European Journal of Biotechnology and Bioscience
ISSN: 2321-9122; Impact Factor: RJIF 5.44
Received: 09-06-2019; Accepted: 12-07-2019
www.biosciencejournals.com
Volume 7; Issue 5; September 2019; Page No. 22-28

Theory and practice of genomic selection: A short review

Asif Ashraf[1*], Ram Vishwakarma[2]
[1, 2] CSIR-Indian Institute of Integrative Medicine (IIIM), Srinagar, Jammu and Kashmir, India

Abstract
Genomic selection (GS) overcomes the drawback of marker assisted selection not being effective in improving complex polygenic traits. GS is a form of MAS that selects favourable individuals based on genomic estimated breeding values. But due to its limited scope in practical applicability and insufficient data on its practical use, Genomic selection is yet to make more advances in the field of plant breeding.
Scope: With the current knowledge of GS algorithms and machine learning methods have been proposed for genomic prediction. With this progress has been made in genomic selection studies and its practical use in animal and plant breeding fields.

Keywords: genomic selection, breeding value, parametric models, LASSO, BLUP

Introduction

Plant breeding is a predictive science; it is driven by new technologies and new knowledge. Everything we do as plant breeders, we are trying to predict something. Over time plant breeders have developed methodologies like family selection, progeny testing improving the heritability and gain from selection (Sorells *et al*, 2010). Prediction accuracy is important at every step in breeding programme and breeding methods are designed to improve accuracy. Plant breeders all around the world are interested in using new strategies and take advantage of the power of using molecular markers. Genomic selection is one such method which overcomes the drawback of marker assisisted selection not being effective in improving complex polygenic traits in part, to its inability to capture small effect quantitative trait loci (QTL; Bernardo, 2008; Xu and Crouch, 2008) [58, 60]. A promising approach, termed genomic selection (GS), attempts to avoid this deficiency by capturing both large and small-effect QTL with dense genome wide molecular marker coverage to predict complex trait values (Meuwissen *et al.*, 2001) [43]. Prediction accuracies reported by GS studies, coupled with the continued advances in high-throughput genotyping technologies, make GS a promising tool to increase plant breeding efficiency (reviewed by Heffner *et al.*, 2009) [25]. The term 'GS' was first introduced by Haley and Visscher at the 6th World Congress on Genetics Applied to Livestock Production at Armidale, Australia in 1998 according to Meuwissen (2007) [42], although it was not used in the main text of Meuwissen *et al.* (2001) [43]. However, the overall MAS programme using GEBV was later referred to as GS.

The general processes of GS and traditional MAS used for quantitative traits (QTs) are shown in Fig. 1. The main frameworks of the two approaches are similar, where both GS and traditional MAS consist of training and breeding phases. In the training phase, phenotypes and genome-wide (GW) genotypes are investigated in a subset of a population, i.e. the training population in GS and the mapping population in traditional MAS. Within populations, significant relationships between phenotypes and genotypes are predicted using statistical approaches. In the breeding phase, genotype data are obtained in a breeding population, before favourable individuals are selected based on the genotype data obtained. Three obvious differences between the two approaches are apparent: (1) in the training phase, quantitative trait loci (QTLs) are identified in traditional MAS while formulae for GEBV prediction are generated in GS, known as GS models; (2) in the breeding phase, genotype data are only required for targeted regions in traditional MAS, whereas GW genotype data are considered to be necessary in GS; (3) in the breeding phase, favourable individuals are selected based on the genotypes of markers in MAS, whereas GEBVs are used for selection in GS. Thus, GS jointly analyses all the genetic variance of each individual by summing the marker effects of GEBV (Heffner *et al.*, 2009) [25], and it is expected to address small effect genes that cannot be captured by traditional MAS (Hayes *et al.*, 2009) [24]. Since GS was first propounded by Meuwissen *et al.* (2001) [43], many reports have indicated the usability of GS for breeding for QTs. However, GS has still not become a popular methodology in the field of plant breeding. We consider that a Major obstacle is the availability of insufficient knowledge of GS for practical use. Indeed, most fields of GS studies have dealt with statistics and simulation that are discussed in terms of formulae, which are often too specific for breeders and molecular biologists to understand. To initiate further discussions on the applicability of GS in plant breeding, here our aim is to discuss GS from a practical breeding viewpoint. The statistical approaches used in GS are briefly explained to understand the essence of this approach. Fig. 1. Schemes of genomic selection (GS) (left) and traditional MAS for the selection of quantitative traits (right). Both GS and traditional MAS contained training and breeding phases. In the training phase, quantitative trait loci (QTLs) are identified in traditional MAS to produce formulae for genomic estimated breeding value (GEBV) prediction, i.e. GS models. In the breeding phase, favourable individuals are selected based on the genotypes of the selected markers in MAS, whereas GEBVs are used for selection in GS.

