International Journal of Science, Environment and Technology, Vol. 6, No 1, 2017, 339 – 347

ISSN 2278-3687 (O)
2277-663X (P)

GENOMIC SELECTION IN DAIRY CATTLES: A REVIEW

[1]Dr. Devasee Borakhatariya*, [2]Dr. Pravin Kandhani and [3]Dr. Vijay Trivedi

[1]Assistant Professor, Department of Veterinary Gynaecology and Obstetrics
College of Veterinary Science and Animal Husbandry
Junagadh Agricultural University, Junagadh

[2,3]M.V.Sc Scholars, Animal Genetics, Kamdhenu University, Gandhinagar, Gujarat

E-mail: Devasee94@gmail.com (**Corresponding Author*)

Abstracts: New technology called genomic selection is revolutionizing dairy cattle breeding. Genomic selection refers to selection decisions based on genomic breeding values (GEBV). The GEBV are calculated as the sum of the effects of dense genetic markers, or haplotypes of these markers, across the entire genome, thereby potentially capturing all the quantitative trait loci (QTL) that contribute to variation in a trait. The QTL effects, inferred from either haplotypes or individual single nucleotide polymorphism markers, are first estimated in a large reference population with phenotypic information. In subsequent generations, only marker information is required to calculate GEBV. Reliabilities of GEBV for young bulls without progeny test results in the reference population were between 20 and 67%. The reliability achieved depended on the heritability of the trait evaluated, the number of bulls in the reference population, the statistical method used to estimate the single nucleotide polymorphism effects in the reference population, and the method used to calculate the reliability. The BLUP method is attractive because the only prior information required is the additive genetic variance of the trait. All countries included a polygenic effect (parent average breeding value) in their GEBV calculation. This inclusion is recommended to capture any genetic variance not associated with the markers, and to put some selection pressure on low-frequency QTL that may not be captured by the markers. The reliabilities of GEBV achieved were significantly greater than the reliability of parental average breeding values, the current criteria for selection of bull calves to enter progeny test teams. The increase in reliability is sufficiently high that at least 2 dairy breeding companies are already marketing bull teams for commercial use based on their GEBV only, at 2 years of age. This strategy should at least double the rate of genetic gain in the dairy industry. Many challenges with genomic selection and its implementation remain, including increasing the accuracy of GEBV, integrating genomic information into national and international genetic evaluations, and managing long-term genetic gain.

Keywords: Genomic selection, Genetic Gain, Genetic Breeding, Selection, Progeny Testing.

Introduction

Genomic selection is the selection based on the prediction of breeding values from the information of dense molecular markers covering the whole genome (Meuwissen *et al*, 2001). Next-generation sequencers are transforming animal breeding, enabling cost-effective markers over the entire genome. Large numbers of SNPs have been discovered in domestic

*Received Dec 29, 2016 * Published Feb 2, 2017 * www.ijset.net*

animals and livestock by performing whole-genome association studies. These studies can detect statistical associations between commercially important traits and one or more SNP markers, supporting the development of comprehensive marker arrays for Genomic selection, a form of MAS using genetic markers covering the whole genome. When the marker effects are known and it is known which markers the animal carries, the breeding value of the animal based on these markers can be calculated and can be used for selection.

Methodology of Genomic Selection

1. Part of the population is genotyped using the dense SNP chip and phenotyped for quantitative traits. This part of population is referred as the Reference Population
2. The establishment of an appropriate Reference Population is one of the key aspects in Genomic Selection
3. For dairy cattle Reference Populations, Saatchi *et al*, (2010) recommended to use (>90%) progeny tested sires from recent generations rather than older bulls.
4. Size of the Reference Population is inversely proportional to the heritability of the trait and directly proportional to the effective population size (Hayes *et al*, 2009).

One of the challenges in small populations, and especially for low heritability traits is to increase the predictive accuracy obtained with genomic evaluations (VanRaden *et al*., 2010). Different international collaboration consortia have emerged to increase the accuracy of genome-enhanced predictions for a successful implementation of genomic selection. The first association appeared between Canada and the United States to share genotypes and technical knowledge in 2008 with an initial population of around 17,000 genotypes (VanRaden *et al.,* 2010). The effect of all the SNPs is estimated in the Reference Population by statistical models; where the association between SNPs and phenotypes is calculated. The rest of the population (other than Reference Population) is genotyped using the same SNP chip and the total genetic value (GEBV) of the animals is predicted by using the prediction equations derived from Reference Population.

In Genomic selection, estimates of the Genomic Estimated Breeding Values (GEBVs) are calculated with help of various statistical models.

1. Least Squares Analysis (LS)
2. BLUP (Best Linear Unbiased Prediction)
3. Bayesian Approach

The ***least squares*** method is a form of mathematical regression analysis that finds the line of best fit for a dataset, providing a visual demonstration of the relationship between the data

points.In least-squares analyses, a stepwise approach can be as follows. QTL is polymorphic and its allelic effects differ, it can be adopted to tackle problems with insufficient degrees of freedom and genes are added to the model if they significantly improve the fit of the existing model. It seems, however, quite arbitrary to set the effects of loci to zero that are just below the significance threshold and include the full effect of those that are above this threshold. A better weighting of the information must be possible.

In statistics, ***Best Linear Unbiased Prediction (BLUP)*** is used in linear mixed models for the estimation of random effects. BLUP was derived by Charles Roy Henderson in 1950 but the term "Best Linear Unbiased Prediction" seems not to have been used until 1962.The use of the term "prediction" may be because in the field of animal breeding in which Henderson worked, the random effects were usually genetic merit, which could be used to predict the quality of offspring. However, the equations for the "fixed" effects and for the random effects are different

Best Linear Unbiased Prediction of allelic effects can be calculated even if there are more effects to be predicted than data points. If we assume that all loci or genes explain *a priori* an equal amount of variance (*i.e.*, the variance per locus is Vg/n, where Vg is the total genetic variance and n is the number of loci), we have only one variance to estimate. But having equal variances explained by all loci seems an unrealistic assumption.

