Advances in genomic selection in domestic animals

ZHANG Zhe, ZHANG Qin & DING XiangDong[*]

Key Laboratory of Animal Genetics and Breeding of the Ministry of Agriculture, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

Genomic selection (GS) is a marker-assisted selection method, in which high density markers covering the whole genome are used simultaneously for individual genetic evaluation via genomic estimated breeding values (GEBVs). GS can increase the accuracy of selection, shorten the generation interval by selecting individuals at the early stage of life, and accelerate genetic progress. With the availability of high density whole genome SNP (single nucleotide polymorphism) chips for livestock, GS is reshaping the conventional animal breeding systems. In many countries, GS is becoming the major genetic evaluation method for bull selection in dairy cattle and GS may soon completely replace the traditional genetic evaluation system. In recent years, GS has become an important research topic in animal, plant and aquiculture breeding and many exciting results have been reported. In this paper, the methods for obtaining GEBVs, factors affecting the accuracy of GEBVs, and the current status of implementation of GS in livestock are reviewed. Some unresolved issues related to GS in livestock are also discussed.

genomic selection (GS), genomic estimated breeding value (GEBV), marker-assisted selection (MAS), domestic animals, breeding

Most economically important traits for livestock animals are quantitative traits. In traditional breeding, breeders improve these traits by recording phenotypes and pedigrees. Elite individuals are selected according to the rank order of their estimated breeding values (EBVs) obtained by using best linear unbiased prediction (BLUP). Over the past decades, this strategy has proved to be successful for most traits. For example, in the American dairy cattle population the average milk yield increased by about 5000 kg in the past 40 years, and 60% of this change has been attributed to genetic improvements. However, for traits with low heritability, sex-limited traits, longevity traits, and traits that are difficult or expensive to measure (such as carcass traits), traditional selection methods have been of little use or are inefficient. With the advance of molecular genetics in the past 20 years, genetic markers have been used in marker-assisted selection (MAS) in animal breeding programs. MAS is expected to increase the accuracy of selection and enhance genetic improvement. However, although many theoretical studies have been carried out, the application of MAS in animal breeding practices has been very limited. This is mainly because, so far, only a few markers have been validated to have significant effects on economically important traits, and these markers only account for a small proportion of the genetic variation [1].

With advances in animal genome sequencing, a large number of SNPs covering whole genomes have been discovered. To take full advantages of the genotypic information from whole genomes in the genetic evaluation of animals, Meuwissen et al. [2] proposed a genomic selection (GS) method and showed that, using genetic markers alone, breeding values can be predicted with high accuracy. In recent years, the GS methodology has been widely used and the main factors affecting the accuracy of GS, GS application strategies, and the implementation of GS in breeding practice have been reported in livestock [1], plant [3] and aquaculture [4] breeding. For example, at the 9th World Congress on Genetics Applied to Livestock Production

*Corresponding author (email: xding@cau.edu.cn)

(WCGALP 2010, Leipzig, Germany), over a third of the presentations focused on GS [5]. Undoubtedly, with GS, animal breeding has entered a new era of development. Figure 1 shows a schematic diagram of selection methods used in animal breeding.

In China, the GS methodology and its applications in the breeding schemes of cattle, pigs, chicken and other livestock species are being investigated. Hence, an overview of the work that has been done in the field of GS will be very useful for Chinese animal breeders. In this paper, a comprehensive review of the developments in the GS methodology, important factors affecting the accuracy of genomic estimated breeding values and the current status of the application of GS to livestock breeding is presented. Some unresolved issues related to GS are also discussed.

1 Principles of genomic selection

Unlike traditional MAS which uses only the few markers that are thought to have large effects on the underlying trait of interest to predict breeding values, GS simultaneously uses high density (HD) markers covering the whole genome to predict the breeding values of each genotyped individual [6]. A breeding value predicted in this way is termed the genomic estimated breeding value (GEBV). In GS, it is assumed that each gene or QTL (quantitative trait locus) affecting the trait of interest is in linkage disequilibrium (LD) with at least one of the markers. Theoretically, all genetic variance can be tracked by markers if the marker density is high enough. Hence, GS could overcome the drawback of traditional MAS and predict breeding values more accurately.

2 Methodology to predict GEBV

Depending on the statistical model used, GS methods can be classified as: either indirect methods or direct methods. Indirect methods estimate marker effects in a reference population and then calculate the GEBVs of genotyped individuals in the candidate population by summing the effects of all

relevant markers. Direct methods calculate the GEBVs directly using mixed model equations (MME) as the traditional BLUP method does. The phenotype information of the reference population and the genotype information for all the markers from both reference and candidate populations are used, and a marker-derived relationship matrix is constructed using the HD markers. The indirect and direct GS methods are illustrated schematically in Figure 2.