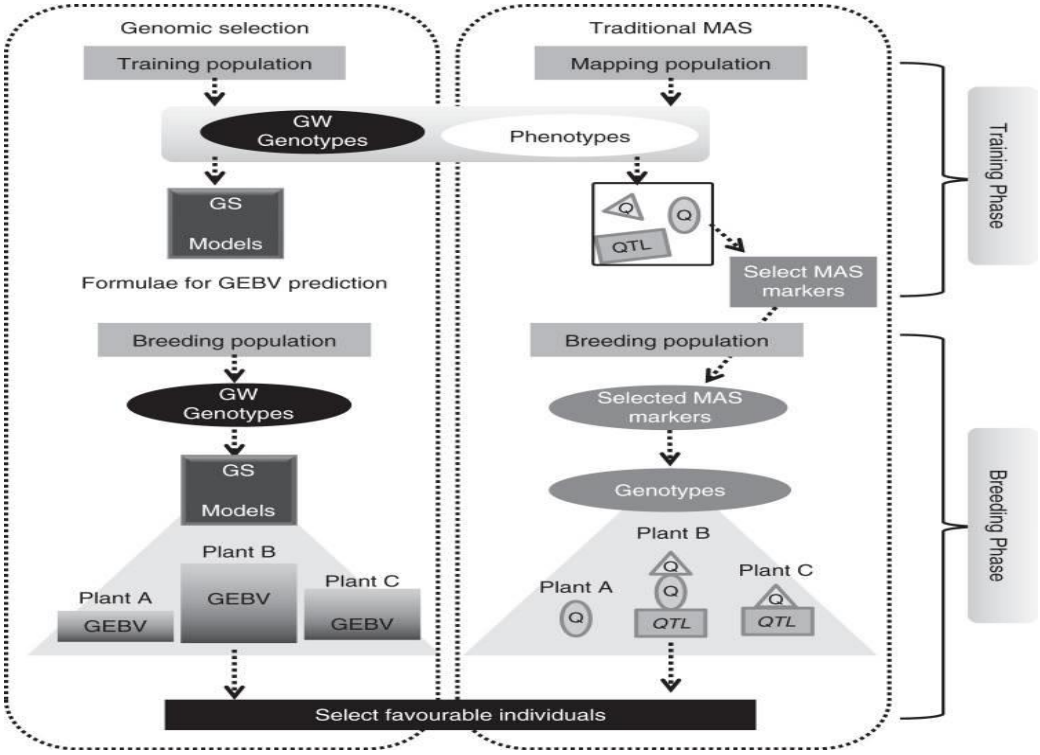

Fig 1

Statistical concepts used in GS

All GS, traditional MAS and pedigree-based phenotypic selection(PS) methods are reliant on a common selection framework, i.e. finding a causal relationship between genetic factors and target traits based on putative genetic factors underlying the phenotypic distribution (in PS) or observed marker genotypes

(in GS and traditional MAS) in a training population. Before describing the statistical approaches used for GEBV prediction, we briefly review the general statistical concepts that are commonly used in PS, traditional MAS and GS. Genetic models and variance decomposition A genetic model of QTs is generally constructed based on an assumption that only effects caused by genetic factors are inherited across the generations. A simple but frequently used genetic model is that the phenotypic value of an individual (P) is expressed as the summation of the genetic value (G) and the residual environmental effect (E): P = G + E, where the genetic value G includes additive genetic effect, dominance effect and epistasis. If we suppose that there is no correlation between G and E (i.e. no G × E effect), the covariance between G and E can be set at zero [Cov (G, E) ¼ 0]. The phenotypic variance, V(P), is then expressed as the summation of the genetic variance, V(G), and the environmental variance, V(E):\ V (P) = V (G) + V(E) + 2Cov(G, E) = V(G) + V(E).

Heritability

Heritability is a measure for evaluating the degree to which the phenotypic characteristics of a population are inherited to the next generation, and it is represented as the ratio of genetic variance to phenotypic variance. Broad sense heritability (H2) focuses on the total genetic effects, G, including the additive, dominance and epistatic effects, whereas narrow sense heritability (h2) counts only additive genetic effects. Therefore, for h2, the genetic model (P ¼ G + E) can be rewritten using the additive genetic effect, A: P = A + E′ Here, E′ represents the residual effects that are not included in the additive genetic effect, A. Note that the dominant and epistatic effects are in E′. If we suppose that there is no correlation between A and E′, then the phenotypic variance V(P) can be broken down into the additive genetic variance, V(A), and the residual effects variance, V(E): V(P) = V(A) + V(E′).Because h2 is defined by the ratio of V(A) to V(P), it is represented as follows: h2 = V(A)/V(P).

