Whole Genome Selection in Livestock

G.K. Gaur, M. Singh

[1]*Livestock Production and Management Section*
Indian Veterinary Research Institute,
Izatnagar 243 122 (Uttar Pradesh), India

1. INTRODUCTION

Genetic programs in livestock have been extensively used to improve the productivity of populations over the last 50 years. The traditional approach of breeding, which is based on sophisticated quantitative genetic principle, has been very successful (Schaeffer, 2006), even without knowing the exact genes that have increased or decreased in frequency in populations (black box approach). The use of effective progeny testing program including wide spread use of proven/elite bulls has made drastic genetic improvement in populations. Several advances in basic approaches for calculating genetic merit of individuals have further resulted in incremental improvements in the process. Recently; livestock producers, geneticists and breeders are focusing on overall development of animal including improved health, production, reproduction and longevity and according all the traits have become the part of selection program (Decker et al., 2009). Advances in DNA sequencing and high-throughput genotyping techniques has led to discovery of large number of single nucleotide polymorphisms (SNP) in livestock (Tassell et al., 2008). Automated methods for SNP genotyping are also now commercially available. Use of dense SNP arrays covering the entire bovine genome explains the majority of genetic variation in important traits of livestock (Meuwissen et al., 2001). Genomic selection (GS) basically refers to selection decisions based on genomic estimated breeding values (GEBV), which are computed through combining SNP/genotypic data with phenotypic and pedigree data with the objective to increase the accuracy of prediction of breeding and genotypic values. Selection based on marker effects alone has dramatically changed standard practices used for the genetic improvement of quantitative traits in livestock. The cost of marker technology is now being continuously reduced drastically with increase in number of available markers. Besides accelerating the selection cycles, GS offers the opportunity to increase the selection gains per unit of time (Meuwissen et al., 2001; Boichard et al., 2006). Genomic selection in livestock allows identification of genetically superior animals at much earlier age. Traditional breeding schemes are time consuming and costly because we have to wait many years to obtain genetic predictions with sufficient accuracy for making selection decisions. Animals that have been DNA tested may receive an accurate GEBV before they reach sexual maturity. Therefore, it is believed that GS will replace traditional breeding systems in near future.

Evaluation of genomic breeding value is not carried out directly on the basis of gene information but it is calculated using marker effects. The valuable genomic information has many advantages. It may predict genetic merit of young animals with up to double the accuracy of traditional parent averages. It has the potential to shorten the generational interval and hence to speed the genetic improvement (Georges and Massey, 1991; Peippo et al., 2007). It enables bull selection program from a much wider genetic pool. It provides earlier information about genetic differences between siblings. It improves the reliability of current progeny testing results for low-heritability or later-expressed traits, including health and fitness traits. Genomic breeding value evaluation and selection is now close to that reality, which was dream in past. Inbreeding is considered as challenge in dairy cattle. It is observed that GS has lower increment of inbreeding as compared to traditional BLUP because it estimates the Mendelian sampling term of breeding values (Daetwyler et al., 2012). The success of genomic selection mainly depends on the extent of linkage disequilibrium between markers and quantitative trait loci, the number of animals in the training set and the heritability of the trait. The extent of linkage disequilibrium depends on the genetic structure of the population and the density of markers.

