

www.stmjournals.com

Application of Genomic Selection in Livestock Improvement

Ajoy Mandal, S. K. Das, M. Karunakaran, C. Bhakat, T. K. Dutta* Eastern Regional Station, National Dairy Research Institute, Kalyani-741235, Nadia, West Bengal, India

Abstract

Livestock contributes significantly to the livelihood of rural poor in our country and has enormous potential to reduce poverty. There is a predicted increase in demand for animal food products in India by 2020 and a rapidly growing market is also coming up. Livestock revolution will provide a better opportunity for poor farmers for benefiting from the rapidly growing market and thereafter providing a better food security for our country's growth. Better livestock production efficiency can be obtained through improving animals genetically, which ultimately leads to enhancement of productivity of the production system in a sustainable manner. Genetic improvement of livestock through traditional selection for increasing livestock productivity has some major limitations. To overcome these limitations, genomic selection (GS) played a crucial role in livestock industry. GS, a marker-assisted selection on a genome-wide scale, uses information on variation in DNA sequences between animals to predict the breeding value of animals more accurately. Moreover, this method provides an accurate assessment of breeding values of animals without measuring the traits at early stages of life. Furthermore, the information is available not just for a single gene or trait, but also for all the genes and all traits at the same time. This is a DNA technology in breeding with true breakthrough potential. Through genomic selection, genetic progress can be increased dramatically.

Keywords: Food security, livestock production, genomic selection, breeding

*Author for Correspondence E-mail: ajoymandal@rediffmail.com

INTRODUCTION

Traditional genetic improvement of livestock is generally done through the selection of superior animals for economically important traits and this selection is mainly based on the phenotypic records of the individual and its relatives to predict the breeding values of animals. Estimated breeding values based on this phenotypic data are commonly calculated by the best linear unbiased prediction (BLUP) [1]. However, the molecular genetics research on livestock helps to utilize the information at DNA level which leads to faster genetic gain than that achieved based on phenotypic data only. Breeding value of animals can be predicted more accurately by using information on variation in DNA sequences between animals. Although the inclusion of marker information into BLUP breeding values is predicted to improve the genetic gain (approx. 8-38% extra) than gain obtained through using the phenotypic data in dairy animals [2], but the implementation of markerassisted selection (MAS) has been limited because of many quantitative traits, such as production and health traits etc., in dairy animals, a large number of loci affect the trait, with any one locus capturing only a limited proportion of the total genetic variance [3, 4]. Consequently, relatively small gains are possible with the limited number of markers available, and the cost of genotyping these markers is high. The complexity of calculating breeding values including marker information is a further barrier to the application of marker assisted selection. To overcome this problem, a new technology called genomic selection has revolutionized the dairy cattle breeding program.

Genomic selection (GS)

Genomic selection may be defined as the simultaneous selection for many (tens or hundreds of thousands) markers, which cover the entire genome in a dense manner so that all genes are expected to be in linkage disequilibrium with at least some of the markers. In a sense, GS is marker-assisted selection on a genome-wide scale. In other words, it simply implies that the selection of animals based on the sum of all haplotypes effects is determined from the mean of the population. Though the methodology for genomic selection was first presented by Meuwissen et al. [5], but the term genomic selection was first introduced by Haley and Visscher at Armidale WCGALP in 1998.

Methodology to Calculate GS

Genomic selection refers to selection decisions based on genomic breeding values (GEBV). The underlying assumption of genomic selection is that haplotypes at some loci are in linkage disequilibrium (LD) with quantitative trait loci (QTL) alleles that affect the traits that are subject to selection. To calculate GEBV, first a prediction equation based on a large number of DNA markers, such as SNP (single polymorphisms) markers, nucleotide is derived. The entire genome is divided into small segments and the effects of these markers are estimated in a reference population in which animals are both phenotyped and genotyped. In this way, the effects of all loci that contribute to genetic variation are captured, even if the effects of the individual loci are very small. In subsequent generations, animals can be genotyped for the markers to determine which chromosome segments they carry, and the estimated effects of the segments the animal carries can then be summed across the whole genome to predict the GEBV. Thus, the effect of the chromosome segments they carry can be summed across the genome:

 $GEBV = \sum_{i}^{n} \hat{g}_{i}$

where *n* is the number of chromosome segments across the genome, Xi is a design matrix allocating animals to the haplotype effects at segment *i*, and \hat{g}_i is the vector of

effects of the haplotypes within chromosome segment *i*.