In GS, V(A) is broken down again into the variances explained by multiple DNA markers, V(A1), V(A2), . . ., V(AM) under the assumption that DNA markers are not correlated with each other (Meuwissen *et al*., 2001) [43].

Breeding value (BV)

The BV of an individual i in a population is defined as follows based on the narrow sense heritability, h2: BVi = m0 + h2 (yi − m0) = m0 + (yi − m0) V(A)/V(P).

Here, yi is the phenotypic value of individual i, while m0 is the mean phenotypic value of the population. Because V(A) cannot be directly observed, h2 has been conventionally estimated by a comparison of the phenotypic values of parents and their offspring. The BVs that are predicted based on an estimated heritability are known as EBVs. By contrast markers as follows: yi = b0 + x1ib1 + x2ib2 . . . + xMibM + 1i = SM, j=0x jib j + 1i (2) where xji is the genotype of the jth marker in the ith individual, coefficient bj is its effect on the phenotype, and x0i ¼ 1 is a dummy variable. Similarly, the coefficients are determined by minimizing an error function, E ¼ SNi ¼ 1 (yi – _Mj ¼ 0 xji bj).

GS in biparental populations

Thus far, we have only discussed GS in the context of population-wide linkage disequilibrium, where the population might be defined as an entire breed of cattle, a market class of a crop (e.g. hard red wheat), or perhaps a

breeding program. Because plants can often produce very large full sibships (an F2 population derived from a single F1 by selfing is an example of such a sibship), however, there is also a tradition of QTL detection, MAS and GS within such sibships [i.e. in F2, recombinant inbred line, or doubled haploid populations; 24, 48–50]. These simulations have almost exclusively used ridge regression. Some interesting results are (i) very low marker densities, on the order of eight per Morgan, can deliver accuracies close to the maximum observed; (ii) using ridge regression, there was a marker density optimum above which accuracy declined [48]; (iii) accuracy assuming true marker variances were known was only marginally higher (0–8%) than assuming all marker variances were equal [48]; (iv) GS can out-perform phenotypic selection even when the biparental population is composed of very few (e.g. 35) individuals [50]. As an overall population improvement strategy, no study has contrasted performing GS within biparental crosses to performing it across a breeding program as a whole. The primary advantage we see to the former approach is its low marker density need.

The primary disadvantages are (i) that it requires separate model training within each cross: it seems suboptimal not to analyze all crosses jointly (as would occur if GS were performed over the breeding program as a whole); and (ii) the first generation of progeny from a cross cannot be selected on the basis of prior information but needs to be phenotyped. This practice slows down the breeding cycle relative to program-wide GS. Numerous methods from statistics and machine learning have been proposed for genomic prediction since, due to the high modeling complexity associated to the large amount of markers available. For instance, modeling the effects of thousands interacting genes (i.e., epistasis) associated to complex quantitative traits is not trivial. There is an increasing number of studies supporting that epistasis may be the most prevalent form of genetic architecture for quantitative traits (Flint and Mackay, 2009; Moore and Williams, 2009; Huang *et al*., 2012). Hence genomic prediction methods which can account for epistatic genetic architectures have been proposed. For example, Gianola *et al*. (2006) and Gianola and van Kaam (2008) first proposed reproducing kernel Hilbert space (RKHS) regression for genomic prediction when dealing with epistatic genetic architectures. Later Howard *et al*. (2014) showed that RKHS and support vector machine regression (SVR), when dealing with an additive genetic architecture, could be almost as competitive as parametric methods such as best linear unbiased predictor (BLUP), least absolute shrinkage and selection operator (LASSO) or Bayesian linear regressions (Bayes A, Bayes B, Bayes C, Bayes Cπ, and Bayesian LASSO). These authors also showed that RKHS regression and SVR, with some other non-parametric methods, clearly outperformed parametric methods for an epistatic genetic architecture. Parametric and nonparametric methods have been developed for purposes of predicting phenotypes. These methods are based on retrospective analyses of empirical data consisting of genotypic and phenotypic scores.