2.1 Indirect GS methods

In the indirect methods, two steps are needed to obtain the GEBV of a genotyped selection candidate. In the first step, the effects on the traits of interest for each of the whole-genome HD markers are estimated in a reference population in which all individuals are phenotyped and genotyped. In the second step, the GEBV of a genotyped animal in the candidate population is obtained by summing up all the pre-estimated marker effects according to its genotypes. The marker effects can be estimated with the following model:

$$y = Xb + \sum_{i=1}^{m} Z_i g_i + e, \tag{1}$$

where y is a vector of the phenotypic values of all individuals in the reference population, b is a vector of fixed effects, X is the design matrix for b, g_i is the effect of the ith marker, m is the total number of markers, z_i is an index vector representing the genotypes of all individuals at marker i in the reference population, and e is a vector of residual errors with variance-covariance matrix $I\sigma_e^2$ (σ_e^2 is residual variance).

Based on Model (1), different methods have been developed to estimate the marker effects. These methods include ridge regression BLUP [2], Bayesian variable selection methods A and B [2], and Bayesian shrinkage [7]. The main difference between these methods is in the assumption of the distribution of marker effects. In RRBLUP, all the markers are assumed to have equal variance. In BayesA, all markers are assumed to have a nonzero effect with different variances that follow an inverse chi-square distribution. In BayesB, markers are assumed to have a zero effect with an
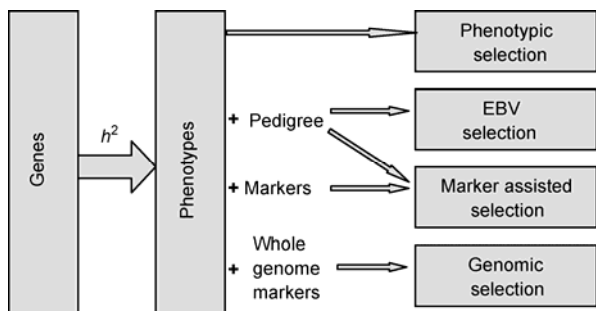

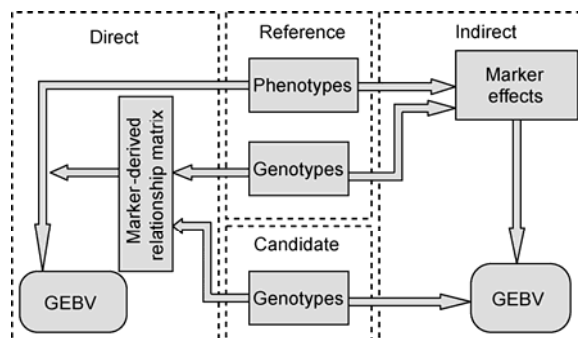**Figure 1**   Schematic diagram of selection methods in animal breeding.

Figure 2   Schematic illustration of the indirect and direct GS methods.

arbitrarily selected probability π; for those with a nonzero effect, the variances follow an inverse chi-square distribution. In BayesS, the assumptions are similar to those of BayesA except that, for markers which have no effects on the trait, their effects and variances are shrunk towards zero during the calculation. Of these methods, BayesB performed better than the others in most simulation studies [2]. A limited number of simulated QTLs, which would favor the BayesB assumptions, might be the main explanation [8].

To avoid the influence of an arbitrarily selected π on the accuracy of predictions, several extensions were made to the BayesB method, and a new method termed BayesC [9,10] was proposed. Verbyla et al. [9] simultaneously estimated the π and marker effects using a stochastic search variable selection method. Meuwissen [10] assumed that marker effects were in a normal distribution with a mixed variance, and the variance, π, and marker effects were simultaneously estimated. Because the Metropolis-Hasting and Gibbs sampling arithmetic used in BayesB is extremely time-consuming, this method is infeasible when frequent predictions are needed. To overcome this difficulty, an iteration conditional expectation arithmetic for BayesB was proposed and termed fBayesB [11]. The fBayesB method can greatly decrease the computation time at the cost of a negligible loss of accuracy.

Besides the methods mentioned above, dimension-reduced methods, such as machine learning [12], principle component analysis [13], least square regression [2], a semi-parametric approach [14], non-parametric additive regression [15] and Bayes LASSO (Least Absolute Shrinkage and Selection Operator) [16] have been used. Based on the RRBLUP method, a non-linear regression method was suggested to vary marker variance in MME [17]. In this method, the BLUP solution could be obtained while assuming different markers with different variances.

With the estimations of marker effect ($\hat{g}_i$), the GEBV of each of the genotyped candidates in the candidate population can be calculated as $\sum_{i=1}^{m} Z_{ji}\hat{g}_i$.

2.2   Direct GS methods

BLUP is a routinely used genetic evaluation method in animal breeding all over the world. The main advantage of BLUP over previous methods, is the use of all information from all known relatives via a numerator relationship matrix (NRM, or A matrix) constructed based on pedigree.

