

<https://helda.helsinki.fi>

---

## Plant Breeding Integrated with Genomic-Enabled Prediction

Bidabadi, Siamak Shirani

2021-09-15

---

Bidabadi , S S , Sharifi , P & Jain , S M 2021 , ' Plant Breeding Integrated with Genomic-Enabled Prediction ' , OBM genetics , vol. 5 , no. 3 . <https://doi.org/10.21926/obm.genet.2103137>

---

<http://hdl.handle.net/10138/342281>

<https://doi.org/10.21926/obm.genet.2103137>

---

cc\_by

publishedVersion

---

*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

Research Article

## Plant Breeding Integrated with Genomic-Enabled Prediction

Siamak Shirani Bidabadi <sup>1</sup>, Parisa Sharifi <sup>2</sup>, S. Mohan Jain <sup>3,\*</sup>

1. Department of Horticulture, College of Agriculture, Isfahan University of Technology, IUT Isfahan, 84156-83111, Iran; E-Mail: [siamakshirani.upm@gmail.com](mailto:siamakshirani.upm@gmail.com)
2. Department of Agricultural Extension and Education, Higher Education Center Shahid Bakeri Miyandoab, Urmia University, Urmia; E-Mail: [pa.sharifi@urmia.ac.ir](mailto:pa.sharifi@urmia.ac.ir)
3. Department of Agricultural Sciences, University of Helsinki, PL-27, Helsinki, Finland; E-Mail: [mohan.jain@helsinki.fi](mailto:mohan.jain@helsinki.fi)

\* **Correspondence:** S. Mohan Jain; E-Mail: [mohan.jain@helsinki.fi](mailto:mohan.jain@helsinki.fi)

**Academic Editor:** Joep Geraedts

**Special Issue:** [Genomic Prediction in Crop Improvement](#)

**OBM Genetics**

**2021, volume 5, issue 3**

**doi:10.21926/obm.genet.2103137**

**Received:** June 09, 2021

**Accepted:** August 30, 2021

**Published:** September 15, 2021

### Abstract

Plant breeding programs have used conventional breeding methods, such as hybridization, induced mutations, and other methods to manipulate the plant genome within the species' natural genetic boundaries to improve crop varieties. However, repeatedly using conventional breeding methods might lead to the erosion of the gene reservoir, thereby rendering crops vulnerable to environmental stresses and hampering future progress in crop production, food and nutritional security, and socio-economic benefits. Integrating innovative technologies in breeding programs to accelerate gene flow is critical for sustaining global plant production. Genomic prediction is a promising tool to assist the rapid selection of premiere genotypes and accelerate breeding gains for climate-resilient plant varieties. This review surveys the annals and principles of genomic-enabled prediction. Based on the problem that is investigated through the prediction, as well as several other factors, such as trait heritability, the relationship between the individuals to be predicted and those used to train the models for prediction, the number of markers, sample size, and



© 2021 by the author. This is an open access article distributed under the conditions of the [Creative Commons by Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is correctly cited.

the interaction between genotype and environment, different levels of accuracy have been reported. Genomic prediction might play a decisive role and facilitate gene flow from gene bank accessions to elite lines in future breeding programs.

### **Keywords**

Breeding values; genomic prediction; genetic gains; machine learning methods

## **1. Introduction**

The selection of high-yielding plants to satisfy the continuously growing human population dates back to 10,000 years, although the process has been fundamentally modified over the last hundred years [1]. Most of the food products available to consumers are produced using modern and conventional breeding programs [2]. In conventional plant breeding programs, various methods are applied to the natural genetic structure of a species [3] for improving cultivars using genes from different plant species. The principles of conventional plant breeding programs are based on the development of new plant cultivars using old equipment and involving natural events [4]. Additionally, feeding the rapidly expanding world population requires the latest agricultural research. According to many researchers, in conventional plant breeding programs, desired traits in the plants are expressed without introducing a new gene in the plant [5, 6]. Progress in the fields of plant genetics and plant nutrition, as well as advances in the chemical industry, have significantly accelerated the advancement of agricultural innovation. The development of modified crops and the use of chemical fertilizers and other modern inputs significantly increased agricultural productivity in the United States and Europe in the second half of the twentieth century [6]. Given that the challenges of climatic change, human population growth, decrease in arable land, and pollution are expected to spread to many agricultural areas, innovative and cost-effective technologies might play a decisive role in dealing with the problem of crop adaptation, especially when global food production is severely limited by agricultural land, water, and energy shortages [1, 7, 8]. Successfully breeding conventional plants to improve or develop new crop varieties is highly time-consuming, as many phases of crossing, selection, and testing need to be performed. Also, traditional plant breeding cannot solve the problems caused by intense agriculture and climate change, e.g., improper use of chemical fertilizers along with unsustainable farming methods. Therefore, innovative molecular tools, marker-assisted selection (MAS), molecular marker-assisted-breeding (MAB), and gene-editing are needed for the rapid development of new cultivars to alleviate food scarcity and promote food security [10]. Due to several stages of crossing, selection, and testing required in the traditional production of new plant varieties, creating a new cultivar might take two decades [9]. Hence, novel molecular techniques, integrated with predictions based on the genome, might provide new strategies to breed plants more efficiently [10]. In this review, the tenets and privileges of genomic-enabled - predictions, some of the complexities of applying such -predictions in crop breeding programs, and the problems that arise due to the interaction between the genotype and the environment have been discussed.

## **2. Plant Breeding Methods**

Plant breeders choose plant varieties with the highest edible quality, followed by cultivation, harvesting, and processing, along with the highest resistance to biotic and abiotic stressors [12]. Plant breeding approaches are divided into three parts. First, high-yielding and sustainable varieties, and those with high-quality oil, in case of rapeseed breeding [11, 12], are selected from the naturally occurring organic or conventional varieties.

Second, controlled pollination in selected plants is performed to recombine desirable genes from different parents. Third, molecular techniques are used to select the marker profiles or preferred genes to detect any alterations in the genome [12]. In the 1970s plant breeding demonstrated a major impact on the accessibility of enhanced varieties of plants for some traits such as high-yielding, semi-dwarf varieties worldwide [13]. However, conventional breeding is based on phenotypic selection and has less impact on low heritable -multi-genic quantitative attributes such as biotic and abiotic stresses, which are significantly affected by gene-environment interactions. It is time-consuming, laborious, requires a large area, expensive, has a large population size, is less precise, reliable, and requires an immediate, rapid, and more efficient selection system [14]. We have discussed important traditional breeding methods, including hybridization, doubled haploid breeding, pure line selection, mutation induction, and molecular marker-assisted breeding (MAB).

