semiparametric methods in different settings of population size and trait heritability, one can refer to Howard et al. (2014) and Budhlakoti et al. (2020c). Recently, some semiparametric (Legarra and Reverter, 2018) and advanced approaches (Tanaka, 2018; Budhlakoti et al., 2020a, 2020b; Majumdar et al., 2020; Sehgal et al., 2020; Tanaka, 2020; Mishra et al., 2021) have also been proposed and implemented in context to genomic selection. In the next section, few most commonly used methods for genomic selection studies have been discussed.

STATISTICAL MODEL FOR GENOMIC SELECTION

The process of selecting the suitable individuals in GS starts with a simple linear model sometimes also called least-squares regression or ordinary least-squares regression (OLS):

$$Y = 1_n \mu + X\beta + \varepsilon$$

where $\mathbf{Y}$ = $n \times 1$ vectors of observations, μ is the mean, $\boldsymbol{\beta}$ = $p \times 1$ vectors of marker effects, ε = $n \times 1$ vectors of random residual effects, $\mathbf{X}$ = design matrix of order $n \times p$ (where each row represents the genotype/individuals/lines (n) and each column corresponds to the marker (p)), and $\varepsilon \sim N(0, \sigma_e^2)$.

One major problem in linear models using several thousands of genome-wide markers is that the number of markers (p) exceeds the number of observations (n), i.e., genotype/individuals/lines, and this creates the problem of over-parameterization (large "p" and small "n" problem (p >> n)). Using a subset of significant markers can be an alternative for dealing with the large "p" and small "n" problem. Meuwissen et al. (2001) used a modification of the least-squares regression for GS. They performed least-squares regression analysis on each marker separately with the following model:

$$Y = X_j \beta_j + \varepsilon$$

where X_j = j^{th} column of the design matrix of the markers and β_j = genetic effect of the jth marker.

Markers with significant effects are selected using the log likelihood of this model, and those are further used for estimation of breeding values. However, it has to be noted that some key information may be lost by selection based on the subset of markers.

Hence, an efficient solution for the over-parameterization problem in linear models is using ridge regression (RR), which is a penalized regression–based approach (Meuwissen et al., 2001). It also solves the problems of multicollinearity at the same time (i.e., correlated predictors, e.g., SNP, or markers). RR shrinks the coefficients of correlated predictors equally toward zero and solves the regression problem using ℓ2 penalized least squares. Here, the goal is to derive an estimator of parameter β with a smaller variance than the least-squares estimator. Similar to RR, the least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996; Usai et al., 2009) is another variant of penalized regression, which uses the ℓ_1 penalized least-squares criterion to obtain a sparse solution. However, sometimes LASSO may not work well with highly correlated predictors (e.g., SNPs in high linkage disequilibrium) (Ogutu et al., 2012). The elastic net (ENET) is an extension of the LASSO that is robust to extreme correlations among the predictors (Friedman et al., 2010), and it is a compromise between ℓ1 penalty (LASSO) and ℓ2 penalty (RR) (Zou and Hastie, 2005).

The RR model considers that each marker contributes to equal variance, which is not the case for all traits. Therefore, the variance of the markers based on the trait's genetic architecture has to be modeled. For this purpose, several Bayesian models have been proposed where it is assumed that there is some prior distribution of marker effects. Furthermore, inferences about model parameters are obtained on the basis of posterior distributions of marker effects. There are several variants of Bayesian models for genomic prediction such as Bayes A, Bayes B, Bayes Cπ, and Bayes Dπ (Meuwissen et al., 2001; Habier et al., 2011) and other derivatives, e.g., Bayesian LASSO and Bayesian ridge regression (BRR). Besides the marker-based models, the best linear unbiased prediction (BLUP) (Henderson et al., 1959) is one of the most commonly used genomic prediction methods. There are many variants of BLUP available for this purpose, e.g., genomic BLUP (GBLUP), single-step GBLUP (ssGBLUP), ridge regression BLUP (RRBLUP), and

FIGURE 2 | Overall summary of the most commonly used models in genomic selection.
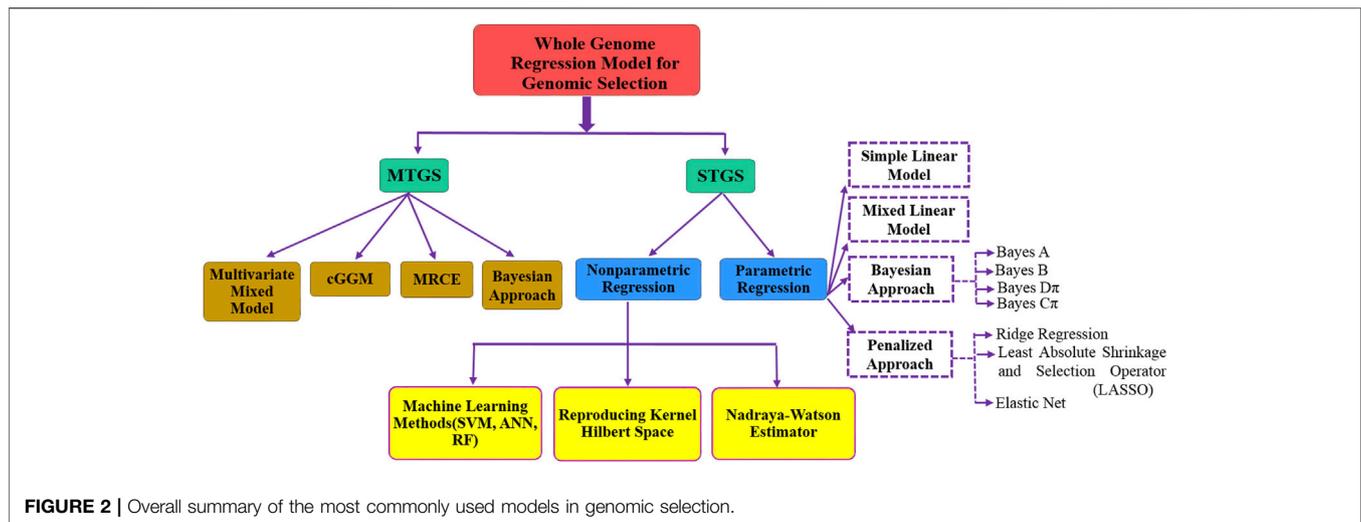
GBLUP with linear ridge kernel regression (rrGBLUP), of which GBLUP is very frequently used. The GBLUP uses the genomic relationships calculated using markers instead of the conventional BLUP which uses the pedigree relationships to obtain the GEBV of the lines or individuals (Meuwissen et al., 2001).

The genomic prediction models discussed so far perform well for traits with additive genetic architecture, but their performance becomes very poor in case of epistatic genetic architectures. Hence, Gianola et al. (2006) first used non-parametric and semiparametric methods for modeling the complex genetic architecture. Subsequently, several statistical methods were implemented to model both additive and epistatic effects for genomic selection (Xu, 2007; Cai et al., 2011; Legarra and Reverter, 2018). There are several non-parametric methods that have been studied in relation to genomic selection, e.g., NW (Nadaraya–Watson) estimator (Gianola et al., 2006), RKHS (reproductive kernel Hilbert space) (Gianola et al., 2006), SVM (support vector machine) (Maenhout et al., 2007; Long et al., 2011), ANN (artificial neural network) (Gianola et al., 2011), and RF (random forest) (Holliday et al., 2012), among them SVM, NN, and RF are based on the machine learning approach.

Methods discussed earlier in this section are based on genomic information where information is available for a single trait, i.e., single-trait genomic selection (STGS). As the performance of STGS-based methods may be affected significantly in case of pleiotropy, i.e., one gene linked to multiple traits, a mutation in a pleiotropic gene may have an effect on several traits simultaneously. It was observed that low heritability traits can borrow information from correlated traits and consequently achieve higher prediction accuracy. However, STGS-based methods consider the information of each trait independently. Hence, we may lose crucial information which may ultimately result in poor genomic prediction accuracy. Nowadays, as we are receiving data on multiple traits, so multi-trait genomic selection (MTGS)-based methods may provide more accurate GEBV and subsequently a higher prediction accuracy. Several MTGS-based methods have been studied in relation to GS, e.g., multivariate

mixed model approach (Jia and Jannink, 2012; Klápště et al., 2020), Bayesian multi-trait model (Jia and Jannink, 2012; Cheng et al., 2018), MRCE (multivariate regression with covariance estimation) (Rothman et al., 2010), and cGGM (conditional Gaussian graphical model) (Chiquet et al., 2017). Jia and Jannink (2012) presented three multivariate linear models (i.e., GBLUP, Bayes A, and Bayes Cπ) and compared them to univariate models, and a detailed comparison of various STGS- and MTGS-based methods has also been studied by Budhlakoti et al. (2019c). A brief structure of different STGS- and MTGS-based methods used in GS studies is given in **Figure 2**.

GS: IMPLICATIONS IN CROP IMPROVEMENT

GS in Cereals

Cereals are an important part of our daily diet as they contribute about 50% of the total dietary energy supply (WHO/FAO, 2003). Wheat, rice, maize, and barley are the major cereal crops, which are being grown on arable land all over the world amounting to a total of 2,817 million tonnes of production (FAO). Production of these crops is being challenged by calamities created by a change in climatic pattern (Reynolds, 2010), and over that, it is being complicated by the rising demand of increasing population (Tester and Langridge, 2010; Furbank and Tester, 2011). To meet the challenges, the production system has to be efficient and sustainable with lower pressure on the ecosystem. High-yielding, resource-efficient crop varieties are an integral component of such production systems which can address the challenges. But the development of such variety is a painstaking endeavor as most of the crop productivity traits are under the control of a complex genetic system (most genes are of minor effect) with the complication of low heritability and high order of epitasis (Mackay, 2001). Though conventional selection methods have resulted in a number of varieties but the genetic gain per unit time is not as much rewarding as GS, it provides an opportunity to hasten the cycle of selection (Bernardo and Yu, 2007; Lorenz

et al., 2011). The potential of GS can be assessed from the fact that it has the ability to select high breeding value individuals rapidly from early-generation populations without the need of extensive phenotyping. This has been shown effectively in cereal crops in the recent past. Wheat, rice, maize, and barley are the first candidate crops where the effectiveness of GS has been studied. GS in these crops leads to the identification of different models which were able to efficiently predict the performance of traits under question and filter out the important breeding material. In the following section, the role of GS in cereal crops has been discussed.

Grain Yield and Related Traits' Improvement

Grain yield is a major trait which is affected directly or indirectly by other traits including thousand grain weight, number of tillers bearing panicle, number of grains per panicle, number of filled grains per panicle etc. Genomic prediction for these traits utilizing different types of training populations and models have been evaluated. The variations in the accuracies of genomic prediction have been attributed to the heritability of the trait, training population, and models used. The genomic prediction accuracy for a very complex and physiological trait–like distribution of weight to the individual grain in the panicle in rice (Yabe et al., 2018) ranged from 0.28 to 0.78 for grain yield in maize (Rio et al., 2019). For the improvement of accuracy, the role of training population also has a significant effect, and it has been reported that prediction based on the training set developed using North Carolina mating design II (0.60) was found at par with that of full diallel matings (0.58) and superior to that of test cross (0.10) (Fristche-Neto et al., 2018). Similarly, better prediction accuracies for grain yield were observed in recombinant inbred lines and doubled haploid populations compared to natural populations (Liu et al., 2018). The accuracy of GS for grain yield is also highly influenced by the size of training populations and genetic relationships between the training and breeding populations (Lozada et al., 2019; Lozada and Carter, 2020). Longin et al. (2014) reported that GS followed by one cycle of phenotypic selection has been reported to facilitate identification of superior parental lines with better combining ability and high annual genetic gain for grain yield in wheat than simple phenotypic selection. However scheme had not considered the cost and time involved in production and nursery screening of these lines, and thus, additional schemes like GSrapid have been proposed which have better selection gain and been recommended for utilization in a hybrid breeding program of different cereal crops (Marulanda et al., 2016). GS could also be potentially used in the prediction of the performance of a large number of hybrid combinations (VanRaden, 2008; Crossa et al., 2017). The earlier GS studies on cereals started with wheat where the DArT marker system was used (Crossa et al., 2010, 2011; Heffner et al., 2011; Burgueño et al., 2012; Pérez-Rodríguez et al., 2012). However, later, other genome-wide SNP platforms became the routine marker in genomic selection owing to their own advantages (Poland et al., 2012; Zhao et al., 2012). Detailed information on GS studies for grain yield and related traits in major cereals, pulses, oilseeds, and horticultural crops with the details of statistical models, marker platforms, types of populations used, and the prediction accuracies of statistical models are listed in **Table 1**.

