Chapter 5

The Usage of Genomic Selection Strategy in Plant Breeding

Mohd Shamshad and Achla Sharma

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.76247

Abstract

Major paradigm shift in plant breeding since the availability of molecular marker technology is that mapping and characterizing the genetic loci that control a trait will lead to improved breeding. Often, one of the rationales for cloning of QTL is to develop the "perfect marker" for MAS, perhaps based on a functional polymorphism. In contrast, an advantage of genomic selection is precisely its black box approach to exploiting genotyping technology to expedite genetic progress. This is an advantage in our view because it does not rely on a "breeding by design" engineering approach to cultivar development requiring knowledge of biological function before the creation of phenotypes. Breeders can therefore use genomic selection without the large upfront cost of obtaining that knowledge. In addition, genomic selection can maintain the creative nature of phenotypic selection which couple's random mutation and recombination to sometimes arrive at solutions outside the engineer's scope. Currently, the lion's share of research on genomic selection has been performed in livestock breeding, where effective population size, extent of LD, breeding objectives, experimental design, and other characteristics of populations and breeding programs are quite different from those of crop species. Nevertheless, a great number of findings within this literature are very illuminating for genomic selection in crops and should be studied and built upon by crop geneticists and breeders. The application of powerful, relatively new statistical methods to the problem of high dimensional marker data for genomic selection has been nearly as important to the development of genomic selection as the creation of high-density marker platforms and greater computing power. The methods can be classified by what type of genetic architecture they try to capture.

Keywords: genomic selection, training population, breeding population, linkage disequilibrium, genomic prediction model, breeding value

IntechOpen

© 2018 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Marker-assisted selection (MAS) is an important scheme in plant breeding since the 1990s, after promising analysis results for tagging genes or mapping QTL [1]. Marker assisted selection and molecular breeding have been used in the identification of underlying major genes in gene pools and their transfer to desirables traits of major plant breeding programs. Using of MAS have shown some shortcomings due to long selection schemes and also the look for vital marker-QTL associations being unable to capture "minor" gene effects. Thus marker-assisted selection (MAS) is difficult to improve traits having complex inheritance such as grain yield and abiotic stresses.

Using whole-genome prediction models, the genomic selection (GS) strategy has paved the way to over-come these limitations. High-density molecular markers using is one of the main features of genomic selection. Therefore, each of the trait loci has the likelihood of being in linkage disequilibrium (LD) with a minimum of one marker locus within the entire breeding population. Genome selection strategy removes the need to mapping of genes and search for linked QTL–marker loci associated individually. Rather, Genomic selection accounts for bunches of predictors simultaneously and is characterized by constraining random estimates towards zero. Moreover, Genomic selection helpful for accelerate breeding cycles in such a way that the rate of annual genetic gain per unit of time and cost can be decreased. Genomic selection has been well established in the field of animal breeding, but is in its beginning in crops plants and forest tree breeding.

Genome-wide selection or genomic selection estimates marker effects across the full ordering of the breeding population (BP) supported the prediction model developed within the training population (TP). Training population could be a group of related individuals (such as half-sibs or lines) that are each phenotypes and genotypes. Breeding population typically is just genotyped not phenotypes. Hence, Genomic selection depends on the degree of genetic similarity between training population and breeding population within the Linkage disequilibrium between marker and trait loci. Breeding values have not been a preferred index in plant breeding, however it is in animal breeding. Once plan of genomic estimated breeding value (GEBV) was planned, it had been considered an unrealistic approach due to lack of enormous scale genotyping technologies. However, currently, it has been a possible approach with recent advances in high throughput genotyping platforms (3rd generation platforms). Generally processes of genomic selection and marker assisted selection used for Quantitative Traits are shown in **Figure 1**.

The main schemes of the two approaches are similar, wherever each marker assisted selection and genomic selection consist of breeding and training phases. In the training phase, phenotypes and genome-wide (GW) genotypes are investigated in an exceedingly set of a population, i.e., the mapping population in marker assisted selection and also the training population in genomic selection. Among populations, important relationships between phenotypes and genotypes are expected utilizing statistical models. Within the breeding phase, genotype data are obtained in an exceedingly breeding population, on the basis of genotypic information favorable individuals are selected. There are three prominent variations