The A matrix represents additive genetic relationship between individuals, and the elements are the expected proportion of identity by descent (IBD) alleles shared by paired animals. The expected proportion derived from pedigree may deviate from the true one because of Mendelian sampling error. With whole-genome HD markers, a more precise IBD proportion can be estimated to replace the expected one [18]. In the direct methods, the A matrix is re-

placed by a relationship matrix constructed from markers, and the GEBVs are predicted in the same way as in the conventional BLUP by solving the MME based on a mixed model as follows:

$$y = Xb + Zu + e, \tag{2}$$

where u is the vector of GEBVs of all individuals in both reference and candidate populations. The covariance matrix of u is $G\sigma_a^2$, where σ_a^2 is the additive genetic variance, and G is the marker-derived relationship matrix (or realized relationship matrix).

The method based on Model (2) was introduced by VanRaden et al. [19] and Habier et al. [20], and was termed GBLUP (genomic BLUP). Compared with RRBLUP, GBLUP has 3 favorable features: (i) the dimension of the genetic effects in MME is reduced from $m \times m$ to $n \times n$ (where n is the number of individuals in the reference and candidate populations) which is computationally less time demanding; (ii) un-genotyped animals can be added into the MME via the pedigree relationship; and (iii) the reliability of individual GEBV can be calculated in the same way as in traditional BLUP [17,21] which is inaccessible for methods based on Model (1). VanRaden [17] proposed and investigated several rules for constructing a G matrix. The equivalence between GBLUP and RRBLUP was shown theoretically by Goddard [22] and Hayes et al. [23] to be the result of the similar assumptions for the genetic basis of the studied trait. Because the G matrix captures not only the expected IBD proportion included in the A matrix but also the Mendelian sampling error, GBLUP has a theoretically higher predicting ability than traditional BLUP [17,20,24]. This advantage has been evidenced in several studies [20,25,26].

The G matrix in GBLUP is constructed based on the infinitesimal model and this might not be optimal in cases where the trait substantially deviated from the infinitesimal model. Zhang et al. [27] suggested that, while constructing the relationship matrix, markers should be weighted according to their contributions to the trait of interest. This matrix was termed the trait-specific marker-derived relationship matrix (TA matrix) because the weights used in TA are trait related. The corresponding BLUP method, with the TA matrix replacing the G matrix, was termed TABLUP. Results from both simulation studies and from practical tests showed that the predicting ability of TABLUP is better than RRBLUP and GBLUP [27,28].

2.3   Extension of the GS methods

The existence of high LD between QTLs and markers is the main assumption of GS methods. If a lower density panel is used in GS, the accuracy of GEBVs decreases because of a lower level of LD [29]. Two strategies were suggested to increase the accuracy of the predictions based on low density panels.

One strategy takes polygenic effects into account in Model (1). The new model can be written as

$$y = Xb + \sum_{i=1}^{m} Z_i g_i + Wu + e, \qquad (3)$$

where u is the vector of polygenic effects with covariance matrix $A\sigma_u^2$ (σ_u^2 is the polygenic variance) and W is the design matrix corresponding to u. Calus et al. [30] showed that Model (3) could increase the accuracy of prediction in cases where the markers are not dense enough. In Model (2), VanRaden [17] suggested using a weighted mixture of the A and G matrices, $wG+(1-w)A$. Because the averaged information was already considered in the G [20], the performance of this extension is yet to be validated.

The second strategy fits haplotype effects rather than single marker effects into Model (1). The performance of different haplotypes has been compared, and the results showed that a longer haplotype is better than a shorter one for a sparse panel [26]. However, the inference of haplotypes, a necessary step before the prediction of haplotype effects, would increase computational complexity and introduce haplotype inference errors. Instead of using haplotypes, Habier et al. [31] estimated probabilities of descent of marker alleles. Their results showed that this strategy gave a relatively high accuracy and only required a cheap low density panel.

In practice, it is often the case that not all animals have been genotyped with HD panels. To overcome this problem, Legarra et al. [32] proposed that all genotyped and ungenotyped animals are put into one relationship matrix. The extended matrix was called the H matrix and this extension of the method was shown to be feasible in a broiler population [33].

3 Accuracy of GEBV

3.1 Prediction of accuracy

The accuracy (r) of GEBV is defined as the correlation between the GEBVs and true breeding values (TBVs) [2]. Accuracy can be obtained directly in simulation studies where the TBVs are known; however, in a real population, TBVs are unknown, therefore for the design of an optimum GS program, a reliable theoretical prediction of the accuracy of the GEBVs will be very useful for efficient breeding practice. Several researchers have contributed to this aspect.

The first formula, shown below, to predict the accuracy of GEBV was derived by Daetwyler et al. [34]:

$$r = \sqrt{N_p h^2 / (N_p h^2 + N_{qtl})}, \qquad (4)$$

where N_p is the total number of phenotypic records in the reference population, h^2 is the heritability of the trait investigated, and N_{qtl} is the number of independent QTLs affecting the trait.

Goddard [22] derived a different form of the formula,

$$r = \sqrt{1 - \frac{\lambda}{2N_p\sqrt{a}} \log\left(\frac{1+a+2\sqrt{a}}{1+a-2\sqrt{a}}\right)}, \qquad (5)$$

where $a = 1+2\lambda/N_p$, $\lambda = (1-h^2)M_e/(h^2\log(2N_e))$ in which N_e is the effective population size in historical population, and M_e is the effective number of chromosome segments estimated from

$$M_e = (2N_e L)/\log(4N_e L), \qquad (6)$$

where L is the length of the genome.