Biotic Stress Tolerance

With the change in weather patterns, emergence/resurgence of new races and biotypes of pathogens and insects is being reported globally (Juarez et al., 2013; Váry et al., 2015; Fones et al., 2020). Hence, identification of resistance genes in the germplasm and their incorporation into the breeding program are required to develop biotic stress–tolerant varieties. MAS has proved to be efficient in breeding for qualitative resistance, but for quantitative resistance which is governed by many genes with smaller effects, MAS has not been so effective. GS has proved its role in improving tolerance against biotic stresses in cereals which are quantitatively controlled, though it has been applied to a very limited extent. Most of the studies on the utility of GS for biotic stress tolerance have been reported from wheat, for a wide array of diseases including three types of rusts, *Fusarium* head blight, septoria tritici blotch, powdery mildew, tan spot, and *Stagonospora nodorum* blotch. The genomic prediction accuracies for these diseases ranged from 0.14 to 0.85 (Rutkoski et al., 2012; Daetwyler et al., 2014; Mirdita et al., 2015; Juliana et al., 2017; Sarinelli et al., 2019). In rice, GS has been utilized to identify blast-tolerant lines (Huang et al., 2019). In maize, GS has been successfully utilized to select lines from natural populations for tolerance to *Stenocarpella maydis* causing ear rot (dos Santos et al., 2016) and from biparental populations for superior yield under heavy infestation of *Striga* (Badu-Apraku et al., 2019). In case of barley, markers and prediction models were utilized for *Fusarium* head blight severity, and the prediction accuracy was quite higher, i.e., 0.72, than that of conventional phenotyping (Lorenz et al., 2012; Sallam and Smith, 2016).

Abiotic Stress Tolerance

The occurrence of drought, high-temperature stress during crop growth stages, flood, etc., is at surge due to climate change, causing significant crop losses (Qin et al., 2011). With the 1˚C increase in global temperature, yield reduction has been predicted up to 6.4% in wheat (Liu et al., 2016). The sustainable and economic options under such situations to cover the losses are changing cropping patterns or developing abiotic stress–tolerant varieties. Identification of tolerant genotypes from the germplasm and their utilization in the breeding program become a prime requirement for development of such varieties (Baenziger, 2016). The major issue in breeding for abiotic stress tolerance is their complex inheritance, low heritability, and high environmental effect on them (Bernardo, 2008).

Conventional breeding methods for abiotic stresses suffer from limitations of accuracy and reproducibility. Though molecular markers have been utilized to identify and transfer yield QTLs under abiotic stress conditions (Ribaut and Ragot, 2007; Almeida et al., 2013), but it may not be effective as QTL from limited genetic resources explain little variation for grain yield under stress and are also highly influenced by the genetic background (Semagn et al., 2013) as well the environment and

TABLE 1 | Genomic prediction for grain yield and related traits in different crops (i.e. Cereals, Pulses, Oilseeds and Horticultural crops).

Crop	Model	Genotyping Techniques	Population type	Trait		Prediction accuracy (PA)	Reference
A. Cereals							
i) Maize	GBLUP	Taqman (ABI 2002)	F1 from half diallel and test crosses	Grain yield (GY)		0.58	Zhao et al. (2012)
	GBLUP	Affymetrix®	F1 from test crosses (TC), North Carolina design II (NCII), and full diallel (FD)	GY		0.10 (TC) 0.58(NCII) 0.60(FD)	Fristche-Neto et al. (2018)
	RRBLUP	55 K SNP array	Natural population (NP), recombinant inbred line (RIL), double haploid (DH), and F2:3	GY		RIL&DH (0.41) > F2:3 (0.36) > NP(0.40)	Liu et al. (2018)
	GBLUP			100 kernel weight		F2:3 (0.77) > RIL&DH (0.65) > NP(0.48) ()	
	Bayes A, Bayes B, Bayes C, LASSO, and RKHS						
	GBLUP and multigroup GBLUP	Genotyping by sequencing (GBS)	TC	GY		0.78	Rio et al. (2019)
		50 K Illumina®		Yield index (YI)		0.73	
		600 K and Affymetrix® Axiom					
	RRBLUP and BSSV (Bayesian stochastic search variable)	DArTSeq™ and Illumina HiSeq2000	Inbred lines	Ear rot	Proportion of rotten kernel	0.87	dos Santos et al. (2016)
					Ear rot incidence	0.24–0.56	
	BLUP	Kompetitive Allele Specific PCR (KASP)	Inbred lines and test cross progenies	Striga resistance		0.58	Badu-Apraku et al. (2019)
				Drought tolerance		0.42–0.65	
	GBLUP	GBS	Breeding lines	Drought tolerance		0.37–0.38	Beyene et al. (2015)
	RRBLUP and GBLUP	KASP	Inbred lines and half diallel population	Water-logging tolerance		0.53–0.84	Das et al. (2020)
	BLUP	KASP	Asian and African inbred lines	Drought tolerance	GY	0.71–0.75	Vivek et al. (2017)
					Anthesis–silking interval (ASI)	0.35–0.43	
	RRBLUP	Infinium Maize SNP50 Bead Chip	Subtropical maize lines	Drought tolerance	ASI	0.93	Shikha et al. (2017)
	Bayes A, Bayes B, and LASSO				100 kernel weight	0.92	
ii) Wheat	RRBLUP, RKHS, and Bayesian LASSO	Diversity Arrays Technology (DArT)	Advanced breeding and germplasm lines	GY		0.49–0.61	Crossa et al. (2010)
	Bayesian LASSO and RKHS	DArT	Breeding lines	GY		0.43–0.79	Crossa et al. (2011)
	Bayes A, Bayes B, Bayes C, and RRBLUP	DArT	Breeding lines	GY		0.48	Heffner et al. (2011)
	Bayesian LASSO	DArT	Breeding lines	GY		0.5–0.6	Burgueño et al. (2012)
	RRBLUP, Bayes A, Bayes B, Bayes C, LASSO, NN, and RKHS	DArT	Breeding lines	GY		0.6–0.7	Pérez-Rodríguez et al. (2012)
	GBLUP	DArT	Breeding lines	GY		0.2–0.4	Poland et al. (2012)
	RRBLUP, Bayes A, Bayes B, and Bayes C	9 K Illumina® Infinium	F1s	GY		0.3–0.6	Zhao et al. (2013)
	RRBLUP	9 K Illumina® and 90 K iSelect	Red winter wheat breeding lines	GY		0.14–0.43	Lozada et al. (2019)
	GBLUP	DArT and KASP	F4:6 population	GY		0.75	Michel et al. (2019)
	GBLUP	GBS	Breeding lines	GY		0.42–0.56	Juliana et al. (2019)
	GBLUP	GBS	Breeding population	GY		0.12–0.34	Sun et al. (2019)
	GBLUP and IBCF:MTME (item-based collaborative filtering: multi-trait multi-environment)	Illumina® 90 K	Winter wheat lines	GY		-0.21 to 0.42	Lozada and Carter (2020)
	GBLUP and BRR	Infinium iSelect 9 K	Germplasm	Leaf rust resistance (LRR)		0.35	Daetwyler et al. (2014)
				Stem rust resistance (SRR)		0.27	
				Yellow rust resistance (YRR)		0.44	
	RR	DArT	Breeding lines	Fusarium head blight (FHB) resistance		0.006–0.463	Rutkoski et al. (2012)
	RKHS					0.118–0.575	
	RF			Deoxynivalenol (DON) resistance			
	Bayesian LASSO and multiple linear regression						
	RRBLUP	Illumina Infinium 9 K and 90 K	Winter wheat breeding lines	FHB resistance		0.6	Mirdita et al. (2015)
	Bayes Cπ and RKHS			Septoria leaf blotch resistance		0.5	
	RRBLUP	GBS	Winter wheat breeding lines	Powdery mildew resistance		0.60	Sarinelli et al. (2019)
				GY		0.64	
				Test weight		0.71	
	RKHS and GBLUP	GBS	Lines from International Bread Wheat Screening Nursery	LRR	Seedling	0.31–0.74	Juliana et al. (2017)
					Adult	0.12–0.56	
				YRR	Seedling	0.70–0.78	
					Adult	0.34–0.71	
				SRR		0.31–0.65	

(Continued on following page)

TABLE 1 | (*Continued*) Genomic prediction for grain yield and related traits in different crops (i.e. Cereals, Pulses, Oilseeds and Horticultural crops).

Crop	Model	Genotyping Techniques	Population type	Trait	Prediction accuracy (PA)	Reference
iii) Rice	Bayesian LASSO	DArT	Inter-related synthetic population	GY	0.309	Grenier et al. (2015)
				Panicle weight	0.327	
	RRBLUP	GBS	Tropical rice breeding lines	GY	0.31	Spindel et al. (2015)
	GBLUP	Illumina HiSeq 2000	128 Japanese rice varieties	Field grain	0.30	Yabe et al. (2018)
				Field grain weight	0.28	
		Illumina HiSeq 4000 and HiSeqX		Variance of field grain	0.53	
	GBLUP, SVM, LASSO, and PLS	GBS	North Carolina design II population	GY	~0.5	Xu et al. (2018)
				Thousand grain weight (TGW)	~0.28	
	GBLUP	Illumina® HiSeq 2000	Hybrid population	GY	0.54	Cui et al. (2020)
				Grain length	0.92	
	GBLUP, RKHS, and Bayes B	GBS	Breeding lines	Panicle weight	0.30	Hassen et al. (2018)
				Nitrogen balance index	0.21	
	GBLUP	SNP	Breeding lines	GY	0.39	Wang et al. (2018)
				TGW	0.88	
	RRBLUP and GBLUP	GBS	Rice population	Blast resistance	0.17–0.73	Huang et al. (2019)
	GBLUP and RKHS	962 K Core SNP dataset	Germplasm	Drought tolerance	0.226–0.809	Bhandari et al. (2019)
iv) Barley	RRBLUP	Illumina GoldenGate	Breeding lines	GY	0.57	Sallam et al. (2015)
				DON	0.72	
				FHB	0.74	
	GBLUP and RKHS	GBS	Breeding lines	Thousand kernel weight (TKW)	0.67	Abed et al. (2018)
	GBLUP	Illumina	Breeding lines	GY	0.362	Tiede and Smith (2018)
				DON resistance	0.367	
B. Pulses						
i) Lentil	RRBLUP	Exome capture	Lentil diversity panel, RIL	Maturity duration	0.58–0.84	Haile et al. (2020)
	GBLUP					
	Bayes A					
	Bayes B					
	Bayes Cπ					
	Bayesian LASSO					
	BRR and RKHS					
ii) Common bean	GBLUP	GBS	RIL, multi-parent advanced generation inter-cross (MAGIC), germplasm	Cooking time	0.22–0.55	Diaz et al. (2021)
	Bayes A					
	Bayes B					
	Bayes C					
	Bayesian LASSO and BRR					
	RKHS	GBS	Breeding lines	Root rot resistance *Fusarium*	0.52	Diaz et al. (2021)
				Pythium	0.72–0.79	
iii) Chickpea	RRBLUP	Whole-genome re-sequencing (WGRS)	Breeding lines	Drought tolerance	0.56–0.61	Li et al. (2018)
	Bayesian LASSO and BRR					
C. Oilseeds						
i) Groundnut	Bayesian generalized linear regression	Affymetrix GeneTitan®	Breeding lines	Yield	0.49–0.60	Pandey et al. (2020)
				Protein	0.41–0.46	
				Rust resistance	0.74–0.75	
				Late leaf spot resistance	0.57–0.65	
ii) *Brassica napus*	RRBLUP	Infinium Array 60 K	Test cross F1s	Seed yield	0.45	Jan et al., 2016
				Oil content	0.81	
				Lodging resistance	0.39	
	GBLUP	Transcriptome GBSt assay	Spring canola lines	Seed yield	0.69	Fikere et al. (2020)
				Oil content	0.64	
	GBLUP	Illumina Infinium 60 K	Double haploid population	Seed yield	0.27–0.55	Xiong et al. (2020)
	LASSO					
iii) Sunflower	and multi-kernel BLUP	GBS	F1s from factorial mating design	Oil content	0.783	Mangin et al. (2017)
iv) Soybean	RRBLUP	iSelect Bead Chip	RILs from interspecific cross	Yield	0.68	Beche et al. (2021)
				Oil content	0.76	
	Bayes B and Bayesian LASSO	BARCSoySNP6K		Protein content	0.76	
	RRBLUP	iSelect Bead Chip	Breeding lines	Oil content	0.30	Stewart-Brown et al. (2019)
		BARCSoySNP6K		Protein content	0.55	

TABLE 1 | (Continued) Genomic prediction for grain yield and related traits in different crops (i.e. Cereals, Pulses, Oilseeds and Horticultural crops).

