The Pharma Innovation Journal 2017; 6(12): 241-245

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating 2017: 5.03
TPI 2017; 6(12): 241-245

www.thepharmajournal.com
Received: 20-10-2017
Accepted: 21-11-2017

Priyanka Priyadarshini
Department of Animal Breeding and Genetics, Chhattisgarh Kamdhenu Vishwavidyalaya, Anjora, Durg, Chhattisgarh, India

Dr. Chinmoy Mishra
Department of Animal Breeding and Genetics, C.V.Sc & A.H. Orissa University of Agriculture and Technology, Bhubaneswar, Odisha, India

Dr. Gangadhar
Department of Animal Breeding and Genetics, C.V.Sc & A.H. Orissa University of Agriculture and Technology, Bhubaneswar, Odisha, India

Krutanjali Swain
Department of Veterinary Parasitology, Chhattisgarh Kamdhenu Vishwavidyalaya, Anjora, Durg, Chhattisgarh, India

Sumitra Panigrahi
Department of Veterinary Public Health and Epidemiology, LUVAS, Hisar, Haryana, India

Correspondence
Priyanka Prakash Padhi
Department of Animal Breeding and Genetics, Chhattisgarh Kamdhenu Vishwavidyalaya, Anjora, Durg, Chhattisgarh, India

Genomic selection: Beginning of a new era in animal breeding

Priyanka Priyadarshini, Dr. Chinmoy Mishra, Dr. Gangadhar Nayak, Krutanjali Swain and Sumitra Panigrahi

Abstract
Traditional marker-assisted selection (MAS) did not result in a widespread use of DNA information in animal breeding. The main reason was that the traits of interest in livestock production were much more complex than expected: they were determined by thousands of genes with small effects on phenotype. These effects were usually too small to be statistically significant and so were ignored. Genomic selection (GS) assumes that all markers might be linked to a gene affecting the trait and concentrates on estimating their effect rather than testing its significance. Three technological breakthroughs resulted in the current wide-spread use of DNA information in animal breeding: the development of the genomic selection technology, the discovery of massive numbers of genetic markers (single nucleotide polymorphisms; SNPs), and high-throughput technology to genotype animals for (hundreds of) thousands of SNPs in a cost-effective manner. This selection method can be used in genetic improvement and genomic prediction in livestock and poultry breeding.

Keywords: livestock, MAS, selection, SNP, trait

1. Introduction
Traditionally in animal breeding, for improvement of economic traits, selection was done on the basis of phenotypic recordings. Best Linear Unbiased Prediction (BLUP), combined individual records and data from relatives to estimate Expected Breeding Value (EBV). From 1990 onwards due to advances in molecular genetics it was known that information at DNA level would lead to more genetic improvement than using only phenotypic records. This lead to development of Marker Assisted Selection (MAS). MAS included mainly two steps; 1) detect and map the genes controlling the trait of interest, i.e. so called Quantitative Trait Loci (QTL); 2) include QTL information into BLUP-EBV [7].

The QTL mapping step (1) was successful in the scene that most mapping studies detected QTL. But the repeatability of the mapping studied was low, i.e. QTL positions moved/(dis)appeared from one study to the next. One reason for this is that the majority of QTL have very small effects. When this is combined with testing a large number of markers, there is a marked "Beavis effect" in which the estimated effect of significant markers is overestimated [1]. Genome Wide Association Studies (GWAS) typically focus on association between Single Nucleotide Polymorphisms (SNPs) and the trait of interest.

These tests can be done only when the largest QTL were known. Large QTL were detected for some traits e.g., DGAT1 affecting fat content in milk [12], CDH1 affecting infectious pancreatic necrosis virus (IPNV) resistance in Atlantic salmon [22]. However, for many other traits, no reliable QTL were found and less than 10% of the variation of the breeding objective was explained by QTL. That means while doing selection more than 90% of the variation between the animals was handled by traditional selection.