### **2.1 Hybridization**

In hybridization programs, plant breeders produce new cultivars with desirable traits of different plants by crossing those plants that usually have traits close to that of the new cultivar [15]. Hybridization is performed by controlled pollination that might be influenced by artificial alternatives [16]. Numerous well-regulated pollination techniques are used to facilitate hybridization in five ways, which include the emasculation of bagged flowers and pollination of succeeding receptive stigmas, emasculation of the stigmas and instant pollination through induced receptivity that is followed by bagging (One Stop Pollination), and pollination using cut styles with no bagging and emasculation (Artificially Induced Protogyny) [3, 17]. Carputo et al. [18] reported that interspecific crosses with wild relatives resulted in the generation of variants with high genetic diversity and resistance against biotic or abiotic stress. For hybridization, significant diversity is present in existing combinations without the need for new genes. [3]. During artificial hybridization, pollen grains are collected from plants with desirable traits for pollination and fertilization, and subsequently, used in plant breeding programs [19]. However, sexual incompatibility often occurs while producing hybrids and is a challenge that needs to be overcome by asexual propagation and in vitro methods, such as protoplast fusion, somatic cell hybridization, or embryo rescue [16]. A widely used technique, free from GMO, in crop breeding programs is somatic hybridization, which is performed for improving crop yield. Somatic hybridization can effectively induce hybrid vigor, through which the hybrid offspring displays greater size, fertility, biomass, and enhanced resistance to diseases. However, certain genomic shocks in progenies, e.g., chromosomal rearrangements, alterations in genome size, transpositions, and other chromatin results, might interfere with somatic hybridization [20]. Several plants, including citrus, benefit from the use of molecular markers during somatic hybridization [21]. Somatic hybrids were used as pollen parents in interploidy sexual crossing. As a result, high-quality triploid plants were

obtained using flow cytometry, embryo rescue, and SSR marker analysis [21]. According to Bashir et al. [20], hybridization could alter the somatic mutation rates in plants.

## **2.2 Doubled Haploid Breeding**

Doubled haploid breeding involves the production of haploid plants by culturing anthers, microspores, ovaries, and interspecific hybridization, followed by doubling the chromosome numbers with colchicine treatment to obtain diploid plants, in which, consequently, every gene is homozygous [22]. Producing completely homozygous plants in the first generation is a very important step in plant breeding programs [23] since it reduces the breeding time and the cost of developing a new cultivar [24, 25]. *Lolium perenne* L. and other essential forage species have been effectively bred using double haploids [26]. However, when creating double haploids, cost efficiency, stabilization of rare and useful alleles, retention of genetic diversity in their lines, as well as distinctiveness, uniformity, and stability of new species must also be considered. Double haploids possess both negative and positive effects regarding the above-mentioned conditions [26]. Rare alleles, obtained by producing doubled haploids, might control various essential characteristics in plant breeding. The double haploid method assists while crossing for desirable genotypes, thus resulting in hybrids with chromosome sets from both parents. Recombination leads to the development of novel genomic amalgamations through the formation of gametes that could then be corrected during the induction of doubled haploids. Therefore, a reasonable level of genetic diversity is needed to preserve rare alleles for future breeding purposes [27, 28]. Doubled haploids could be used in conjunction with molecular breeding to enhance gene frequencies and strengthen qualitative traits. By adding markers to haploid plants, highly efficient molecular breeding can be achieved by developing doubled haploids [29, 30]. As an example, for haploid recognition, novel haploid inducers were obtained by combining novel molecular marker systems, such as the high oil marker and the red root marker, that were used to create a doubled haploid technique from germplasm [31].

## **2.3 Pure Line Selection**

Self-pollinated plant species are usually improved via pure line selection [1]. Due to self-pollination, the genetic structures of pure line varieties (inbred lines) are homogeneous and homozygous [32]. They have a very narrow genetic base and are mostly used as parents while developing new cultivars [3, 15, 16]. Pure line selection is rapid and easy to perform, especially for selected genotypes obtained from a variable population [33]. Farmers perform pure line selection and then multiply the plants as a new cultivar under natural conditions [34]. To determine the performance of the selected plant, progeny testing is essential in pure line selection. If the selected line is superior, the population is released as a variety [34]. Pure line selection was used effectively to assess genetic and phenotypic components and select preferable genotypes that expressed traits of interest that performed well in a heterogeneous population of soybean cultivars [35].

## **2.4 Mutation Breeding**

Plant breeders have been using mutation breeding to improve plants since the 1920s [36]. Thereafter, most induced mutations were performed during the 1950s and the 1960s, and to date, a significant number of mutant plants have been developed for desired modified traits [37]. Induced mutations have been effectively used to improve plants. Several mutants have also been detected and reported in many countries such as China, India, the USA, Japan, the Netherlands, and Russia. Induced mutants are produced randomly; therefore, an accurate prediction of the desired outcome is relatively difficult [38-40]. However, the genomic aspects of mutations, from a molecular point of view, could be useful in selecting the appropriate mutation induction techniques for gene function analysis. Thus, creating plant mutant reservoirs could help basic and applied research, e.g., gene mapping and functional genomics [40].

## **2.5 Problems While Breeding for Crop Improvement and Sustainability**

Scientists have focused on being able to feed the ever-growing global population for a long time, but in recent years, sustainable food production has received increased attention due to the impact of climate change [41, 42]. Fortunately, the development of improved cultivars has led to faster global food production, which has led to a steady decline in food prices. The continuous production of high levels of biomass to minimize external inputs of non-renewable resources such as water, fuel, agricultural land, and fertilizers, to potentially reduce greenhouse gas emissions during cultivation, along with the production of plants that have high resistance to biotic and abiotic stressors that are well-adapted to potential climate change, are some of the challenges for future plant growers [42, 44]. Plant breeding programs should be one of the highest priorities of the government to ensure the availability of high quality and quantity of food worldwide [43]. Hybridization using different parents, and the subsequent selection following consecutive generations, has several limitations, including a long period required to develop novel varieties and inadequacies in multifaceted traits with low heritability [44].

Thus, to overcome the limitations of conventional plant breeding methods, molecular marker-assisted breeding techniques, which require less phenotypic information for indirectly selecting desirable traits in plants, can be used [45]. Performing whole-genome predictions is a good strategy to conduct phenotypic selection or marker-assisted breeding [46].

## **3. Molecular Marker-Assisted Breeding**

Plant breeders face two main challenges: efficient selection and accelerating the breeding phase. Jiang [47] proposed a new method for improving plants using molecular marker-assisted breeding (MAB). Molecular breeding was shown to greatly enhance selection efficacy, elevating genetic diversity, explaining the genetic architecture, and adjusting gene function [48]. Marker-assisted selection (MAS) can be used in genotyping and choosing complementary parents for early crossing, and selecting superior genotypes. For this, tissue culture allows the production of disease-free stock plants with a high rate of multiplication, which plays a critical role in the crops' yield and quality [49]. Marker-assisted selection could improve the efficacy of plant breeding [46]. Using F<sub>2</sub> and backcross populations, recombinant inbred lines, and doubled haploids, molecular-assisted breeding allows the selection of various characters of interest [50]. Genomic prediction

and selection are beneficial for the attributes that are governed by fewer quantitative trait loci (QTLs). Genomic prediction patterns compute the impact of all molecular markers at all loci and capture small QTL effects [46, 51]. The invention of various marker-QTL systems has improved the quantitative attributes of economic importance. Although QTLs can successfully detect simple traits, there are no reports of them improving complex quantitative attributes. Due to the interaction between QTL and the environment, finding the same QTL across multiple environments seems difficult. Due to low rates of recombination in chromosomes, linkage analysis for QTL mapping in parental populations has failed. As the cost of high-throughput SNP (single nucleotide polymorphism) genotyping for QTL is low, it is possible to use high-density SNP arrays to develop statistical models for predicting marker-trait associations, which depend on the genetic architecture of the predicted traits [52].

