Int.J.Curr.Microbiol.App.Sci (2019) *8*(6): 1529-1534

International Journal of Current Microbiology and Applied Sciences
***ISSN: 2319-7706* Volume 8 Number 06 (2019)**
Journal homepage: http://www.ijcmas.com

Review Article https://doi.org/10.20546/ijcmas.2019.806.184

Genomic Selection in Crop Improvement – An Overview

Nenavath Krishna Kumar Rathod[1*], Noor E. Mujjassim[2], M. Mallik[3] and S.D. Nehru[4]

[1]Division of Genetics, Indian Agricultural Research Institute, Pusa Campus, New Delhi, India
[2]Department of Genetics and Plant Breeding, University of Agricultural Sciences, GKVK, Bengaluru, Karnataka, India
[3]Division of Genetics, Indian Agricultural Research Institute, Pusa Campus, New Delhi, India
[4]AICRP on Sunflower, UAS, GKVK, Bengaluru, Karnataka, India

**Corresponding author*

Keywords

Genomic selection, Linkage disequilibrium, Polymorphism

Article Info

Accepted:
12 May 2019
Available Online:
10 June 2019

A B S T R A C T

Genomic selection (GS) is a form of marker-assisted selection in which genetic markers covering the whole genome are used so that all quantitative trait loci (QTL) are in linkage disequilibrium with at least one marker. With advances in molecular biology and marker technology, a new era of molecular breeding has emerged that has greatly accelerated the pace of plant breeding. Whole-genome prediction models estimate all marker effects in all loci and capture small QTL effects. Here, we review difference between MAS and Genomic selection (GS) along with GS models with respect to both the prediction accuracy and genetic gain from selection.

Introduction

Molecular mapping used to identify a molecular locus that resides very near or in the gene of interest. This locus can be used as a molecular marker for indirectly selecting the trait of interest. Various markers have been developed starting from first marker RFLP (Restriction Fragment length polymorphism) which was developed by Alec Jeffreys. Most of the economic traits are influenced by polygenes, tracking a small number of these genes through DNA markers will only explain a small proportion of the genetic variance. In addition, individual genes are likely to have small effects and so a large amount of data is needed to accurately estimate their effects. Selection for the desirable allele of a gene/quantitative trait locus (QTL) on the basis of molecular marker(s) linked to it in the place of phenotype generated by this allele is known as marker-assisted selection (MAS). Current MAS methods are better suited for manipulating a few major effect genes than many small-effect genes (Dekkers and Hospital, 2002).

Genomic selection (GS) is specialized form of MAS based on the principle that information from a large number of markers distributed across the genome can be used to capture diversity in that genome with the help of genomic estimated breeding values (GEBVs) of different individuals/lines, which form the basis of selection without knowledge of where specific genes are located. The major difference between Genomic selection and Marker-assisted selection is MAS not suitable for the improvement of QTL with small effects and also it is based on markers showing significant association. GS was first described in 2001 by Meuwissen and colleagues to rectify the deficiency of MAS and MARS. The review here contains current GS methods and their advantages with its limitations (Fig. 1).

The GEBV of an individual is the sum total of effects associated with all the marker alleles present in the individual and included in the GS model applied to the population under selection. The breeding value (BV) of an individual/linere presents the expected phenotype of its progeny. The BV is, therefore, determined by progeny testing and is based only on the additive genetic effects. Breeder's equation;

$$\Delta G = i\, r\, \sigma_A / L$$

ΔG- Genetic gain;i- Selection intensity (directly proportional to population size andproportion selected); r - accuracy (proportional to the reference population size), σ_A- genetic variation within a population; L- generational interval.

The use of GS allows the breeder to:

Increase i through the use of larger breeding populations

Increase r through the use of larger reference populations

Increase σ_A due to a more robust quantification of variation

Decrease L due to the reduction of the time required to obtain individuals carrying

All these factors contribute to a higher genetic gain and, consequently, a better yield of an individual carrying the desired qualities.

MAS limitations

Tightly linked, widely applicable, and reliably diagnostic markers are available for only a limited number of target traits.

The marker-trait associations discovered in one population have to be independently confirmed and validated in unrelated germplasm; this is particularly relevant for QTLs. QTL introgression is often problematic since relatively large genomic regions need to be selected for. This is because the confidence intervals for QTLs span ~10 cM.