Crop	Model	Genotyping Techniques	Population type	Trait	Prediction accuracy (PA)	Reference
D. Horticultural crops						
i) Apple	RRBLUP	HiScan Illumina and Infinium Array 20 K	Germplasm and biparental families	Firmness	0.81	Roth et al. (2020)
	RRBLUP and Bayesian LASSO	Infinium® II 8 K	F1 from factorial mating design	Fruit firmness Soluble solids	0.83 0.89	Kumar et al. (2012)
ii) Citrus	GBLUP	GBS	F1 from different parental lines	Fruit weight Sugar content Acid content	0.650 0.519 0.666	Imai et al. (2019)
iii) Apricot	GBLUP	Illumina HiSeq 2000	Varieties and their full sib families	Fruit weight distribution	0.89	Minamikawa et al. (2017)
	GBLUP	GBS	F1 pseudo-testcross population	Glucose content	0.31	Nsibi et al. (2020)
	RRBLUP; Bayes A, Bayes B, Bayes C, Bayesian LASSO, and BRR			Ethylene content	0.78	
iv) Pear	GBLUP	GBS	Full sib families	Crispness Sweetness	0.32 0.62	Kumar et al. (2019)
v) Capsicum	GBLUP, Bayesian LASSO, Bayes B, Bayes C, and RKHS	GBS	Core collection and RIL population	Fruit length Fruit width Fruit shape Fruit weight	0.32 0.50 0.34 0.48	Hong et al. (2020)
vi) Tomato	RRBLUP	Infinium Assay	Germplasm	Fruit weight Firmness Soluble solids Sugar content Acidity Biochemical profile	0.814 0.614 0.714 0.649 0.619 0.126–0.705	Duangjit et al. (2016)

there interactions. GS is superior to MAS, and the prediction efficiency is also higher for abiotic stress tolerance (Cerrudo et al., 2018). The usefulness of GS has been shown in wheat, maize, and rice for drought and heat tolerance.

Beyene et al. (2015) have reported a gain of 0.086 t/ha for grain yield, following the rapid cycling GS strategy in eight biparental populations of maize under drought conditions, and a final gain of 0.176 t/ha after three cycles of selection. This increased the genetic gain as the time required for selection was reduced significantly as compared to that of the conventional breeding scheme, where it was three times higher with phenotypic selection. Similarly, Das et al. (2020) reported a genetic gain of 0.110 and 0.135 t/ha/yr for grain yield under drought and 0.038 and 0.113 t/ha/yr under water logging in two maize populations, viz., Maize Yellow Synthetic 1 and Maize Yellow Synthetic 2, respectively, following rapid cycling genomic selection. Vivek et al. (2017) compared the performances of second cycle selection through phenotypic and rapid cycle genomic selection and found 10–20% superiority using the latter. Genomic prediction accuracies using multi-environment models for drought stress tolerance were higher than those using single-environment models in rice and wheat (Sukumaran et al., 2018; Bhandari et al., 2019). Prediction accuracies were higher for heat and drought stress in case of wheat when secondary traits contributing to yield were considered under stress rather than yield *per se* using genomic prediction (Rutkoski et al., 2016). Comparative analysis among different models leads to the conclusion that multi-trait models are superior when selection is carried out in severe drought conditions, while the random regression model was better than the repeatability model and multi-trait model under normal drought conditions and also use of secondary high-throughput traits in genomic prediction improved accuracies by ~70% (Sun et al., 2017).

Quality Improvement

Quality traits have varied genetic architectures, some being controlled oligogenically like grain color, while others are polygenic in nature, viz., grain size and protein content (Battenfield et al., 2016). GS has been carried out in wheat extensively for quality-related traits, viz., milling and flour quality, and when prediction accuracies were compared in biparental and multi-family populations, it was concluded that the prediction accuracies in multi-family populations were better (Heffner et al., 2011).

Protein content is known to be negatively correlated with yield due to physiological compensation (Lam et al., 1996). Michel et al. (2019) employed multi-trait genomic selection for grain yield, protein content, and dough rheological traits for efficient selection with optimized yield and protein content with better quality. The prediction accuracy for the quality traits depends on variability in the germplasm, the relationship among training and prediction populations, etc. (Crossa et al., 2014; Zhao et al., 2015). Joukhadar et al. (2021) used Bayesian regression and BRR for rapid improvement of grain yield as well as mineral content to biofortify wheat and reported Bayesian regression was better in predicting mineral content with an accuracy of 0.55. In rice, grain length and width are important quality parameters, and the

prediction accuracy for these traits ranged from 0.35 to 0.45 and 0.5 to 0.7, respectively, in 110 Japanese rice cultivars employing various GS models (Onogi et al., 2015). In barley, the prediction for quality traits like malting quality (prediction accuracy: 0.4–0.8) has shown the prospects of GS for screening large populations without the need of cost-intensive phenotyping (Schmidt et al., 2016).

GS in Oilseeds

Oilseeds are a source of livelihood to the smallholder farmers in developing countries of Asia and Africa. The yield potential is still to be realized by bridging the yield gap *via* inducing tolerance to biotic and abiotic stresses and improvement in quality (Janila et al., 2016). Different traits related to biotic and abiotic stresses have been mapped, but most of them are qualitative in nature, and the report of GS is limited in such potential crops. Oil quality and yield traits are influenced by the environment and GxE interactions (Patil et al., 2020). Hence, it is important to use the appropriate GS models to account for the GxE effects for accurate selection. Pandey et al. (2020) employed GS in groundnut with different models and validation schemes to account for GxE interaction effects. The model having genomic information generated from the SNP (G), genotypic effect of the line (L), environment effect (E), and their interactions (LxE and GxE) had better mean accuracy (0.58) for all the traits compared to other models. Jan et al. (2016) employed the RRBLUP model for GS in *Brassica* using 950 cross combinations derived from utilizing 475 lines and two testers, for the improvement of oil-specific traits, and the accuracy for oil content and oil yield was 0.81 and 0.75, respectively. Hence, they concluded that the GS model is helpful in pre-selecting superior cross combinations before extensive field evaluation over location and years saving resources. Fikere et al. (2020) employed GS for 22 traits related to yield, disease resistance, and quality in *B. napus* and reported prediction accuracy was highest for yield (0.69) followed by oil content (0.64) using GBLUP. They also evaluated genomic prediction for compositional fatty acid estimated under rainfed and irrigated conditions and concluded that the prediction accuracies for these traits were lower under non-irrigated conditions. Xiong et al. (2020) employed various prediction models, viz., LASSO, GBLUP, OLS, and OLS post-LASSO, for different traits in *B. napus* and reported the two-stage method OLS post-LASSO to be the most accurate (0.90 and 0.55 for oil content and single plant yield, respectively) with the provision of incorporating GxE interactions. For oil content in sunflower which is highly heritable and additive in nature, Mangin et al. (2017) reported that accuracy based on general combining ability (GCA) and GS were on par, and in case if there is no knowledge about one of the parents of hybrid combination, GS excels the GCA-based predictions. Similar inferences had been made by Reif et al. (2013) for the prediction of hybrid performance in sunflower.

From a cross between cultivated and wild progenitors of soybean (*G. max* X *G. sojae*), Beche et al. (2021) reported that the yield-related alleles were associated with the cultivated elite line, but the protein content alleles were from the wild progenitor. The difference in the distribution of trait-contributing alleles in such crosses has a greater impact on

their predictive accuracy. When each allele is distributed equally in the population, the predictive accuracy for both the alleles is the same. In such cases, it is obvious that the less frequent allele's prediction is biased downward. Contiguous breeding programs are very common where new cross combinations are added each year. In such cases, using nested association mapping (NAM) population is better in terms of prediction accuracy (for yield 0.68 and oil and for protein content 0.76) than biparental population, showing the potential of NAM where connectedness is there among the population on the basis of the common parent (Beche et al., 2021). Similarly, Stewart-Brown et al. (2019) have reported that, for better predictions in soybean, it is important to have good relatedness among training and breeding populations. They have observed that the size of the training population has a larger effect on the prediction accuracy, compared to the marker density, but increasing the training population sizes beyond a limit had a diminishing return on the prediction accuracy. Hu et al. (2011) applied GS for biological process, i.e., embryogenesis capacity in soybean, and reported a good prediction accuracy (0.78).

GS in Pulses

In lentil, Haile et al. (2020) showed that if large-effect QTLs were present in the population, multi-trait–based Bayes B is the best GS model, while single-trait GS (STGS) is suitable in their absence. They also reported that, for low heritable traits with GxE interactions, MTGS improves predictability. Considering quality traits in *Phaseolus*, i.e., cooking time for screening of fast culinary genotypes, Diaz et al. (2021) evaluated GS using different populations (RIL, MAGIC, Andean, and Mesoamerican breeding lines). The trait was highly heritable (0.64–0.89), and genomic prediction accuracies for cooking time using MAGIC population were promising and high (0.55) compared to those of Mesoamerican genotypes (0.22).

Under the circumstance of less connectedness in the training and prediction populations, markers generated using the whole genome re-sequencing (WGRS) platform increase the prediction accuracy; however, Li et al. (2018) proposed first identifying causal variants and then utilizing them into the prediction. The prediction accuracy was 0.148–0.186 for yield under drought when using all the SNP from WGRS, but when filtered yield-related causal SNPs were employed, it was observed that prediction accuracy significantly improved (0.56–0.61). Diaz et al. (2021) employed GS for root rot resistance and reported high prediction accuracies (0.7–0.8) for both rots (*Pythium* and *Fusarium*) in *Phaseolus* and proposed it to be promising for improving quantitative tolerance.

GS in Horticultural Crops

Fruit and vegetables are indispensable in achieving nutritional security. However, the problem associated with their breeding, especially of fruits, has its own limitations, viz., long juvenile phase and highly heterozygous nature. Therefore, genetic gain is not much as per the Lush equation. In such crops, GS can be a perfect tool where prediction of performance for quality- and yield-related traits which are under a complex genetic system can

be utilized to improve selection accuracy and efficiency in developing varieties. The success of GS in annual crops has led the horticultural crop breeder to utilize its potential in perennial fruit as well as annual fruit and vegetable crops. Roth et al. (2020) evaluated 537 genotypes in apple for fruit texture traits and performed GS and reported the accuracy up to 0.81. It was suggested to have a large training population from which a tailored training population with a priori genetic relatedness information and ample variation can be formed and utilized to predict the performance of population under consideration. Kumar et al. (2012) have shown high prediction accuracy in apple for different quality traits utilizing a factorial mating design (0.70–0.90). Imai et al. (2019) reported that ssGBLUP predicts with higher accuracy (0.650, 0.519, and 0.666) than GBLUP (0.642, 0.432, and 0.655) for quality traits in citrus, viz., fruit weight, sugar content, and acid content from population where some individuals are not genotyped using information from genotyped related individuals, hence reducing the cost at hand.

As fruits are perishable produce and the post-harvest attribute of the fruits plays an important role in storability, attempts have been made to employ GS for such traits. In apricot, Nsibi et al. (2020) reported prediction accuracy ranging from 0.31 to 0.78 for glucose content and ethylene production. Minamikawa et al. (2017) compared different models of GS for fruit weight distribution among two groups of fruit sizes and reported that, among a large fruit size group, rrGBLUP (0.89) was superior to GBLUP (0.74) and the same was in the case of a small fruit size group, i.e., rrGBLUP (0.32) and GBLUP (0.30). Also, it was proposed to have breeding population or combined parental and breeding population as training population to have better accuracy than only having parental as training population which was consistent for all the quality-related traits. Kumar et al. (2019) employed GS in pear for various fruit quality traits ranging from texture to taste and observed the prediction accuracy ranged from 0.32 to 0.62 averaging to 0.42 and also suggested that training population should be multi-generational and evaluated rigorously over location and time, to have better prediction accuracy. Various GS models have been evaluated for different fruit-related traits in capsicum and reported that RKHS had better accuracy ranging from 0.75 to 0.82 and positively correlated with the number of markers (Hong et al., 2020). GS is also performed to evaluate the accuracy of prediction of different biochemical parameters important for fruit quality in tomato which ranged from 0.13 to 0.70 for aspartate content and also for other traits, viz., fruit weight (0.81), firmness (0.61), soluble solids (0.71), sugar content (0.65), and acidity (0.62) (Duangjit et al., 2016).