#### **4. Benefits and Challenges of Genomic-Enabled Prediction in Plant Breeding**

The increase in crop productivity is due to both conventional and marker-assisted breeding (MAB) programs [11]. High-density single nucleotide polymorphism markers are used in genomic prediction across the whole genome to predict genetic values beneficial to plant breeding programs [46, 53-55]. They not only improve selection efficiency but also reduce breeding costs and unify breeding methods, biological discovery, as well as, tools and methods of selection [56]. This method, besides helping in identifying plants with complex traits, also makes plant breeding programs extremely efficient [14]. Meuwissen et al. [53] presented the idea of genomic selection as an efficient technique to predict complicated traits in plants and animals and has been widely used in breeding crops. The use of genetic prediction has allowed the selection of new genotypes in plant breeding programs based on genomic data without the need for phenotyping [52, 57]. However, it requires the construction of a prediction pattern by integrating marker information with a phenotypic database in a model training step [53] and has led to fundamental changes in plant breeding programs [58]. Genetic progression is improved by genomic selection in crop breeding programs via phenotypic predictions to select ideal phenotypes, based only on their genomic information [59].

The cost of genotyping, especially for next-generation sequencing, is usually much lower than phenotyping [46, 54]. Moreover, the use of genomic prediction does not acquit phenotyping but replaces many of the steps of selection that are associated with phenotyping, based on whole-genome prediction [46, 54]. Most studies on genomic prediction have preferentially predicted the validation set using cross-validation to predict the validation complex [46, 54]. However, two factors, which include the cost of genotyping and the ambiguous protocols to efficiently apply genome selection restrict the use of genomic prediction in breeding programs. Genomic prediction focuses on two procedures, A) predicting additive impacts in the early generations of a breeding process for a fast selection cycle at a short interval and B) predicting the perfect genetic values of individuals, considering both additive and non-additive effects, to estimate the efficiency of the desired cultivars [52]. Generally, genetic gains are enhanced by increasing the selection intensity and boosting the breeding cycle using genomic prediction [60-62].

## 5. Accuracy of Genomic Prediction and Genomic Selection-Assisted Genetic Gains

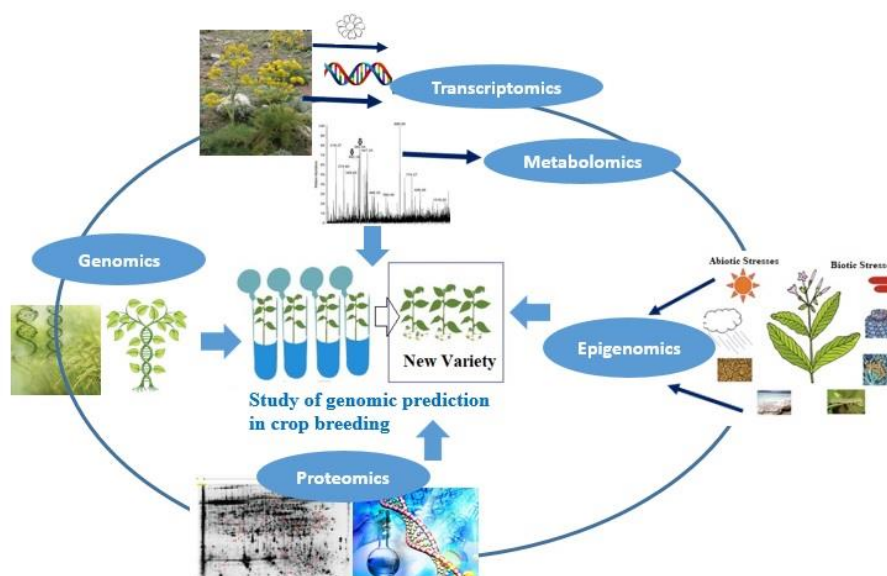
Several genetic and statistical limitations make genomic prediction inefficient. These limiting factors arise from the size and diversity of the training population and the heritability of the attributes to be predicted. Statistical limitations arise due to the high dimensionality of marker data, where the number of markers is considerably more than the number of observations and due to the multicollinearity among markers, as adjacent markers are highly correlated [52]. Different levels of genomic prediction accuracy have been reported in crops, which depend on the evaluation of the prediction problem, character heritability, the correlation between the individuals to be anticipated and those applied to train the patterns for prediction, the number of markers, sample size, and the genotype-environment interaction [62]. Crossa et al. [62] performed genetic predictions for maize and wheat. They concluded that pedigree (population structure) accounted for a significant portion of prediction accuracy when the assessment for the global population demonstrated a predictive problem. The precision of genomic prediction might be enhanced by incorporating data from the associated environments while modeling the relationship between genotype and environment.

Several studies have reported the usefulness of genomic prediction in plant breeding programs [63-65]. Genomic prediction can improve selection, lower costs, and provide a way to combine reproductive techniques and biological discovery [56]. Some important characters, including grain yield, biomass accumulation, resistance against diseases, and events that are involved in the flowering processes under various environmental conditions, have been predicted with various degrees of accuracy, which depend on the heritability of the character, the size of the training population, the number of markers, the correlation between the training and testing cycles, and the interaction between genotype and environment [14, 46, 52, 60, 61, 66-73]. Windhausen et al. [60] applied marker effects, estimated in 255 diverse maize (*Zea mays* L.) hybrids, to anticipate grain yield, anthesis date, and anthesis-silking interval within the diversity panel and testcross progenies of 30 F<sub>2</sub>-derived lines from each of five populations. They reported that predictions were made based on the differences in the average performance of breeding populations and were less related to the relationship between training and accreditation sets or linkage disequilibrium, with causal variants underlying the predicted traits. Potential uses of genomic prediction in maize hybrid breeding include having a clear breeding event in which genomic prediction should be applied, a detailed analysis of the population structure before cross-validation, and large training sets with a strong genetic relationship to the validation set [60]. In maize, up to three generations are conceivable per year, selection speed is high and, consequently, genomic selection is very promising for breeding programs [61]. As a cost-effective and efficient breeding approach, genomic selection and prediction can have small and large impacts on genetic factors, and therefore, achieve higher genetic gains for complex traits, such as the yield of the oil content in groundnut [73]. Integration of genomic prediction with rapid generation advancement technology, such as speed breeding, can significantly and positively impact breeding programs, including those for maize and wheat [62, 72], groundnut [73], sorghum [74], and apple [75]. Roth et al. [75] identified key genetic parameters to consider while performing genomic prediction for the texture of apples. Zhang et al. [76] provided beneficial information to maize breeders regarding genomic selection activities for future breeding programs. There are few reports on genomic prediction for some traits of soybean, including seed yield [77], seed weight [78], and



resistance to biotic stresses [79, 80]. Recently, Haikka [81] reported the applicability of genomic prediction in the breeding program of oat and barley. The study predicted that the yield of those crops was improved by using multi-trait prediction. Haikka [81] also estimated the benefits of trait-assisted prediction for oat and barley, and the prediction of grain yield was improved by 4% and 9%, respectively, with multi-trait prediction, and by 9-14% and 11-28% with trait-assisted prediction, compared to the prediction of grain yield alone. Batista et al. [82] developed and applied a methodology for genomic prediction for autotetraploids and even for higher (and mixed) ploidy levels. They used the models to build covariance matrices of both additive and digenic dominance effects that were subsequently applied in genomic prediction models.