Genomic selection can proceed using single markers [6], haplotypes of markers or using the derive identical-by-descent (IBD) [7] approach. A number of approaches have been proposed for estimating the single marker or haplotype effects across chromosome segment effects for genomic selection. A key difference between these approaches is the assumption they make about the variances of haplotype or single marker effects across chromosome segments. Different methods have been applied for the calculation of genomic estimated breeding values (GEBV), ranging from least-squares, BLUP [8] Bayesian methods such as BayesB [7], and machine learning techniques [9]. However, among these methods, Bayes obtained the highest accuracy of selection which is comparable to that of progeny testing.

Implementation of Genomic Selection

Genomic selection in breeding programs requires a discovery dataset where a large number of single nucleotide polymorphism (SNP) are to be assayed on a moderate number of animals that have phenotypes for all the relevant traits. A prediction equation that uses markers as input and predicts breeding value (BV) is to be derived from this data. There would then be a validation sample where a larger number of animals are recorded for the traits and genotyped at least for the markers that are proposed to be used commercially. The prediction equation is tested to assess its accuracy on this independent sample. Then selection candidates are genotyped for the markers and the prediction equation estimated in the discovery data is to be used to calculate GEBV. The combined discovery and validation datasets is considered as the "reference" population. Hence, in brief, to implement genomic selection, a reference population is needed. The granddaughter design as it is commonly used in dairy cattle QTL (quantitative trait loci) mapping studies fits the purpose. Once the animals in the reference population are typed for the markers, genomic selection can immediately be applied to all traits for which estimated breeding values are available.



In traditional marker-assisted selection, we relied on individual loci that were identified significantly using a genome-wide significance test and the total amount of variance explained by the QTL identified was limited. In genomic selection, we immediately explain most of the variance for all of the traits, without intermediate time-consuming QTL hunting.

The major limitation to the implementation of genomic selection has been the large number of markers required and the cost of genotyping these markers. Recently, both these limitations have been overcome in most livestock species following the sequencing of the livestock genomes, the subsequent availability of hundreds of thousands of single nucleotide polymorphisms (SNP), and dramatic developments in SNP genotyping technology which allow genotyping an SNP for as little as 1 US cent per animal.

Factors affecting the accuracy of genomic selection

The accuracy of genomic selection depends on following several factors:

(i) Number of markers required based on the extent of linkage disequilibrium (LD): The haplotypes or single markers must be in sufficient LD with the QTL such that the haplotype or single markers will predict the effects of the QTL across the population. For genomic selection to be successful, the level of LD between adjacent markers should be $r^2 > = 0.2$ as observed by Ref. [7].

(ii) Number of phenotypic records required in the initial experiment: The accuracy of genomic selection depends on the number of haplotype effects at the chromosome segments, and the number of phenotypic records per unique haplotype, or per marker allele if single markers are used. The more phenotypic records available, the more observations there will be per haplotype and the higher the accuracy of genomic selection.

(iii) Role of non-additive effects: Generally breeding values by definition should include only additive effects; in some cases it may be desirable to predict genetic merit which better predict an animal's actual phenotype; for example, through the inclusion of dominance and epistatic effects. If phenotypes are used in the estimation of chromosome segment effects, inclusion of epistatic and dominance effects in the model could improve the accuracy of estimating the additive effect of the chromosome segment effects. Further, dominance and epistatic effects can be exploited to produce sets of progeny with maximum genetic merit, through mate selection [10].

Traditional Selection vs. Genomic Selection

Traditional selection method successfully improves a large number of traits in animal breeding schemes, but it requires widespread (all selection candidates), reasonably accurate and preferably early in life recording of the traits. In traditional breeding, the accuracy of selection increased with the traits having medium to high heritability and when the selection is carried out based on animal's own record or progeny data. But traditional breeding becomes less effective when traits are lowly heritable and only expresses in one sex (sex-limited traits) or can only be measured later in life and for selection of young animals having no records or no progeny. Genomic selection may overcome these problems. Genomic selection could be used to predict total breeding values for juvenile selection candidates that are genotyped for the markers have no phenotypic information. but Moreover, this breeding methodology could be successfully applied for lowly heritable traits provided large sets of data are available. The GS may be able to considerably decrease the cost of animal breeding program by shortening the generation interval, increasing the accuracy of selection at younger age and may even by replacing the traditional testing of progeny and sibs. The accuracy/reliability obtained through the traditional PTA (predicted transmitting ability) and genomic PTA of different economic traits has been presented in Table 1 [7] reported that inclusion of genomic information in breeding scheme yields a considerable increase in selection responses for juvenile animals that do not have phenotypic records as compared to traditional breeding practice and potentially can reduce the costs of a breeding program up to 92% [11].