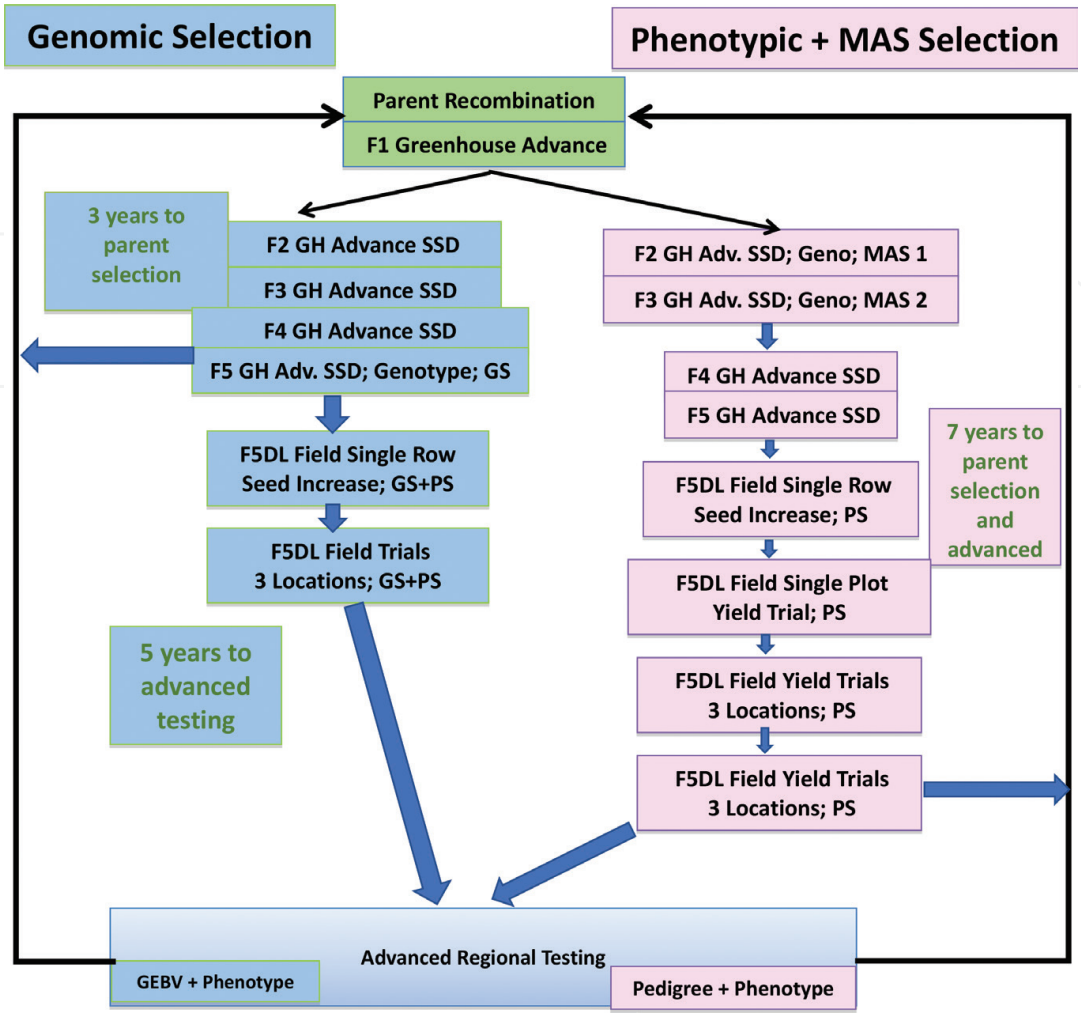

Figure 1. Self-pollinated crop genomic selection vs. phenotypic/MAS selection timeline.

between the two approaches: (1) within the training section, quantitative trait loci (QTLs) are known in marker assisted selection whereas formulae for genetic estimation of breeding value prediction are generated in genomic selection, called genomic selection models; (2) within the breeding section, genotype data are solely needed for targeted regions in marker assisted selection, whereas genomic selection genotype data are considered to be mandatory in genomic selection (3) within the breeding phase, favorable individuals are selected on the bases of the linked markers in marker assisted selection, whereas GEBVs are used for selection in GS. Thus, GS collectively analyses all the genetic variance of every individual by summing the marker impacts of GEBV and it is expected to deal with little effect genes that cannot be captured by traditional MAS.

The statistical ways employed by GS are comparatively new the plant-breeding community. The ways of marker-assisted selection (MAS) or marker-assisted recurrent selection (MARS) assume that the user is aware of that alleles are favorable, and what their average effects on the phenotype are. This assumption is viable for major-gene traits however not for quantitative traits that are influenced by several loci of little impact and the environment. To deal with

quantitative traits, new statistical approaches that might account for this uncertainty were required to get the most effective predictions potential. Finding problem with locus identification, entailed that the consequences for all marker loci be at the same time estimated. Once a prediction based on allele effects, the allele becomes the unit of analysis. Alleles are so the units that need to be replicated inside and across environments. However that replication will occur in spite of the particular lines carrying the alleles such lines themselves no longer need to be replicated. Within the breeding context, removing the requirement for line replication opens the likelihood of dramatically increasing the amount of lines pushed through the pipeline of a breeding program, and successively of accelerating selection intensity.

2. Genomic selection scheme

Genomic selection is to assemble a training population for individuals for which both genotypes and phenotypes are available and use those data to create a statistical model that relates variation in observed genotypes marker loci to variation in the observed phenotypes of the individuals. Multiple generations of parents and progenies provided powerful training population than a single generation individuals and larger number of individual's generations and markers provide more powerful training population (TP). The statistical model obtained from genotype and phenotype is then applied to a prediction population comprised of individuals for which genotypes are available, but phenotypes are not. GS is based on similarity between the training population (TP) and breeding population (BP) in the LD between marker loci and trait loci. This similarity may exist because breeding population is selected from training population or descended from training population or because density of markers is so high that every trait locus is in disequilibrium with at least one marker locus across the entire population of the target species. The training population is genotyped and phenotyped to train the genomic selection (GS) prediction model. In Genomic selection main role of phenotyping is to calculate effect of markers & cross validation. Genotypic information from the breeding material is then fed into the model to calculate genome estimated breeding values (GEBV) for these lines (**Figure 2**).