Many explanations for missing heritability problem was published [21]. The most reliable explanation may be due to small effect of genes controlling the traits which can't pass the particular statistical tests despite of having large number of genotypes in the individual. All these small effects of the genes together explain the most of the variation for traits [31]. It was predicted 50-100 genes affected dairy traits, which was considered a high estimate at that time [15]. Based on current GWAS and genomic selection result approximately 2,000-10,000 genes controlling the dairy traits. Thus the number of genes controlling the complex trait like dairy trait was increased up to approximately many folds in just 16-17 years.

Recently due to more application of DNA information other methods were developed like (1) the Genomic Selection (GS) methodology [9], (2) the identification of many thousands of

SNP markers, and (3) SNP-chip genotyping technologies that render the genotyping of all these SNPs cost effective. In GS, the effects of all these SNPs are estimated simultaneously without any significance testing. Hence, the assumption that all SNPs have an effect may be approximately valid, and we should change our focus from significance testing to estimating the effects of all markers.

In GS, a reference population is genotyped and recorded for the trait to estimate SNP effects. Then selected candidates are genotyped and their effects are estimated. By combining these genotypes and estimated effects, genomic expected breeding values (GEBV) estimated for the selected candidates. It may be noted, that the GS approach does not require pedigree recording, which was essential to traditional BLUP-EBV, and in the elite breeding animals, i.e., the selection of candidates, are not necessarily trait recorded.

Genomic selection methods

All SNP effects are simultaneously estimated in a reference population, which is genotyped and phenotyped using the statistical model (assuming 50,000 SNPs):

$y_i = \mu + X_{1i} \times b_1 + X_{2i} \times b_2 + \ldots + X_{50000i} \times b_{50000} + e_i$

where y_i is phenotype of animal i; μ is the overall mean; X_{1i} is the genotype of animal i for marker 1; and e_i is the residual. Since usually we have < 50,000 reference animals, we cannot estimate 50,000 SNP effects if they are treated as fixed effects, i.e., using traditional statistical methods. This problem is solved in GS by treating the SNP effects as random effects drawn from a known distribution. This can be viewed as a Bayesian approach, where prior information on the SNP effects is added to make all effects estimable.

GBLUP method

In traditional BLUP, EBV are estimated using phenotypes and family relationships, which are based on the pedigree of the animals. In GBLUP (Genomic Best Linear Unbiased Prediction), GEBV are estimated using phenotypes and genomic relationships, which are based on genome-wide dense marker data. The genomic relationship between animals 1 and 2 is calculated as the correlation between their SNP genotypes X_{j1} and X_{j2} across all the SNPs j. The GBLUP method is thus very similar to traditional BLUP, except that pedigree relationships are replaced by genomic relationships.

The pedigree relationship between two full sibs is 0.5, which means that two full sibs are expected to have 50% of their alleles in common. However, in real life, two full sibs may share 60% of their alleles or 40%, and this deviation from the pedigree-based expectation of 50% will be detected by dense marker genotyping. Thus, GBLUP is more accurate than traditional BLUP because genomic relationships are more accurate than pedigree-based relationships. The latter requires genomic relationship estimates to be based on a sufficiently large number of SNPs. For livestock and relationships within a breed, 50,000 SNPs distributed across the entire genome seems to sufficient [9]. Relationships across breeds are small and require a larger number of SNPs to be used. The statistical model for the GBLUP method is:

$y_i = \mu + u_i + e_i$

where u_i is the breeding value of animal i. The GBLUP and SNP-BLUP breeding values are equivalent if we define ui as:

$u_i = X_{1i} \times b_1 + X_{2i} \times b_2 + \ldots + X_{50000i} \times b_{50000}$

The SNP-BLUP and GBLUP model imply the same covariance between animals, and thus also identical regression coefficients of the records on the genetic value of animals. The latter implies that, when parameters are carefully adjusted, SNP-BLUP and GBLUP yield identical GEBV, i.e., the methods are said to be equivalent. More formal derivations of the equivalence of GBLUP and SNP-BLUP can be found in the literature [13, 26, 10].