Daetwyler et al. [35] subsequently modified the formula by introducing M_e into Formula (4). They deduced that GBLUP would not be affected by N_{qtl} but BayesB would be affected. Therefore, they proposed two different formulas, one for GBLUP and another for BayesB.

For GBLUP

$$r = \sqrt{N_p h^2 / (N_p h^2 + M_e)}. \qquad (7)$$

For BayesB

$$r = \sqrt{N_p h^2 / (N_p h^2 + \min(M_e, N_{qtl}))}. \qquad (8)$$

Because several assumptions were made in the derivation of these formulas [22,34], any violation of the assumptions might cause deviations between the observed and predicted accuracies. Simulation studies showed that although the predicted accuracy was slightly higher than the observed accuracy, they both respond in similar ways to changes in the relevant parameters [35,36].

3.2 Main factors affecting accuracy of GS

Although the formulas above do not exactly give the accuracy of the GEBV for a particular GS breeding scheme, they indicate the important factors that affect the accuracy of GS. These factors can be classified into two categories according to their alterabilities.

(i) Unalterable factors. Unalterable factors include the length of the genome (L) which is species specific, the effective population size (N_e) which is population specific and depends on the evolution history, and the genetic architecture of the trait of interest which can be described based on heritability and the number of QTLs affecting the trait. These factors are fixed for a given trait and a given population.

According to Formula (6), the accuracy of GEBVs decreases with increases of M_e which depends on the product of L and N_e. In addition, N_e contributes much to the intensity of LD (measured with r^2) between paired markers. Sved [37] approximately estimated r^2 for two linked loci at a distance of d as

$$r^2 = 1/(1 + 4N_e d). \qquad (9)$$

From this formula, it is clear that r^2 decreases with increasing N_e. In such a case the accuracy of GEBVs also decreases.

Formulas (4) and (7) indicate that, with a fixed N_{qtl} or M_e, the number of phenotypic records needed for a trait with $h^2 = 0.1$ is 5 fold more than for a trait with $h^2 = 0.5$.

(ii) Alterable factors. Alterable factors are changeable or designable in any GS breeding program. These factors mainly include the size and structure of the reference population, marker density, and the approach used to predict the GEBVs.

Generally, the size of the reference population is equal to the number of phenotypic records (N_p) available. The importance of N_p is obvious from all of the formulas above and this has been evidenced in both real data and in simulation studies [2]. Muir et al. [25] also investigated the affect of the number of generations in the reference population on GS and concluded that, when the reference population size is fixed, the more generations in the reference population, the better is the accuracy and the persistency of accuracy.

Formula (9) indicates that the denser the markers are the higher the r^2 (LD) between them are, resulting in a higher accuracy of the GEBV. This has been clearly shown in several studies [10,26,29], and will not be discussed further in this review.

Although the accuracy of the GEBVs depends on marker density, it does not mean that the accuracy of the GEBVs persists across all generations of the selection candidates that were genotyped with the same marker panels. Usually, the accuracy of the GEBVs decreases in subsequent generations [2,20]. This loss of accuracy is mainly because of the change of LD between markers and QTLs and is attributable to recombination, selection, and migration in domestic animals. To retain the accuracy, breeders can use denser panels, enlarge the size of the reference population, or re-estimate marker effects frequently.

Different GS methods have been compared in various scenarios in several studies [8,38]. None of the methods seem to dominate all scenarios. The performance of a method depends not only on the theoretical properties of the method but also on the situations in which it is applied (e.g., different species, different populations and different traits) [35]. The specific situation determines whether or not the assumptions behind a particular method hold. Therefore, method validation is strongly recommended before its implementation. If this is not done, some of the advantages of GS could be lost because of improper implementation procedures.

4 Current status of the implementation of genomic selection

A simulation study on the economics of using GS showed that dairy cattle breeding organizations could save up to 92% of breeding costs if the traditional progeny test system was replaced by a GS breeding program [39]. This saving is mainly attributed to the dramatic reduction of generation interval and increase of selection accuracy for bull dams [6]. This promising prospect, which is definitely attractive to breeders and researchers, has driven to a worldwide upsurge in the implementation of GS in animal breeding programs in the past few years.

4.1 GS in dairy cattle

GS has been implemented in national and international dairy cattle breeding programs in many countries [40,41]. The application of GS in dairy cattle has been reported in many countries, including USA and Canada [41], Australia [9,38,42], Norway [43], New Zealand [44], The Netherland [45], Denmark [46], Germany [47], and Ireland [48]. The traits currently evaluated using genomic information include all the target traits included in the traditional breeding programs. Genotyped individuals ranged from tested bulls, milking cows to heifers. A recent survey conducted by Interbull (http: //www.interbull.org/) [49] reported on the current situation of GS implementation in dairy cattle; the results of the survey are summarized in Table 1. It should be noted that the reference population size reported in the various studies might change as the GS projects proceed.