STATISTICAL TOOLS FOR IMPLEMENTING GENOMIC SELECTION

Several tools and packages have been developed for the evaluation of genomic prediction and implementation of GS, some of which are discussed below.

GMStool

It is a genome-wide association study (GWAS)-based tool for genomic prediction using genome-wide marker data. It searches for the optimum number of markers for prediction using appropriate statistical and machine learning/deep learning–based models and chooses the best prediction model (Jeong et al., 2020). Furthermore, it identifies SNP markers with the lowest p-values (e.g., top 100 markers) in the GWAS and then chooses the relevant markers set to be included in the final prediction model. GMStool is R-based and freely available through the GitHub repository at https://github.com/ JaeYoonKim72/GMStool. The whole process or its algorithm is basically divided into three steps: data preparation, marker selection, and final prediction model. The detailed procedure of GMStool is discussed below.

Step 1: Input data are divided into training and test sets (user defined)

Step 2: The training set is further divided into small datasets for performing cross validation (i.e., k-folds, for example, five or ten folds) followed by marker selection in each group or fold. The process of marker selection is performed in each fold/group simultaneously.

Step 3: The selected marker from each fold is integrated into the final marker set for updating the model. Appropriate statistical and machine learning–based models are then used for genomic prediction.

solGS

It is an open-source tool based on the Linux operating system. The workflow of the tool is broadly divided into two steps, i.e., training of the prediction model and obtaining GEBV. However, there are three approaches available for training the prediction model, i.e., trait-based approach, trial approach, and custom lists approach. Here, model input and output could be visualized graphically and can be interactively explored or downloaded. It is designed to store a large amount of genotypic, phenotypic, and experimental data. In the background, it basically uses two R-based packages, i.e., nlme (Pinheiro et al., 2017) for data preprocessing and rrBLUP (Endelman, 2011) for statistical modeling. solGS was earlier used by the NEXTGEN Cassava project (http://nextgencassava. org) and implemented at the Cassavabase website (http:// cassavabase.org/solgs).

rrBLUP

It is an R package based on BLUP, which is a mixed linear model framework (Endelman, 2011). It is one of the most widely used packages for genomic prediction in animal and plant breeding. This package estimates the marker effects from training datasets and ultimately estimates the GEBV for the selection candidates. The *mixed.solve* function, a linear mixed model equation which estimates marker effects and GEBV, is one of the most commonly used functions of this package. An additive relationship matrix of individuals can be calculated using genotypic data for the estimation of GEBV using GBLUP. rrBLUP is an open-source package and can be accessed at https://CRAN.R-project.org/ package=rrBLUP.

BWGS

It is an integrated pipeline based on R and freely available at https://CRAN.R-project.org/package=BWGS. The BWGS (i.e., BreedWheat Genomic Selection) pipeline (Charmet et al., 2020) basically consists of three modules: i) missing data imputation, ii) dimension reduction, i.e., reducing the number of markers as it could enhance the speed of computation on large datasets, and iii) estimation of GEBV. It has a wide choice of totally 15 parametric and non-parametric statistical models for estimation of GEBV for selection candidates. It could be used for estimation of GEBV for a wide range of genetic architectures. This tool comprises mainly two functions: *bwgs.cv* and *bwgs.predict*. The former is used for missing value imputation, dimension reduction, and cross validation, while the later is used for model calibration and estimation of GEBV for selection candidates.

BGLR

This package is basically an extension of the BLR package (Perez and Campos, 2014). It can be used to implement several Bayesian models and also provides flexibility in terms of prior density distribution. Here, the response to be considered could be continuous or categorical (either binary or ordinal). It is freely available in the public domain through the CRAN mirror at https://CRAN.R-project.org/package=BGLR.

GenSel

The GenSel software program was developed and implemented under the BIGS (Bioinformatics to Implement Genomic Selection) project (Fernando and Garrick, 2009). It is used for estimation of molecular marker–based breeding values of animals for the trait of interest. This can serve the purpose through the command line (MAC or Linux) interface or as a user-friendly tool. The jobs are submitted and assigned in the queue for analysis. The software uses the Bayesian approach in the background for estimation of marker effects from the training data and further for estimation of GEBV for breeding candidates. This software program can be accessed at https://github.com/austin-putz/GenSel.

GSelection

This is an R-based package and is freely available at https://CRAN.R-project.org/package=GSelection. The package comprises of a set of functions to select the important markers and estimates the GEBV of selection candidates using an integrated model framework (Majumdar et al., 2019). The motivation behind this package is that not a single method performs best in case of all crop plants or animal breeding programs as they may have diverse genetic architectures, i.e., additive and non-additive genetic effects. This package has been developed by integrating the best performing model from each category of additive and non-additive genetic models.

lme4GS

lme4GS is an R-based package freely available and can be accessed through the GitHub repository at https://github.com/perpdgo/lme4GS. It is an extension of the lme4 R package, which is the standard package for fitting linear mixed models. lme4GS package is basically motivated from existing R packages pedigreemm (Vazquez et al., 2010) and lme4qtl (Ziyatdinov et al., 2018). lme4GS package can also be considered an extension of the rrBLUP (Endelman, 2011) package. Further, lme4GS package can be used for fitting mixed models with covariance structures defined by the user, bandwidth selection, and genomic prediction.

STGS

It is an R-based package developed for genomic predictions by estimating marker effects, and the same is further used for calculation of genotypic merit of individuals, i.e., GEBV. GS may be based on single-trait or multi-trait information. This package performs genomic selection only for a single trait, hence named STGS, i.e., single-trait genomic selection (Budhlakoti et al., 2019a). STGS is a comprehensive package which gives a single-step solution for genomic selection based on most commonly used statistical methods (i.e., RR, BLUP, LASSO, SVM, ANN, and RF). It is freely available through the CRAN server at https://CRAN.R-project.org/package=STGS.

MTGS

It is an R-based package developed for genomic predictions by estimating marker effects based on information available on multiple traits. Currently, STGS methods could not utilize additional information available when using multi-trait data. The package MTGS performs genomic selection using multi-trait information (Budhlakoti et al., 2019b). MTGS is a comprehensive package which gives a single-step solution for genomic selection using various MTGS-based methods (MRCE, MLASSO, i.e., multivariate LASSO, and KMLASSO, i.e., kernelized multivariate LASSO). It is freely available through the CRAN server at https://CRAN.R-project.org/package=MTGS.

FACTORS AFFECTING GENOMIC PREDICTION: EFFECTS OF MARKER DENSITY, POPULATION SIZE, TRAIT ARCHITECTURE, AND HERITABILITY

In general, increased marker density enhances the prediction accuracy using most of the GS models such as BLUP, LASSO, machine learning–based, or deep learning–based methods. However, there may be a chance of slow convergence in methods like Bayesian (Bayes A, Bayes B, Bayes Cπ, and Bayes Dπ), where convergence in terms of MCMC (i.e., Markov chain Monte Carlo) iteration is required (Arruda et al., 2016; Zhang et al., 2017; Norman et al., 2018; Zhang et al., 2019). Sometimes, low-density markers of a few hundreds to thousands also enable high prediction accuracies in breeding populations provided that there is a strong LD among the markers; however, it may be trait specific

and may vary with the architecture and heritability of studied traits (Lorenz et al., 2011; Werner et al., 2018). Also sometimes keeping a very high density of markers may have economic constraints as incorporation of such aspects into evaluation of GS strategies is also necessary for a profitable and efficient GS. Therefore, it is always difficult to give a benchmark for the number of markers to be used in such genomic studies; however, it is advisable to keep a moderate density, at least 2000 SNPs, so that prediction accuracy could not be significantly hampered (Abed et al., 2018). However, the cost of genotyping can also be significantly reduced by increasing the level of multiplexing without paying any penalty in terms of genomic prediction accuracy (e.g., genotyping a single line by GBS (96-plex) can cost 3.75 and 4.25 times less than using 9 K and 50 K arrays, respectively, in barley) (Abed et al., 2018). The position of SNPs and how they are placed in genomic arrangements over the chromosome may have a key role, for example, SNPs located in the intergenic space are slightly better at capturing the underlying haplotype diversity related to SNPs located in the genic space as the intergenic space is a playground of many important regulatory sequences, such as promoters and enhancers (Barrett et al., 2012; Abed et al., 2018). The use of high-quality SNP genotyping data (i.e., minor allele frequency (MAF)>0.1) could also be suggested to achieve a good prediction accuracy.

Population size has a significant role in the prediction accuracy whether it is conventional MAS or genomic selection, especially training population. If the population size or training population size is small, it is obvious that a decrease in accuracy is expected because the model will poorly estimate the marker effects and hence prediction accuracy. However, as an idea or estimate for the size of training population as $2*N_e*L$ (where N_e is the effective population size and L is the genome size in Morgan) and the number of markers as $10*N_e*L$ to achieve a prediction accuracy of 0.9 and reducing the size of the training population to $1*N_e*L$ results in a prediction accuracy of 0.7, provided that training population and breeding population are unrelated or both separated by many generations (Meuwissen, 2009). However, for most of the cases, training population and breeding population are related, so high genomic prediction accuracy could be achieved with a training population size much smaller than that referred above (Meuwissen, 2009).

Apart from these factors, prediction accuracy can also be affected by trait heritability especially for lower heritability ($h^2 < 0.4$) (Hayes et al., 2009). Numerous studies up-to-date showed that genomic selection accuracy is strongly influenced by trait heritability, i.e., the fraction of the phenotypic variance to the genetic variance of studied traits. Generally, it is assumed that the target trait with high heritability has good prediction accuracies and vice versa. However, as most of the agricultural traits have low to moderate heritability, it poses a challenge to genomic selection studies, especially in plants. However, low heritability traits would require a larger training population in order to attain the same prediction

accuracy as in the case of traits with moderate to high heritability. However, to achieve this goal, sometimes cost may be a limiting factor, especially in developing countries. Moreover, it could be observed from the available literature that even for low heritable and complex traits, the performance of BLUP and its derivatives (e.g., GBLUP and RRBLUP), Bayesian methods (Bayes A, Bayes B, Bayes Cπ, and Bayes Dπ), and RKHS seems to be robust as compared to their counterparts (Crossa et al., 2010; Crossa et al., 2011; Heffner et al., 2011; Poland et al., 2012; Zhao et al., 2013; Spindel et al., 2015; Crossa et al., 2017; Wang et al., 2018; Xu et al., 2018; Juliana et al., 2019; Lozada et al., 2019; Michel et al., 2019), and at the same time, most of the models work fine with highly heritable traits, although the most suitable method is usually case-dependent. Sometimes missing observations also poses a challenge in estimating GEBV. However, the issue of low heritable trait and missing observation could be handled simultaneously, provided that data are available on multiple traits. In multiple traits, if we have few traits with low heritability and at the same time we have a good correlation with other highly heritable traits, i.e., by using the appropriate MTGS-based model, we can borrow information from other traits. In such scenarios, by using the MTGS model, we can estimate the GEBV more precisely and accurately.

CONCLUSION AND PROSPECTS

Genomic selection has shown its potential in plant and animal breeding research by increasing genetic gains in the last two decades. Revolution in terms of cheaper NGS technologies has made it possible to sequence the crop and animal genomes at a relatively low cost. It resulted in a number of completely sequenced crop and animal genomes with high-density SNP genotyping chips and their availability in the public domain, which may further boost the predictive ability of a GS model. Even after more than a decade in the field of genomic selection studies, still there is a lot of scope for improvement in this area. Methodological refinements (such as imputation of missing genotypic value, implementation of GxE interaction, information on epigenetic regulation, haplotypes, and including multi-trait information into prediction models) will be definitely helpful for a successful implementation of GS in plant and animal breeding programs. Consistent updation of the training set for GS is highly desirable by including the new markers in each generation. Evaluation of the training populations should be done in controlled and well-managed conditions as it significantly affects the performance of prediction models. There is a need for a structured program in the field of genomic selection including human resource development, advanced data recording methodologies, and trait phenotyping in order to come out with fruitful outcomes.