Available resources limit the size of mapping populations and, consequently, the accuracy of QTL position and effect estimates (Dekkers and Hospital, 2002; Schon *et al.*, 2004). Also, allelic diversity and genetic background effects that are present in a breeding program will not be captured with a single biparental population. Therefore, multiple mapping populations are needed, QTL positions require validation, and QTL effects must be re-estimated by breeders in their specific germplasm (Bernardo, 2001).

QTLs introgressed into different genetic backgrounds generally show unpredictable expression. Often QTL introgression produces discouraging improvement in complex traits like yield, and sometimes even unfavorable

effects may be obtained. MAS increases the amount of data generated in a breeding program by about seven-fold, which increases the workload of breeders. The decision making in breeding programs becomes more involved and more frequent, e.g., three to four times per year in the case of MAS as compared to one to two times in the conventional breeding programs. Marker genotyping has to be accomplished in a short growing season, for which high throughput genotyping facilities may become necessary.

In general, MAS increases the overall cost of the breeding program. Marker genotyping cost is the chief factor limiting the widespread adoption of MAS by plant breeding programs, especially in the developing countries.

In the Genomic Selection first step, they were unable to estimate all marker effects simultaneously with simple regression due to the lack of degrees of freedom. Therefore, they proposed selecting the most significant markers from the previous generation via multiple linear regressions and then re-estimating effects of the selected markers in the current generation with independent multiple regressions (Lande and Thompson, 1990) (Table 1).

GS - prediction models

Shrinkage models

SR, RR-BLUP, G-BLUP

Dimension reduction methods

Partial least square Regression

Principal component regression

Least absolute shrinkage and selection operator (LASSO)

Variable selection models

Bayes A & B, BayesCπ, BayesDπ

Kernel Regression and machine learning methods

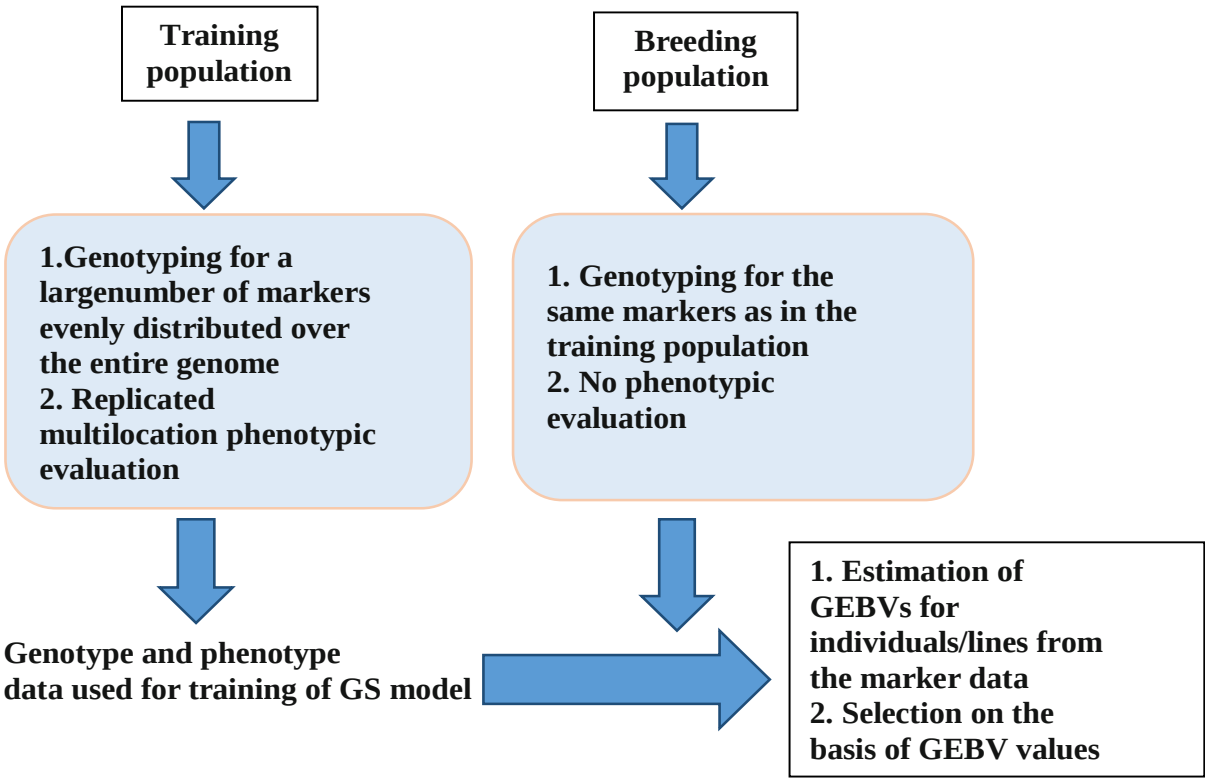
Support vector machine regression (SVM)