2.1. Need of genomic selection

Traditional marker assisted selection, whereas helpful for merely transmitted traits controlled by few loci, loses effectiveness because the number of loci will increase. This is often true for individual quantitative traits or once multiple traits are below selection. Quantitative traits like grain yield, abiotic stress have verified hard to enhance with marker-assisted selection. The main limitations are (i) tiny population sizes and traditional statistical strategies that have inadequate power to find and accurately estimate effects of small-effect quantitative trait loci (QTL) and (ii) gene x gene interactions (epistasis) and (iii) genotype x environment interactions (G.E) that have restricted the exchangeability of quantitative trait loci result estimates across populations and environments. The Beavis effect is a statistical phenomenon in biology that refers to the overestimation of the effect size of quantitative trait loci (QTL) as a result of small sample sizes in QTL studies.

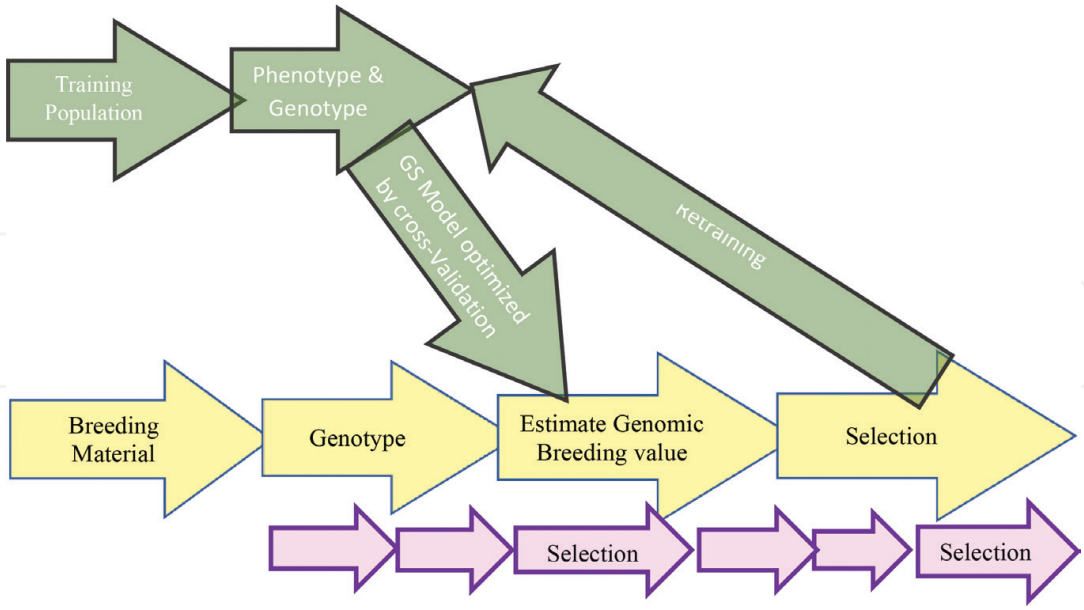

Figure 2. Genomic selection scheme. Information on phenotype and genotype for a training population allows estimating parameters for the model. (Modified: Castro et al. 2012) [12].

The availability of low cost and extensive molecular markers in plants has allowed breeders to raise however molecular markers might best be used to win breeding progress. Additionally advances in high-throughput genotyping have markedly reduced the value per data point of molecular markers and increasing genome coverage. This reduction was in the main the results of three parallel developments [2] (i) the invention of huge numbers of single nucleotide polymorphism (SNP) markers in several species; (ii) development of high-throughput technologies, like multiplexing and gel-free deoxyribonucleic acid arrays, for screening SNP polymorphisms; and (iii) automation of the marker-genotyping method, together with efficient procedures for deoxyribonucleic acid extraction [2]. Phenotyping prices are increased Genotyping prices are being reduced and marker densities are being increased speedily.

Statistical strategies are inadequate for improving polygenic traits controlled by several loci of small impact. There will be more markers (explanatory variables) than lines (observations) that introduce statistical issues. Drawback of small p (number of traits) and enormous m (number of markers) ends up in a lack of degrees of freedom. The foremost acceptable statistical model is required to at the same time estimate several marker effects from a limited range of phenotypes. In so-called "large p, small m" problems, standard multiple linear regression cannot be used without variable selection, that conflicts with the initial goal of avoiding marker selection. To overcome these issues, a range of ways, e.g., best linear unbiased prediction, ridge regression, Bayesian regression, kernel regression and machine learning methods are projected to develop prediction models for genomic selection.

The most economical use of GS is to exchange expensive and long phenotyping by a prediction of the genetic worth of the character below selection (or any multi trait index). Thus, the foremost expected advantage is to shorten selection cycles. However, to learn from shorter

cycles, the genetic gain per selection cycle ought to be near that predicted from phenotypic or combined MAS + phenotypic selection. Progeny testing schemes have a high accuracy of selection, however the time interval is also additional, takes long term to perform a cycle of selection that decreases the genetic gain. The univariate breeder's equation was used for the GS-BPs as a result of they include just one stage of selection [3]. Selection accuracy is adequate to the correlation between selection criteria and breeding value (i.e., correlation between phenotypes or GEBVs and true breeding values [TBVs]). In oxen, Schaeffer [4] determined that the time and value savings exploitation GS with GEBV accuracy of 0.75 would increase genetic gain twofold and supply a price savings of ninety two in comparison to the present ways.