AUTHOR CONTRIBUTIONS

NB, AK, AR, and DM contributed to conceptualization. NB, AK, KC, AK, AP, RK, and UK reviewed and edited the paper. PJ, DM, and SK contributed to the final editing and correction. All authors contributed to the manuscript and approved the submitted version.

REFERENCES

Abed, A., Pérez-Rodríguez, P., Crossa, J., and Belzile, F. (2018). When Less Can Be Better: How Can We Make Genomic Selection More Cost-Effective and Accurate in Barley? *Theor. Appl. Genet.* 131, 1873–1890. doi:10.1007/S00122-018-3120-8

Almeida, G. D., Makumbi, D., Magorokosho, C., Nair, S., Borém, A., Ribaut, J.-M., et al. (2013). QTL Mapping in Three Tropical maize Populations Reveals a Set of Constitutive and Adaptive Genomic Regions for Drought Tolerance. *Theor. Appl. Genet.* 126, 583–600. doi:10.1007/S00122-012-2003-7

Arelli, P. R., Young, L. D., and Concibido, V. C. (2009). Inheritance of Resistance in Soybean PI 567516C to LY1 Nematode Population Infecting Cv. Hartwig. *Euphytica* 165, 1–4. doi:10.1007/S10681-008-9760-Z

Arelli, P. R., Young, L. D., and Mengistu, A. (2006). Registration of High Yielding and Multiple Disease-Resistant Soybean Germplasm JTN-5503. *Crop Sci.* 46, 2723–2724. doi:10.2135/cropsci2005.12.0471crg

Arruda, M. P., Lipka, A. E., Brown, P. J., Krill, A. M., Thurber, C., Brown-Guedira, G., et al. (2016). Comparing Genomic Selection and Marker-Assisted Selection for Fusarium Head Blight Resistance in Wheat (*Triticum aestivum* L.). *Mol. Breed.* 36, 84. doi:10.1007/s11032-016-0508-5

Badu-Apraku, B., Talabi, A. O., Fakorede, M. A. B., Fasanmade, Y., Gedil, M., Magorokosho, C., et al. (2019). Yield Gains and Associated Changes in an Early Yellow Bi-parental maize Population Following Genomic Selection for Striga Resistance and Drought Tolerance. *BMC Plant Biol.* 19, 129. doi:10.1186/S12870-019-1740-Z

Baenziger, P. S. (2016). "Wheat Breeding and Genetics," in *Reference Module in Food Science*. Editor C. Beddows (Amsterdam, Netherlands: Elsevier). doi:10.1016/B978-0-08-100596-5.03001-8

Barrett, L. W., Fletcher, S., and Wilton, S. D. (2012). Regulation of Eukaryotic Gene Expression by the Untranslated Gene Regions and Other Non-coding Elements. *Cell. Mol. Life Sci.* 69, 3613–3634. doi:10.1007/S00018-012-0990-9

Battenfield, S. D., Guzmán, C., Gaynor, R. C., Singh, R. P., Peña, R. J., Dreisigacker, S., et al. (2016). Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program. *Plant Genome* 9, 1–12. doi:10.3835/PLANTGENOME2016.01.0005

Beche, E., Gillman, J. D., Song, Q., Nelson, R., Beissinger, T., Decker, J., et al. (2021). Genomic Prediction Using Training Population Design in Interspecific Soybean Populations. *Mol. Breed.* 41, 1–15. doi:10.1007/S11032-021-01203-6

Ben Hassen, M., Bartholomé, J., Valè, G., Cao, T.-V., and Ahmadi, N. (2018). Genomic Prediction Accounting for Genotype by Environment Interaction Offers an Effective Framework for Breeding Simultaneously for Adaptation to an Abiotic Stress and Performance under normal Cropping Conditions in rice. *G3 Genes, Genomes, Genet.* 8, 2319–2332. doi:10.1534/g3.118.200098

Bernardo, R. (2008). Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. *Crop Sci.* 48, 1649–1664. doi:10.2135/CROPSCI2008.03.0131

Bernardo, R., and Yu, J. (2007). Prospects for Genomewide Selection for Quantitative Traits in maize. *Crop Sci.* 47, 1082–1090. doi:10.2135/CROPSCI2006.11.0690

Beyene, Y., Semagn, K., Mugo, S., Tarekegne, A., Babu, R., Meisel, B., et al. (2015). Genetic Gains in Grain Yield through Genomic Selection in Eight Bi-parental Maize Populations under Drought Stress. *Crop Sci.* 55, 154–163. doi:10.2135/CROPSCI2014.07.0460

Bhandari, A., Bartholomé, J., Cao-Hamadoun, T.-V., Kumari, N., Frouin, J., Kumar, A., et al. (2019). Selection of Trait-specific Markers and Multi-Environment Models Improve Genomic Predictive Ability in rice. *PLOS ONE* 14, e0208871. doi:10.1371/JOURNAL.PONE.0208871

Budhlakoti, N., Mishra, D. C., Rai, A., and Chaturvedi, K. K. (2019a). Package 'STGS'. 1–11. Available at: https://cran.r-project.org/web/packages/STGS/STGS.pdf.

Budhlakoti, N., Mishra, D. C., and Rai, A. (2019b). Package 'MTGS'. 1–6.

Budhlakoti, N., Mishra, D. C., Rai, A., Lal, S. B., Chaturvedi, K. K., and Kumar, R. R. (2019c). A Comparative Study of Single-Trait and Multi-Trait Genomic Selection. *J. Comput. Biol.* 26, 1100–1112. doi:10.1089/CMB.2019.0032

Budhlakoti, N., Rai, A., and Mishra, D. C. (2020a). Effect of Influential Observation in Genomic Prediction Using LASSO Diagnostic. *Indian J. Agric. Sci.* 90, 1155–1159.

Budhlakoti, N., Rai, A., Mishra, D. C., Jaggi, S., Kumar, M., and Rao, A. R. (2020c). Comparative Study of Different Non-parametric Genomic Selection Methods under Diverse Genetic Architecture. *Ijgpb* 80, 395–401. doi:10.31742/IJGPB.80.4.4

Budhlakoti, N., Rai, A., and Mishra, D. C. (2020b). Statistical Approach for Improving Genomic Prediction Accuracy through Efficient Diagnostic Measure of Influential Observation. *Sci. Rep.* 10, 1–11. doi:10.1038/s41598-020-65323-3

Burgueño, J., de los Campos, G., Weigel, K., and Crossa, J. (2012). Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction Using Pedigree and Dense Molecular Markers. *Crop Sci.* 52, 707–719. doi:10.2135/CROPSCI2011.06.0299

Cai, X., Huang, A., and Xu, S. (2011). Fast Empirical Bayesian LASSO for Multiple Quantitative Trait Locus Mapping. *BMC Bioinformatics* 12, 211–213. doi:10.1186/1471-2105-12-211/FIGURES/5

Cerrudo, D., Cao, S., Yuan, Y., Martinez, C., Suarez, E. A., Babu, R., et al. (2018). Genomic Selection Outperforms Marker Assisted Selection for Grain Yield and Physiological Traits in a maize Doubled Haploid Population across Water Treatments. *Front. Plant Sci.* 9, 366. doi:10.3389/FPLS.2018.00366/BIBTEX

Charmet, G., Tran, L.-G., Auzanneau, J., Rincent, R., and Bouchet, S. (2020). BWGS: A R Package for Genomic Selection and its Application to a Wheat Breeding Programme. *PLOS ONE* 15, e0222733. doi:10.1371/JOURNAL.PONE.0222733

Cheng, H., Kizilkaya, K., Zeng, J., Garrick, D., and Fernando, R. (2018). Genomic Prediction from Multiple-Trait Bayesian Regression Methods Using Mixture Priors. *Genetics* 209, 89–103. doi:10.1534/GENETICS.118.300650/-/DC1

Chiquet, J., Mary-Huard, T., Robin, S., and Robin, S. (2017). Structured Regularization for Conditional Gaussian Graphical Models. *Stat. Comput.* 27, 789–804. doi:10.1007/s11222-016-9654-1

Clark, S. A., Hickey, J. M., and Van Der Werf, J. H. (2011). Different Models of Genetic Variation and Their Effect on Genomic Evaluation. *Genet. Sel. Evol.* 43, 18. doi:10.1186/1297-9686-43-18

Cooper, M., Podlich, D. W., Micallef, K. P., Smith, O. S., Jensen, N. M., Chapman, S. C., et al. (2002). "Complexity, Quantitative Traits and Plant Breeding: a Role for Simulation Modelling in the Genetic Improvement of Crops," in *Quantitative Genetics, Genomics and Plant Breeding*. Editor M. S. Kang (Wallingford, UK: CAB International), 143–166. doi:10.1079/9780851996011.0143

Crossa, J., Campos, G. d. l., Pérez, P., Gianola, D., Burgueño, J., Araus, J. L., et al. (2010). Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. *Genetics* 186, 713–724. doi:10.1534/GENETICS.110.118521

Crossa, J., Pérez, P., de los Campos, G., Mahuku, G., Dreisigacker, S., and Magorokosho, C. (2011). Genomic Selection and Prediction in Plant Breeding. *J. Crop Improvement* 25, 239–261. doi:10.1080/15427528.2011.558767

Crossa, J., Pérez, P., Hickey, J., Burgueño, J., Ornella, L., Cerón-Rojas, J., et al. (2014). Genomic Prediction in CIMMYT maize and Wheat Breeding Programs. *Heredity* 112, 48–60. doi:10.1038/HDY.2013.16

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., et al. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends Plant Science* 22, 961–975. doi:10.1016/J. TPLANTS.2017.08.011

Cui, Z., Dong, H., Zhang, A., Ruan, Y., He, Y., and Zhang, Z. (2020). Assessment of the Potential for Genomic Selection to Improve Husk Traits in Maize. *G3: Genes, Genomes, Genet.* 10, 3741–3749. doi:10.1534/G3.120.401600

Daetwyler, H. D., Bansal, U. K., Bariana, H. S., Hayden, M. J., and Hayes, B. J. (2014). Genomic Prediction for Rust Resistance in Diverse Wheat Landraces. *Theor. Appl. Genet.* 127, 1795–1803. doi:10.1007/s00122-014-2341-8

Daetwyler, H. D., Hickey, J. M., Henshall, J. M., Dominik, S., Gredler, B., van der Werf, J. H. J., et al. (2010). Accuracy of Estimated Genomic Breeding Values for Wool and Meat Traits in a Multi-Breed Sheep Population. *Anim. Prod. Sci.* 50, 1004–1010. doi:10.1071/AN10096

Das, R. R., Vinayan, M. T., Patel, M. B., Phagna, R. K., Singh, S. B., Shahi, J. P., et al. (2020). Genetic Gains with Rapid-cycle Genomic Selection for Combined Drought and Waterlogging Tolerance in Tropical maize (Zea May S L.). *Plant Genome* 13, 1–15. doi:10.1002/tpg2.20035

de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., and Calus, M. P. L. (2013). Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics* 193, 327–345. doi:10.1534/GENETICS.112.143313

De Roos, A. P. W., Hayes, B. J., and Goddard, M. E. (2009). Reliability of Genomic Predictions across Multiple Populations. *Genetics* 183, 1545–1553. doi:10.1534/ GENETICS.109.104935

Diaz, S., Ariza-Suarez, D., Ramdeen, R., Aparicio, J., Arunachalam, N., Hernandez, C., et al. (2021). Genetic Architecture and Genomic Prediction of Cooking Time in Common Bean (Phaseolus vulgaris L.). *Front. Plant Sci.* 11, 2257. doi:10. 3389/FPLS.2020.622213/BIBTEX

dos Santos, J. P. R., Pires, L. P. M., de Castro Vasconcellos, R. C., Pereira, G. S., von Pinho, R. G., and Balestre, M. (2016). Genomic Selection to Resistance to Stenocarpella Maydis in maize Lines Using DArTseq Markers. *BMC Genet.* 17, 86. doi:10.1186/S12863-016-0392-3

Duangjit, J., Causse, M., and Sauvage, C. (2016). Efficiency of Genomic Selection for Tomato Fruit Quality. *Mol. Breed.* 36, 29. doi:10.1007/S11032-016-0453-3

Endelman, J. B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome* 4, 250–255. doi:10.3835/ PLANTGENOME2011.08.0024

Fernando, R., and Garrick, D. (2009). GenSel- User Manual for a Portfolio of Genomic Selection Related Analyses. Available at: http://taurus.ansci.iastate. edu/Site/Welcome_files/GenSel%20 Manual%20v2.pdf.