In parametric regression models for dense molecular markers (e.g. Meuwissen *et al*., 2001) [43], gi is a parametric regression on marker covariates, xik with k=1, p indexing markers. The linear model takes the form: yi=gp k=1xikbk+ei, where bk is the regression of yi on xik. Often, p>>n and some shrinkage estimation method such as ridge regression (Hoerl & Kennard, 1970a, 1970b) [28] or LASSO (Least Absolute Shrinkage and Selection Operator, Tibshirani, 1996) [51], or their Bayesian counterparts, are used to estimate marker effects. Among the latter, those using markers specific shrinkage such as the Bayesian LASSO of Park & Casella (2008) [45] or methods BayesA or BayesB of Meuwissen *et al*. (2001) [43] are the most commonly used. In linear regressions, dominance and epistasis may be accommodated by adding appropriate interactions between marker covariates to the model; however, the number of predictor variables is extremely large and modelling interactions is only feasible to a limited degree. However, issues of high dimensionality, multicollinearity, and the inability of parametric models deal effectively with epistasis can jeopardize accuracy and predictive ability. To overcomes such issues new non – parametric models like: pRKHS, which combines the features of supervised principal component analysis (SPCA) and reproducing kernel Hilbert spaces (RKHS) regression, with versions for traits with no/low epistasis, pRKHS-NE, to high epistasis, pRKHS-E has been proposed.

Recent Progress in GS Studies

Most important factor determining the success of GS is the accurate prediction of GEBVs. The accuracy of the predicted GEBVs is often estimated based on the correlation between the observed phenotypic value and GEBVs. To produce accurate GEBVs, several studies have applied comparative statistical approaches to GEBV prediction. In addition, simulations studies have been widely used to investigate the affect of the number of QTLs, markers, individuals and other variables. These studies were reviewed recently by Heffner *et al*. (2009) [25] and Jannink *et al*. (2010) [26], and so are not described further in this section. Instead, we focus on recent progresses in GS based on empirical data to understand better the practical use of GS.

Animal science

Studies of GS are more common in the field of animal science than plant science. The BV concept was used in animal breeding long before the emergence of GS, so the GS approach was more readily accepted by animal scientists. In addition, the lower diversity of the targeted species and fewer effects of environmental factors during the growing stage might have contributed to the rapid introduction of GS in animal science. The first empirical GS study in animal science was reported by Legara *et al*. (2008). A total of 1884 individuals were generated from eight inbred lines and genotyped using 10 946 single nucleotide polymorphism (SNP) markers, before predicting the GEBVs for four traits related to body sizes. A comparison of the predictive ability and accuracy of GEBVs generated with or without SNP genotypes and polygenetic effects demonstrated that GW genetic evaluation and selection provided better accuracy and predictive ability than the classical polygenic model. The most advanced progress in GS has been observed in dairy cattle. In Table 1, the results of three GS studies in dairy cattle are summarized (Hayes *et al*., 2009; Luan *et al*., 2009; Van Raden *et al*., 2009) [24, 53]. In addition to the three reports in Table 1, seven empirical GS studies of dairy cattle were also reported and reviewed by Hayes *et al*. (2009) [24], Moser *et al*. [44] (2009) and Calus (2010) [6]. Of the three cattle studies

in Table 1, a total of 500–5335 individuals were used for GEBV prediction using 18 991–38 416 SNPs. GEBVs for various QTs related to milk production, cattle body size and fertility were predicted using several different methods, where the accuracy of GEBVs ranged from 0.14 to 0.69. Rolf *et al*. (2010) [48] and Mujibi *et al*. (2011) [41] reported

Table 1: Only studies that investigated the accuracy of GEBVs based on the correlation between observed phenotypic values and GEBVs are listed. Number of individuals used for GEBV prediction (training population) versus that used for validation (validating population).

Species	Population type	Size of test population	SNPs	GEBV	Models for GEBV prediction	Traits	Reference
Mice	A heterogeneous population derived from eight inbred lines	1884	10 946	0.1 60.27	Linear mixed model with SNP genotypes, not polygenetic effects	Weight, growth slope, body length, body mass index	Legarra *et al*. (2008) [37]
Dairy cattle; Australian Holstein-Friesian	Bull progeny tested by Genetics Australia	730	38 259	0.14,0.55	BayesA	Breeding value, profit ranking, selection value, protein yield and protein percentage	Hayes *et al*. (2009) [24]
Dairy cattle; Norwegian Red	34 sires and 466 sons	500	18 991	0.6	G-BLUP	Milk yield, fat yield, protein yield, clinical mastitis, calving ease	Luan *et al*. (2009)
Dairy cattle; North American Holstein	American Holstein bulls born between 1952 and 2002	5335	38 416	0.33,0.69	Linear mixed model	27 traits for milk production, body size, shape and fertility	Van Raden *et al*. (2009) [53]
Beef cattle; Angus	Parental identified steers and sires	2405	41 028	0.23,0.45	Genomic relationship matrices	Average daily feed intake, residual feed intake, average daily gain	Rolf *et al*. (2010) [48]
Beef cattle; Angus, Charolais,	Admixture population	721	37 959	-.07,0.48	RR-BLUP (with top 200 SNPs)	Average daily gain, dry matter intake, residual feed intake	Mujibi *et al*. (2011) [41]
University of Alberta hybrid bulls Chicken; blown-egg layer line	Five consecutive generations in a single line	2708	23 356	0,2	G-BLUP and Bayes-C-p	13 traits for eggs and 3 traits for chicken bodies	Wolc *et al*. (2011) [57]

