

Era of Genomic Selection in Plant Breeding

Rhitisha Sood

Ph. D. Scholar, Department of Genetics and Plant Breeding, CSKHPKV, Palampur, H.P

ARTICLE ID: 106

Abstract

Because of the world population boom and climate change, the demand for increased food production has risen substantially. Breeding based on genomics could help farmers enhance their genetics, resulting in improved crop yields. One of the most promising strategies for improving polygenic variables, such as crop output, is genomic selection (GS). The number of polymorphic markers available to plant breeders and molecular biologists has expanded dramatically as a result of the advent of several molecular marker systems. Because chromosomes with low recombination frequency have little power for finding market-trait connections, linkage analysis for QTL mapping is done on biparental populations. As a result, in the early 2000s, association mapping (AM) was developed to overcome the poor power of linkage analysis, allowing for the discovery of marker-trait relationships in non-biparental populations and fine-mapping chromosome segments with high recombination rates. The fundamental issue with fine-AM is that it has a limited detection capability for rare mutations that could be linked to economically important characters. For features that are complicated and influenced by many genes, each with a minor cumulative effect, traditional marker-assisted selection has been inefficient. However, with the introduction of GS, that paradigm has evolved. Instead of attempting to identify individual loci that are strongly related with a characteristic, GS now analyses all marker data as predictors of performance, resulting in more accurate predictions. Selection can be based on GS predictions, potentially resulting in faster and lower-cost breeding gains. GS, unlike QTL and AM, uses all molecular markers for GP of candidate performance for selection. As a result, GS's goal is to forecast breeding and/or genetic values. The genomic estimated breeding values (GEBVs) of individuals in a testing population (TST) that have been genotyped but not phenotyped are obtained using GS, which combines molecular and phenotypic data in a training population (TRN). The key advantages of GS over phenotypic based selection in breeding are that it lowers the cost per cycle and shortens the time it takes to generate a variety. It develops a model that takes

genotypic data from a candidate population of untested individuals and produces GEBVs using a training population of genotyped and phenotyped individuals. These GEBVs don't reveal anything about the function of the underlying genes, but they're an excellent selection criterion. Untested individuals in plant breeding would be part of a larger population defined as a crop market class or the breeding programme as a whole. GEBVs based merely on an individual's genotype have been surprisingly accurate in simulated studies. In the face of declining genotyping costs and static or rising phenotyping prices, and the ability to select individuals much earlier in the breeding cycle, GS is revolutionizing both animal and plant breeding. There's no doubt that GS will transform plant breeding techniques and efficiency, and that it's in the midst of a period of vigorous scientific research. Even while GS is still in its early stages of implementation, its practice is already outpacing its theory. Different GS approaches function differently depending on the underlying genetic architecture, which is defined at least by the number of QTL and the distributions of their allelic effects and frequencies. As GS is used in plant breeding, future empirical study will provide both intriguing insight into that design and critical reality checks. GS has been shown to boost genetic gain in crop breeding in both simulated experiments and real breeding operations. In a traditional breeding method, data on yield-related features is thought to be more useful for increasing crop yields. Information on yield-related features (i.e. secondary traits) could also help GS predict grain yield more accurately and efficiently (target trait). The incorporation of secondary trait information into the GS method will significantly boost the genetic gain of complex target traits.

Introduction

Haley and Visscher coined the name 'GS' in 1998 at the 6th World Congress on Genetics Applied to Livestock Production in Armidale, Australia. However, Meuwissen in 2001 introduced GS in his publication "Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819-29" which is just recently being utilised in applied crop breeding to predict complicated traits of plants. The basis of selection in GS is genotyping data on marker alleles encompassing the complete genome, which is a specialised version of MAS. As a result, the effects associated with all of the marker loci, regardless of whether the effects are substantial or not, are estimated throughout the entire genome. The genomic estimated breeding values (GEBVs) of

distinct individuals/lines, which constitute the basis of selection, are calculated using the marker effect estimations. The sum amount of effects associated with the entire marker alleles present in the individual and incorporated in the GS model applied to the population under selection is the GEBV of that individual. The predicted phenotype of an individual/progeny lines is represented by the breeding value (BV) of that individual/line which is in turn estimated by progeny screening and is solely based on additive genetic influences. The genotypic value of an individual/line, on the other hand, is the phenotype predicted by its genotype & is determined by both additive and non-additive genetic factors. Individuals' phenotypic data and their relatives have been used to determine estimated breeding values (EBVs) since the 1980s.