Fernando, R., and Grossman, M. (1989). Marker Assisted Selection Using Best Linear Unbiased Prediction. *Genet. Selection Evol.* 21 (421), 467–477. doi:10. 1186/1297-9686-21-4-467

Fikere, M., Barbulescu, D. M., Malmberg, M. M., Maharjan, P., Salisbury, P. A., Kant, S., et al. (2020). Genomic Prediction and Genetic Correlation of Agronomic, Blackleg Disease, and Seed Quality Traits in Canola (Brassica Napus L.). *Plants* 9, 719–19. doi:10.3390/PLANTS9060719

Fones, H. N., Bebber, D. P., Chaloner, T. M., Kay, W. T., Steinberg, G., and Gurr, S. J. (2020). Threats to Global Food Security from Emerging Fungal and Oomycete Crop Pathogens. *Nat. Food* 1, 332–342. doi:10.1038/s43016-020-0075-0

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* 33, 1–22. doi:10.18637/jss.v033.i01

Fristche-Neto, R., Akdemir, D., and Jannink, J.-L. (2018). Accuracy of Genomic Selection to Predict maize Single-Crosses Obtained through Different Mating Designs. *Theor. Appl. Genet.* 131, 1153–1162. doi:10.1007/s00122-018-3068-8

Furbank, R. T., and Tester, M. (2011). Phenomics - Technologies to Relieve the Phenotyping Bottleneck. *Trends Plant Sci.* 16, 635–644. doi:10.1016/J. TPLANTS.2011.09.005

Gianola, D., Fernando, R. L., and Stella, A. (2006). Genomic-Assisted Prediction of Genetic Value with Semiparametric Procedures. *Genetics* 173, 1761–1776. doi:10.1534/GENETICS.105.049510

Gianola, D., Okut, H., Weigel, K. A., and Rosa, G. J. (2011). Predicting Complex Quantitative Traits with Bayesian Neural Networks: a Case Study with Jersey Cows and Wheat. *BMC Genet.* 12, 87. doi:10.1186/1471-2156-12-87

Gopalakrishnan, S., Sharma, R. K., Anand Rajkumar, K., Joseph, M., Singh, V. P., Singh, A. K., et al. (2008). Integrating Marker Assisted Background Analysis with Foreground Selection for Identification of superior Bacterial Blight Resistant Recombinants in Basmati rice. *Plant Breed.* 127, 131–139. doi:10. 1111/J.1439-0523.2007.01458.X

Grenier, C., Cao, T.-V., Ospina, Y., Quintero, C., Châtel, M. H., Tohme, J., et al. (2015). Accuracy of Genomic Selection in a Rice Synthetic Population Developed for Recurrent Selection Breeding. *PloS one* 10, e0136594. doi:10. 1371/JOURNAL.PONE.0136594

Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the Bayesian Alphabet for Genomic Selection. *BMC Bioinformatics* 12, 186–197. doi:10.1186/1471-2105-12-186/FIGURES/2

Habier, D., Fernando, R. L., and Dekkers, J. C. M. (2009). Genomic Selection Using Low-Density Marker Panels. *Genetics* 182, 343–353. doi:10.1534/GENETICS. 108.100289

Habier, D., Fernando, R. L., and Garrick, D. J. (2013). Genomic BLUP Decoded: a Look into the Black Box of Genomic Prediction. *Genetics* 194, 597–607. doi:10. 1534/GENETICS.113.152207

Haile, T. A., Heidecker, T., Wright, D., Neupane, S., Ramsay, L., Vandenberg, A., et al. (2020). Genomic Selection for Lentil Breeding: Empirical Evidence. *Plant Genome* 13, 1–30. doi:10.1002/tpg2.20002

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., and Goddard, M. E. (2009). Invited Review: Genomic Selection in Dairy Cattle: Progress and Challenges. *J. Dairy Sci.* 92, 433–443. doi:10.3168/JDS.2008-1646

Heffner, E. L., Jannink, J.-L., Sorrells, M. E., Heffner, E. L., Sorrells, M. E., Univ, C., et al. (2011). Genomic Selection Accuracy Using Multifamily Prediction Models in a Wheat Breeding Program. *The Plant Genome* 4, 65–75. doi:10.3835/ plantgenome.2010.12.0029

Heffner, E. L., Lorenz, A. J., Jannink, J. L., and Sorrells, M. E. (2010). Plant Breeding with Genomic Selection: Gain Per Unit Time and Cost. *Crop Sci.* 50, 1681–1690. doi:10.2135/cropsci2009.11.0662

Heffner, E. L., Sorrells, M. E., and Jannink, J.-L. (2009). Genomic Selection for Crop Improvement. *Crop Sci.* 49, 1–12. doi:10.2135/CROPSCI2008.08.0512

Henderson, C. R., Kempthorne, O., Searle, S. R., and von Krosigk, C. M. (1959). The Estimation of Environmental and Genetic Trends from Records Subject to Culling. *Biometrics* 15, 192. doi:10.2307/2527669

Holliday, J. A., Wang, T., and Aitken, S. (2012). Predicting Adaptive Phenotypes from Multilocus Genotypes in Sitka Spruce (Picea Sitchensis) Using Random Forest. *Using Random For.* 2, 1085–1093. doi:10.1534/g3. 112.002733

Hong, J. P., Ro, N., Lee, H. Y., Kim, G. W., Kwon, J. K., Yamamoto, E., et al. (2020). Genomic Selection for Prediction of Fruit-Related Traits in Pepper (Capsicum spp.). *Front. Plant Sci.* 11, 570871. doi:10.3389/FPLS.2020. 570871/BIBTEX

Howard, R., Carriquiry, A. L., and Beavis, W. D. (2014). Parametric and Nonparametric Statistical Methods for Genomic Selection of Traits with Additive and Epistatic Genetic Architectures. *G3: Genes, Genomes, Genet.* 4, 1027–1046. doi:10.1534/G3.114.010298/-/DC1

Hu, Z., Li, Y., Song, X., Han, Y., Cai, X., Xu, S., et al. (2011). Genomic Value Prediction for Quantitative Traits under the Epistatic Model. *BMC Genet.* 12, 15. doi:10.1186/1471-2156-12-15

Huang, M., Balimponya, E. G., Mgonja, E. M., McHale, L. K., Luzi-Kihupi, A., Wang, G.-L., et al. (2019). Use of Genomic Selection in Breeding rice (Oryza Sativa L.) for Resistance to rice Blast (Magnaporthe Oryzae). *Mol. Breed.* 39, 1–16. doi:10.1007/S11032-019-1023-2

Imai, A., Kuniga, T., Yoshioka, T., Nonaka, K., Mitani, N., Fukamachi, H., et al. (2019). Single-step Genomic Prediction of Fruit-Quality Traits Using Phenotypic Records of Non-genotyped Relatives in Citrus. *PLOS ONE* 14, e0221880. doi:10.1371/JOURNAL.PONE.0221880

Jan, H. U., Abbadi, A., Lücke, S., Nichols, R. A., and Snowdon, R. J. (2016). Genomic Prediction of Testcross Performance in Canola (Brassica Napus). *PLOS ONE* 11, e0147769. doi:10.1371/JOURNAL.PONE.0147769

Janila, P., Variath, M. T., Pandey, M. K., Desmae, H., Motagi, B. N., Okori, P., et al. (2016). Genomic Tools in Groundnut Breeding Program: Status and Perspectives. *Front. Plant Sci.* 7, 289. doi:10.3389/FPLS.2016.00289/BIBTEX

Jeong, S., Kim, J.-Y., and Kim, N. (2020). GMStool: GWAS-Based Marker Selection Tool for Genomic Prediction from Genomic Data. *Sci. Rep.* 10, 1–12. doi:10. 1038/s41598-020-76759-y

Jia, Y., and Jannink, J.-L. (2012). Multiple-Trait Genomic Selection Methods Increase Genetic Value Prediction Accuracy. *Genetics* 192, 1513–1522. doi:10.1534/GENETICS.112.144246

Joukhadar, R., Thistlethwaite, R., Trethowan, R. M., Hayden, M. J., Stangoulis, J., Cu, S., et al. (2021). Genomic Selection Can Accelerate the Biofortification of spring Wheat. *Theor. Appl. Genet.* 134, 3339–3350. doi:10.1007/S00122-021-03900-4

Juarez, M., Legua, P., Mengual, C. M., Kassem, M. A., Sempere, R. N., Gómez, P., et al. (2013). Relative Incidence, Spatial Distribution and Genetic Diversity of Cucurbit Viruses in Eastern Spain. *Ann. Appl. Biol.* 162, 362–370. doi:10.1111/AAB.12029

Juliana, P., Poland, J., Huerta-Espino, J., Shrestha, S., Crossa, J., Crespo-Herrera, L., et al. (2019). Improving Grain Yield, Stress Resilience and Quality of Bread Wheat Using Large-Scale Genomics. *Nat. Genet.* 51, 1530–1539. doi:10.1038/s41588-019-0496-6

Juliana, P., Singh, R. P., Singh, P. K., Crossa, J., Rutkoski, J. E., Poland, J. A., et al. (2017). Comparison of Models and Whole-Genome Profiling Approaches for Genomic-Enabled Prediction of Septoria Tritici Blotch, Stagonospora Nodorum Blotch, and Tan Spot Resistance in Wheat. *Plant Genome* 10, 1–16. doi:10.3835/PLANTGENOME2016.08.0082

Klápště, J., Dungey, H. S., Telfer, E. J., Suontama, M., Graham, N. J., Li, Y., et al. (2020). Marker Selection in Multivariate Genomic Prediction Improves Accuracy of Low Heritability Traits. *Front. Genet.* 11, 499094. doi:10.3389/FGENE.2020.499094/FULL

Krishnan, S. G., Singh, A. K., Rathour, R., Nagarajan, M., Bhowmick, P. K., Ellur, R. K., et al. (2019). Rice Variety Pusa Samba 1850. *Indian J. Genet.* 79, 109–110.

Kumar, S., Chagné, D., Bink, M. C. A. M., Volz, R. K., Whitworth, C., and Carlisle, C. (2012). Genomic Selection for Fruit Quality Traits in Apple (Malus×domestica Borkh.). *PLoS One* 7, e36674. doi:10.1371/JOURNAL.PONE.0036674

Kumar, S., Kirk, C., Deng, C. H., Shirtliff, A., Wiedow, C., Qin, M., et al. (2019). Marker-trait Associations and Genomic Predictions of Interspecific Pear (Pyrus) Fruit Characteristics. *Sci. Rep.* 9, 1–10. doi:10.1038/s41598-019-45618-w

Lam, H.-M., Coschigano, K. T., Oliveira, I. C., Melo-Oliveira, R., and Coruzzi, G. M. (1996). The Molecular-Genetics of Nitrogen Assimilation into Amino Acids in Higher Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47, 569–593. doi:10.1146/annurev.arplant.47.1.569

Legarra, A., and Reverter, A. (2018). Semi-parametric Estimates of Population Accuracy and Bias of Predictions of Breeding Values and Future Phenotypes Using the LR Method. *Genet. Sel Evol.* 50, 53–18. doi:10.1186/S12711-018-0426-6/FIGURES/3

Li, Y., Ruperao, P., Batley, J., Edwards, D., Khan, T., Colmer, T. D., et al. (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. doi:10.3389/FPLS.2018.00190/BIBTEX

Liu, B., Asseng, S., Müller, C., Ewert, F., Elliott, J., Lobell, D. B., et al. (2016). Similar Estimates of Temperature Impacts on Global Wheat Yield by Three Independent Methods. *Nat. Clim Change* 6, 1130–1136. doi:10.1038/NCLIMATE3115

Liu, X., Wang, H., Wang, H., Guo, Z., Xu, X., Liu, J., et al. (2018). Factors Affecting Genomic Selection Revealed by Empirical Evidence in maize. *Crop J.* 6, 341–352. doi:10.1016/J.CJ.2018.03.005