GEBV prediction in beef cattle. Parentally identified steers and sires of 2405 Angus cattle were genotyped using 41 028 SNPS in a study by Rolf *et al*. (2010) [48], while an admixture population consisting of Angus, Charolais and hybrid bulls was genotyped using 37 959 SNPs for 721 individuals in a study by Mujibi *et al*. (2011) [41]. GEBVs for traits related to daily gain and daily intake were investigated, and the estimated accuracies ranged from 20.07 to 0.48. In chickens, Wolc *et al*. (2011) [57] tested 16 traits related to eggs and chicken body size with 23 356 SNP genotypes using 2708 individuals derived from a single blown egg-layer line. The accuracy of GEBVs estimated ranged from 0.2 to 0.7. The reported studies used different materials and statistical methods for GEBV prediction, but many of these studies showed that the accuracy of GEBV was higher than that of traditional EBV and it was increased with a larger population size, larger numbers of genotyped SNPs, and higher heritability of the targeted traits. The details are not described here, but some of the studies compared different statistical methods for GEBV prediction. Note that the best approaches with the highest accuracy of GEBVs were different in each case (Table 1). The accuracy of GEBVs estimated in empirical studies fell below 0.7 (Table 1), which was lower than that suggested by many simulation studies such as 0.85 in Meuwissen *et al*. (2001) [43]. Calus (2010) [6]. indicated that the distribution of QTL effects in real data is generally lower than that assumed in simulation studies.

Plant science

Plant breeding targets a diversity of species with different reproduction systems, generation times, genome structures and utilized organs. Thus, various methods are used in conventional breeding, i.e. PS and traditional MAS, to adapt to the demands of different targeted species and breeding objectives. Like conventional breeding, GS should be adapted to the fit different types of plant species and breeding objectives. Reports on plant species that specified 'genomic selection' or 'genome wide selection' have been published since 2007. Piyasatin *et al*. (2007) simulated the efficiency of GS in a cross of inbred lines, which is common in plant breeding but not in animal breeding. However, no specific plant species was considered as the targeted species in this paper. Simulation studies of specific species were firstly published for maize (Bernardo and Yu 2007) [4], where a comparison between GS and marker-assisted recurrent selection (MARS) was demonstrated for three cycles of selection of doubled haploid lines (DHLs). The response of GS was 18 – 43 % greater than that of MARS with different numbers of QTLs (20, 40 and 100). Moreover, simulation studies using maize. were performed to determine the advantages of using DHLs compared with F_2 populations in GS and MARS (Mayor and Bernardo, 2009) [3], and to develop a methodology for the rapid introgression of exotic germplasms in an adapted line of maize via GS (Bernardo, 2009) [3]. In addition to maize, two GS simulations were performed with the oil palm, which is an outcrossing species that requires 19 years for one cycle of (PS) (Wong and Bernardo, 2008) [58], and with a self-pollinated crop, barley (Bernardo, 2010) [2]. These studies simulated biparental cross populations, three studies also reported GS simulation using multiple inbred lines in barley

based on real genotype data obtained mainly from SNPs and diversity array technology (DART) (Zhong *et al*., 2009; Jannink, 2010; Iwata and Jannink, 2011) [61, 26]. Zhong *et al*. (2009) [61]. compared the accuracy of four GS prediction methods that were affected by marker density, level of linkage disequilibrium (LD), QTL number, and sample size, where the level of replication in populations was generated using 42 multiple inbred lines of two-row spring barley with the genotypes of 1933 loci obtained from SNP, DArT and classical markers. They concluded that the GS prediction method with the highest accuracy changed with different levels of LD between the marker and QTLs, QTL effects, and generations of individuals. Moreover, Iwata and Jannink (2011) simulated the accuracy of GS using more large-scale data, consisting of 1325 SNPs in 863 breeding lines of barley derived from nine breeding programmes in the USA. Seven methods were used for GEBV prediction and the mean of the predictions in all methods was more accur-ate than predictions based on any single method under medium and high heritability. Jannink (2010) [26]. simulated the dynamics of long-term GS using 192 breeding lines from an elite six-row spring barley programme with genotypes identified by 983 polymorphic markers. The results suggested that losing favourable alleles with weak LD with markers during selection cycles was inevitable, while placing additional weight on low-frequency favourable alleles was important for long-term GS.