In animal breeding experiments, the profits via selection showed a significant increase when data on markers known to be connected to known QTLs were paired with morphological data for generating EBVs. When existing QTLs were included in the GS framework, the targeted QTLs accumulated at substantially higher frequency than when ordinary ridge regression was utilized. Gene-assisted genomic selection is a GS model that makes use of knowledge about known QTLs

Procedure

The GS approach is based on a training (TP) and a breeding population (BP), which are two different but related populations. The training population is used to train the GS model and obtain estimates of marker-associated effects, which are required for estimating GEBVs of individuals/lines in the breeding population. The breeding population, is indeed the population that has undergone GS in order to achieve the necessary improvement and to isolate outstanding lineages for use as new varieties/parents of new enhanced hybrids (Fig. 1).

Establishment of training population appropriate for the breeding population is the first step for conducting GS methodology. The training population's individuals/lines are successfully amplified for a significant number of markers that are equally dispersed across the entire genome at a sufficient density. In repeating trials across locations and, preferably, years, the individuals/lines in the training population are subjected to critical phenotypic evaluation for the target trait(s). Model training is the process of computing the GS set of parameters using the phenotypic and marker genotype data. Model training can also be



repeated as needed to incorporate data from new markers and attributes. The GS model parameter estimations are saved for further applicability towards the breeding population. The breeding population is assessed using the same markers as were used in the training population to estimate model parameters. The breeding population has not been evaluated phenotypically. The GEBVs of breeding population individuals/lines are determined using marker genotype data and marker-associated effects inferred from the training population. On the basis of their GEBV calculations, superior individuals/lines are chosen from the breeding population. To attain the required results, this GS approach might be paired with a suitable breeding programme.

Incorporation of Genomic Selection in Breeding Programs

GS follows a simple select-and-inter cross procedure i.e. easily incorporated into any breeding programme which allows one selection cycle to be completed in a single generation such that the time required for completing a given number of selection cycles can be greatly reduced. **e.g.** Using off-season nursery/greenhouse facilities, three generations of a crop like barley can be grown each year. Every year, one selection (F_2 generation) – inter mating (amongst identified F_2 plants) – selfing (F_1 generation rearing) cycle can be accomplished. As a recombination (selection–inter mating–selfing) cycle population, BP can be maintained indefinitely. Also, to maintain variety, new germplasm lines can be added to the BP and TP. However, because of the minimal population-wide LD in a population made up of varied individuals/lines, including unadapted germplasm, would need the employment of extremely high marker density. To account for the increased genetic variety, a greater effective population size (N_e) would be required. In Morgans, the quantity of loci required to achieve GEBV, accuracy of 0.9 has been estimated to be around $10 N_e$ genome size. Although N_e values can be calculated using computer software using marker data, they are rarely used in plant breeding operations. As a result, genome-wide LD estimates might be a little more relevant.

A two-step GS-based breeding technique can be applied in crop species like maize. Plants from a segregating population, including F_2 , are genotyped for a sufficient number (e.g., 250–500) of alleles covering the whole genome during first step of this strategy. In multilocation replicated productivity experiments, these plants are also tested for testcross performance. Breeding values associated for each marker allele are estimated using marker

polymorphism and testcross performance data, and these estimates are used to forecast GEBVs of F_2 plants. The F_2 plants with the greatest GEBV estimates are chosen in the second stage and crossed in all conceivable combinations to create the populace for another selection period.