The computational requirements of GBLUP and SNP-BLUP may be very different. SNP-BLUP requires the estimation of 50,000 SNP effects, and thus the solving of a set of 50,000 equations, whereas GBLUP requires the estimation of N GEBV and solving of N equations, where N is the number of animals. Since usually the number of genotyped animals is smaller than 50,000, the GBLUP method is (computationally) preferred.

Sequence data

Genomic selection based on SNP chip genotypes relies on linkage disequilibrium (LD) between the QTL and the SNPs. Increasing the density of SNPs increases the probability that any QTL has a SNP that is in perfect LD with it. The ultimate density is to replace SNP genotypes with whole-genome sequence (WGS) data.

Recently, it was demonstrated small (2–5%) increases in the accuracy of GEBV with sequence data [2]. That current WGS data do not result in substantial improvements in accuracies of GEBV may be explained as follows. First, the GBLUP method is expected to yield little improvement when going from 777k to WGS data since the genomic relationships are accurately estimated with 777k data and WGS will hardly improve the accuracy of the relationships and thus GEBV. However, the nonlinear GS methods attempt to identify the causal SNPs and are expected to benefit substantially from WGS data. Second, current WGS data are not very accurate, either due to imperfect genotype calling, the extensive reliance on SNP imputation (see next section), or structural genomic variations, which are difficult to assess by short reads of sequences. The inaccuracies in the WGS data may compensate for the benefits of higher SNP density. Third, long-range LD may be extensive in the reference population animals, causing large chromosomal segments or haplotypes to be common. Despite current issues with the efficient use of WGS data, it is expected that WGS data will be the future's genotype data because, if the sequencing costs continue to fall, WGS may become the most effective genotyping method [11].

Imputation of missing genotypes

After SNP-chip genotyping, some of the genotypes will be missing. This is solved by a process called genotype imputation. Based on the known genotypes of the animals, the haplotype that the animal carries are recognized since the same haplotype was also observed in other animals. Thus, the missing genotype can be read from the genotype of these other animals, which carry the same haplotype. Software for imputation includes Beagle [3] Fimpute[24], and Alphaimpute [16]. The same strategy is employed to obtain WGS data on many animals: the 1,000-bull-genome project [4] collects a set of sequenced bulls across breeds, which is used to identify (hopefully all) bovine haplotypes and their sequences. Next, many animals are genotyped with SNP chips, the bovine haplotypes that they carry are recognized, and their WGS data are imputed. Another option is to sequence the descendants at low coverage. In this case, the low coverage sequence should be just enough to recognize the haplotypes [11] [5] [25].

Ungenotyped animals

In genomic selection, many (probably most) animals are not genotyped, but we need to include their phenotypic information in the breeding value estimation. At least, traditional selection would use such information. One way to do this is by multiple-step GS: in step 1, pseudo-phenotypes are calculated for the genotyped animals where the pseudo-phenotype of animal includes information (records) on its ungenotyped relatives; in step 2, genomic prediction is performed using the pseudo-records and their genotypes; and in step 3, the traditional EBV and GEBV are combined into a total EBV [26]. As an example of a pseudo-record, the average production of the daughters of a bull can be used. Here, the bull is genotyped but not phenotyped whereas his daughters are phenotyped but not genotyped.

In single-step GBLUP (ssGBLUP), all data are accounted for in a single estimation step [19]. When moving from traditional BLUP to GBLUP, the entire matrix of pedigree relationships is replaced with genomic relationships. However, if genotyping shows that e.g., some animals in different families are more related than expected based on pedigree, then other ungenotyped animals in these families are probably also more related than expected. The correct relationship matrix can be obtained by starting with the genotyped animals and then using the pedigree to calculate relationships involving ungenotyped descendants of these genotyped animals which is going down the pedigree and accounting for the marker-based relationships of the ancestors of the pedigree [9][19].