Investigations of the accuracy of GEBV predictions using em-pirical data have been reported for maize, barley, wheat and Arabidopsis thaliana (Table 2). It was first demonstrated by Lorenzana and Bernardo (2009) [3] for maize, A. thaliana and barley. All the test populations were generated from biparental crosses where the number of test progeny and markers ranged from 119 to 415 and 69 to 1339, respectively. Arabidopsis thaliana had the highest accuracy of GEBVs, although the number of polymorphic markers used for genotyping was the lowest. This study was followed by demonstrations of GS using empirical data in maize by Piepho (2009), Crossa *et al*. (2010) [9] and Guo *et al*. (2011), as shown in Table 2. Piepho (2009) compared the performance of nine models using a series of experiments with DHLs derived from a single cross conducted in five environments, and suggested the need to appropriately model genotype – environment interactions and to employ an independent estimate of error. Crossa *et al*. (2010) [9] demonstrated GS using a genetically diverse population [300 lines bred in CIMMYT (The International Maize and Wheat Improvement Center)] and 1148 SNPs, with a predicted accuracy of GEBVs ranging from 0·42 to 0·79 by ridge regression BLUP. The largest-scale analysis of maize was performed by Guo *et al*. (2011) [22], which used 4699 progeny derived from 25 nested association mapping populations with genotypes for 1106 SNPs. While a common line, 'B73', was used as the maternal line across the 25 mapping populations, the paternal lines were all different.

Table 2: Ranges of GEBVs accuracy investigated in empirical plant GS studies.

Species	Training population ratio*	No. of genotyped markers	Accuracy of GEBVs	Models for GEBV prediction	Traits	Reference
Maize	0.43, 0.65, 0.80	1339 SSRs and RFLPs	0.48–0.73	BLUP	8 morphological traits, 3 chemical components, grain moisture	Lorenzana and Bernardo (2009) [3]
Maize	0.20, 0.40, 0.60, 0.80	1106 SNPs	0.26–0.57	RRBLUP	Three flowering traits	Guo *et al*. (2011) [22]
Wheat	0.14, 0.28, 0.55	574 DArTs	0.41–0.73	RRBLUP	8-grain quality	Piepho (2009)
Beef Cattle	not shown	37959 SNPs	-0.07-0.48	RRBLUP	Average daily gain, dry matter intake, residual feed intake	Mujibi *et al*. (2011) [41]
Loblolly pine	not shown	3938 SNPs	0.64–0.77	BLUP	Diameter at breast height, total height	Resende *et al*. (2011) [47]
Loblolly pine	Not shown	3406 SNPs	0.3–0.83	Pedigree model	Growth and quality traits	Isik *et al*. (2011) [32]

Interestingly, the accuracy of the predicted GEBVs was different in the 25 crosses, although the study used almost the same SNPs, targeted traits and population sizes. Interestingly, the ranges of accuracies in empirical studies were higher in plant studies than animal studies, although most plant studies employed lower numbers of genotyping markers. This might be due to the lower genetic diversity caused by a small number of parental lines and a greater bottle-neck in the breeding materials. Note that the numbers of markers used for woody species was higher than that used for annual plant species. Empirical plant GS studies show that GS is a potential method for plant breeding and that it can be performed with realistic sizes of populations and markers when the populations used are carefully chosen.

Computer package for GS modelling

An R-Package for GS is available on http://www.r-project.org/. No user-friendly software has yet been developed, such as QTL Cartographer (Wang *et al*., 2011) and Map QTL (Van Ooijen, 2004) that are used in QTL analysis. The development of a user-friendly software package is required to enhance the general application of GS.

Future perspectives in GS

Genomic selection is one such method which overcomes the drawback of marker assisted selection not being effective in improving complex polygenic traits an traits with low heritability (in the narrow sense) traits. Narrow-sense heritability is defined as the ratio of the genetic variance of additive genetic effects to the phenotypic variance. Low heritability traits are caused by the high variance of non-additive genetic effects, such as environmental factors, G × E interactions, and dominant and epistatic genetic effects. Thus, accurate prediction of GEBVs is a central and

recurring challenge in plant and animal breeding. Numerous methods from statistics and machine learning viz best linear unbiased predictor (BLUP), least absolute shrinkage and selection operator (LASSO), reproducing kernel Hilbert spaces (RKHS) regression, with versions for traits with no/low epistasis, pRKHS-NE, to high epistasis, pRKHS-E have been proposed for genomic prediction. Current GS algorithms have not been connected with previous and current studies of genetics and genomics, such as QTLs and (candidate) gene identification. By integrating the essence of GS with other fields of genetics and genomic studies, it might be possible to escape the black box. Meuw issen (2007) [42].