Random Forest (RF)

Table.1 Main features of genome-wide prediction models

RR-BLUP	**Assumes that all markers have equal variances with small but non-zero effect.** **Applies homogeneous shrinkage of predictors towards zero, but allows for markers to have uneven effects.** **Computed from a realized-relation matrix based on markers.** **Some QTL are in LD to marker loci, whereas others are not.**	**9, 12**
LASSO	LASSO Combines both shrinkage and variable selection methods. RR-BLUP does not use variable selection, but outsmarts LASSO when there is multi-collinearity between the predictors.	8, 16
BRR	Induces homogeneous shrinkage of all marker effects towards zero and yields a Gaussian distribution of marker effects Similar to RR-BLUP, there is a problem of QTL linkages to the marker loci.	[5]
Bayes A	Utilizes an inverse chi-square (x^2) on marker variances yielding a scaled t-distribution for marker effects. Similar to BL and in contrast to BRR, it shrinks tiny marker effects towards zero and larger values survive. Has a higher peak of mass density zero compared with the DE distribution.	14, 7
Bayes B	Similar to Bayes A, uses an inverse x^2 resulting in a scaled t-distribution. Unlike Bayes A, utilizes both shrinkage and variable selection methods When p = 0, then it is similar to Bayes A.	14, 20
Bayes C	Applies both shrinkage and variable selection methods. Characterized by a Gaussian distribution Bayes B and Bayes C consist of point of mass at zero in their slab priors.	7,6

Fig.1 A simple schematic representation of genomic selection (GS) scheme



Stepwise Regression (SR)

Select most significant markers on the basis of arbitrary significant thresholds and non-significant markers effect equals to zero (Lande and Thompson, 1990). Estimate the effect of significant markers using multiple regression and only a portion of the genetic variance will be captured. This stepwise approach to set non-significant marker effects to zero is critical for maintaining model estimability (Lande and Thompson, 1990). Significance thresholds that may maximize response to selection cannot be determined analytically, although guidelines have been established through simulation (Hospital *et al.,* 1997; Moreau *et al.,* 1998).

Limitations

Detects only large effects and that cause over-estimation of significant effects (Goddard and Hayes, 2007; Beavis, 1994) SR resulted in low GEBV accuracy due to limited detection of QTLs (Meuwissen *et al.,* 2001).

Ridge Regression-BLUP (RR-BLUP)

The ridge regression BLUP (RR-BLUP) method can simultaneously estimate all marker effects for GS (Meuwissen *et al.,* 2001; Whittaker *et al.,* 1997). Rather than categorizing markers as either significant or as having no effect, ridge regression shrinks all marker effects toward zero (Breiman, 1995; Whittaker *et al.,* 1997). The method makes the assumption that markers are random effects with a common variance (Meuwissen *et al.,* 2001). Simultaneously select all marker effects rather than categorizing into significant or non-significant. Ridge regression shrinks all marker effects towards zero. The method makes the assumption that markers are

random effects with an equal variance (Meuwissen *et al.,* 2001). Ridge Regression-BLUP are Superior to Stepwise Regression (SR).

Limitation

RR-BLUP incorrectly treats all effects equally which is unrealistic (Xu *et al.,* 2003).

Bayesian Regression (BR)

The simplifying assumption of equal and fixed marker effect variances allows RR-BLUP parameters to be efficiently computed using maximum likelihood methods (Meuwissen *et al.,* 2001). While RR-BLUP can provide a conservative EBV by shrinking all marker effects equally (Muir, 2007), the presumably incorrect assumption that underlies it can lead to over shrinking of large effects. Bayesian methods have been adopted to relax this assumption and better model marker effects of differing sizes (Hayes, 2007). Here, a separate variance is estimated for each marker, and the variances are assumed to follow a specified prior distribution (Meuwissen *et al.,* 2001).

Marker variance treated more realistically by assuming specified prior distribution. BayesA: uses an inverted chi-square to regress the marker variance towards zero. All marker effects are > 0 (Bayes A)

BayesB: assume a prior mass at zero, thereby allowing for markers with no effects. Some marker effects can be = 0 (Bayes B).