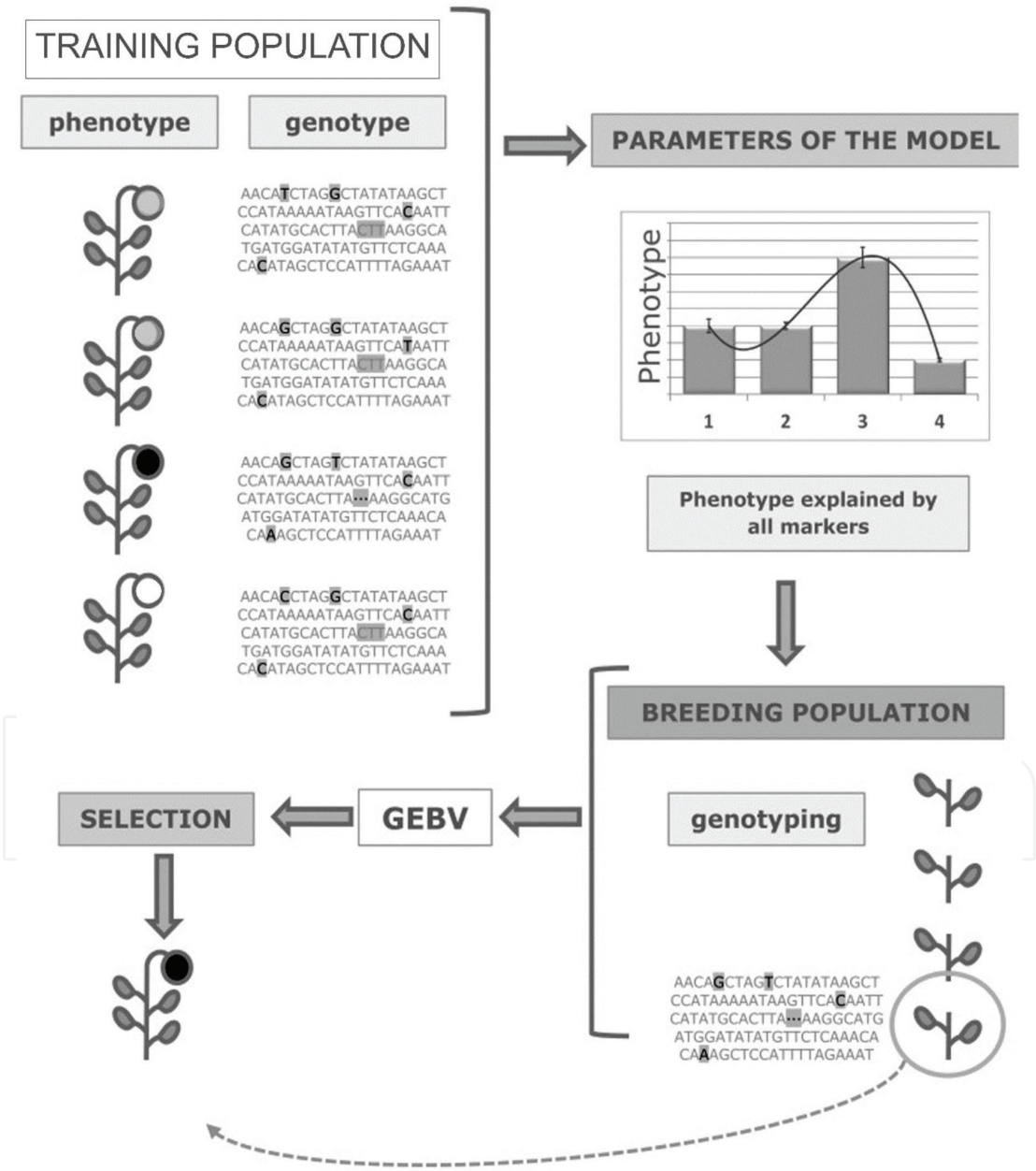

Figure 3. Genomic selection scheme.

The power to calculate extremely correct GEBVs and also the potential to drastically cut back makeup analysis frequency and selection cycle time expedited a speedy adoption of genomic selection and is revolutionizing the oxen breeding trade (**Figure 3**).

2.2. Model for genomic selection

The basic model may be denoted as

$$Y_i = g(x_i) + e_i \tag{1}$$

where Y_i is an observed phenotype of individual i (i = 1 … n) and x_i is a *1 x p* vector of SNP genotypes on individual *i*, *g(xi)* is a function relating genotypes to phenotypes, and e_i a residual term. The GEBV is generally equal to g (x_i). Further similarities among GS models can be seen by recognizing that they all seek to minimize a certain cost function. In least squares analysis, the well-known cost function is simply the sum of squared residuals.

2.3. Cross-validation

Evaluating GEBV accuracy through cross validation (CV). CV entails splitting the data into training and validation set. The ratio of observations in each set varies, but often a fivefold CV is used, that is, the data set is randomly divided into five sets, with four sets being combined to form the training set and the remaining set designated as the validation set. Each subset of the data is used as the validation set once, before applying of the prediction model to the breeding population, the accuracy of the model should be tested. For this, most of the training population is used to create a prediction model, which is then used to estimate the genomic estimation breeding values of the remaining individuals in the training population, using genotypic data only. This permits researchers to "test" and refine the prediction model to make sure the prediction accuracy is high enough that future predictions are often relied upon. Once valid, the model is often applied to a breeding population to calculate GEBVs of lines that genotypical, however no phenotypical, information is available.