In general, for the analysis of genomic prediction in crop breeding, the incorporation of multi-omics data, comprising epigenomics, genomics, proteomics, transcriptomics, and metabolomics, appears to be critical to predict complex traits using molecular genetic information (Figure 1). Many types of -omic data, such as transcriptomics, epigenomics, proteomics, and metabolomics, have become readily available. Therefore, evaluating the utility of this huge amount of information in predicting complex traits would be interesting [83].



**Figure 1** Multi-omics data integration in the genomic prediction analysis of breeding programs.

### **5.1 Comparison between Single- and Multiple-Trait Genomic Models in Genomic Prediction**

Multiple-character models for genomic prediction were recently reported [84-88]. A multiple-character genomic model (MTGM) had higher accuracy of prediction than a single-trait genomic model (STGM) [89, 90]. Comparing the results of the single-character genomic model (STGM) with those of the multiple-character genomic model (MTGM) for genomic prediction of high and low heritability traits demonstrated that both MTGM and STGM had similar validity for the genomic predictions for the high heritability traits, while for the low heritability traits, MTGM was more efficient than STGM [89].

## 5.2 Genomic Prediction for Interaction between Genotype and Environment

The interaction between genotype and environment ( $G \times E$ ) strongly affects the yield of crop plants. In genomic prediction models,  $G \times E$  can boost the accuracy of prediction and assist the selection of lines across target environments [91-94]. Genomic predictions have sufficiently improved the yield and other important traits in different crops when applied during breeding programs [51, 60, 95]. However,  $G \times E$  complicates the screening of fixed pure lines and negatively affects the heritability of the characters. It is represented as a change in the plant line's performance from one environmental condition to another. Hence modeling for  $G \times E$  during genomic prediction could be beneficial for breeders to select the lines with optimal overall performance across environments, as well as in specific target environments [92]. One of the models can link genotypes in different locations, even in cases where plant lines are not present in all locations. This model has the advantage of applying genetic covariance matrices in  $G \times E$  mixed models [96]. In all genomic prediction models that incorporated  $G \times E$ , the accuracy of single-environment analyses was reported to be higher than the models without the  $G \times E$  factor in bread wheat, maize, barley, and legumes [91, 93, 96-100]. The main patterns applied to evaluate genomic prediction accuracy by incorporating  $G \times E$  are as follows:

Burgueno et al. [91] were the first to report the application of marker-based and pedigree-based linear models of genomic prediction for investigating  $G \times E$ . Crop-modeling data was also incorporated to investigate  $G \times E$  by Heslot et al. [101]. Several models, possessing the main effects of environment and lines, were coordinated to determine prediction accuracies in chickpea. The results demonstrated that accurate and precise prediction depends on improving genomic prediction models and applying various genotyping platforms, which leads to the selection of valuable candidates [92]. Bohlouli et al. [102] assessed the accuracy of genomic prediction across several scenarios applying single-character and multiple-character genomic models to detect  $G \times E$ . They concluded that  $G \times E$  contributes to perceptual variability in quantitative characters and enhances the accuracy of genomic prediction. Hence, the interaction must be considered while selecting different genotypes. With the appearance of molecular markers, linear mixed models, associated with the differential sensitivity of genotypes to environments to particular regions of the plant genome and specific biological mechanisms, were used in new procedures of  $G \times E$  analysis [103, 104]. Hassen et al. [105] reported the first study that explored the feasibility of breeding rice, adapted to alternate wetting and drying, using genomic prediction methods that accounted for genotype by environment interactions. Mageto et al. [106] evaluated genomic prediction with  $G \times E$  analysis for the level of Zn in the kernel of tropical maize germplasm and demonstrated the capability of genomic prediction to boost the breeding of plants for enhanced Zn levels in the kernel of the selected superior genotypes. Multi-environment analysis can influence  $G \times E$  by applying genetic and remaining covariance roles, markers and environmental covariates, or marker with  $G \times E$  [91, 97, 107]. Keller et al. [57] demonstrated the potential of genomic selection to enhance the genetic gain in agronomic traits of common bean (*Phaseolus vulgaris* L.) bred under environmental stress. However, change in the responses of each cultivar with climatic conditions is a major challenge for breeders and can have a stronger effect on heritability and genotype ranking compared to the effect of the different environments [96].

## 6. Prediction Methods for Additive Genetic Effects

Genomic prediction using marker and phenotype information from observed populations is useful to make connections between markers and phenotypes. Genomic prediction applies genome-wide markers to measure the impacts of all loci and ultimately predicts the genetic values of untested populations to perform more inclusive and reliable selection and to accelerate genetic progress in crop breeding. However, genomic prediction models might have a problem when the number of markers is much higher than phenotypic observations. To solve this problem and increase the accuracy of genomic prediction, many models, such as parametric procedures, as well as non-parametric methods such as machine learning algorithms, including gradient boosting machine (GBM) and extreme gradient boosting method (XgBoost), reproducing kernel Hilbert space (RKHS), random forest (RF), and radial basis function neural network (RBFNN), have been used for genomic prediction in breeding programs [108].

In genomic best linear-unbiased prediction (GBLUP) method, the impacts of all the loci possess a joint discrepancy that makes these loci more fit for characters affected by many minor genes [108, 109]. The Bayesian method is efficacious and pliable for phenotypic prediction and breeding value estimation. In the Bayesian method, variance components are calculated with genomic breeding values. It is based on differences calculated from data, and therefore, are novel and specific to the analyzed character, allowing skepticism of the variance components to be included in the breeding values [110, 111]. Several Bayesian methods of phenotype and genomic prediction, according to multi-locus aggregation patterns, have been considered [111]. The least absolute shrinkage and selection operator (LASSO) is a regression model presented by Tibshirani [112]. This method is a constrained version of the usual least squares. It is slightly ineffective for the strongly correlated markers and tends to pick one trait and ignore the others [108]. Machine learning, as a non-parametric method, has also been successfully adopted in genomic prediction [108, 113]. Machine learning methods have been recently adopted in genome-wide association studies for detecting candidate genes and epistasis, analysis of the gene network pathway, and genomic prediction of phenotypic values [114]. Non-parametric methods are more efficient than parametric methods when the underlying genetic architecture is quiet because of epistasis. The non-parametric methods have smaller genomic prediction errors and enhanced prediction accuracy of phenotypic values [115]. Among machine learning methods, RF is a commonly used tree-based ensemble procedure for regressing multiple variables [116-118]. Gradient boosting machine (GBM) is an efficacious method that can refine SNPs and reduce complex models in the analyses of multivariate phenotype genome-wide association studies [114, 119]. Another machine learning method known as extreme gradient boosting (XgBoost), introduced by Chen and He [120], is based on a similar principle as GBM, but applies a more regularized model than GBM to control over fitting.

Analysis of genome-wide loci and their interaction involves many variables [95], for which, the above-mentioned statistical methods of omics data analysis have been associated with benefits in genomic prediction. However, using an inappropriate model in breeding practices might obscure some genetic variances, resulting in a biased and unreliable genomic prediction. Accurately analyzing the interactions among many genes and examining the intensity of their effects is an issue that needs to be addressed in the future [108].