The results of these implementations indicated no obvious differences in the GEBV accuracy for the different GS methods for most traits [38,41]. An exception was fat percentage for which the accuracy was higher for methods with informative prior, (e.g. BayesB) compared to the methods with flat prior, (e.g. GBLUP) [9,41]. A reasonable explanation is that the bovine *DGAT1* gene located on chromosome 14 explains a considerable proportion of the genetic variance for fat percentage. This would result in an obvious deviation from the infinitesimal model for this trait which might favor BayesB. The accuracy of GBLUP is higher than the variable selection methods when the reference population is small [42], but decays quickly when the relationship between the reference and candidate populations becomes weaker [42,47]. Results from within and between population comparisons showed that the larger the reference population, the more accurate the GEBV becomes [41,43]. Accuracy decreases as marker density [43], but is less sensitive to marker density than to the size of the reference population [41].

Proven bulls were used to form the reference populations in all the countries mentioned above. The highly reliable EBVs (average reliability > 0.9) of the bulls that were employed to estimate the marker effects greatly improved the performance of GS. In China, a project aimed at implementing GS in a Chinese Holstein cattle population has been running since 2008. A reference population with about 2100 milking cows has been genotyped with the Illumina

Table 1   Summary of the implementation of genomic selection in dairy cattle from the Interbull survey, 2009

Country (Company)	Reference population[a]	No. markers (panel)	Method
USA [41]	18080 (H), 2828(J) 1331 (B)	43382 (Illumina 54 K)	GBLUP, nonlinear-BLUP
Canada [41]	8800 (H)	43382 (Illumina 54 K)	GBLUP, nonlinear-BLUP
Australia [9]	1098 (H)	39048 (Illumina 54 K)	BayesA, BayesB, SSVS, BLUP
Australia [38]	1239	7372 (Affy 25K)	FR-LS, RRBLUP, BayesS, SVR, PLSR
Australia [42]	781 (H), 287 (J)	39048 (Illumina 54 K)	GBLUP, BayesA, SSVS
Norway [43]	500 (R)	18991 (Affymetrix 25 K)	RRBLUP, BayesB, BayesC
New Zealand [44]	2626(H), 1639 (J), 642 (HxJ)	44146 (Illumina 54 K)	GBLUP
Netherland [45]	16173 (H)	46529 (CRV Illumina 60 K)	Bayes
Denmark/Sweden/Finland [46]	10217 (H)	38055 (Illumina 54 K)	SSVS, BLUP
France	20918 (H), 6835 (M), 4970 (N)	(Illumina 54 K)	–
Germany [47]	17477 (H)	45181 (Illumina 54 K)	GBLUP, BayesB, BLUP
Ireland [48]	4300 (H)	42598 (Illumina 54 K)	GBLUP

a) Swedish brown cattle (B), Holstein (H), Jersey (J), Montbeliarde (M), Normande (N), Norway red cattle (R).

BovineSNP50 BeadChip. Preliminary results showed that the GEBV accuracy for milk production traits ranged from 0.6 to 0.75 (unpublished data).

4.2   GS in other livestock species

Although the theoretical issues of GS in swine [50,51], sheep [52,53], and chicken [54−57] have been discussed, and HD SNP panels for swine (Illumina PorcineSNP60), sheep (Illumina OvineSNP50), horse (Illumina Equine SNP50), and chicken (Illumina iSelect 18K Custom genotype) are now available, the reported implementation of GS in these species is still very limited. In sheep, a theoretical calculation revealed that, compared with traditional selection methods, GS could increase the overall response for a terminal sire index by about 30% and for a fine wool merino index by about 40% . Daetwyler et al. [58] reported that using a multi-breed reference population with 7180 sheep, the accuracies of GEBV ranged from 0.15 to 0.79 for wool traits in Merino sheep and from −0.07 to 0.57 for meat traits. In broilers, Gonzalez-Recio et al. [57] found that GS could increase the accuracy of selection for food conversion rates up to an almost 4-fold relative to a pedigree index. Applying GS in layers has been shown to increase the accuracy of selection by up to two-fold at an early age and by up to 88% when applied again at a later age [56].

An important reason for the limited application on GS in these species is that the cost of GS is relatively high when compared with GS in dairy cattle. Consequently, to reduce breeding cost, a low density panel, which might be more cost effective, was suggested to replace HD panels. Using simulations, Zhang et al. [36] concluded that with only a small proportion of the selected markers, 95% of accuracy obtained from HD panels could be obtained. Further, they found that a low density panel with markers pre-selected based on their estimated effects performs better than a panel with evenly spaced markers. The feasibility and performance of GS with low density panels has been investigated in chicken [55] and cattle [59] populations.