2.4. Genomic selection prediction accuracies

The prediction accuracy of the GEBVs is evaluated by the correlation between the GEBVs and empirically estimated breeding values, r(GEBV: EBV), where the EBV can be obtained in a number of ways, most simply, as a phenotypic mean. This correlation provides an estimate of selection accuracy and thus directly relates GEBV prediction accuracy to selection response [2]. Other statistics such as mean-square error (MSE) are used occasionally [3]. Genomic selection accuracy is defined as the correlation between GEBV and the true breeding value (TBV), that is, r(GEBV:TBV). Since we can only measure r(GEBV:EBV), this measure needs to be converted to an estimate of r(GEBV:TBV). To do so, it is assumed that r(GEBV:EBV) = r(GEBV:TBV) X r(EBV:TBV). This assumption is correct if the only component common between the GEBV and the EBV is the TBV itself. In other words, the assumption holds if GEBV = TBV + e1 and EBV = TBV + e2, where e1 and e2 are uncorrelated residuals. The assumption could be violated if the training and validation data were collected in the same

environment. In that case, genotype by environment (G.E) interaction would generate a common component of error in both GEBV and EBV, biasing their correlation upward. Thus, training and validation data should be collected in different environments to ensure sound estimates of GEBV prediction accuracy. The correction, r(EBV:TBV), accounts for the fact that the EBV in the validation set is not free of error. When the EBVs are phenotypes, r(EBV:TBV) is equal to the square root of heritability (h) within the validation set [2].

Accurate GEBV predictions offer the possibility that future elite and parental lines will be selected on GEBV rather than phenoypic data from extensive field testing. Immediate impact would be a great increase in speed of breeding cycle increasing selection gains per unit time. Thus, GS could radically change the practice of field evaluation for breeders. Of course, regardless of the breeding method used, final field evaluations of varieties across the target environments will be needed before they are distributed to farmers.

Breeding cycle time is shortened by removing phenotypic evaluation of lines before selection as parents for the next cycle. Model training and line development cycle length will be crop and breeding program specific. In a GS breeding schema, genome-wide DNA markers are used to predict which individuals in a breeding population are most valuable as parents of the next generation of offspring. The prediction model is additionally continuously rejuvenated as genotypical and phenotypic data from elite lines derived from the collaborating breeding programs is incorporated into the prediction models. In this manner, new germplasm may be infused into the system at any point. As lines derived from the recently infused germplasm advance within the breeding process, their genotypical and phenotypic data may be incorporated into the prediction models.

The purpose of phenotyping now is to pick the best lines from a segregating population and to judge fewer lines with larger replication in every cycle of selection. However during a GS driven breeding cycle, the aim of phenotyping is to estimate or re-estimate marker effects. It is far from clear at this point whether or not it will be advantageous to evaluate solely the best lines or to evaluate few lines with high replication. So separates the germplasm improvement cycle from the prediction model improvement cycle. Indeed, if we tend to use the rules for optimum QTL linkage mapping, analysis should include not just the best however the best and the the} worst lines **Figure 3** also emphasizes the requirement for model updating and re-evaluation. Marker effects might amendment as a results of allomorph frequency changes [6] or of epistatic gene action. Model updating with every breeding cycle should mitigate reduced gains from GS caused by these mechanisms. Thus, GS may radically change the practice of field evaluation for breeders. Of course, despite the breeding technique used, final field evaluations of varieties across the target environments are going to be needed before they are distributed to farmers.

Accuracy declines as generation number between the last model update and selection candidates increases [4–6], because selection causes changes in variances, allele frequencies, and LD relationships between markers and QTL [4]. Under random mating, simulations have shown model accuracy to decrease by about 5% per generation [5, 6], but accuracy decrease was much more rapid under selection.

3. Factors responsible for the estimation accuracy of GS models

The response of genomic selection is that the output of varied factors responsible for estimation accuracy of GEBVs. These factors are interconnected in an exceedingly advanced and comprehensive manner. They include model performances, sample size and relatedness, marker density, gene effects, heritability and genetic design, and therefore the extent and distribution of linkage disequilibrium between markers and QTL.

3.1. Population size

The most important characteristic of the population is its effective size. An obvious measure of population size is its census: how many individuals it contains. But populations with the same census size can behave quite differently. For a population of a given rate of inbreeding, the effective size is equal to the census size of a randomly mating ("ideal") population that would have that same rate.

Accuracy due to genetic relationships can represent from a small minority to a large majority proportion of the overall accuracy. The combination of long-distance LD due to pedigree relatedness (e.g., full sibs and half sibs) and short-distance ancestral LD due to small effective population size are among the key features of our training population. With improved marker technology, large TPs that use a representative sample of germplasm in a given breeding program may be a good strategy for long-term accuracy over a broad range of families. It has been observed to be monotonic increase in the prediction accuracy for grain yield with increasing population size without any substantial decrease in the slope (**Figure 4**). Studies in this the size of the training population is of crucial importance in genomic selection. The impact of the population size on the accuracy of genomic selection is less pronounced for fewer characters like grain moisture, which might be due to presence of larger variance among populations

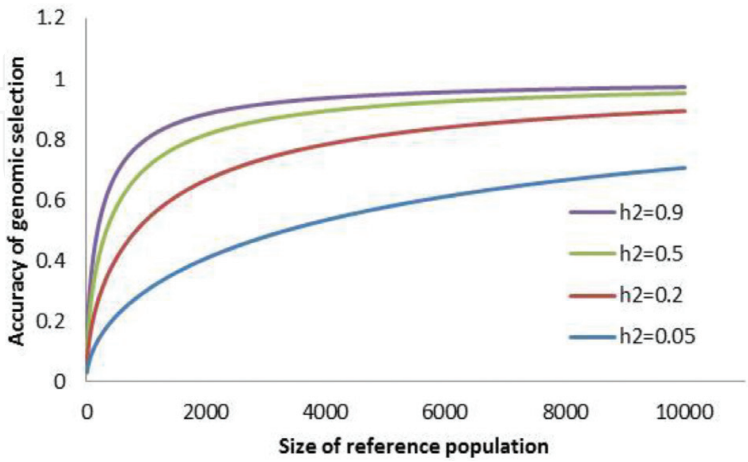

Figure 4. Relation between number of plants in the training population and accuracy of genomic selection for traits with different heritabilities.