2. HIGH DENSITY AND HIGH THROUGHPUT GENOTYPING

SNP markers are composed of one genetic letter and there are two different forms for every marker in the entire population. Many hundreds or thousands of SNP markers are known in different livestock which are shared equally in the entire genome. All chromosomes can now be extracted for all the markers simultaneously in one step process, which is known as typing or high-density and high-throughput genotyping. This is done in laboratory through a unit called chip carrying 10K (in 10K chip), 50K (in 50K chip) or 700K (in 700K chip) markers to be tested simultaneously. Very little animal's genetic material is used for the typing. The result of typing indicates n times AA or BB or AB where n is number of markers in chip and AA/BB/AB represents genotype of the marker locus. Out of equal spread of many markers over the entire genome, it is believed that there may be at lest one marker close to every gene, which influences the transmission trait. This indicates that the marker is also transmitted with the appropriate combination of genes. The first high-density and high-throughput genotyping assay was 10K single nucleotide polymorphism (SNP) chip commercialized by Affymetrix (The Bovine HapMap Consortium, 2009). However, the density of SNPs in this panel was insufficient for many genomic studies (including genomic selection and genome-wide association analyses), which resulted in need of high density chip. Accordingly, Illumina BovineSNP50 chip was developed by a consortium of animal scientists using SNP discovery populations in Holstein, Angus and mixed breeds of beef cattle (Van Tassell et al., 2008) and provided much higher density (~50, 000 SNPs per animal). This assay became the international standard for genomic selection and genome-wide association studies in cattle and applied to other species to resolve the evolutionary relationship among the horned ruminants

(Decker et al., 2009; MacEachern et al., 2009). Though, Illumina BovineSNP50 was proven to be extremely useful for many genomic studies, however, higher density assays were realized to build models for genomic selection across the breeds. A new high-density SNP genotyping chip was introduced in 2010 by Illumina utilizing bead technology and single base extension chemistry similar to Bovine SNP50K chip. This chip is capable to genotype approximately 700K SNPs simultaneously.

SNP discovery for Illumina BovineSNP50 assay was performed using pools of DNA samples from Angus, Holsteins and important US beef breeds. As a result, there may be bias in the assay towards SNPs that have high minor allele frequency in Angus and Holsteins and accordingly the assay may perform slightly better for genomic selection and genome-wide association studies in these breeds. However, SNP discovery for the design of Illumina 700K panels was performed by sequencing a large number of animals from many different breeds (including Bos taurus and Bos indicus) to minimize the bias in SNP information across the breeds.

3. PROCESS OF WHOLE GENOME SELECTION

Different genes and their effects on the performance traits are not known. To find out the SNPs effects, i.e. association between markers and trait; the SNP samples are to be compared with known (genetic) production traits. Animals with known genetic production may be daughter proven bulls. Comparison of SNP samples with genetic production of bulls may determine the influence of SNP on the trait. This population included in the analysis is called training sample. The training sample/set is therefore, a reference population in which the marker effects are estimated. It contains phenotypic information from breeding germplasm evaluated over a range of environmental conditions, molecular marker scores and pedigree information or kinship. Marker effects are estimated based on the training set information using certain advanced statistical methods. It must be ensured that the reference population is updated over the generations; otherwise predictive ability will decrease over time (Meuwisen et al., 2001). The number of genotyped animals depends on heritability of the trait and increase in genetic gain per genotyped animal. It is advisable that the best and worst, both animals should be genotyped for a given trait. The marker effects conveyed from the training sample are then used for calculation of genomic breeding values of other (usually younger) animals without their own reliable breeding value information. These animals refer target/validation sample/set and contains the selection candidates (derived from the reference population) that have their genotypic information (but no phenotypic information) and selected based on marker effects estimated in the training set.

Statistical analysis of genomic selection experiments (Meuwissen et al., 2001) aimed to know the effect of all alleles/markers on performance traits (breeding goal traits). Classical statistics however encounters the problem of shortage of degree of freedom (more number of markers/alleles as