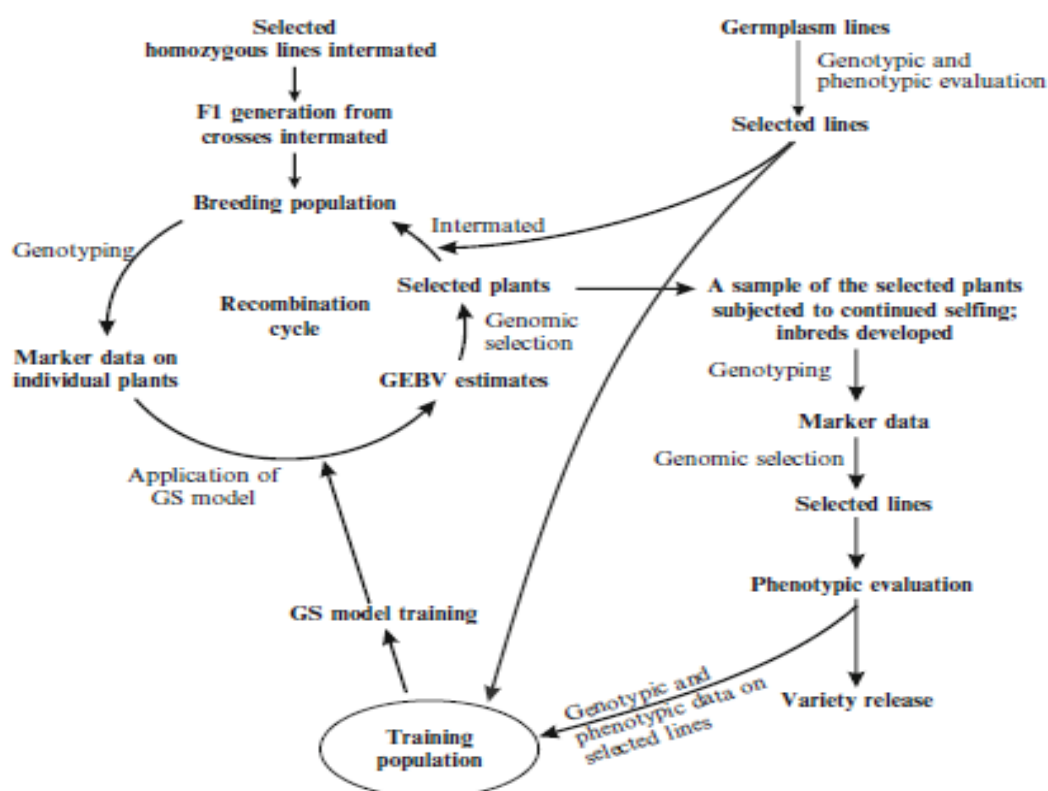


Fig.1 Scheme of GS in self-pollinated crops. It represents incorporation of new germplasm, use and devp. Of purelines, variety release, maintenance of GEBV by regular monitoring and updating BP and TP.

1. QTL discovery, mapping, and other methods are not necessary because the marker effects are calculated from the trained population and used straight for GS in the breeding population in question.
2. GS outperforms phenotypic selection in terms of gains per unit of time and is better at predicting the performance of breeding lines than pedigree data. **e.g.** maize, especially for variables with low heritability; also enhances rice breeding efficiency.
3. Phenotyping for each selection phase in BP is not required which narrows down the breeding cycle significantly, especially in perennial plants **e.g. oil palm.**

4. In compared to selection based on breeding values determined from phenotype data, GS would tend to lower the rate of inbreeding and the degradation of genetic variability; this would be accomplished without compromising selection gains.
5. Breeders may be able to choose parents for hybridization programmes using GS from among lines that have not been tested in the target environment. The GEBVs of these lines, which are evaluated for their responsiveness to the testing environment, would be used to make this decision. This would make it easier to exchange material and use it in breeding operations and, possibly, for the development of hybrid varieties.
6. GS can incorporate data on marker genotype and attribute phenotype gathered over time in diverse evaluation programmes encompassing a variety of contexts into GEBV estimations for different individuals/lines. This would enable GEBV estimation for features which have never been evaluated before.
7. This technique has been found successful for various crops (Fig.2).