In traditional selection methods, such as sib and best linear unbiased prediction (BLUP) selection, which increased genetic gain by increasing the accuracy of evaluation, have also led to an increased rate of inbreeding per generation. On the contrary, genome-wide selection utilizes the Mendelian sampling term and increases genetic gain by increasing accuracy of evaluation without increasing the inbreeding rate per generation. As the genomic selection has the potential to achieve a more balanced selection response (balanced over production and functional traits), and substantially reduced inbreeding rates, so its main effect on future breeding schemes may be an increased sustainability. Moreover, the ability to predict the total genetic merit of livestock using molecular markers would allow the opportunity to completely redesign animal breeding and management programs.

 Table 1: Comparison of Reliability between Traditional PTA and Genomic PTA for Different Traits of Dairy Cattle.

2 an y cannot			
Traits	Realized reliability		
	Traditional PTA	Genomic PTA	Change
Milk yield	35%	58%	+23%
Fat yield	35%	68%	+ 33%
Protein yield	35%	57%	+22%
Fat %	35%	78%	+ 43%
Protein %	35%	69%	+ 34%
Productive life	27%	45%	+18%
Somatic cell score	30%	51%	+21%
Daughter fertility	25%	41%	+ 16
Net merit	30%	53%	+ 23%

(Source: VanRaden [12]).

Advantage of Genomic Selection

The genomic selection has the following advantages in livestock breeding:

- a. It increases the genetic gain by increasing accuracy of selection. If the parent average is 40%, then the genomic breeding value will be higher.
- b. GS reduces the generation interval of animals and thus one can select the animals before they are of productive and/or reproductive age and it ultimately reduces the cost of animal rearing by reducing/eliminating the need for progeny testing.
- c. It lowers the rate of inbreeding per generation [13] and thus one can move from family selection to individual selection
- d. Once marker effects are estimated, they can be used for a few generations, but the accuracy will reduce in each generation.
- e. Selection can be made on novel traits and expensive phenotyping traits.
- f. GS helps to formulate the new breeding strategies in livestock inbreeding.

Disadvantages/Limitations of Genomic Selection

There are some disadvantages/limitations in implementing the genomic selection in animal breeding:

- a. GS is a somewhat newer selection method in animal breeding which is not yet fully proven and tested.
- b. In this method, there is a need to genotype a sufficiently large set of animals for accurate marker estimates. So, a larger number of markers are required for genomic selection. Moreover, if the trait is lowly heritable, then more records are needed for genomic selection.
- c. Marker estimates must be estimated in population that they will be used in as across-breed accuracy is low in this methodology.
- d. Genotyping of animals is still costly to afford for many breeding industries.
- e. Some species do not have dense marker maps yet, required for genomic selection approach.
- f. When generation intervals of animals are already low, then genetic gain due to genomic selection will be less.



g. In large litters, when the accuracy can be gained from information on sibs, then GS will be less advantageous.

Potential Application/Benefits of Genomic Selection in Dairying

Genomic selection has the potential to radically alter the structure of livestock breeding programs. It can be used initially to select young bulls for progeny testing. If it is successful, then sires of sons and sires of replacement dams will be selected based on genetic markers and formal progeny testing will disappear, which will potentially reduce the rearing cost of the animals. In breeding programs without extensive recording, it is more important to rely on direct markers and when the genes have large economic effects, the genomic selection approach will be really useful. As the whole genome is analyzed simultaneously in this selection method so. there is no need for OTL or gene identification. This method yields with high accuracies of estimated breeding values (EBV) based on genotypic information in newborn without phenotypic records. individuals Moreover, this high accuracy could then be maintained, with only minor loss, over subsequent generations when neither offspring nor parent had records. Pre-selecting young dairy sires with MAS increased genetic gain and offered a method to select within families. Thus, the genome-wide prediction offers the possibility that an individual's Mendelian sampling term can be estimated with great accuracy early in its life. In a study of Meuwissen et al. [5], it was observed that the accuracy of EBV derived from genome-wide sense marker maps could be as high as 0.80 at birth for moderately heritable traits. Furthermore, because of the greater accuracy of EBVs, greater selection intensity of germplasm can be adopted which, when coupled with the higher accuracy of selection and reduced generation intervals, results in increased genetic gains and reduced costs associated with progeny testing of young sires. This will lead to the use of reproductive technology to decrease the age at first breeding and increase number of offspring (e.g., juvenile in vitro embryo transfer). Hence, genome-wide selection has greater potential than nucleus, multiple ovulation and embryo transfer (MOET), or marker-assisted schemes