compared to animals). There are three ways to get around the huge shortage of degree of freedom i.e. Least Squares Method, Best Linear Unbiased Prediction (BLUP) and Bayesian Estimation Method. In least squares method, all the markers are tested one by one for their statistical significance and the effect of non-significant markers is set to zero. Subsequently, significant markers are taken in model simultaneously for analysis. Best linear unbiased prediction method fits the allelic effects as random instead of fixed. The fitting of random effect does not require degree of freedom and thus all allelic effects are estimated simultaneously. Random effect however requires an estimate of variance of the allelic effect. The method assumes similar variance for all the markers. Bayesian estimation method is similar to BLUP except that the variance of allelic effects is assumed different for every marker and is estimated using a prior distribution for the variance. The prior distribution of variance of marker i (Vai) is assumed as Vai = 0 with probability p; Vai ~ χ -2(v, s) with probability (1-p); where p depends on the mutation rate at the marker locus and χ -2(v, s) denotes the inverse-chi squared distribution with v degrees of freedom and scale parameter S. In brief, process of genomic selection may be looked in to three simple steps i.e. use of markers to deduce the genotype of each animal at each quantitative trait loci (QTL), estimating effects of each QTL genotype on the trait and summing of all the QTL effects for selection candidates to obtain their genomic EBV (GEBV). Accordingly, marker estimated breeding value (MEBV) or genomic estimated breeding value (GEBV) is predicted by MEBVi/GEBVi = $\sum$Mij, where Mij is the estimate of the effect of the jth allele/heplotype of animal i. The development of statistical models dealing with GS to get the accurate estimates, however need to be considered further. The Gen Sel Program (Fernando and Garrick, 2009) which is based on Bayesian approach is being extensively used for estimation of GEBVs of animals under selection, incorporating SNP data of large number of markers and the phenotype of interest.

4. GENOMIC SELECTION IN LIVESTOCK

Genomic selection has modified concept of livestock breeding plans during last several years. Although, idea of molecular markers for improving genetic gain in livestock is not very new (Smith, 1967; Soller and Beckman, 1983), but its application has been observed to be limited due to several reasons. A large number of loci affect the quantitative traits with any one locus capturing only limited proportion of total genetic variance (Hayes and Goddard, 2001; Raden et al., 2009). Furthermore, small gains were only expected with few markers having high genotyping cost. Computation of genetic worth using marker information further restricted application of marker-assisted selection in livestock. Sequencing of bovine genome, discovery of markers in form of SNPs and reduction in cost of genotyping ultimately brought revolution in genomic selection. Accordingly, making very accurate selection decisions based on dense marker data alone became possible (Meuwissen et al., 2001). The elite bulls in traditional dairy cattle breeding programs are selected through progeny testing scheme. Awaiting progeny testing results prolongs the generation

interval and ultimately increases the over all cost. Schaeffer (2006) suggested that GS should be used for the selection of young bulls and progeny tests should be given less priority. Additionally, nucleus breeding herd especially cows need to be selected by GS. Lillehammer et al. (2011) performed computer simulation of GS cattle breeding schemes and found an increase in genetic improvement by 13%. Pryce et al. (2010) observed that genetic gain through GS may increase between 60 and 120%, depending on age of breeding males and females.