that can be efficiently utilized by few individuals per population. Parameters such as effective population size and QTL number strongly influence marker densities and TP sizes required for acceptable accuracy. Indeed, simulations similar to those of Meuwissen et al. [6] have shown that marker density needs to scale with effective population size [7]. Until very low marker densities were reached, marker number had very little, if any, effect on prediction accuracies within families from various plant species [8]. Likewise, GEBV accuracy of several traits in cattle, including net merit, was hardly affected when as many as 75% of the original markers were masked. Adequate marker density and TP size depend on QTL number and trait heritability. Calus and Veerkamp [9] used the average r^2 between adjacent markers as a measure of marker density relative to decay of linkage disequilibrium. They found that for a highly heritability trait, average adjacent marker $r^2$of 0.15 was sufficient, but for a low heri - tability trait, increasing the r^2 to 0.20 improved prediction accuracy. Heritability dramatically affects TP sizes required for successful GS, especially at h^2 less than 0.40 [3].

3.2. Linkage disequilibrium

Genetic drift is an important cause in generating LD, the non-independence of alleles at different loci. This non-independence allows marker alleles to predict the allelic state of nearby QTL, enabling marker genotypes to predict the phenotype. The LD intensity decays with greater distance between two markers. Decay rates which vary widely across species, populations, and genomes due to forces of mutation, recombination, population size, population mating marker density must increase with increases in Ne*c, where Ne is the effective population size and c is the recombination rate between loci. LD patterns. Marker density can be inferred by the rate of LD decay across the genome as inferred by the relationship b/w inter marker coefficient of determination r^2 and genetic distance. LD estimates can be used to determine target marker densities for GS at equilibrium, drift generating LD is balanced by recombination, causing it to decay, such that nearby loci are expected to be in higher LD than faraway loci. LD has a major effect on the operability of GS, so it has to be well understood before performing GS. LD is defined as the non-random association of alleles at different loci. It has been found that for high heritability trait average adjacent marker $r^2$of 0.15 is sufficient but for low heritability trait increasing r^2 value to 0.2 improve accuracy of GEBV predictions.

4. Types of marker platforms

Since, then marker of choice is very important to accurate estimate GEBVBs, different platforms are available

4.1. SNP chips

Single nucleotide polymorphisms (SNPs) differentiate individuals based on variations detected at the level of a single nucleotide base in the genome. SNPs have become the marker of choice for crop genetics and breeding applications because of their high abundance in genomes, and the availability of a wide array of genotyping platforms with various multiplex capabilities for

SNP analysis [10]. Recent breakthroughs in next generation sequencing (NGS) technologies enabled millions of sequences reads to be generated from a single run at a more affordable cost. The ability to perform GS requires routine genotyping at a high number of loci. Single nucleotide polymorphisms (SNPs) differentiate individuals based on variations detected at the level of a single nucleotide base in the genome. SNPs have become the marker of choice for crop genetics and breeding applications because of their high abundance in genomes, and the availability of a wide array of genotyping platforms with various multiplex capabilities for SNP analysis [10]. Recent breakthroughs in next generation sequencing (NGS) technologies enabled millions of sequences reads to be generated from a single run at a more affordable cost. The resulting large amount of data provided sequence depth adequate for de novo sequence assembly, which has made the SNP discovery on a large scale a feasible task, particularly for species without completed genome sequences. Successful results on large-scale discovery of SNPs based on NGS methods have been reported in several plant species, including both and polyploid species, and more are on the way. The development of highly parallel SNP assays, such as Illumina's Golden Gate assay with 1536-plex platform enabled the genome-wide studies previously not feasible for economically important crops. Using these techniques, SNP-based high-density genetic maps are now available in several crop plants such as soybean, maize, barley and wheat. Thus, genotyping lines for use in GS using SNP and direct resequencing with next-generation.

4.2. Genotyping by sequencing (GBS)

Advances in next generation technologies have driven the costs of DNA sequencing down to the point that genotyping-by sequencing (GBS) is now feasible for high diversity, large genome species. GBS is a highly multiplexed approach is based on high-throughput, next-generation sequencing of genomic subsets targeted by restriction enzymes (REs). Genotyping by sequencing (GBS) in any large genome species requires reduction of genome complexity. Genotyping-by-sequencing can be applied to different populations or even different species without any prior genomic knowledge as marker discovery is simultaneous with the genotyping of the population. GBS sequence allows access to any sequence within low copy genomic regions or regulatory regions controlling the expression of plant genes responsible for agronomically important phenotypes are often located in non-coding DNA. The use of GBS for GS, therefore, should be applicable to a range of model and non-model crop species to implement genomics-assisted breeding. Genotyping-by sequencing combines marker discovery and genotyping of large populations, making it a superb marker platform for breeding applications even within the absence of a reference genome sequence or previous polymorphism discovery. Additionally, the pliability and low price of GBS create this an ideal approach for genomics-assisted breeding.