5   Impact of genomic selection on traditional breeding systems

5.1   Effect on current breeding schemes

In traditional genetic evaluation systems, the relationships between individuals are inferred through pedigree, while in GS the relationship can be inferred using whole-genome markers. This may have an influence on the livestock breeding system. In swine and poultry, the breeding system is usually composed of 3 levels of herds: breeding herds, multipliers and production herds. The performance records of hybrid animals in production herds are seldom used for evaluating the animals in breeding herds because no pedigrees are available to set up the relationships between animals in the production and breeding herds. However, the connection between the animals from different levels of herds can be easily built using HD markers and making use of information from as many different levels of herds as possible [60]. In dairy cattle, it usually takes 6 or more years before a tested bull can be used for breeding in the progeny testing breeding system. Such a breeding scheme is not cost effective. With GS, early stage selection of young bulls based on their GEBVs can be conducted, resulting in a shortened generation interval and decreased breeding costs [61].

5.2   Joint breeding of multiple populations and multiple breeds

To evaluate the reliability of genomic prediction across multiple populations, de Roos et al. [62] simulated reference populations composed of two cattle populations that

diverged 6, 30, and 300 generations ago. Toosi et al. [63] and Ibánz-Escriche et al. [60] investigated GS in admixed and crossed populations by simulating purebred and F_1, F_2, 3- and 4-way crosses. Their results indicated that GS across multiple populations or multiple breeds is feasible. However, a sufficient number of HD markers are required for the marker–QTL LD to be consistent across populations. GS can be conducted in crossbred populations and models that fit the breed-specific effects of SNP alleles may not be required, especially with HD markers [60]. Kizilkaya et al. [64] carried out a multiple breeds GS investigation with simulated phenotypes and real 50K SNP genotypic data. Hayes et al. [42] investigated the accuracies of GEBVs in Holstein and Jersey dairy cattle with the reference population consisting of a single breed or multi-breed. They reported that a multi-breed reference population performed better than a single-breed reference population. A similar tendency was also found by them in simulation studies.

5.3   Cooperation of breeding organizations

Because it is different from traditional breeding systems, GS makes the cooperation of breeding organizations feasible. The cost of genotyping is still very high, making it hard for a single breeding organization to build the large reference population that is essential to ensure high accuracy of GS. It is plausible that breeding organizations may agree to collaborate with each other. At the moment, the ongoing joint GS project in North America [41] and Europe (EuroGenomics) [65] has demonstrated the benefits of such a collaboration. Even in a joint project, each breeding organization can maintain the independency and security of its genotyping information because the original SNP genotype information in a G matrix is invisible [66]. A new technical platform to integrate data from different sources is being built by Interbull (http://www.interbull.org/). This platform should allow all participants to benefit from this project without worrying about the loss of commercial secrets.

Moreover, Daetwyler et al. [24] reported that GS can simultaneously slow down the increase of inbreeding while improving the accuracy of selection in the comparison with the traditional BLUP selection. The main reason for this is that the differences between sibs are enlarged as the Mendelian sampling deviation is taken into account in GS. This will decrease the probability of sibs being selected together, and thus decrease inbreeding.

6   Challenges

Although comprehensive studies on GS have been carried out and its successful implementation in breeding practices have been reported, this is a new field and challenges still exist.

6.1   Computational challenges

Higher density SNP panels like the bovine 800K SNP panel designed by Illumina (780000 SNPs) are now available and, almost simultaneously, new high-throughput genotyping technology has been developed. Recently, the next generation sequencing technology has been used to genetic analysis [67,68]. More SNP markers will become available if this fast and cheap genotyping technology is applied to livestock breeding. However, with increasing numbers of SNP markers, the computational demands increase tremendously. For example, if a 800K SNP panel was used in place of the 54K SNP panel for GS, the memory requirement for RRBLUP and the computing time for the Bayesian methods would increase by about 200 and 15 times, respectively. In practice, such increases make the use of the larger panels infeasible.

In practice, multiple trait phenotypes and repeated measures are routinely recorded. These records are useful and the data that they contain are perfectly fitted for use in the traditional BLUP model to evaluate the genetic merit of each selection candidate using a repeatability model, a random regression model, or a multi-trait model. However, for the present GS models and especially for Model (1), only single trait and single record data can be fitted. GS using repeated measures and multiple traits deserves further investigation.

6.2   Marker-derived relationship matrix

Although the dimension of the equations related to the genetic effects can be reduced from m to n by using the direct GS methods based on Model (2), challenges still exist in the construction of a marker-derived relationship matrix and in calculating its inversion. The numerator relationship matrix is a sparse matrix and its inverse can be easily constructed from a pedigree. The marker-derived relationship matrix, on the other hand, is a dense matrix and the rules for constructing its inverse are still unknown. New rules or parallel arithmetic might be important in this regard.

6.3   Other challenges

Genomic selection has been widely implemented in national and international genetic evaluation in the dairy cattle industry. However, breeding individuals for other species (in particular males) is not as economically important as the breeding of dairy bulls. The high price of genotyping and the cost of HD panels (around US$200 per sample) may mean that GS is not cost effective in other livestock species. Hence, the implementation of GS in these species is likely to be delayed. The availability of low density SNP panels would reduce costs and may enable other species to benefit from the promising GS method [31,36,69]. Until the cost of genotyping with HD panel becomes low enough to make it profitable enough to be used by breeding organizations, GS

using low density panels could be a good option.