The extent of accuracy of GEBV (correlation between true breeding value and GEBV) has been worked out for dairy cattle, beef cattle, sheep, and pigs for a range of traits. Meuwissen et al. (2001) demonstrated that accuracies of predicted breeding values from markers alone may be achieved to the level of 0.85. Similarly, Kolbehdari et al. (2006) obtained high correlation between GEBV and true breeding values (0.80) in a simulation study using different heritability and evenly or randomly spaced QTLs. The simulation results suggest that the accuracy of the GEBV for a bull calf at birth may be as high as the accuracy of an EBV of progeny test. Consequently, genomic selection may lead to double rate of genetic gain (Schaeffer, 2006) as compared to 5-19 % improvement in genetic response through marker assisted selection (Boichard et al., 2006). Furthermore, avoiding progeny testing, bull breeding programs may save up to 92% of their costs (Schaeffer, 2006), however, some of these savings may be offset by investment on genotyping. Spelman et al. (2006) reported accuracy of DNA-based BVs as measured by their correlation with progeny test BVs varying from 0.45 to 0.60 for the production traits in validated population of Holstein-Friesian breed. They further predicted that selection based on genomic information will increase the rate of genetic gain in New Zealand by 50-70%. Berry (2007) in Irish Holstein – Friesian dairy cows expected genetic gain through genomic selection @ 0.539 genetic standard deviations per year with accuracy of selection of 0.75 against 0.199 genetic standard deviations per year under the current scheme of progeny testing. Harris et al. (2008) reported reliabilities of GEBV in New Zealand dairy cattle in an experiment performed by the Livestock Improvement Corporation. Reliabilities of GEBV for young bulls with no daughter information ranged from 50 to 67% for milk production traits, live body weight, fertility, somatic cell count and longevity; compared with an average 34% for parental average breeding values. Furthermore, Bayesian methods gave slightly greater (2 to 3%) reliabilities than the BLUP approach, whereas the regression methods performed poorly. Hayes et al (2009) reviewed results of genomic selection in Australia. Reliability of genomic breeding values at the time of birth ranged from 0.18 (fertility) to 0.53 (Australian Profit Ranking) in BLUP method and from 0.14 (fertility) to 0.55 (Australian Profit Ranking) in Bayesian method. The reliability of the fertility GEBV was substantially lower than the reliability of other traits. The Bayesian method gave small increases in reliability for all traits except fertility, in order of 2 to 7%. Raden et al. (2009) reported reliabilities of GEBV for US and Canadian young bulls. Across the traits, the GEBV had reliability of 50%, compared with 27%

from the parent average alone. Using BLUP in place of Bayesian approach gave slightly (1%) reduced reliability. Fernanda et al. (2011) observed that accuracy of GEBV increased significantly with the increase in number of bulls in the training set (480, 960 or 1920), trait h2 (0.10, 0.25 or 0.40) and marker densities (40 k or 800 k). Hayes et al. (2009) reviewed results of genomic selection in Netherlands. The increase in reliability of GEBV over parent average EBV at the time of birth was 33% (fat percentage), 19% (kilograms of protein), 15% (feet and legs), 13% (udder depth, SCS), and 9% (fertility). They concluded that having a larger number of bulls in their reference population would increase the reliability of GEBV in their selection candidates substantially. The effect of different breeding structures was investigated by Duthie et al. (2010). It was realized that including genomic information increases the economic response to selection in both breeding structures.

Genomic selection is an attractive strategy for dairy sheep, because rams can be selected for milk production and other traits' GEBV, which are expressed earlier in life. Duchemin et al. (2012) in French Lacaune dairy sheep noticed that for young rams with no progeny test records, accuracies of EBV for milk yield, fat content, and somatic cell scores may be improved by 18–25% by including DNA marker information. The advantages of genomic selection are relatively less for meat and wool sheep. However, additional genetic gain @ 20–30% may be obtained through application of GS (Banks and Werf, 2009). Accuracies of genomic breeding values were high (>0.6) for wool traits such as fleece weight and fiber diameter (Daetwyler et al., 2012) but lower for traits like lean-meat yield. These lower accuracies were however still higher than those estimated based on pedigree alone (Daetwyler et al., 2012). Accuracies of genomic selection in beef cattle, evaluated using cross-validation technique for growth and carcass traits were reported up to 0.42 and 0.65 (Weber et al., 2012). GS offered great opportunities for selection in pigs. Target trait for the genomic selection in pigs is feed conversion ratio (FCR) because this trait is expensive to measure. Christensen et al. (2012) estimated accuracy of GEBV for FCR using 60, 000 SNPs in Danish Duroc pigs and concluded that GEBVs were significantly accurate than pedigree EBV. Forni et al. (2010) in litter size of pigs noticed 68% increase in accuracy of the breeding values of the training population through GS over traditional selection (BLUP). Simianer (2009) showed the potential of GS for litter size in two breed scheme and reported that the genetic progress per year could be increased by 37% compared to the conventional scenario.