## **7. Factors Affecting the Accuracy of Genomic Prediction for Agricultural Economic Traits**

The accuracy of predicting attributes that possess complex genetic architecture can be improved by enhancing the marker density [121]. In cases where simple attributes are governed by one or several genes with large effects, higher marker density might lower the accuracy of the prediction if the quantitative trait nucleotide (QTN) is included but lead to a higher prediction accuracy if the QTN is excluded [121]. Some important agents, including marker density, statistical method, minor allele frequency (MAF), trait heritability, training population size, and genetic architecture, have a significant impact on the accuracy of genomic prediction [76, 121, 122]. The accuracy of genomic prediction for agricultural economic traits is enhanced with marker density, although, in some cases, it may be accompanied by a decrease [121].

In agriculture science, an outlier may be generated because of wrong data imputation. Although statistical procedures are reported in the literature for the identification of outliers [123], the identification of true outliers is still a challenge, especially for high-dimensional genomic data. The predictive efficiency of genomic prediction techniques might be negatively affected in the presence of outliers. [124]. Budhlakoti et al. [124] efficiently detected outliers in high-dimensional genomic data and observed a significant improvement in genomic prediction. Zhu et al. [125] measured the estimated values of genomic reproduction by providing different subsets of mononucleotide polymorphisms (SNPs) based on different marker densities and MAF to evaluate the effect of marker density and MAF on predictability. They concluded that a low-density chip, by adapting low-frequency markers with large SNP effect sizes, should be designed for genomic prediction. It works better for traits with high heritability than for traits with low heritability [121]. The size and composition of the training population affect the efficacy of genomic prediction in breeding programs. By comparing optimization techniques to select individuals from the training population with higher predictive ability, Berro et al. [126] demonstrated benefits in considering the training population when designing an optimal training population for genomic prediction. Furthermore, they found that a weighted relationship matrix with stratified sampling was the best procedure for submitting quantitative character predictions in different populations several generations apart. The accuracy of prediction for characters possessing a complex genetic architecture can be enhanced by the density of a specific marker. For simple characters that are governed by one or several genes with large effects, higher marker density can also increase the prediction accuracy if the QTN is unspecified and not included [121]. Thus, it can be concluded that a combination of all the above-mentioned factors that affect prediction accuracy will always generate a high and fixed prediction accuracy with acceptable breeding and computational costs.

## **8. Perspectives of Genomic Prediction in Plant Breeding**

Advancement in plant breeding programs is essential to boost genomic prediction. In addition to reducing the cost of creating new variability, the combined application of genomic prediction associated with high-throughput phenotype in plant breeding must be determined. Furthermore, the use of genomic prediction in germplasm enhancement, and the ability to anticipate the performance of a line in an unobserved environment is very important in breeding programs and needs to be addressed in future plant breeding plans [52]. Omics data have indicated the role of genomic prediction in future plant breeding programs. Therefore, aligning genomic predictions for targeted plant breeding programs in the era of climate resilience, where predictions are also

based on historical weather data available at the time of prediction, significantly improves predictive accuracy; moreover, the inclusion of  $G \times E$  causes selection for stress-tolerant lines to maintain food security under changing climatic conditions. However, regardless of the interaction of environmental factors with the genotype ( $G \times E$ ), the accuracy of genomic predictions for low heritability attributes and characters with only a small amount of phenotype data could be improved by MTGM. The estimated response variables for the breeding values obtained from the multiple-character and single-trait models insignificantly affected the reliability of the estimated breeding values for characters with no missing genomic data. However, for traits with missing data, the response variable obtained from the multiple-trait model gave better genomic predictions than the response variable obtained from the single-trait model. Furthermore, the use of various genomic prediction models seems promising for genomic prediction in populations, although choosing a suitable model significantly affected genomic prediction in the ongoing plant breeding programs.

## 9. Conclusion and Future Prospects

The successful implementation of genomic prediction to establish an effective plant breeding program requires a strategic approach in designing breeding programs for sustainable production, better cooperation, a new association between the public and private sectors, and acquiring new skills. Furthermore, the effect of genomic prediction must be extended beyond the creation of new crop varieties and follow an integrative framework. Adjusting genomic prediction to mitigate the risks of genetic losses in breeding programs might represent the biggest incentive for the enhanced coordination of funding to support modern plant breeding programs globally.

Studies to determine the implementation of genomic prediction for the selection of future phenotypes are inadequate. Hence, predictions for the whole genome must be further highlighted in plant breeding. Detailed studies to integrate  $G \times E$  into proper models might determine whether genomic prediction could be a good strategy to determine efficiency under changing climatic conditions to feed the ever-growing global population.

## Author Contributions

Siamak Shirani Bidabadi and Parisa Sharifi have participated in literature reviewing and writing the review manuscript. Mohan Jain supervised the writing of this review manuscript and also applied his corrections and comments to improve the manuscript.

## Competing Interests

The authors have declared that no competing interests exist.

## References

1. Kaiser N, Douches D, Dhingra A, Glenn KC, Herzig PR, Stowe EC, et al. The role of conventional plant breeding in ensuring safe levels of naturally occurring toxins in food crops. *Trends Food Sci Technol.* 2020; 100: 51-66.
2. Das B, Ahmed N, Singh P. *Prunus* diversity-early and present development: A review. *Int J Biodivers Conserv.* 2011; 3: 721-734.

3. Acquaah G. Conventional plant breeding principles and techniques. In: *Advances in plant breeding strategies: Breeding, biotechnology and molecular tools*. Switzerland: Springer; 2015. p. 115-158.
4. Jain HK, Kharkwal MC. *Plant breeding: Mendelian to molecular approaches*. Dordrecht: Springer Netherlands; 2004.
5. Jakowitsch J, Mette M, van Der Winden J, Matzke M, Matzke A. Integrated pararetroviral sequences define a unique class of dispersed repetitive DNA in plants. *PNAS*. 1999; 96: 13241-13246.
6. Qaim M. *Genetically modified crops and agricultural development*. New York: Palgrave Macmillan; 2016.
7. Davies FT, Garrett B. Technology for sustainable urban food ecosystems in the developing world: Strengthening the nexus of food-water-energy-nutrition. *Front Sustain Food Syst*. 2018; 2: 84.
8. Voss-Fels KP, Stahl A, Hickey LT. Q&A: Modern crop breeding for future food security. *BMC Biol*. 2019; 17: 1-7.
9. Ahmar S, Gill RA, Jung KH, Faheem A, Qasim MU, Mubeen M, et al. Conventional and molecular techniques from simple breeding to speed breeding in crop plants: Recent advances and future outlook. *Int J Mol Sci*. 2020; 21: 2590.
10. Varshney RK, Singh VK, Hickey JM, Xun X, Marshall DF, Wang J, et al. Analytical and decision support tools for genomics-assisted breeding. *Trends Plant Sci*. 2016; 21: 354-363.
11. Breseghello F, Coelho AS. Traditional and modern plant breeding methods with examples in rice (*Oryza sativa* L.). *J Agric Food Chem*. 2013; 61: 8277-8286.
12. Abbadi A, Leckband G. Rapeseed breeding for oil content, quality, and sustainability. *Eur J Lipid Sci Technol*. 2011; 113: 1198-1206.
13. Oury FX, Godin C, Mailliard A, Chassin A, Gardet O, Giraud A, et al. A study of genetic progress due to selection reveals a negative effect of climate change on bread wheat yield in France. *Eur J Agron*. 2012; 40: 28-38.
14. Bhat JA, Ali S, Salgotra RK, Mir ZA, Dutta S, Jadon V, et al. Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Front Genet*. 2016; 7: 221.
15. Acquaah G. *Principles of plant genetics and breeding*. 2nd ed. Oxford: Wiley; 2012.
16. Brown J, Caligari P. *An introduction to plant breeding*. Oxford: Blackwell Publishing; 2008.
17. Horsley TN, Johnson SD, Myburg AA. Comparison of different control-pollination techniques for small-flowered eucalypts. *New For (Dordr)*. 2010; 39: 75-88.
18. Carputo D, Barone A, Frusciante L. 2n gametes in the potato: Essential ingredients for breeding and germplasm transfer. *Theor Appl Genet*. 2000; 101: 805-813.
19. Borojevic S. *Principles and methods of plant breeding*. New York: Elsevier; 1990.
20. Bashir T, Chandra Mishra R, Hasan M, Mohanta TK, Bae H. Effect of hybridization on somatic mutations and genomic rearrangements in plants. *Int J Mol Sci*. 2018; 19: 3758.
21. Guo W, Xie K, Zheng B, Wang S, Fang Y, Xu Q, et al. Molecular marker-assisted cell engineering and citrus improvement. *Acta Hortic*. 2016: 265-269.
22. Baenziger PS. Reflections on doubled haploids in plant breeding. In: *In vitro haploid production in higher plants*. Dordrecht: Springer; 1996. p. 35-48.