In most studies, increase in reliability due to single step, over a pure genomic model, is small [18]. A more important feature of single-step models may be that they can account for pre-selection of young genotyped bulls, which could otherwise cause bias in the GEBV [27]. Until recently, the requirement that the G matrix must be inverted directly limited the size of the dataset to which ssBLUP could be applied. The Ancestor, Proven, Young Bull algorithm (APY) uses recursion to build a large component of the G-1 matrix directly, overcoming this limitation and expanding the application of ss-BLUP to millions of genotyped animals [8] but at the expense of some approximation in G-1. For the future, there is a clear need for a single-step method that uses a nonlinear statistical method on sequence level data.

Implementation of genomic selection in livestock

Genomic selection in dairy cattle breeding

The accuracy of genomic prediction in dairy cattle exceeds 0.8 for production traits and 0.7 for fertility, longevity, somatic cell count, and other traits [29] [21]. These high accuracies reflect the large reference populations for each breed that have been assembled to enable genomic predictions and the fact that many of the in the reference populations are progeny-tested bulls with highly accurate phenotypes from average daughter performance. In addition, the GEBV are often used to predict close relatives of animals in the reference population. A feature of dairy genomic prediction is collaboration between countries to assemble these large reference sets, with the consortiums established (Eurogenomics, including the Netherlands, Germany, France, the Nordic countries, Spain, and Poland; The North American Consortium including USA, Canada, Italy, and Great Britain; and a "rest of the world" consortium consisting of a number of remaining countries).

The high accuracies of genomic prediction and relatively low cost of obtaining the genomic predictions from low-density genotyping followed by imputation, has resulted in very large numbers of selection candidates being genotyped. Worldwide, approximately 2 million dairy cattle have now been genotyped for the purposes of genomic prediction.

Implementing genomic selection in dairy cattle has resulted in increased genetic gain, which has now been demonstrated by genetic trend analysis in a number of countries. There is also some suggestion that genomic selection has increased the rate of inbreeding per year [24]. Maximizing genetic gain from genomic selection while constraining the rate of inbreeding will therefore be an important topic for future research. Interestingly, the majority of the genotyped animals in many countries are now heifer calves. While genotyping young bull calves results in the greatest genetic gain, genotyping is now sufficiently cheap that genotyping heifer calves for the purposes of choosing which heifers to retain in the herd is profitable [23] [28]. The genotypes of the heifers can also be used when choosing bulls to which to mate them so that inbreeding of the resulting calf can be minimized. When the selected heifers enter the herd and have herd recording data, they can be used in the reference population for genomic prediction [29]. When the aim is to increase the size of the reference population to improve accuracy of genomic prediction, genotyping mature cows with good phenotypic records can help [17].

Genomic selection in pig breeding

In pig breeding, the most important selection step is the selection of elite boars in the nucleus herd. The boar test recordings come generally before the selection of the elite boars, so extra gains due to a reduction of the generation interval are limited. The implementation of GS in pig breeding is therefore mainly directed at traits whose recording is invasive such as slaughter quality, maternal traits that cannot be recorded on the boars, and crossbred performance, which cannot be recorded on the purebred animals.

Male selection accuracies of ~50% can be achieved for the selection for maternal traits [20]. The selection for maternal traits competes with the selection for production traits such as growth rate and feed conversion efficiency, resulting in substantial increase in genetic gain for maternal traits accompanied with a somewhat reduced rate of gain for the production traits. Rates of gain for total merit increase moderately, but the direction of rate of gain complies much more closely to the direction as indicated by the breeding goal. The substantially increased progress for the maternal traits thus results in a more balanced and thus sustainable selection response.

With respect to slaughter traits, sibs of the test boars may be slaughtered and recorded for these traits before the test boars are selected. Thus, GS can be based on a reference population that is very close to the selection candidates, and thus high selection accuracies can be achieved. The extra gains will partly be at the expense of gains for the traditional production traits, but the direction of genetic change will comply much closer to the breeding goal, and thus can be sustained over a longer period into the future.