Long, N., Gianola, D., Rosa, G. J. M., and Weigel, K. A. (2011). Application of Support Vector Regression to Genome-Assisted Prediction of Quantitative Traits. *Theor. Appl. Genet.* 123, 1065–1074. doi:10.1007/S00122-011-1648-Y

Longin, C. F. H., Reif, J. C., and Würschum, T. (2014). Long-term Perspective of Hybrid versus Line Breeding in Wheat Based on Quantitative Genetic Theory. *Theor. Appl. Genet.* 127, 1635–1641. doi:10.1007/S00122-014-2325-8

Lorenz, A. J., Chao, S., Asoro, F. G., Heffner, E. L., Hayashi, T., Iwata, H., et al. (2011). Genomic Selection in Plant Breeding. *Adv. Agron.* 110, 77–123. doi:10.1016/B978-0-12-385531-2.00002-5

Lorenz, A. J., Smith, K. P., and Jannink, J. L. (2012). Potential and Optimization of Genomic Selection for Fusarium Head Blight Resistance in Six-Row Barley. *Crop Sci.* 52, 1609–1621. doi:10.2135/cropsci2011.09.0503

Lorenzana, R. E., and Bernardo, R. (2009). Accuracy of Genotypic Value Predictions for Marker-Based Selection in Biparental Plant Populations. *Theor. Appl. Genet.* 120, 151–161. doi:10.1007/s00122-009-1166-3

Lozada, D. N., and Carter, A. H. (2020). Genomic Selection in Winter Wheat Breeding Using a Recommender Approach. *Genes* 11, 1–14. doi:10.3390/GENES11070779

Lozada, D. N., Mason, R. E., Sarinelli, J. M., and Brown-Guedira, G. (2019). Accuracy of Genomic Selection for Grain Yield and Agronomic Traits in Soft Red winter Wheat. *BMC Genet.* 20, 1–12. doi:10.1186/s12863-019-0785-1

Luan, T., Woolliams, J. A., Lien, S., Kent, M., Svendsen, M., and Meuwissen, T. H. E. (2009). The Accuracy of Genomic Selection in Norwegian Red Cattle Assessed by Cross-Validation. *Genetics* 183, 1119–1126. doi:10.1534/GENETICS.109.107391

Lush, J. L. (1943). *Animal Breeding Plans.* Edn 2. Charleston, South Carolina: Bibliolife DBA of Bibilio Bazaar.

Mackay, T. F. C. (2001). The Genetic Architecture of Quantitative Traits. *Annu. Rev. Genet.* 35, 303–339. doi:10.1146/annurev.genet.35.102401.090633

Madhavi, K. R., Rambabu, R., Abhilash Kumar, V., Vijay Kumar, S., Aruna, J., Ramesh, S., et al. (2016). Marker Assisted Introgression of Blast (Pi-2 and Pi-54) Genes in to the Genetic Background of Elite, Bacterial Blight Resistant Indica rice Variety, Improved Samba Mahsuri. *Euphytica* 212, 331–342. doi:10.1007/S10681-016-1784-1

Maenhout, S., De Baets, B., Haesaert, G., and Van Bockstaele, E. (2007). Support Vector Machine Regression for the Prediction of maize Hybrid Performance. *Theor. Appl. Genet.* 115, 1003–1013. doi:10.1007/s00122-007-0627-9

Majumdar, S. G., Rai, A., and Mishra, D. C. (2020). Integrated Framework for Selection of Additive and Nonadditive Genetic Markers for Genomic Selection. *J. Comput. Biol.* 27, 845–855. doi:10.1089/CMB.2019.0223

Majumdar, S. G., Rai, A., and Mishra, D. C. (2019). Package 'GSelection', 1–14. Available at: https://rdrr.io/cran/GSelection/man/GSelection-package.html.

Mangin, B., Bonnafous, F., Blanchet, N., Boniface, M.-C., Bret-Mestries, E., Carrère, S., et al. (2017). Genomic Prediction of sunflower Hybrids Oil Content. *Front. Plant Sci.* 8, 1633. doi:10.3389/FPLS.2017.01633/BIBTEX

Marulanda, J. J., Mi, X., Melchinger, A. E., Xu, J.-L., Würschum, T., and Longin, C. F. H. (2016). Optimum Breeding Strategies Using Genomic Selection for Hybrid Breeding in Wheat, maize, rye, Barley, rice and Triticale. *Theor. Appl. Genet.* 129, 1901–1913. doi:10.1007/s00122-016-2748-5

Meuwissen, T. H. (2009). Accuracy of Breeding Values of 'unrelated' Individuals Predicted by Dense SNP Genotyping. *Genet. Sel Evol.* 41, 35–39. doi:10.1186/1297-9686-41-35/TABLES/3

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-wide Dense Marker Maps. *Genetics* 157, 1819–1829. doi:10.1093/GENETICS/157.4.1819

Michel, S., Löschenberger, F., Ametz, C., Pachler, B., Sparry, E., and Bürstmayr, H. (2019). Combining Grain Yield, Protein Content and Protein Quality by Multi-Trait Genomic Selection in Bread Wheat. *Theor. Appl. Genet.* 132, 2767–2780. doi:10.1007/S00122-019-03386-1

Minamikawa, M. F., Nonaka, K., Kaminuma, E., Kajiya-Kanegae, H., Onogi, A., Goto, S., et al. (2017). Genome-wide Association Study and Genomic Prediction in Citrus: Potential of Genomics-Assisted Breeding for Fruit Quality Traits. *Sci. Rep.* 7, 1–13. doi:10.1038/s41598-017-05100-x

Mirdita, V., He, S., Zhao, Y., Korzun, V., Bothe, R., Ebmeyer, E., et al. (2015). Potential and Limits of Whole Genome Prediction of Resistance to Fusarium Head Blight and Septoria Tritici Blotch in a Vast Central European Elite winter Wheat Population. *Theor. Appl. Genet.* 128, 2471–2481. doi:10.1007/S00122-015-2602-1

Mishra, D. C., Budhlakoti, N., Majumdar, S. G., and Rai, A. (2021). Innovations in Genomic Selection : Statistical Perspective. 101–111. Available at: https://ssca.org.in/media/9_Spl_Proceedings_2021_006072021_Dwijesh_Mishra_Final.pdf.

Mohan, M., Nair, S., Bhagwat, A., Krishna, T. G., Yano, M., Bhatia, C. R., et al. (1997). Genome Mapping, Molecular Markers and Marker-Assisted Selection in Crop Plants. *Mol. Breed.* 3, 87–103. doi:10.1023/A:1009651919792

Moore, J. H., and Williams, S. M. (2009). Epistasis and its Implications for Personal Genetics. *Am. J. Hum. Genet.* 85, 309–320. doi:10.1016/J.AJHG.2009.08.006

Nakaya, A., and Isobe, S. N. (2012). Will Genomic Selection Be a Practical Method for Plant Breeding? *Ann. Bot.* 110, 1303–1316. doi:10.1093/AOB/MCS109

Neeraja, C. N., Maghirang-Rodriguez, R., Pamplona, A., Heuer, S., Collard, B. C. Y., Septiningsih, E. M., et al. (2007). A Marker-Assisted Backcross Approach for Developing Submergence-Tolerant rice Cultivars. *Theor. Appl. Genet.* 115 (6), 767–776. doi:10.1007/s00122-007-0607-0

Norman, A., Taylor, J., Edwards, J., and Kuchel, H. (2018). Optimising Genomic Selection in Wheat: Effect of Marker Density, Population Size and Population

Structure on Prediction Accuracy. *G3 Genes|Genomes|Genetics* 8, 2889–2899. doi:10.1534/G3.118.200311

Nsibi, M., Gouble, B., Bureau, S., Flutre, T., Sauvage, C., Audergon, J.-M., et al. (2020). Adoption and Optimization of Genomic Selection to Sustain Breeding for Apricot Fruit Quality. *G3 Genes|Genomes|Genetics* 10, 4513–4529. doi:10.1534/G3.120.401452

Ogutu, J. O., Schulz-Streeck, T., and Piepho, H.-P. (2012). Genomic Selection Using Regularized Linear Regression Models: ridge Regression, Lasso, Elastic Net and Their Extensions. *BMC Proc.* 6, S10. doi:10.1186/1753-6561-6-S2-S10

Onogi, A., Ideta, O., Inoshita, Y., Ebana, K., Yoshioka, T., Yamasaki, M., et al. (2015). Exploring the Areas of Applicability of Whole-Genome Prediction Methods for Asian rice (Oryza Sativa L.). *Theor. Appl. Genet.* 128, 41–53. doi:10.1007/S00122-014-2411-Y

Pandey, M. K., Chaudhari, S., Jarquin, D., Janila, P., Crossa, J., Patil, S. C., et al. (2020). Genome-based Trait Prediction in Multi- Environment Breeding Trials in Groundnut. *Theor. Appl. Genet.* 133, 3101–3117. doi:10.1007/S00122-020-03658-1/TABLES/5

Pérez, P., and de los Campos, G. (2014). Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics* 198, 483–495. doi:10.1534/genetics.114.164442

Pérez-Rodríguez, P., Gianola, D., González-Camacho, J. M., Crossa, J., Manès, Y., and Dreisigacker, S. (2012). Comparison between Linear and Non-parametric Regression Models for Genome-Enabled Prediction in Wheat. *G3: Genes, Genomes, Genet.* 2, 1595–1605. doi:10.1534/G3.112.003665/-/DC1

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., and Van Willigen, B. (2017). Package 'nlme'. Available at: https://cran.r-project.org/web/packages/nlme/nlme.pdf.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., et al. (2012). Genomic Selection in Wheat Breeding Using Genotyping-by-Sequencing. *Plant Genome* 5, 1–11. doi:10.3835/PLANTGENOME2012.06.0006

Qin, F., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2011). Achievements and Challenges in Understanding Plant Abiotic Stress Responses and Tolerance. *Plant Cel Physiol.* 52, 1569–1582. doi:10.1093/PCP/PCR106

Rai, K. N., Hash, C. T., Singh, A. K., and Velu, G. (2008). Adaptation and Quality Traits of a Germplasm-Derived Commercial Seed Parent of Pearl Millet. *Plant Genet. Resour. Newsl.* 154, 20–24.

Reif, J. C., Zhao, Y., Würschum, T., Gowda, M., and Hahn, V. (2013). Genomic Prediction of sunflower Hybrid Performance. *Plant Breed* 132, 107–114. doi:10.1111/pbr.12007

Reynolds, M. P., and Ortiz, R. (2010). "Adapting Crops to Climate Change: a Summary," in *Climate Change and Crop Production*. Editor M. P. Reynolds (Wallingford,UK: CABI), 1–8. doi:10.1079/9781845936334.0001

Ribaut, J.-M., and Ragot, M. (2007). Marker-assisted Selection to Improve Drought Adaptation in maize: the Backcross Approach, Perspectives, Limitations, and Alternatives. *J. Exp. Bot.* 58, 351–360. doi:10.1093/JXB/ERL214

Rio, S., Mary-Huard, T., Moreau, L., and Charcosset, A. (2019). Genomic Selection Efficiency and A Priori Estimation of Accuracy in a Structured Dent maize Panel. *Theor. Appl. Genet.* 132, 81–96. doi:10.1007/s00122-018-3196-1

Roth, M., Muranty, H., Di Guardo, M., Guerra, W., Patocchi, A., and Costa, F. (2020). Genomic Prediction of Fruit Texture and Training Population Optimization towards the Application of Genomic Selection in Apple. *Hortic. Res.* 7, 1–14. doi:10.1038/s41438-020-00370-5

Rothman, A. J., Levina, E., and Zhu, J. (2010). Sparse Multivariate Regression with Covariance Estimation. *J. Comput. Graphical Stat.* 19, 947–962. doi:10.1198/JCGS.2010.09188

Rutkoski, J., Benson, J., Jia, Y., Brown-Guedira, G., Jannink, J.-L., and Sorrells, M. (2012). Evaluation of Genomic Prediction Methods for Fusarium Head Blight Resistance in Wheat. *The Plant Genome* 5, 51–61. doi:10.3835/PLANTGENOME2012.02.0001

Rutkoski, J., Poland, J., Mondal, S., Autrique, E., Pérez, L. G., Crossa, J., et al. (2016). Canopy Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat. *G3: Genes, Genomes, Genet.* 6, 2799–2808. doi:10.1534/G3.116.032888