S.no.	Species	NGS marker platform	Trait	Population size	Total SNP markers	Prediction accuracy	Model	Software packages	Reference
1	Rice	GBS	Grain yield, flowering time	363	73,147	0.31-0.63	RR-BLUP	R package mBLUP	Spindel et al., 2015
2	Rice	DAV/seq	Grain yield, plant height	343	8,336	0.54	G-BLUP, RR-BLUP	BGLR and ASReml R packages	Grenier et al., 2015
3	Wheat	GBS	Stem rust resistance	365	4,040	0.61	G-BLUP B	R package GAPIT	Rutkoski et al., 2014
4	Wheat	GBS	Grain yield, plant height, heading date and pre-harvest sprouting	365	38,412	0.54	BLUP	R package mBLUP	Heslot et al., 2013
5	Wheat	GBS	Grain yield	254	41,371	0.28-0.45	BLUP	ASReml 3.0	Poland et al., 2012
6	Wheat	GBS	Yield and yield related traits, protein content	1127	38,893	0.20-0.59	BLUP	mBLUP version 4.2	Isidro et al., 2015
7	Wheat	GBS	Fusarium head blight resistance	273	19,992	0.4-0.90	RR-BLUP	R package GAPIT	Aruda et al., 2016
8	Wheat	GBS	Grain yield, protein content and protein yield	659	-	0.19-0.51	RR-BLUP	R package mBLUP	Michel et al., 2016
9	Wheat	GBS	Grain yield	1477	81,999	0.50	G-BLUP	R package mBLUP	Lado et al., 2016
10	Wheat	DAV/seq	Grain yield	803	-	0.27-0.36	G-BLUP	BGLR and ASReml R packages	Pierre et al., 2016
11	Wheat	GBS	Grain yield, Fusarium head blight resistance, softness equivalence and flour yield	470	4858	0.35-0.62	BLUP	BGLR R-package	Hofstetter et al., 2016
12	Wheat	GBS	Hail and drought stress	10819	40000	0.18-0.65	G-BLUP	BGLR R-package	Crossa et al., 2016
13	Maize	GBS	Drought stress	3273	58,731	0.40-0.50	G-BLUP	BGLR R-package	Zhang et al., 2015
14	Maize	GBS	Grain yield, anthesis date, anthesis-silking interval	504	158,281	0.51-0.59	PGBLUP, PRKHS	R Software	Crossa et al., 2013
15	Maize	GBS	Grain yield, anthesis date, anthesis-silking interval	296	235,265	0.62	PGBLUP, PRKHS	R software	Crossa et al., 2013
16	Maize	DAV/seq	Ear rot disease resistance	238	23,154 Dart-seq markers	0.25-0.59	RR-BLUP	R package mBLUP	dos Santos et al., 2016
17	Soybean	GBS	Yield and other agronomic traits	301	52,349	0.43-0.64	G-BLUP	MissForest R package, TASSEL 5.0	Jarquín et al., 2014b
18	Cancola	DAV/seq	Flowering time	182	18,804	0.64	RR-BLUP	R package GAPIT	Raman et al., 2015
19	Alfalfa	GBS	Biomass yield	190	10,000	0.66	BLUP	R package, TAASEL software	Li et al., 2015
20	Alfalfa	GBS	Biomass yield	278	10,000	0.50	SVR	R package mBLUP, R package BGLR, R package RandomForest	Annicchiarico et al., 2015
21	Miscanthus	RADseq	Phenology, biomass, cell wall composition traits	138	20,000	0.57	BLUP	R package mBLUP	Slavov et al., 2014
22	Switchgrass	GBS	Biomass yield	540	16,669	0.52	BLUP	gimnet R package, R package mBLUP	Lipka et al., 2014
23	Grapevine	GBS	Yield and related traits	800	90,000	0.50	RR-BLUP	R package BLR, R package mBLUP	Fodor et al., 2014
24	Intermediate wheatgrass	GBS	Yield and other agronomic traits	1126	3883	0.67	RR-BLUP	R package mBLUP, BGLR R-package	Zhang et al., 2016
25	Perennial ryegrass	GBS	Plant herbage dry weight and <i>in vitro</i> digestibility	211	10,885	0.16-0.56	RR-BLUP	R software	Faville et al., 2016

Fig.2. Attempt of GS in different crop species showing range of prediction frequency

Limitations

1. GS has yet to gain traction in the plant breeding community, owing to a lack of proof for its practical utility.
2. Due to changes in allele frequency and epistatic interactions, the marker influences and GEBV estimations may alter. This would demand the GS model being updated after each breeding cycle.
3. Additive genetic variance is used in the majority of simulation models. These models don't take into account epistatic effects, which doesn't appear to be realistic.
4. Capacity to construct adequate GS models to attain optimum prediction accuracy is limited due to a lack of understanding about the genetic basis of polygenic traits.
5. The need to assay a substantial percentage of marker loci in each generation of selection raises the expense significantly.