5. SUMMARY

Major breakthrough in molecular biology including discovery of DNA-markers, sequencing and genome mapping of farm animal species created new avenues for identifying major genes, genetic defects, quantitative trait loci and ultimately applying genomic selection in livestock. Genomic selection is considered as extension of marker assisted selection technique, wherein high density markers in form of SNPs covering the entire genome are used simultaneously for genetic

evaluation of individual. SNPs are especially useful when they occur on or close to gene/s that contributes to an important trait. Use of genomic information in genetic evaluation programs has brought revolutionary change in livestock selection. Whole genome selection has the potential to increase the accuracy of selection, shorten the generation interval and to accelerate the rate of genetic progress in many traits. The technique is more effective when there is limited information on females and young bulls and the trait/s are of low heritability. Conventional livestock breeding systems are being reshaped due to the availability of high density whole genome SNP chips. GS has become the important method in many countries for dairy bull selection. It is believed that GS may soon completely replace all the traditional genetic evaluation systems of livestock.

REFERENCES

[1] R G Banks and JHJ Werf (2009), Economic evaluation of whole genome selection using meat sheep as a case study, *Proc. Assoc. Adv. Anim. Breed. Genet*, **18**, pp. 430–33.

[2] D Berry (2007), Report on the potential of genomic selection in Irish dairy cattle, Draft Paper I (available on net).

[3] D Boichard, S Fritz, M N Rossignol, F Guillume, J.J Colleu and T Druet (2006), Proceedings of the 8th World Congress on Genetics Applied to Livestock Production (available on net).

[4] P C Madsen, B Nielsen, T Ostersen and G Su (2012), Single-step methods for genomic evaluation in pigs, Animal, **5**, pp. 1–7.

[5] H D Daetwyler, K E Kemper, J H Werf and B J Hayes (2012), Components of the accuracy of genomic prediction in a multi-breed sheep population, *J. Anim. Sci.,* **90,** 10, pp. 3375–3384.

[6] J E Decker, J C Pires, G C Conant, S D McKay, M P Heaton, K Chen, A Cooper and J J F Taylor (2009), Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics, *Proc. Natl. Acad. Sci.,* **106**, 44, pp. 18644-18649.

[7] S I Duchemin, C Colombani, A Legarra, G Baloche and H Larroque (2012), Genomic selection in the French Lacaune dairy sheep breed, *J. Dairy Sci*, **95**, 5, pp. 2723–2733.

[8] C Duthie, R Sawalha, E Navajas, R Roehe and T Roughsedge (2010), Economic response that can be achieved from including genomic information to the terminal sire index of beef cattle. Presented to the British Society of Animal Science Annual Conference, April 12 to 14, 2010, Queen's University, Belfast.

[9] V B Fernanda, J B Neto, M Sargolzaei, J A Cobuci and F S Schenkel (2011), Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle, *BMC Genetics*, **12,** pp.80.

[10] R Fernando and D Garrick (2009), GenSel- User Manual for a portfolio of Genomic Selection related Analyses (http://taurus. ansci. iastate. Edu /Site /Welcome_files /GenSel %20Manual % 20v2. pdf)

[11] S Fornis, I Aguilar, I Misztal and N Deeb (2010), Genomic relationships and biases in the evaluation of sow litter size. Proc 9th World Congress on Genetics Applied to Livestock Production, Liepzig, Germany, Aug 2-6, pp. 266.

[12] M Georges and J M Massey (1991), Velogenetics, or the synergistic use of marker assisted selection and germ-line manipulation, *Theriogenology*, **35,** pp.151–159.

[13] B L Harris, D L Johnson and R J Spelman (2008), Genomic selection in New Zealand and the implications for national genetic evaluation. Proc. Interbull Meeting, Niagara Falls, Canada (available on net).

[14] B J Hayes and M E Goddard (2001), The distribution of the effects of genes affecting quantitative traits in livestock, *Genet. Sel. Evol.,* **33**, pp. 209–229.