Pork is produced by crossbred pigs, but the elite breeding nucleus animals are selected for purebred performance in a favourable environment. The relationship between purebred production in very good environments and crossbred performance on less favourable environments varies between 0.4 to 0.7 [6]. This implies that only 40-70% of the genetic improvement realized in the nucleus will also result in

improved performance in practice, By genotyping crossbred pigs and recording their performance in the commercial environment, GS can be used to improve purebred nucleus animals for crossbred performance under commercial circumstances. This requires across breed and crossbred genomic selection, which has not yet been demonstrated.

Genomic selection in poultry breeding

In layers, there has actually been an experiment to test if genomic selection can achieve more rapid gains than traditional selection. A layer population was split into two sublines; one was submitted to conventional phenotypic selection, and one was selected based on genomic prediction [30]. The experiment ran for 3 years, in which time, four cycles of genomic selection and two of phenotypic selection were conducted. At the end of the 3-years experiment, the two sublines were compared for multiple performance traits that are relevant for commercial egg production. The genomic selection line outperformed the phenotypic selection line for most of the 16 traits that were included in the index used for selection. Although the two programs were designed to achieve the same rate of inbreeding per year, it was found that the realized inbreeding per year assessed from pedigree was higher in the genomic-selected line than in the conventionally selected line [30]. In broilers or meat poultry, the case for GS is not as obvious as in layers because most traits can be recorded on both sexes at an early age. However, the breeding companies are actively investigating the use of GS. Possible uses are for selection to improve crossbred performance in a commercial environment and for traits that cannot be recorded in the nucleus such as disease challenge tests.

Future prospective

Genomic selection offers two opportunities, which have so far not been fully utilized. First, GS combined with reproductive technology could greatly decrease generation length, and in combination with multiple ovulation and embryo transfer (MOET), GS may be used to pick the best embryos to produce the next generation of animals (instead of random embryos). Second, we could train the prediction equation on commercial animals, rather than stud animals, which have been measured for the commercially relevant traits. For instance, commercial animals are often crossbreds, run under a harsher environment than the purebred stud animals. In addition, we can gather information on traits not measured at the stud level such as meat quality and disease resistance. This implies a reduction of costs at the stud level due to less phenotypic and pedigree recording and an increase in costs to generate the reference dataset. This change may also be expected as the costs of genotyping and practical trait recording keep on falling. A commercial mechanism to fund this change is not yet apparent. It seems likely that the genomic selection in animal breeding will eventually lead to structural changes in the genetic improvement industry, but it may be too early to nominate what these changes might be. One possibility is that the number of businesses breeding cattle, sheep, and pigs will decrease as has already happened in poultry.