Future Perspectives

The majority of current GS models solely consider additive gene effects. The dominant gene effects are likely to be accommodated effectively in GS models in the near future, but the incorporation of epistatic associations remains a challenge. To allow the incorporation of epistatic interactions in regression models, semi-parametric GS models and machine learning approaches are being developed. According to the research, including epistatic gene effects improves GS accuracy significantly. Furthermore, the current GS models ignore the impact of genetic heritage on QTL activity.

There is essentially no evidence on the implications of such interactions, which are present in training and/or breeding populations, on GEBV estimates' accuracy. Significant QTL genetic background interactions are likely to lower the precision of both QTL consequences and GEBV estimates. The idea of GS is now being proven, and it appears to be a powerful and appealing breeding approach. The use of GS is likely to make it easier to design optimum genotypes and create them through focused GS. The GS approach appears to have a major flaw in that it ignores the biological events that underpin the emergence of the phenotypes in question. This is due to the fact that the GS algorithms ignore key studies from genetics and genomics, such as the detection of QTLs and the genes they represent. Developing such GS algorithms that reap the benefits of genetics and genomics insights could be beneficial. The beginning of the next stage of MAS appears to be GS. As numerous

concerns linked to GS are sufficiently handled and talks on GS become less theoretical and mathematical, it is expected to be implemented into many plant breeding initiatives. Furthermore, the decreasing cost of marker genotyping will hasten the application of GS in traditional plant breeding projects.

References

- Abbertson M, Batley J, Bentley A, Bryant J, Cai H, Cockram J, de Oliveira CA, Cswkw LJ, Bernardo R and Yu J. 2007. Prospects for genome-wide selection for quantitative traits in maize. *Crop Sci* 41:1–4
- Bernardo R. 2016. Bandwagons I, too, have known. *Theoretical and applied genetics* 129: 2323-2332
- Bernardo R. 2010. Genome-wide selection with minimal crossing in self-pollinated crops. *Crop Sci* 50:624–627
- Bhat JA, Sajad Ali S, Salgotra RK, Mir ZA, Dutta S, Jadon V, Tyagi A, Mushtaq M, Jain N, Singh PK, Singh GP and Prabhu KV. 2016. Genomic Selection in the Era of Next Generation Sequencing for Complex Traits in Plant Breeding. *Front. Genet.* <https://doi.org/10.3389/fgene.2016.00221>
- Crossa J, Campos G, Perez P et al. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186:713–724
- Crossa J, Perez P, Campos G et al. 2011. Genomic selection and prediction in plant breeding. *J Crop Improv* 25:239–261
- Dempewolf H, de Pace C and Edwards D. 2016. Global agricultural intensification during climate change: a role for genomics. *Plant Biotechnology Journal* 14: 1095-1098
- Heffner EL, Sorrells ME and Jannink JL. 2009. Genomic selection for crop improvement. *Crop Sci* 49:1–12
- Meuwissen THE and Goddard ME. 2001. Prediction of identity by descent probabilities from marker haplotypes. *Genet Sel Evol* 33:605–634
- Meuwissen THE. 2009. Accuracy of breeding values of ‘unrelated’ individuals predicted by dense SNP genotyping. *Genet Select Evol* 41:35. doi:10.1186/1297-9686-41-35
- Rutkoski JE, Heffner EL, and Sorrells ME. 2010. Genomic selection for durable stem rust resistance in wheat. *Euphytica* 179:161–173

- Spindel J, Begum H, Akdemir D et al. 2015. Genomic selection and association mapping in rice (*oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet* 11: e1004982. doi:10.1371/journal.pgen.1004982
- Yabe S, Takashi H, Mariko U, Hiroyuki E, Tatsuro K, Satoru N, Yasuo Y, Ryo O and Hiroyoshi I. 2018. Potential of GS in mass selection breeding of an allogamous crop: an empirical study to increase yield of common buckwheat. *Frontiers in Plant Science* 9: 276