Limitations of genomic selection

GS has still not become popular with plant breeding community primarily due to insufficient evidence for its practical usefulness. In fact, most discussions on its usefulness are largely statistical treatments and simulations that are not easily appreciated by plant breeders.

The accuracy of GEBV estimates has been evaluated using simulation models based on additive genetic variance. These models ignore epistatic effects, which does not seem to be realistic. It has been argued that since epistasis makes only a small contribution to the breeding value, the use of only additive genetic models for GS may be expected to maximize selection gains (Heffner *et al.,* 2009). GS is more effective than phenotypic selection on per unit time basis only when off-season/greenhouse facilities are used to grow up to three generations per year. The usefulness and the cost-effectiveness of GS would be doubtful where such facilities are not available.

Using substantially different means, various GS methods deal with the issues of increasing dimensionality and computational complexity, and thereby capture different aspects of the association between genotype and phenotype. For this reason, the performances of different methods depend on the genetic architecture underlying the specific trait. For crop improvement, most quantitative traits are influenced by polygenes. Among the general GS methods, GBLUP is recommended for breeding practice because of its robustness, and computational efficiency (Xin Wang *et al.,* 2018). Some machine learning algorithms, such as RKHS, have also been successful in GS, and likely represent avenues for future research.

References

Beavis W. D. 1994. QTL analysis: power, precision and accuracy In: A.H. Paterson (ed.), Molecular Dissection of Complex Traits. CRC Press, Boca Raton, pp. 145–162.

Bernardo, R. and Yu, J. 2007. Prospects for genome wide selection for quantitative traits in maize. *Crop Sci*., 47: 1082–1090.

De Los Campos, G. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*, 182: 375–385.

De Los Campos, G. 2013. Prediction of complex human traits using the genomic best linear unbiased predictor. *PLoS Genet*., 9: e1003608.

De Los Campos, G. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, 193: 327–345.

Dekkers J. C. 2004. Commercial application of marker and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci*., 82 E-Suppl, E313–E328.

Dekkers, J. C. M. and Hospital, F. 2002. The use of molecular genetics in the improvement of agricultural populations. *Nature Rev. Genet*., 3: 22-32.

Friedman, J. 2010. Regularization paths for generalized linear models *via* coordinate descent. *J. Stat. Softw*., 33: 1–22.

Goddard, M. E. and Hayes, B. J. 2007. Genomic selection. *J. Anim. Breed. Genet*, 124: 323–330.

Hayes, B. J., Chamberlain, A. C., McPartlan, H., McLeod, I., Sethuraman, L. and Goddard, M. E. 2007. Accuracy of marker assisted selection with single markers and marker haplotypes in cattle. *Genome. Res*., in press.

Heffner, E. L. 2009. Genomic selection for crop improvement. *Crop Sci*., 49: 1–12.

Heffner, E. L. 2011. Genomic selection accuracy for grain quality traits in bi-parental wheat populations. *Crop Sci*., 51: 2597–2606.

Heffner, E.L. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genet*, 4: 65–75.

Lande, R. and Thompson, R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, 124(3): 743-756.

Laurence Moreau, Alain Charcosset, Frederic Hospital and Andre Gallais. 1998. Marker-assisted selection efficiency in populations of finite size. *Genetics*, 148(3): 1353-65.

Li, Z. and Sillanpaa, M. J. 2012. Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection. *Theor. Appl. Genet*., 125: 419–435.

Meuwissen, T. H. E., Hayes, B. J., Goddard, M. E. 2001. Prediction of total genetic value using genome wide dense marker maps. *Genetics*, 157:1819–1829.

Whittaker, J.C., Haley, C. and Thompson, R. 1997. Optimal weighting of information in marker-assisted selection. *Genet. Res*., 69: 137–144.

Xin Wanga, B. C., Yang Xua, Zhongli Huc, Chenwu Xua. 2018. Genomic selection methods for crop improvement Current status and prospects. *Crop journal*, 6: 330 – 340.

Xu, S. 2003. Estimating polygenic effects using markers of the entire genome. *Genetics*, 163: 789–801.

How to cite this article:

Nenavath Krishna Kumar Rathod, Noor E. Mujjassim, M. Mallik and Nehru, S.D. 2019. Genomic Selection in Crop Improvement – An Overview. *Int.J.Curr.Microbiol.App.Sci.* 8(06): 1529-1534. doi: https://doi.org/10.20546/ijcmas.2019.806.184