for making genetic change. Costs of genotyping are also likely to decrease over time, which would make genome-wide selection more affordable to implement in many developing countries.

Relevancy of Implementation of GS in Indian Context

Developing countries are often limited by the absence of programs that record phenotypes on pedigreed animals and the lack of evaluation or national testing programs to assess the genetic value of germplasms. But in some developing countries like India there are many independent breeding units where trait recording and genetic evaluation programs are in practice. Sire evaluation programs for traits of interest have taken place through sib- and Considering the long BLUP selection. generation interval, the high value of each animal, the fact that nearly all economic traits are expressed in females, it would seem that inclusion of marker information in livestock breeding program should be an ideal option. The additional value of gene markers will be greatest in breeding programs that already use intensive pedigree and performance recording, and it will help to shift the selection pressure towards traits that are hard to improve based on phenotypic (BLUP) selection (i.e., traits such as fertility, disease resistance and carcass quality). In India, there will be DNA marker data as well as phenotypes and pedigrees on potential selection candidates in the near future. It might be desirable to combine all this data to estimate improved EBV. But the amount of marker data available on each animal will be highly variable; with most animals having none is the major difficulty in this country. However, to overcome these difficulties, different methods can be adopted to calculate EBV using such data. Though at present the cost of genotyping is high to afford for genomic selection but it may likely decrease over the time, which would make genome-wide selection more affordable to implement in our country. Moreover, genomic approaches should help in identifying critical populations for preservation together with some local well-adapted breeds that could be further utilized to breed valuable animals through a combination of selection and crossbreeding. Of course, as with genomics, one can manage only what one can measure, and collecting a minimum number of phenotypes in the field will remain one of the critical and challenging steps to further deployment of genomic selection in developing countries.

REFERENCES

- 1. CR Henderson. *Applications of Linear Models in Animal Breeding*. Univ. Guelph, Guelph, Ontario, Canada: 1984.
- THE Meuwissen, ME Goddard. The use of marker haplotypes in animal breeding schemes. *Genet. Select. Evol.* 1996; 28:161–176p.
- 3. BJ Hayes, ME Goddard. The distribution of the effects of genes effecting quantitative traits in livestock. *Genet. Select. Evol.* 2001; 33:209–229p.
- PM VanRaden, CP Van Tassell, GR Wiggans, et al. Invited review: Reliability of genomic predictions for North American Holstein Bulls. J Dairy Sci. 2009; 92:16–24p.
- 5. THE Meuwissen, BJ Hayes, ME Goddard. Prediction of total genetic value using genome wide dense marker maps. *Genetics* 2001; 157:1819–1829p.
- 6. S Xu. Estimating polygenic effects using markers of the entire genome. *Genetics* 2003; 163:789–801p.

- THE Meuwissen, ME Goddard. Prediction of identity by descent probabilities from marker-haplotypes. *Genet. Sel. Evol.* 2001; 33:605–634p.
- 8. D Kolbehdari, LR Schaeffer, JAB Robinson. Estimation of genome-wide haplotype effects in half-sib designs. J. Anim. Breed. Genet. 2007; 124:356–361p.
- N Long, D Gianola, GJM Rosa, 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– 389p.
- 10. BP Kinghorn. Mate selection by groups. J Dairy Sci. 1998; 81(2): 55–63p.
- 11. LR Schaeffer. Strategy for applying genome-wide selection in dairy cattle. J. *Anim. Breed. Genet.* 2006; 123:218–223p.
- 12. PM VanRaden. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 2008; 99:4414–4423p.
- HD Daetwyler, B Villanueva, P Bijma, et al. Inbreeding in genome-wide selection. *J. Anim. Breed. Genet.* 2007; 124:369–376p.