References

1. Beavis WD. QTL analysis: Power, precision and accuracy. In: A.H. Paterson, editor, Molecular dissection of complex traits. CRC, Boca Raton, FL, 1994
2. Brondum RF, Su G, Janss L, Sahana G, Guldbrandtsen B, Boichard D, Lund MS. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. Journal of Dairy Science. 2015; 98(6):4107-4116.
3. Browning S, Browning B. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. American Journal of Human Genetics. 2007; 81:1084-1097.
4. Daetwyler HD, Capitan A, Pausch H, Stothard P, Binsbergen RV, Brondum RF, Liao X. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nature Genetics. 2014; 46(8):858-865
5. Druet T, Macleod IM, Hayes BJ. Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. Heredity (Edinb). 2014; 112:39-47.
6. Esfandyari H, Sorensen AC, Bijma P. A crossbred reference population can improve the response to genomic selection for crossbred performance. Genetics Selection Evolution. 2015; 47:76.
7. Fernando R, Grossman M. Marker assisted selection using Best Linear Unbiased Prediction. Genetics Selection Evolution. 1989; 21:467-477.
8. Fragomeni BO, Lourenco DA, Tsuruta S, Masuda Y, Aguilar I, Legarra A, Lawlor TJ, Misztal I. Hot topic: Use of genomic recursions in single-step genomic best linear unbiased predictor (BLUP) with a large number of genotypes. Journal of Dairy Science. 2015; 98(6):4090-4094.
9. Goddard M, Hayes B, Meuwissen T. Using the genomic relationship matrix to predict the accuracy of genomic selection. Journal of Animal Breeding Genetics. 2011; 128:409-421.
10. Goddard M. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica (The Hague). 2009; 136:245-257.
11. Gorjanc G, Cleveland MA, Houston RD, Hickey JM. Potential of genotyping- by-sequencing for genomic selection in livestock populations. Genetics Selection Evolution. 2015; 47:12.
12. Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Research. 2001; 12:222-231.
13. Habier D, Fernando RL, Dekkers JCM. The impact of genetic relationship information on genome-assisted breeding values. Genetics. 2007; 177(4):2389-2397.
14. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. Springer Series in Statistics, 2009.
15. Hayes B, Goddard ME. The distribution of the effects of genes affecting quantitative traits in livestock. Genetics Selection Evolution. 2001; 33(3):209-229.
16. Hickey J, Kinghorn B, Tier B, Werf JVD, Cleveland M. A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. Genetics Selection Evolution. 2012; 44(1):9.
17. Kemper KE, Reich CM, Bowman PJ, Vander Jagt CJ, Chamberlain AJ, Mason BA, Hayes BJ, Goddard ME. Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to

greater accuracy of a cross-breed genomic predictions. Genetics Selection Evolution. 2015; 47:29.
18. Koivula M, Stranden I, Su G, Mantysaari EA. Different methods to calculate genomic predictions–comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP). Journal of Dairy Science. 2012; 95(7):4065-4073.
19. Legarra A, Christensen O, Aguilar I, Misztal I. Single Step, a general approach for genomic selection. Livestock Science. 2014; 166:54-65.
20. Lillehammer M, Meuwissen TH, Sonesson AK. Genomic selection for maternal traits in pigs. Journal of Animal Science. 2011; 89(12):3908-3916.
21. Lund MS, Roos AP, Vries AG, Druet T, Ducrocq V, Fritz S, *et al.* Finding the missing heritability of complex diseases. Nature. 2009; 461(7265):747-753.
22. Moen T, Torgersen J, Santi N, Davidson WS, Baranski M, Odegard J. Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in atlantic salmon. Genetics. 2015; 200(4):1313-1326.
23. Pryce J, Hayes B. A review of how dairy farmers can use and profit from genomic technologies. Animal Production Science. 2012; 52:180-184.
24. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 2014; 15:478.
25. Van Binsbergen R, Bink MC, Calus MP, van Eeuwijk FA, Hayes BJ, Hulsegge I, Veerkamp RF. Accuracy of imputation to whole-genome sequence data in Holstein Friesian cattle. Genetics Selection Evolution. 2014; 46:41.
26. Van Raden. Efficient methods to compute genomic predictions. Journal of Dairy Science. 2008; 91:4414-4423.
27. Vitezica ZG, Aguilar I, Misztal I, Legarra A. Bias in genomic predictions for populations under selection. Genetics Research. 2011; 93(5):357-366.
28. Weigel KA, Hoffman PC, Herring W, Lawlor TJ. Potential gains in lifetime net merit from genomic testing of cows, heifers, and calves on commercial dairy farms. Journal of Dairy Science. 2012; 95(4):2215-2225.
29. Wiggans GR, Vanraden PM, Cooper TA. The genomic evaluation system in the United States: Past, present, future. Journal of Dairy Science. 2011; 94(6):3202-3211.
30. Wolc A, Zhao HH, Arango J, Settar P, Fulton JE, O'Sullivan NP, *et al.* Response and inbreeding from a genomic selection experiment in layer chickens. Genetics Selection Evolution. 2015; 47:59
31. Yang J, Beben B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, *et al.* Common SNPs explain a large part of the heritability for human height. Nature Genetics. 2010; 42:565-569.