23. Humphreys DG, Knox RE. Doubled haploid breeding in cereals. In: *Advances in plant breeding strategies: Breeding, biotechnology and molecular tools*. Switzerland: Springer; 2015. p. 241-290.
24. Guzy-Wróbelska J, Szarejko I. Molecular and agronomic evaluation of wheat doubled haploid lines obtained through maize pollination and anther culture methods. *Plant Breed*. 2003; 122: 305-313.
25. Barkley AP, Chumley FG. A doubled haploid laboratory for Kansas wheat breeding: An economic analysis of biotechnology adoption. *Int Food Agribusiness Manag Rev*. 2012; 15: 99-120.
26. Begheyn RF, Lübberstedt T, Studer B. Haploid and doubled haploid techniques in perennial ryegrass (*Lolium perenne* L.) to advance research and breeding. *Agronomy*. 2016; 6: 60.
27. Thomas W, Forster B, Gertsson B. Doubled haploids in breeding. In: *Doubled haploid production in crop plants*. Berlin/Heidelberg: Springer; 2003. p. 337-349.
28. Langridge P, Fleury D. Making the most of 'omics' for crop breeding. *Trends Biotechnol*. 2011; 29: 33-40.
29. Liu S, Rudd J, Guihua B, Scott H. Molecular markers linked to genes important for hard winter wheat production and marketing in the U.S. great plains. *Crop Sci*. 2014; 54: 1304-1321.
30. Randhawa HS, Asif M, Pozniak C, Clarke JM, Graf RJ, Fox SL, et al. Application of molecular markers to wheat breeding in Canada. *Plant Breed*. 2013; 132: 458-471.
31. Chaikam V, Molenaar W, Melchinger AE, Boddupalli PM. Doubled haploid technology for line development in maize: Technical advances and prospects. *Theor Appl Genet*. 2019; 132: 3227-3243.
32. Briggs F, Knowles PF. *Introduction to plant breeding*. London: Reinhold Publishing Corporation; 1967
33. Poehlman JM, Sleper DA. *Breeding field crops*. Iowa: Iowa State Univ Press; 1995.
34. Osei M, Asante M, Agyeman A, Adebayo M, Adu-Dapaah H. Plant breeding: A tool for achieving food sufficiency. In: *Sustainable horticultural systems*. Switzerland: Springer; 2014. p. 253-274.
35. de Oliveira Amaral L, Bruzi AT, de Resende PM, Silva KB. Pure line selection in a heterogeneous soybean cultivar. *Crop Breed Appl Biotechnol*. 2019; 19: 277-284.
36. Stadler LJ. Genetic effects of X-rays in maize. *Proc Natl Acad Sci U S A*. 1928; 14: 69-75.
37. Lundqvist U. Scandinavian mutation research in barley-a historical review. *Hereditas*. 2014; 151: 123-131.
38. Jain SM. Major mutation-assisted plant breeding programs supported by FAO/IAEA. *Plant Cell Tissue Organ Cult*. 2005; 82: 113-123.
39. Taheri S, Abdullah TL, Jain SM, Sahebi M, Azizi P. TILLING, high-resolution melting (HRM), and next-generation sequencing (NGS) techniques in plant mutation breeding. *Mol Breed*. 2017; 37: 40.
40. Prasanna S, Jain SM. Mutant resources and mutagenomics in crop plants. *Emir J Food Agric*. 2017; 29: 651-657.
41. Flachowsky G, Meyer U. Challenges for plant breeders from the view of animal nutrition. *Agriculture*. 2015; 5: 1252-1276.
42. Lenaerts B, Collard BC, Demont M. Improving global food security through accelerated plant breeding. *Plant Sci*. 2019; 287: 110207.

43. Bruins M. The evolution and contribution of plant breeding to global agriculture. Proceedings of the Second World Seed Conference; 2009 September 8-10; Rome.
44. Tuberosa R. Phenotyping for drought tolerance of crops in the genomics era. *Front Physiol.* 2012; 3: 347.
45. Collard BC, Mackill DJ. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos Trans R Soc Lond B Biol Sci.* 2008; 363: 557-572.
46. Desta ZA, Ortiz R. Genomic selection: Genome-wide prediction in plant improvement. *Trends Plant Sci.* 2014; 19: 592-601.
47. Jiang GL. Molecular marker-assisted breeding: A plant breeder's review. In: *Advances in plant breeding strategies: Breeding, biotechnology and molecular tools.* Switzerland: Springer; 2015. p. 431-472.
48. Moose SP, Mumm RH. Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiol.* 2008; 147: 969-977.
49. Bisognin DA. Breeding vegetatively propagated horticultural crops. *Crop Breed Appl Biotechnol.* 2011; 11: 35-43.
50. Jiang GL. Molecular markers and marker-assisted breeding in plants. In: *Plant breeding from laboratories to fields.* Norderstedt, Germany: Books on Demand; 2013. p. 45-83.
51. Zhao Y, Mette M, Gowda M, Longin C, Reif J. Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. *Heredity.* 2014; 112: 638-645.
52. Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, De Los Campos G, et al. Genomic selection in plant breeding: Methods, models, and perspectives. *Trends Plant Sci.* 2017; 22: 961-975.
53. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 2001; 157: 1819-1829.
54. Rincent R, Laloë D, Nicolas S, Altmann T, Brunel D, Revilla P, et al. Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: Comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). *Genetics.* 2012; 192: 715-728.
55. de Los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MP. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics.* 2013; 193: 327-345.
56. Hickey JM, Chiurugwi T, Mackay I, Powell W. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat Genet.* 2017; 49: 1297-1303.
57. Keller B, Ariza-Suarez D, de la Hoz J, Aparicio JS, Portilla-Benavides AE, Buendia HF, et al. Genomic prediction of agronomic traits in common bean (*Phaseolus vulgaris* L.) under environmental stress. *Front Plant Sci.* 2020; 11: 1001.
58. Crossa J, Pérez P, de los Campos G, Mahuku G, Dreisigacker S, Magorokosho C. Genomic selection and prediction in plant breeding. *J Crop Improv.* 2011; 25: 239-261.
59. Fikere M, Barbulescu DM, Malmberg MM, Maharjan P, Salisbury PA, Kant S, et al. Genomic prediction and genetic correlation of agronomic, blackleg disease, and seed quality traits in canola (*Brassica napus* L.). *Plants.* 2020; 9: 719.
60. Windhausen VS, Atlin GN, Hickey JM, Crossa J, Jannink JL, Sorrells ME, et al. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *G3-Genes Genom Genet.* 2012; 2: 1427-1436.