5. Advantages of genomic selection

1. The marker effects are calculable from the training population and used directly for GS within the involved breeding population, and QTL discovery, mapping, etc., are not needed.

2. Each simulation and empirical studies reveal that GS produces larger gains per unit time than constitution selection. For instance, a simulation study in maize showed GS to be superior to MARS, notably for traits having low heritability. Further, GS will predict the performance of breeding lines additional accurately than that supported pedigree data, and GS appears to be an efficient tool for rising the potency of rice breeding.

3. The selection index approach integrates appropriately weighted data from multiple traits into an index that is the premise for concurrent selection for the concerned traits. The genome-wide marker data is integrated into a range index either alone or in conjunction with phenotype data on one or additional traits. Simulation studies show that the on top of combined selection index approach of GS can increase the effectiveness of selection, considerably for low heritability traits.

4. GS would tend to cut back the speed of inbreeding and also the loss of genetic variability as compared to selection based on breeding values calculable from phenotype data; this may be achieved while not sacrificing selection gains. This might be notably vital in species that show severe inbreeding depression.

5. Genomic selection scheme consist of phenotyping for each selection cycle within the breeding population is not needed. This greatly reduces the length of breeding cycle, notably in perennial species. For instance, GS was calculable to reduce the selection cycle time from 19 years to simply 6 years just in case of oil palm (*Elaeis guineensis*). Further, GS was calculable to outperform MARS and phenotypic selection even with a population size of fifty once selection gain was considered on per unit time and price, however not on per selection cycle, basis. (The selection the choice) cycle is reduced as a result of GS does not need analysis of interbreeding performance of the plants being subjected to selection that is critical within the case of phenotypic selection. In perennial species, GS is anticipated to facilitate commercialization of improved genotypes at abundant shorter intervals of time than phenotypic selection.

6. Genomic selection would possibly enable breeders to pick out parents for crossbreeding programs from among those lines that have not been evaluated within the target environment. This selection would be supported GEBVs of these lines estimated for their adaptation to the target environment. This could facilitate germplasm exchange and their utilization in breeding programs.

7. Genotype X environment interaction could be a vital a part of phenotype and its estimation is sort of demanding. GS can utilize information on marker genotype and trait phenotype accumulated over time in varied analysis programs covering a variety of environments and integrate an identical in GEBV estimates of the various individuals/lines. This could enable GEBV estimation even for traits that they have never been tested.

8. Theoretically, GEBV estimates is employed for the selection of parents for crossing programs and, possibly, for the development of hybrid varieties. These applications, however, ought to expect validation of the concept in apply.

6. Limitations of genomic selection

1. GS has still not become popular plant breeding community primarily due to low evidence for its sensible utility. In fact, most discussions on its utility are for the most part statistical treatments and simulations that are not simply appreciated by plant breeders.
2. The potential value of GS should be assessed with caution because GS has been mostly evaluated in simulation studies. There is an imperative have to be compelled to judge genomic selection in crop breeding situations to demonstrate its practical utility.
3. The marker effects and, as a result, GEBV estimates would possibly modification attributable to changes in gene frequencies and epistatic interactions. This is often ready to necessitate amendment of the GS model with every breeding cycle therefore the gains from GS are not reduced.
4. The accuracy of GEBV estimates has been evaluated exploitation simulation models based on additive genetic variance. These models ignore epistatic effects that does not seem to be realistic. It has been argued that since biological process makes alone a small contribution to the breeding value, the employment of solely additive genetic models for GS is additionally expected to maximize selection gains. However, this argument are planning to be entirely valid only for self-fertilizing species, where homozygous lines are used as parents as well as varieties.
5. Our information concerning the genetic design of quantitative traits is severely restricted, that limits our ability to develop applicable models of GS to realize the most prediction accuracy.
6. The selection response declines at a faster rate under GS than with pedigree selection. This may be reduced by continually together with new markers for the prediction of GEBVs. The long response under GS can also be raised by putting higher weights on the low-frequency favorable alleles, considerably within the start of GS program.GS is simpler than phenotypic selection on per unit time basis only if off-season/greenhouse facilities are accustomed grow up to three generations per annum. The utility and also the cost-effectiveness of GS would be uncertain wherever such facilities are not offered.
7. The necessity for genotyping of an oversized variety of marker loci in every generation of selection adds considerably to the price of breeding programs. It has been projected that, inside the future, a bigger stress are going to be placed on the use of marker information than on composition information. Such a shift, however, would need the value of one marker information to be merely 1/5000 the price of phenotyping one entry.
8. Implementation of GS would need intensive infrastructure and completely different resources, which might get on the so much aspect the reach of most moderate size public sector breeding programs, considerably within the developing countries. To boot, planning and execution of GS is kind of exigent, and additionally the breeders would be required to reorient their approach to the breeding programs.

7. Conclusions

Currently, the lion's share of research on GS has been performed in livestock breeding, where effective population size, extent of LD, breeding objectives, experimental design, and other characteristics of populations and breeding programs are quite different from those of crop species. Nevertheless, a great number of findings within this literature are very illuminating for GS in crops and should be studied and built upon by crop geneticists and breeders. The application of powerful, relatively new statistical methods to the problem of high dimensional marker data for GS has been nearly as important to the development of GS as the creation of high-density marker platforms and greater computing power. The methods can be classifiedby what type of genetic architecture they try to capture. Somewhat surprisingly, RR-BLUP, which makes the ostensibly unrealistic assumption that genetic effects are uniformly spread across the genome, often performs as well as more sophisticated models. Exceptions do exist, though, and there is abundant evidence that BayesB is superior for traits with strong QTL effects. Additionally, since BayesB better identifies markers in strong LD with QTL than RR-BLUP, it maintains accuracy for more generations. Finally, the question of whether or not to model epistasis remains open. If epistasis is important for a particular trait in a particular population, the kernel methods and machine-learning techniques such as SVM may be preferable. It is important for the practitioner to consider such issues or test methods on a relevant data set before a method for GEBV calculation is chosen. Although the increasing marker density, training population size, and trait heritability are obvious ways to improve GEBV accuracy; these options add cost to the program. Implementing algorithms for markers imputation and training population design holds the potential for essentially free additional accuracy, leading to greater overall GS efficiency.