References

1. Barendse W, Harrison BE, Bunch RJ, Thomas MB, Turner LB. Genome wide signatures of positive selection: the comparison of independent samples and the identification of regions associated to traits.BMC Genomics. 2009; 10:178. http://dx. doi.org/10.1186/1471-2164-10-178.
2. Bernardo R. Genome wide selection for rapid introgression of exotic germplasm in maize. Crop Science. 2010; 49:419-425.
3. Bernardo R. Genome wide selection with minimal crossing in self pollinated crops. Crop Science. 2009; 50:624–627.
4. Bernardo R, Yu J. Prospects for genome wide selection for quantitative trait in maize. Crop Science. 2007; 47:1082-1090.
5. Calenge F, Legarra A, Beaumont C. Genomic selection for carrierstate resistance in chicken commercial lines. BMC Proceedings. 2011; 5(Suppl. l4):S24. http://dx.doi.org/10.1186/1753-6561-5-S4-S24.
6. Calus MP. Genomic breeding value prediction: methods and procedures Animal. 2010; 4:157-164.
7. 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.
8. Cleveland MA, Forni S, Deeb N, Maltecca C. Genomic breeding value prediction using three Bayesian methods and application to reduced density marker panels. BMC Proceedings. 2010; 4(Suppl 1):S6. http:// dx.doi. org/10.1186/1753-6561-4-S1-S6.
9. Crossa J, Campos G, de L, Pe´rez P, *et al*. Prediction of genetic values of quantitative traits in plant, 2010.
10. Dudley JW, Johnson GR. Epistatic models improve prediction of performance in corn. Crop Science. 2009; 49:763-770.
11. Elshire RJ, Glaubitz JC, Sun Q, *et al*. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species, 2011. PLoS ONE 6: pe19379. http://dx.doi.org/10.1371/journal.pone.0019379.
12. Fisher RA. The correlation between relatives on the supposition on Mendelian inheritance. Transactions of the Royal Society of Edinburgh. 1918; 52:399–433.
13. Flint J, Mackay TF. Genetic architecture of quantitative traits in mice, flies, and humans. Genome research. 2009; 19(5):723-73.
14. Garris AJ, McCouch SR, Kresovich S. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza sativa* L.) Genetics 165:759-769.
15. Gaut BS, Long AD. The lowdown on linkage disequilibrium. The Plant Cell. 2003; 15:1502-1506.
16. Gianola D, Fernando RL. Bayesian methods in animal breeding theory. Journal of Animal Science. 1986; 63:217-244.
17. Gianola D, Van Kaam JBCHM. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics. 2008; 178:2289-2303.
18. Goddard M. Genomic selection: prediction of accuracy and maximization of long term selection. Genetica. 2009; 136:245-257.
19. Goddard ME, Hayes BJ. Genomic selection. Journal of Animal Breeding and Genetics. 2007; 124:323-330.
20. Gonza´lez-Martı´nez SC, Brown GR, Ersoz E, *et al*. 2004. Nucleotide diversity, linkage disequilibrium and adaptive variation in natural populations of loblolly pine. Plant & Animal Genomes XII Conference, 2004, 10-14, San Diego, CA, W3.
21. Grattapaglia D, Resende MDV, Resende MR, *et al*. Genomic selection for growth traits in Eucalyptus: accuracy within and across breeding Populations. BMC Proceedings 5 (Suppl. 7): 2011, O16.
22. Guo Z, Tucker DM, Lu J, Kishore V, Gay G. Evaluation of genome wide selection efficiency in maize nested association mapping populations. Theoretical and Applied Genetics. 2011; 124:261-275.
23. Gupta PK, Pawan S, Kulwal PL. Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Molecular Biology. 2005; 57:461-485.
24. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Genomic selection in dairy cattle: progress and challenges. Journal of Dairy Science. 2009; 92:433-443.
25. Heffner EL, Sorrells ME, Jannink JL. Genomic selection for crop improvement. Crop Science. 2009; 49:1-12.
26. Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME. Plant breeding with genomic selection: gain per unit time and cost. Crop Science. 2010; 50:1681-1690.
27. Henderson CR. Best linear unbiased estimation and prediction under a selection model. Biometrics. 1975; 31:423.
28. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for non-orthogonal problems. Technometrics. 1970a; 12:55-67.
29. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for non-orthogonal problems. Technometrics. 1970b; 12:69-82.
30. Hu Z, Li Y, Song X, *et al*. Genomic value prediction for quantitative traits under the epistatic model. BMC Genetics. 2011; 12:15. http://dx. doi.org/10.1186/1471-2156-12-15.
31. Huang W, Richards S, Carbone MA, Zhu D, Anholt R, R, Ayroles JF, Warner CB. Epistasis dominates the genetic architecture of Drosophila quantitative traits. Proceedings of the National Academy of Sciences. 2012; 109(39):15553-15559.
32. Isik F, Whetten R, Zapata-Valenzuela J, Ogut F, McKeand S. Genomic selection in loblolly pine – from lab to field. BMC Proceedings. 2011; 5(Suppl. 7):I8.
33. Isobe S, Nakaya A, Tabata S. Genotype Matrix