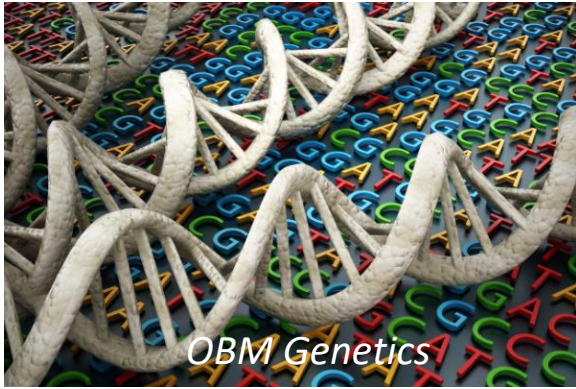


61. Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin FH, et al. Accuracy of genomic selection in European maize elite breeding populations. *Theor Appl Genet.* 2012; 124: 769-776.
62. Crossa J, Beyene Y, Kassa S, Pérez-Rodríguez P, Hickey JM, Chen C, et al. Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3 (Bethesda).* 2013; 3: 1903-1926.
63. Bernardo R, Yu J. Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 2007; 47:1082-1090.
64. Lorenzana RE, Bernardo R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor Appl Genet.* 2009; 120: 151-161.
65. Heffner EL, Jannink JL, Sorrells ME. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome.* 2011; 4: 65-75.
66. de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, et al. Predicting quantitative traits with regression model for dense molecular markers and pedigree. *Genetics.* 2009; 182: 375-385.
67. Crossa J, de los Campos G, Pérez P, Gianola D, Burgueño J, Araus JL, et al. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics.* 2010; 186: 713-724.
68. Burgueno J, Crossa J, Miguel Cotes J, San Vicente F, Das B. Prediction assessment of linear mixed models for multi-environment trials. *Crop Sci.* 2011; 51: 944-954.
69. González-Camacho J, de Los Campos G, Pérez P, Gianola D, Cairns J, Mahuku G, et al. Genome-enabled prediction of genetic values using radial basis function neural networks. *Theor Appl Genet.* 2012; 125: 759-771.
70. Hickey JM, Crossa J, Babu R, de los Campos G. Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. *Crop Sci.* 2012; 52: 654-663.
71. Thistlethwaite FR, Ratcliffe B, Klápště J, Porth I, Chen C, Stoehr MU, et al. Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. *BMC Genom.* 2017; 18: 1-16.
72. Juliana P, Singh RP, Poland J, Mondal S, Crossa J, Montesinos-López OA, et al. Prospects and challenges of applied genomic selection—a new paradigm in breeding for grain yield in bread wheat. *Plant Genome.* 2018; 11. doi: 10.3835/plantgenome2018.03.0017.
73. Pandey MK, Chaudhari S, Jarquin D, Janila P, Crossa J, Patil SC, et al. Genome-based trait prediction in multi-environment breeding trials in groundnut. *Theor Appl Genet.* 2020; 133: 3101-3117.
74. Sapkota S, Boatwright JL, Jordan K, Boyles R, Kresovich S. Multi-trait regressor stacking increased genomic prediction accuracy of sorghum grain composition. *Agronomy.* 2020; 10: 1221.
75. Roth M, Muranty H, Di Guardo M, Guerra W, Patocchi A, Costa F. Genomic prediction of fruit texture and training population optimization towards the application of genomic selection in apple. *Hort Res.* 2020; 7: 1-14.
76. Zhang A, Wang H, Beyene Y, Semagn K, Liu Y, Cao S, et al. Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 biparental tropical maize populations. *Front Plant Sci.* 2017; 8: 1916.

77. Jarquín D, Lemes da Silva C, Gaynor RC, Poland J, Fritz A, Howard R, et al. Increasing genomic-enabled prediction accuracy by modeling genotypex environment interactions in Kansas wheat. *Plant Genome*. 2017; 10: 1-15.
78. Zhang J, Song Q, Cregan PB, Jiang GL. Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycinemax*). *Theor Appl Genet*. 2016; 129: 117-130.
79. Bao Y, Vuong T, Meinhardt C, Tiffin P, Denny R, Chen S, et al. Potential of association mapping and genomic selection to explore PI 88788 derived soybean cyst nematode resistance. *Plant Genome*. 2014; 7. doi: 10.3835/plantgenome2013.11.0039.
80. Bao Y, Kurle JE, Anderson G, Young ND. Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. *Molecular Breed*. 2015; 35: 1-14.
81. Haikka H. Genomic prediction in practical breeding program: A case study in oat and barley. Helsinki, Finland: University of Helsinki; 2021.
82. Batista L, Mello VH, de Souza AP, Margarido GR. Genomic prediction with allele dosage information in highly polyploid species. *bioRxiv*. 2021. doi: 10.1101/2021.06.22.449437.
83. Hu Y, Morota G, Rosa GJ, Gianola D. Prediction of plant height in *Arabidopsis thaliana* using DNA methylation data. *Genetics*. 2015; 201: 779-793.
84. Calus MP, Veerkamp RF. Accuracy of multi-trait genomic selection using different methods. *Genet Sel Evol*. 2011; 43: 1-14.
85. Jia Y, Jannink JL. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics*. 2012; 192: 1513-1522.
86. Hayashi T, Iwata H. A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. *BMC Bioinform*. 2013; 14: 1-14.
87. Gebreyesus G, Lund MS, Buitenhuis B, Bovenhuis H, Poulsen NA, Janss LG. Modeling heterogeneous (co) variances from adjacent-SNP groups improves genomic prediction for milk protein composition traits. *Genet Sel Evol*. 2017; 49: 1-13.
88. Cheng H, Kizilkaya K, Zeng J, Garrick D, Fernando R. Genomic prediction from multiple-trait Bayesian regression methods using mixture priors. *Genetics*. 2018; 209: 89-103.
89. Guo G, Zhao F, Wang Y, Zhang Y, Du L, Su G. Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet*. 2014; 15: 1-7.
90. Karaman E, Lund MS, Anche MT, Janss L, Su G. Genomic prediction using multi-trait weighted GBLUP accounting for heterogeneous variances and covariances across the genome. *G3-Genes Genom Genet*. 2018; 8: 3549-3558.
91. Burgueño J, de los Campos G, Weigel K, Crossa J. Genomic prediction of breeding values when modeling genotypex environment interaction using pedigree and dense molecular markers. *Crop Sci*. 2012; 52: 707-719.
92. Roorkiwal M, Jarquin D, Singh MK, Gaur PM, Bharadwaj C, Rathore A, et al. Genomic-enabled prediction models using multi-environment trials to estimate the effect of genotypex environment interaction on prediction accuracy in chickpea. *Sci Rep*. 2018; 8: 1-11.
93. Gillberg J, Marttinen P, Mamitsuka H, Kaski S. Modelling Gx E with historical weather information improves genomic prediction in new environments. *Bioinformatics*. 2019; 35: 4045-4052.