Another challenge to the application of GS in species other than dairy cattle is the possible lack of highly reliable EBVs. In dairy cattle, bulls with highly reliable EBVs can be used to construct accurate GEBVs where the accuracies can be calculated as the correlations between GEBVs and the EBVs. In other species, such data for the males may not be available, and, therefore, a larger reference population will be required to obtain the desired accuracy of GEBVs. Furthermore, the performance of GS in these populations cannot be directly validated.

1　Goddard M E, Hayes B J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet, 2009, 10: 381–391

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

3　Heffner E L, Sorrells M E, Jannink J L. Genomic selection for crop improvement. Crop Sci, 2009, 49: 1–12

4　Sonesson A K, Meuwissen T H. Testing strategies for genomic selection in aquaculture breeding programs. Genet Sel Evol, 2009, 41: 37

5　Habier D. More than a third of the WCGALP presentations on genomic selection. J Anim Breed Genet, 2010, 127: 336–337

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

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

8　Calus M P L. Genomic breeding value prediction: Methods and procedures. Animal, 2010, 4: 157–164

9　Verbyla K L, Hayes B J, Bowman P J, et al. Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. Genet Res, 2009, 91: 307–311

10　Meuwissen T H. Accuracy of breeding values of 'nrelated' individuals predicted by dense SNP genotyping. Genet Sel Evol, 2009, 41: 35

11　Meuwissen T H, Solberg T R, Shepherd R, et al. A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. Genet Sel Evol, 2009, 41: 2

12　Long N, Gianola D, Rosa G J, et al. Machine learning classification procedure for selecting SNPs in genomic selection: Application to early mortality in broilers. J Anim Breed Genet, 2007, 124: 377–389

13　Solberg T R, Sonesson A K, Woolliams J A, et al. Reducing dimensionality for prediction of genome-wide breeding values. Genet Sel Evol, 2009, 41: 29

14　Gianola D, Fernando R L, Stella A. Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics, 2006, 173: 1761–1776

15　Bennewitz J, Solberg T, Meuwissen T. Genomic breeding value estimation using nonparametric additive regression models. Genet Sel Evol, 2009, 41: 20

16　de los Campos G, Naya H, Gianola D, et al. Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics, 2009, 182: 375–385

17　VanRaden P M. Efficient methods to compute genomic predictions. J Dairy Sci, 2008, 91: 4414–4423

18　Visscher P M, Medland S E, Ferreira M A, et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet, 2006, 2: 316–325

19　VanRaden P M, Tooker M E. Methods to explain genomic estimates

20　Habier D, Fernando R L, Dekkers J C M. The impact of genetic relationship information on genome-assisted breeding values. Genetics, 2007, 177: 2389–2397

21　Strandén I, Garrick D J. Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. J Dairy Sci, 2009, 92: 2971–2975

22　Goddard M E. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica, 2009, 136: 245–257

23　Hayes B J, Visscher P M, Goddard M E. Increased accuracy of artificial selection by using the realized relationship matrix. Genet Res, 2009, 91: 47–60

24　Daetwyler H D, Villanueva B, Bijma P, et al. Inbreeding in genome-wide selection. J Anim Breed Genet, 2007, 124: 369–376

25　Muir W M. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J Anim Breed Genet, 2007, 124: 342–355

26　Calus M P, Meuwissen T H, de Roos A P, et al. Accuracy of genomic selection using different methods to define haplotypes. Genetics, 2008, 178: 553–561

27　Zhang Z, Ding X D, Liu J F, et al. TA-BLUP: A New Genetic Evaluation Method for Genomic Selectionon. In: 9th World Congress on Genetics Applied to Livestock Production, 2010. Leipzig, Germany. 215

28　Zhang Z, Liu J F, Ding X D, et al. Best linear unbiased prediction of genomic breeding values using trait-specific marker-derived relationship matrix. PLoS One, 2010, 5: e12648

29　Solberg T R, Sonesson A K, Woolliams J A, et al. Genomic selection using different marker types and densities. J Anim Sci, 2008, 86: 2447–2454

30　Calus M P L, Veerkamp R F. 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. J Anim Breed Genet, 2007, 124: 362–368

31　Habier D, Fernando R L, Dekkers J C M. Genomic selection using low-density marker panels. Genetics, 2009, 182: 343–353

32　Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. J Dairy Sci, 2009, 92: 4656–4663

33　Chen C Y, Misztal I, Aguilar I, et al. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. J Anim Sci, 2011, 89: 23–28

34　Daetwyler H D, Villanueva B, Woolliams J A. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One, 2008, 3: e3395

35　Daetwyler H D, Pong-Wong R, Villanueva B, et al. The impact of genetic architecture on genome-wide evaluation methods. Genetics, 2010, 185: 1021–1031

36　Zhang Z, Ding X D, Liu J F, et al. Accuracy of genomic prediction using low density marker panels. J Dairy Sci, 2011, Doi: 10.3168/jds.2010-3917