Sallam, A. H., Endelman, J. B., Jannink, J. L., and Smith, K. P. (2015). Assessing Genomic Selection Prediction Accuracy in a Dynamic Barley Breeding Population. *Plant Genome* 8, 20. doi:10.3835/PLANTGENOME2014.05.0020

Sallam, A. H., and Smith, K. P. (2016). Genomic Selection Performs Similarly to Phenotypic Selection in Barley. *Crop Sci.* 56, 2871–2881. doi:10.2135/CROPSCI2015.09.0557

Sarinelli, J. M., Murphy, J. P., Tyagi, P., Holland, J. B., Johnson, J. W., Mergoum, M., et al. (2019). Training Population Selection and Use of Fixed Effects to Optimize Genomic Predictions in a Historical USA winter Wheat Panel. *Theor. Appl. Genet.* 132, 1247–1261. doi:10.1007/S00122-019-03276-6

Schmidt, M., Kollers, S., Maasberg-Prelle, A., Großer, J., Schinkel, B., Tomerius, A., et al. (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, 203–213. doi:10.1007/s00122-015-2639-1

Sehgal, D., Rosyara, U., Mondal, S., Singh, R., Poland, J., and Dreisigacker, S. (2020). Incorporating Genome-wide Association Mapping Results into Genomic Prediction Models for Grain Yield and Yield Stability in CIMMYT Spring Bread Wheat. *Front. Plant Sci.* 11, 197. doi:10.3389/FPLS.2020.00197

Semagn, K., Beyene, Y., Warburton, M. L., Tarekegne, A., Mugo, S., Meisel, B., et al. (2013). Meta-analyses of QTL for Grain Yield and Anthesis Silking Interval in 18 maize Populations Evaluated under Water-Stressed and Well-Watered Environments. *BMC Genomics* 14, 313–316. doi:10.1186/1471-2164-14-313/TABLES/4

Shikha, M., Kanika, A., Rao, A. R., Mallikarjuna, M. G., Gupta, H. S., and Nepolean, T. (2017). Genomic Selection for Drought Tolerance Using Genome-wide SNPs in Maize. *Front. Plant Sci.* 8, 1–12. doi:10.3389/fpls.2017.00550

Singh, A. K., Gopala Krishnan, S., Ellur, R. K., Bhowmick, P. K., Nagarajan, M., Vinod, K. K., et al. (2017a). Notification of Basmati rice Variety, Pusa Basmati 1728. *Indian J. Genet.* 77, 584.

Singh, A. K., Gopala Krishnan, S., Nagarajan, M., Bhowmick, P. K., Ellur, R. K., Haritha, B., et al. (2017b). Notification of Basmati rice Variety Pusa Basmati 1637. *Indian J. Genet.* 77, 583–584.

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., et al. (2015). Correction: Genomic Selection and Association Mapping in Rice (Oryza Sativa): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. *Plos Genet.* 11, e1005350. doi:10.1371/JOURNAL.PGEN.1005350

Stewart-Brown, B. B., Song, Q., Vaughn, J. N., and Li, Z. (2019). Genomic Selection for Yield and Seed Composition Traits within an Applied Soybean Breeding Program. *G3 Genes, Genomes, Genet.* 9, 2253–2265. doi:10.1534/g3.118.200917

Sukumaran, S., Jarquin, D., Crossa, J., and Reynolds, M. (2018). Genomic-enabled Prediction Accuracies Increased by Modeling Genotype × Environment Interaction in Durum Wheat. *Plant Genome* 11, 170112. doi:10.3835/PLANTGENOME2017.12.0112

Sun, J., Poland, J. A., Mondal, S., Crossa, J., Juliana, P., Singh, R. P., et al. (2019). High-throughput Phenotyping Platforms Enhance Genomic Selection for Wheat Grain Yield across Populations and Cycles in Early Stage. *Theor. Appl. Genet.* 132, 1705–1720. doi:10.1007/S00122-019-03309-0

Sun, J., Rutkoski, J. E., Poland, J. A., Crossa, J., Jannink, J. L., and Sorrells, M. E. (2017). Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield. *Plant Genome* 10, 1–12. doi:10.3835/PLANTGENOME2016.11.0111

Tanaka, E. (2020). Simple Outlier Detection for a Multi-environmental Field Trial. *Biometrics* 76, 1374–1382. doi:10.1111/BIOM.13216

Tanaka, E. (2018). Simple Robust Genomic Prediction and Outlier Detection for a Multi-Environmental Field Trial. arXiv preprint arXiv:1807.07268, 1–25.

Tester, M., and Langridge, P. (2010). Breeding Technologies to Increase Crop Production in a Changing World. *Science* 327, 818–822. doi:10.1126/SCIENCE.1183700

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B (Methodological)* 58, 267–288. doi:10.1111/J.2517-6161.1996.TB02080.X

Tiede, T., and Smith, K. P. (2018). Evaluation and Retrospective Optimization of Genomic Selection for Yield and Disease Resistance in spring Barley. *Mol. Breed.* 38, 55. doi:10.1007/S11032-018-0820-3

Usai, M. G., Goddard, M. E., and Hayes, B. J. (2009). LASSO with Cross-Validation for Genomic Selection. *Genet. Res.* 91, 427–436. doi:10.1017/S0016672309990334

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91, 4414–4423. doi:10.3168/JDS.2007-0980

Varshney, R. K., Mohan, S. M., Gaur, P. M., Chamarthi, S. K., Singh, V. K., Srinivasan, S., et al. (2014a). Marker-Assisted Backcrossing to Introgress Resistance to Fusarium Wilt Race 1 and Ascochyta Blight in C 214, an Elite Cultivar of Chickpea. *Plant Genome* 7, 35. doi:10.3835/plantgenome2013.10.0035

Varshney, R. K., Pandey, M. K., Janila, P., Nigam, S. N., Sudini, H., Gowda, M. V. C., et al. (2014b). Marker-assisted Introgression of a QTL Region to Improve Rust Resistance in Three Elite and Popular Varieties of Peanut (*Arachis hypogaea* L.). *Theor. Appl. Genet.* 127, 1771–1781. doi:10.1007/S00122-014-2338-3

Varshney, R. K., Singh, V. K., Kumar, A., Powell, W., and Sorrells, M. E. (2018). Can Genomics Deliver Climate-Change Ready Crops? *Curr. Opin. Plant Biol.* 45, 205–211. doi:10.1016/J.PBI.2018.03.007

Váry, Z., Mullins, E., Mcelwain, J. C., and Doohan, F. M. (2015). The Severity of Wheat Diseases Increases when Plants and Pathogens Are Acclimatized to Elevated Carbon Dioxide. *Glob. Change Biol.* 21, 2661–2669. doi:10.1111/GCB.12899

Vasistha, N. K., Balasubramaniam, A., Mishra, V. K., Srinivasa, J., Chand, R., and Joshi, A. K. (2017). Molecular Introgression of Leaf Rust Resistance Gene Lr34 Validates Enhanced Effect on Resistance to Spot Blotch in spring Wheat. *Euphytica* 213, 1–10. doi:10.1007/s10681-017-2051-9

Vazquez, A. I., Bates, D. M., Rosa, G. J. M., Gianola, D., and Weigel, K. A. (2010). Technical Note: An R Package for Fitting Generalized Linear Mixed Models in Animal Breeding1. *J. Anim. Sci.* 88, 497–504. doi:10.2527/JAS.2009-1952

Viswanatha, K. P., Patil, R., Upadhyaya, H. D., Khan, H., Gururaj, S., ., S., et al. (2020). Genetic Diversity, Association and Principle Component Analyses for Agronomical and Quality Traits in Genomic Selection Training Population of Groundnut (*Arachis hypogaea* L.). *Ijgpb* 80, 282–290. doi:10.31742/IJGPB.80.3.7

Vivek, B. S., Krishna, G. K., Vengadessan, V., Babu, R., Zaidi, P. H., Kha, L. Q., et al. (2017). Use of Genomic Estimated Breeding Values Results in Rapid Genetic Gains for Drought Tolerance in Maize. *Plant Genome* 10. doi:10.3835/PLANTGENOME2016.07.0070/FORMAT/PDF

Wang, X., Xu, Y., Hu, Z., and Xu, C. (2018). Genomic Selection Methods for Crop Improvement: Current Status and Prospects. *Crop J.* 6, 330–340. doi:10.1016/j.cj.2018.03.001

Werner, C. R., Voss-Fels, K. P., Miller, C. N., Qian, W., Hua, W., Guan, C. Y., et al. (2018). Effective Genomic Selection in a Narrow-Genepool Crop with Low-Density Markers: Asian Rapeseed as an Example. *Plant Genome* 11, 170084. doi:10.3835/plantgenome2017.09.0084

WHO/FAO (2003). *Diet, Nutrition and the Prevention of Chronic Diseases: Recommendations for Preventing Excess Weight Gains and Obesity*. Geneva, Switzerland: WHO, 1–148.

Xiong, S., Wang, M., Zou, J., Meng, J., and Liu, Y. (2020). A Two-Stage Method for Improving the Prediction Accuracy of Complex Traits by Incorporating Genotype by Environment Interactions inBrassica Napus. *Discrete Dyn. Nat. Soc.* 2020, 1–12. doi:10.1155/2020/7959508

Xu, S. (2007). An Empirical Bayes Method for Estimating Epistatic Effects of Quantitative Trait Loci. *Biometrics* 63, 513–521. doi:10.1111/J.1541-0420.2006.00711.X

Xu, Y., Wang, X., Ding, X., Zheng, X., Yang, Z., Xu, C., et al. (2018). Genomic Selection of Agronomic Traits in Hybrid rice Using an NCII Population. *Rice (N Y)* 11, 32–10. doi:10.1186/S12284-018-0223-4/FIGURES/5

Yabe, S., Yoshida, H., Kajiya-Kanegae, H., Yamasaki, M., Iwata, H., Ebana, K., et al. (2018). Description of Grain Weight Distribution Leading to Genomic Selection for Grain-Filling Characteristics in rice. *PLOS ONE* 13, e0207627. doi:10.1371/JOURNAL.PONE.0207627

Yuan, Y., Cairns, J. E., Babu, R., Gowda, M., Makumbi, D., Magorokosho, C., et al. (2019). Genome-wide Association Mapping and Genomic Prediction Analyses Reveal the Genetic Architecture of Grain Yield and Flowering Time under Drought and Heat Stress Conditions in maize. *Front. Plant Sci.* 9, 1919. doi:10.3389/FPLS.2018.01919/FULL

Zhang, H., Yin, L., Wang, M., Yuan, X., and Liu, X. (2019). Factors Affecting the Accuracy of Genomic Selection for Agricultural Economic Traits in maize, Cattle, and Pig Populations. *Front. Genet.* 10, 189. doi:10.3389/FGENE.2019.00189/BIBTEX

Zhang, X., Pérez-Rodríguez, P., Burgueño, J., Olsen, M., Buckler, E., Atlin, G., et al. (2017). Rapid Cycling Genomic Selection in a Multiparental Tropical maize Population. *G3: Genes, Genomes, Genet.* 7, 2315–2326. doi:10.1534/G3.117.043141

Zhao, Y., Gowda, M., Liu, W., Würschum, T., Maurer, H. P., Longin, F. H., et al. (2012). Accuracy of Genomic Selection in European maize Elite Breeding Populations. *Theor. Appl. Genet.* 124, 769–776. doi:10.1007/S00122-011-1745-Y

Zhao, Y., Mette, M. F., and Reif, J. C. (2015). Genomic Selection in Hybrid Breeding. *Plant Breed* 134, 1–10. doi:10.1111/PBR.12231

Zhao, Y., Zeng, J., Fernando, R., and Reif, J. C. (2013). Genomic Prediction of Hybrid Wheat Performance. *Crop Sci.* 53, 802–810. doi:10.2135/CROPSCI2012.08.0463

Ziyatdinov, A., Ve1;zquez-Santiago, M., Brunel, H., Martinez-Perez, A., Aschard, H., and Soria, J. M. (2018). lme4qtl: Linear Mixed Models With Flexible Covariance Structure For Genetic Studies Of Related Individuals. *BMC Bioinformat.* 19, 68. doi:10.1186/s12859-018-2057-x

Zou, H., and Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *J. R. Stat. Soc B* 67, 301–320. doi:10.1111/j.1467-9868.2005.00503.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.