Mapping (GMM): searching for QTL interactions in genetic variation in complex traits. DNA Research. 2007; 14:217-225.
34. Iwata H, Jannink JL. Marker genotype imputation in a low-marker-density panel with a high-marker density reference panel: accuracy evaluation in barley breeding lines. Crop Science. 2010; 50:1269-1278.
35. Jannink JL. Dynamics of long-term genomic selection. Genetics Selection Evolution. 2010; 42:35.
36. Jannink JL, Lorenz AJ, Iwata H. Genomic selection in plant breeding: from theory to practice. Briefings in Functional Genomics and Proteomics. 2010; 9:166-177.
37. Legarra A, Robert-Granie´ C, Manfredi E, Elsen JM. Performance of genomic selection in mice. Genetics. 2008; 180:611-618.
38. Mayor PJ, Bernardo R. Genome wide selection and marker-assisted recurrent selection in doubled haploid versus F2 populations. Crop Science. 2009; 49:1719-1725.
39. Mei HW, Li KZ, Shu QY, *et al*. Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two backcross populations. Theoretical and Applied Genetics. 2005; 110:649-659.
40. Moore JH, Williams SM. Epistasis and its implications for personal genetics. The American Journal of Human Genetics. 2009; 85(3):309-320.
41. Mujibi FDN, Nkumah JD, Durunna ON, *et al*. Accuracy of genomic breeding values for residual feed intake in crossbred beef cattle. Journal of Animal Science. 2011; 89:3353-3361.
42. Meuwissen THE. Genomic selection: marker assisted selection on an genome wide scale. Journal of Animal Breeding and Genetics. 2007; 124:321-322.
43. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001; 157:1819-1829.
44. Moser G, Tier B, Crump RE, Khatkar MS, Raadsma HW. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genetics Selection Evolution. 2009; 41:56.
45. Park T, Casella G. The Bayesian LASSO. Journal of the American Statistical Association. 2008; 103:681-686.
46. Remington DL, Thornsberry JM, Matsuoka Y, *et al*. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proceedings of the National Academy of Sciences of the USA. 2001; 98:11479-11484.
47. Resende MFR, Valle PRM, Acosta JJ, Resende MDV, Grattapaglia D, Kirst M. Stability of genomic selection prediction models across ages and environments. BMC Proceedings. 2011; 5 (Suppl. 7):O14.
48. Rolf MM, Taylor JF, Schnabel RD, *et al*. Impact of reduced marker set estimation of genomic relationship matrices on genomic selection for feed efficiency in Angus cattle. BMC Genetics. 2010; 11:24. http://dx. doi.org/ 10.1186/1471-2156-11-24.
49. Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE. Genomic selection using different marker types and densities. Journal of Animal Science. 2008; 86:2447-2454.
50. Thomas D. Gene-environment-wide association studies: emerging approaches. Nature Reviews Genetics; 11:259-272.
51. Tibshirani R. Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society. 1996; B 58:267-288.
52. Van Ooijen. MapQTLw 5, software for the mapping of quantitative trait loci in experimental populations. Kyazma B.V., Wageningen, The Netherlands, 2004.
53. Van Raden PM, Van Tassell CP, Wiggans GR, *et al*. Invited review: reliability of genomic predictions for North American Holsein bulls. Journal of Dairy Science. 2009; 92:16-24.
54. Wang S, Basten CJ, Zeng ZB. Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC USA, 2011.
55. Van Der Werf J. Animal breeding and the black box of biology. Animal Breeding and Genetics. 2007; 124:101.
56. Williams KS, Cummings MR. Concepts of genetics, 5th edn. Englewood Cliffs, NJ: Prentice-Hall, 1997.
57. Wolc A, Stricker C, Arango J, *et al*. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. Genetics Selection Evolution. 2011; 43:5.
58. Wong CK, Bernardo R. Genome wide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theoretical and Applied Genetics. 2008; 116:815-824.
59. Xu S, Jia Z. Genome wide analysis of epistatic effects for quantitative traits in barley. Genetics. 2007; 175:1955-1963.
60. Xu Y, Crouch JH. Marker-assisted selection in plant breeding: from publication to practice. Crop Science. 2008; 48:391-407.
61. Zhong S, Dekkers JCM, Fernando RL, Jannink JL. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics. 2009; 182:355–364.