37　Sved J A. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor Popul Biol, 1971, 2: 125–141

38　Moser G, Tier B, Crump R E, et al. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genet Sel Evol, 2009, 41: 56

39　Schaeffer L R. Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet, 2006, 123: 218–223

40　Hayes B J, Bowman P J, Chamberlain A J, et al. Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci, 2009, 92: 433–443

41　VanRaden P M, Van Tassell C P, Wiggans G R, et al. Reliability of genomic predictions for North American Holstein bulls. J Dairy Sci, 2009, 92: 16–24

42　Hayes B, Bowman P, Chamberlain A, et al. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet Sel Evol, 2009, 41: 51

43　Luan T, Woolliams J A, Lien S, et al. The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. Genetics, 2009, 183: 1119–1126

44 Harris B L, Montgomerie W A. Current status of the use of genomic information in the national genetic evaluation in New Zealand. In: Proc of the Interbull International Workshop in Genomic Information in Genetic Evaluations, 2009. Interbull: Uppasla, Sweden. 35–38

45 de Roos A P W, Schrooten C, Mullaart E, et al. Genomic selection at CRV. Interbull Bull, 2009, 39: 47–50

46 Lund M S, Su G. Genomic selection in the Nordic countries. Interbull bulletin, 2009, 39: 39–42

47 Habier D, Tetens J, Seefried F R, et al. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet Sel Evol, 2010, 42: 5

48 Berry D, Kearney F, Harris B. Genomic selection in Ireland. Interbull Bulletin, 2009, 39: 29–34

49 Loberg A, Dürr J W. Interbull survey on the use of genomic information. Interbull Bull, 2009, 39: 3–14

50 Tuggle C K, Dekkers J C M. Genotyping: How useful is it for producers? Pig Prog, 2009, 25: 9

51 Simianer H. The potential of genomic selection to improve litter size in pig breeding programmes. In: European Association for Animal Production 2009 Annual Meetings, 2009. Barcelona, Spain

52 van der Werf J H J. Potential benefit of genomic selection in sheep. In: The Association for the Advancement of Animal Breeding and Genetics 18th Conference, 2009. Barossa Valley, South Australia. 38–41

53 McEwan J C. Current and future impact of DNA technologies on the New Zealand sheep industry. In: Proceedings of the New Zealand Society of Animal Production, 2009. Canterbury, New Zealand. 165–168

54 Chen C Y, Misztal I, Aguilar I, et al., Genome wide marker assisted selection in chicken: Making the most of all data, pedigree, phenotypic, and genomic in a simple one step procedure. In: 9th World Congress on Genetics Applied to Livestock Production, 2010. Leipzig, Germany

55 Legarra A, Calenge F, Mariani P, et al. Use of a reduced set of single nucleotide polymorphisms for genetic evaluation of resistance to Salmonella carrier state in laying hens. Poul Sci, 2011, 90: 731–736

56 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. Genet Sel Evol, 2011, 43: 5

57 Gonzalez-Recio O, Gianola D, Rosa G J, et al. Genome-assisted prediction of a quantitative trait measured in parents and progeny: Application to food conversion rate in chickens. Genet Sel Evol, 2009, 41: 3

58 Daetwyler H D, Hickey J M, Henshall J M, et al. Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. Anim Prod Sci, 2010, 50: 1004–1010

59 Weigel K A, de los Campos G, Vazquez A I, et al. Accuracy of direct genomic values derived from imputed single nucleotide polymorphism genotypes in Jersey cattle. J Dairy Sci, 2010, 93: 5423–5435

60 Ibánz-Escriche N, Fernando R L, Toosi A, et al. Genomic selection of purebreds for crossbred performance. Genet Sel Evol, 2009, 41: 12

61 König S, Simianer H, Willam A. Economic evaluation of genomic breeding programs. J Dairy Sci, 2009, 92: 382–391

62 de Roos A P, Hayes B J, Goddard M E. Reliability of genomic predictions across multiple populations. Genetics, 2009, 183: 1545–1553

63 Toosi A, Fernando R L, Dekkers J C. Genomic selection in admixed and crossbred populations. J Anim Sci, 2010, 88: 32–46

64 Kizilkaya K, Fernando R L, Garrick D J. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. J Anim Sci, 2010, 88: 544–551

65 Lund M S, Roos A P W D, Vries A G D, et al. Improving genomic prediction by EuroGenomics collaboration. In: 9th World Conference of Genetics Applied on Livestock Production, 2010. Leipzig, Germany. 880

66 Harris B L, Johnson D L, Montgomerie W A. National genomic evaluations without genotypes. Interbull Bull, 2009, 40: 189–192

67 Tian F, Bradbury P J, Brown P J, et al. Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet, 2011, 43: 159–162

68 Huang X H, Wei X H, Sang T, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet, 2010, 42: 961–967

69 Weigel K A, de los Campos G, Gonzalez-Recio O, et al. Predictive ability of direct genomic values for lifetime net merit of Holstein sires using selected subsets of single nucleotide polymorphism markers. J Dairy Sci, 2009, 92: 5248–5257