94. Tsai HY, Cericola F, Edriss V, Andersen JR, Orabi J, Jensen JD, et al. Use of multiple traits genomic prediction, genotype by environment interactions and spatial effect to improve prediction accuracy in yield data. *PloS one*. 2020; 15: e0232665.
95. Xu S, Zhu D, Zhang Q. Predicting hybrid performance in rice using genomic best linear unbiased prediction. *PNAS*. 2014; 111: 12456-12461.
96. Monteverde E, Rosas JE, Blanco P, de Vida FP, Bonnacarrère V, Quero G, et al. Multienvironment models increase prediction accuracy of complex traits in advanced breeding lines of rice. *Crop Sci*. 2018; 58: 1519-1530.
97. Lopez-Cruz M, Crossa J, Bonnett D, Dreisigacker S, Poland J, Jannink JL, et al. Increased prediction accuracy in wheat breeding trials using a marker× environment interaction genomic selection model. *G3-Genes Genom Genet*. 2015; 5: 569-582.
98. Cuevas J, Crossa J, Soberanis V, Pérez-Elizalde S, Pérez-Rodríguez P, Campos G, et al. Genomic prediction of genotype× environment interaction kernel regression models. *Plant Genome*. 2016; 9: 1-20.
99. Saint Pierre C, Burgueño J, Crossa J, Dávila GF, López PF, Moya ES, et al. Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. *Sci Rep*. 2016; 6: 1-11.
100. Cuevas J, Crossa J, Montesinos-López OA, Burgueño J, Pérez-Rodríguez P, de Los Campos G. Bayesian genomic prediction with genotype× environment interaction kernel models. *G3-Genes Genom Genet*. 2017; 7: 41-53.
101. Heslot N, Jannink JL, Sorrells ME. Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data. *Crop Science*. 2013; 53: 921-933.
102. Bohlouli M, Alijani S, Javaremi AN, König S, Yin T. Genomic prediction by considering genotype× environment interaction using different genomic architectures. *Ann Anim Sci*. 2017; 17: 683-701.
103. Boer MP, Wright D, Feng L, Podlich DW, Luo L, Cooper M, et al. A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. *Genetics*. 2007; 177: 1801-1813.
104. van Eeuwijk FA, Bink MC, Chenu K, Chapman SC. Detection and use of QTL for complex traits in multiple environments. *Curr Opin Plant Biol*. 2010; 13: 193-205.
105. Hassen MB, Bartholomé J, Valè G, Cao TV, Ahmadi N. 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 Genom Genet*. 2018; 8: 2319-2332.
106. Mageto EK, Crossa J, Pérez-Rodríguez P, Dhliwayo T, Palacios-Rojas N, Lee M, et al. Genomic prediction with genotype by environment interaction analysis for kernel zinc concentration in tropical maize germplasm. *G3-Genes Genom Genet*. 2020; 10: 2629-2639.
107. Jarquín D, Crossa J, Lacaze X, Du Cheyron P, Daucourt J, Lorgeou J, et al. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor Appl Genet*. 2014; 127: 595-607.
108. Wang X, Xu Y, Hu Z, Xu C. Genomic selection methods for crop improvement: Current status and prospects. *Crop J*. 2018; 6: 330-340.

109. Habier D, Fernando RL, Dekkers JC. The impact of genetic relationship information on genome-assisted breeding values. *Genetics*. 2007; 177: 2389-2397.
110. Hallander J, Waldmann P, Wang C, Sillanpää MJ. Bayesian inference of genetic parameters based on conditional decompositions of multivariate normal distributions. *Genetics*. 2010; 185: 645-654.
111. Kärkkäinen HP, Sillanpää MJ. Back to basics for Bayesian model building in genomic selection. *Genetics*. 2012; 191: 969-987.
112. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol*. 1996; 58: 267-288.
113. Neves HH, Carvalheiro R, Queiroz SA. A comparison of statistical methods for genomic selection in a mice population. *BMC Genet*. 2012; 13: 1-17.
114. Li Y, Raidan F, Vitezica ZG, Reverter A. Using Random Forests as a prescreening tool for genomic prediction: impact of subsets of SNPs on prediction accuracy of total genetic values. *Proceedings of the 11th World Congress on Genetics Applied to Livestock Production (WCGALP)*; 2018 February 11-16; Auckland, New Zealand.
115. Waldmann P. Genome-wide prediction using Bayesian additive regression trees. *Genet Sel Evol*. 2016; 48: 1-12.
116. Chen X, Ishwaran H. Random forests for genomic data analysis. *Genomics*. 2012; 99: 323-329.
117. Alarcon F, Perduca V, Nuel G. Is it possible to detect  $G \times E$  interactions in GWAS when causal exposure is unobserved? *J Epidemiol Res*. 2015; 2: 109-117.
118. Li J, Malley JD, Andrew AS, Karagas MR, Moore JH. Detecting gene-gene interactions using a permutation-based random forest method. *BioData Min*. 2016; 9: 1-17.
119. Lubke G, Laurin C, Walters R, Eriksson N, Hysi P, Spector T, et al. Gradient boosting as a SNP filter: An evaluation using simulated and hair morphology data. *J Data Min Genom Proteom*. 2013; 4: 1-7.
120. Chen T, He T. Higgs boson discovery with boosted trees. *PMLR*. 2015; 42: 69-80.
121. Zhang H, Yin L, Wang M, Yuan X, Liu X. Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. *Front Genet*. 2019; 10: 189.
122. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*. 2011; 6: e19379.
123. Rajaratnam B, Roberts S, Sparks D, Yu H. Influence diagnostics for high-dimensional lasso regression. *J Comput Graph Stat*. 2019; 28: 877-890.
124. Budhlakoti N, Rai A, Mishra DC. Statistical approach for improving genomic prediction accuracy through efficient diagnostic measure of influential observation. *Sci Rep*. 2020; 10: 1-11.
125. Zhu B, Zhang JJ, Niu H, Guan L, Guo P, Xu LY, et al. Effects of marker density and minor allele frequency on genomic prediction for growth traits in Chinese Simmental beef cattle. *J Integr Agric*. 2017; 16: 911-920.
126. Berro I, Lado B, Nalin RS, Quincke M, Gutiérrez L. Training population optimization for genomic selection. *Plant Genome*. 2019; 12: 190028.



Enjoy *OBM Genetics* by:

1. [Submitting a manuscript](#)
2. [Joining in volunteer reviewer bank](#)
3. [Joining Editorial Board](#)
4. [Guest editing a special issue](#)

For more details, please visit:

<http://www.lidsen.com/journals/genetics>