The current drops in genotyping costs, while phenotyping costs remain constant or increase, suggest that efforts to understand how to choose which lines to phenotype on the basis of their genotype, that is, how to design training populations, will be rewarding. Combining training populations from different populations is another way to boost accuracy when individual populations lack sufficient size and assuming that the marker densities required are available. With respect to maximizing long-term selection, we discussed several promising approaches that strive to retain favorable, low-frequency alleles while minimizing loss of short-term gain. Both simulation and empirical results for GS have been quite impressive. Empirical results of GS accuracy in crops, however, are not yet available for the public sector, except in the form of CV within families. Further empirical studies of the effects of statistical models, marker density, TP size and composition, and different selection criteria for the effectiveness of GS in breeding populations are urgently needed. In addition, while the CV approach can be instructive, an important caveat should be mentioned. In CV, the training and validation sets belong to the same population. But in GS, the selected candidates will rarely belong to the same population as the training set and may well be several generations removed from it. Recombination during meiosis between generations erodes the association between marker and QTL, systematically reducing accuracy. The effect of selection on allele frequencies and the Bulmer effect can also have detrimental effects on accuracy. In order to realistically evaluate GS for crops, studies designed for this purpose should be performed.

Clearly, exciting times are ahead of us as public breeding programs launch GS efforts. This review compiles several immediately useful results for breeders wanting to maximize gains through GS.Knowledge of breeding program parameters (effective population size, extent of LD, and trait heritability) allows marker density and training population size to be determined using analytical formulae. The greatest impact of GS on gain per unit time will come from shortening the breeding cycle [11]. Therefore, redesigning crossing and population development schemes incorporate GS as early as possible will likely be the most effective. Consequently, phenotyping resources will need to be shifted from early generation, evaluation for selection to evaluation for model training. The importance of epistasis will need to be assessed for each trait. A major paradigm in plant breeding since the availability of molecular marker technology is that mapping and characterizing the genetic loci that control a trait will lead to improved breeding. Often, one of the rationales for cloning of QTL is to develop the "perfect market" for MAS, perhaps based on a functional polymorphism. In contrast, an advantage of the GS is precisely its black box approach to exploiting genotyping technology to expedite genetic progress. This is an advantage in our view because it does not rely on a "breeding by design" engineering approach to cultivar development requiring knowledge of biological function before the creation of phenotypes. Breeders can therefore use GS without the large upfront cost of obtaining that knowledge. In addition, GS can maintain the creative nature of phenotypic selection which couples random mutation and recombination to sometimes arrive at solutions outside the engineer's scope.

Acknowledgements

Support from the teaching section Department of plant breeding and Genetics, Punjab Agricultural University, Ludhiana India.

Nomenclature

MAS	marker-assisted selection
GS	genomic selection
BV	breeding value
GEBV	genomic estimated breeding value
CV	cross-validation
TP	training population
BP	Breeding population
LD	linkage disequilibrium

Author details

Mohd Shamshad* and Achla Sharma

*Address all correspondence to: shamshad.ratan@gmail.com

Punjab Agricultural University, Ludhiana, India

References

[1] Coors JG. Who are plant breeders, what do they do, and why? Plant Breed Arnel R Hallauer International Symposium. 2008:51-60

[2] Sun CQ, Huang ZZ, Wang YL, et al. Overcoming pre-fertilization barriers in the wide cross between *Chrysanthemum grandiflorum* (Ramat.) *Kitamura* and *C. Nankingense* (Nakai) Tzvel. By using special pollination techniques. Euphytica. 2011;**178**:195-202

[3] Hayes BJ, Bowman PJ, Chamberlain AC, et al. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genetics, Selection, Evolution. 2009;**41**:1-9

[4] Muir WM. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. Journal of Animal Breeding and Genetics. 2007;**124**:342-355

[5] Habier D, Fernando RL, Dekkers JCM. The impact of genetic relationship information on genome-assisted breeding values. Genetics. 2007;**177**:2389-2397

[6] Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;**157**:1819-1829

[7] Solberg TR, Sonesson AK, Woolliams JA, et al. Genomic selection using different marker types and densities. Journal of Animal Science. 2008;**86**:2447-2454

[8] Lorenzana RE, Bernardo R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theoretical and Applied Genetics. 2009;**120**:151-161

[9] Calus MPL, Veerkamp RF. Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. Journal of Animal Breeding and Genetics. 2007;**124**:362-368

[10] Rafalski A. Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology. 2002;**5**:94-100

[11] Heffner EL, Sorrells ME, Jannink JL. Genomic selection for crop improvement. Crop Science. 2009;**49**:1-12

[12] Perez-de-Castro AM, Vilanova S, Canizares J, et al. Application of genomic tools in plant breeding. Current Genomics. 2012;**13**:179-195