[15] B J Hayes, P J Bowman, A J Chamberlain and M E Goddard (2009), Invited review: Genomic selection in dairy cattle: Progress and challenges, *J. Dairy Sci.*, **92**, pp. 433-443.

[16] D Kolbehdari, Z Wang, J R Grant, B Murdoch, A Prasad, Z Xiu, E Marques, P Stothard and S S Moore (2009), A whole genome scan approach to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls, *J. Anim. Breed. Genet*., **12**, 63, pp. 216-272.

[17] M Lillehammer, T H Meuwissen and A K Sonesson (2011), A comparison of dairy cattle breeding designs that use genomic selection, *J. Dairy Sci.,* **94**, pp. 493–500.

[18] S MacEachern, J McEwan, A McCulloch, A Mather, K Savin and M Goddard (2009), Molecular evolution of the Bovini tribe (Bovidae, Bovinae): is there evidence of rapid evolution or reduced selective constraint in Domestic cattle, *BMC Genom,* **10**, pp. 179.

[19] L K Matukumalli, C T Lawley, R D Schnabel, J F Taylor, M F Allan, M P Heaton, J O'Connell, S S Moore, T P L Smith, T S Sonstegard and C P Van Tassell (2009), Development and characterization of a high density SNP genotyping assay for cattle, PLoS One, **4**, pp. 5350.

[20] T H E Meuwissen, B J Hayes and M E Goddard (2001), Prediction of total genetic value using genome-wide dense marker maps, *Genetics*, **157**, pp. 1819–1829.

[21] T H Meuwissen and M E Goddard (2004), Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data, *Genet. Sel. Evol.*, **36**, pp. 261–279.

[22] J Peippo, S Viitala, J Virta, M Räty and N Tammiranta (2007), Birth of correctly genotyped calves after multiplex marker detection from bovine embryo microblade biopsies, *Mol. Reprod. Dev.,* **74**, 11, pp. 1373–1378

[23] J E Pryce, M E Goddard, H W Raadsma and B J Hayes (2010), Deterministic models of breeding scheme designs that incorporate genomic selection, *J. Dairy Sci.*, **93**, 11, pp. 5455–5466.

[24] L R Schaeffer (2006), Strategy for applying genome-wide selection in dairy cattle, *J. Anim. Breed. Genet*., **123**, pp. 218–223.

[25] H Simianer (2009), The potential of genomic selection to improve litter size in pig breeding programs. Proc 60th Annual meeting of the European Association of Animal Production, Barcelona, Spain, pp. 210.

[26] C Smith (1967), Improvement of metric traits through specific genetic loci, *Anim. Prod*., **9**, pp. 349–358.

[27] M Soller and J S Beckmann (1983), Genetic polymorphism in variety identification and genetic improvement, *Theor. Appl. Genet*., **67**, pp. 25–33.

[28] R J Spelman, J Arias, M Keehan and V Obolonkin (2006), Application of genomic information in dairy cattle, Draft Paper (available on net).

[29] P M Van Raden, C P V Tassell, G R Wiggans, T S Sonstegard, R D Schnabel, J F Taylor and F Schenkel (2009), Invited review: Reliability of genomic predictions for North American Holstein bulls, *J. Dairy Sci*., **92**, pp. 16–24.

[30] C P Van Tassell, T P L Smith, L K Matukumalli, J F Taylor, R D Schnabel, C T Lawley, C D Haudenschild, S S Moore, W C Warren and T S Sonstegard (2008), Simultaneous SNP discovery and allele frequency estimation by high throughput sequencing of reduced representation genomic libraries, *Nat. Meth*., **5**, pp. 247-252.

[31] K L Weber, R M Thallman, J W Keele, W M Snelling and G L Bennett (2012), Accuracy of genomic breeding values in multi breed beef cattle populations derived from deregressed breeding values and phenotypes, *J. Anim. Sci.*, **90**, pp. 4177–4190.