Hayes and Goddard (2001) used **GW-BLUP (Genome-Wide BLUP)** to estimate the effect of every SNP and thus estimate the CSE. Difference from traditional BLUP is evident in the assumption. In traditional BLUP,

$\text{Var}(y) = A\sigma_a^2 + I\sigma_e^2$

Where, σ_a^2indicates the additive genetic variance and σ_e^2indicates the error variance; (A) and (I) are the additive genetic variance covariance matrix and identity matrix respectively.

However, in GW-BLUP, $\text{Var}(y) = XX'\sigma_m^2 + I\sigma_e^2$.

Where, σ_m^2indicates the variance due to marker effects. Prior distribution is assumed to be normal and with constant variance.

In **Bayesian statistics**, parameters such as variance explained by the ith locus, Vg_i, are assumed to come themselves from a prior distribution, $p(Vg_i)$. Hence, the variance can vary across loci, and combining of the information from the prior distribution and that of the data yields an estimate of Vg_i. This Bayesian approach, where the variance due to each locus can vary, seems more realistic than assuming that the variance due to each locus is fixed at Vg/n, as is the case in the BLUP method.

Comparison of Methods

Theoretically, if only a few genes and dense SNPs are used, Bayesian method should perform the best. However, if large numbers of genes are available with SNPs of lower density, GW-BLUP should perform the best (Boichard, 2010).As evident from simulation studies, GW-BLUP performs to the lowest extent followed by Bayes methods have been shown to be the best in such studies (Meuwissen *et al*., 2001). Distribution of gene effects is equal for GW-BLUP while, unequal under Bayesian method (Meuwissen *et al.,* 2001).Reliability of GEBVs of Bayesian method is only 1% higher than GW-BLUP, however, reliability of Bayesian method is substantially higher for traits influenced by large QTLs (VanRaden, 2008). Reliabilities of various SNP chips were reported by VanRaden (2010) showing reliabilities for 3k, 50k and 700k chips to be of 70%, 83% and 84% with no imputation of missing genotypes while reliabilities of 80%, 83% and 84% with imputation.

How does Genomic Selection alter Selection?

The local selection pressure will depend on estimated marker breeding values. The alleles with large favourable effects will more often be selected. However, selection will still be on the basis of total breeding value of the animal. Thus, SNP will become the unit of selection (Schaeffer, 2006).Genetic change can be two times greater than the current progeny testing schemes and the savings in logistical costs could be 92% of today's costs. The company that adopts this strategy the earliest will have a major start over other companies. 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.

There will be an initial start-up period for a company in which animals will need to be genotyped so as to estimate the haplotype interval effects. For breeds that are less numerous than Holsteins, for example, Ayrshires, Jerseys, Brown Swiss, Guernsey and Milking Shorthorn in Canada, if the funds were available, all animals in these breeds could be genotyped. AI bulls in these breeds usually take longer than 6 years to prove, and only one or two bulls are proven per year. The accuracy of first bull proofs is often only slightly above 0.50. Thus, progeny testing should be abandoned altogether in these breeds. Bulls should be selected based on their GEBV as calves, and a number of bulls can be chosen to meet the demand for number of services. Bulls only need to be used for one year, and thereafter only new ones are selected that are unrelated to the previous group, but which have higher GEBVs

for the economically important traits. The bulls' GEBVs will have greater accuracy than the first proofs that are available today through progeny testing.

Factors affecting the usefulness of Genomic Selection

The factors which affect the usefulness of Genomic selection include the population structure and its history, the availability of dense marker maps and the availability of many genotyped individuals with records.

Accuracy of Genomic Selection

Accuracy of GEBVs depends on the size of the reference population (VanRaden, 2008 and Hayes *et al*., 2009). Accuracy of Genomic selection increases in general with increase in the reference population and their records. Accuracy of GEBVs also depends on the marker approach, i.e. whether single markers or haplotypes (identical by descent/ identical in state) are used. Calus *et al*. (2008) reported highest accuracy in Genomic selection using single markers followed by haplotype approaches. Accuracy of GEBVs also depends on the amount of linkage disequilibrium between the marker and the QTL. Higher the marker density, higher will be the linkage disequilibrium and consequently the accuracy of Genomic selection will also increase (Calus *et al*., 2008).

Increased Genetic Gain from Genomic Selection

Various authors reported increased genetic gain in various species is as under

- Dairy Cattle - 60-120% (Pryce and Daetwyler, 2011)
- Meat sheep - 21% (Van der Werf, 2011)
- Wool sheep - 38% (Van der Werf, 2011)
- Beef cattle - 29-158% (Van Eenennaam *et al.,* 2011)
- Layers - 40% (Dekkers, 2009)

Impact of Genomic Selection on Dairy Cattle Breeding

As reported by Schaeffer (2006), Genomic selection will reduce the necessity of progeny testing a bull, thus young bulls can be selected at any time since birth. Costs of progeny testing a bull (around 40,000 $/bull) are saved. Generation interval can be reduced by a factor of two, i.e. the generation interval can be halved the original. Annual savings through consideration of costs required for rearing and management for progeny testing a bull can be reduced by 92% i.e. $23 million. Thus, Genomic selection can have a massive impact on dairy cattle breeding programmes provided; updated marker information of the reference population is continuously available.

Advantages of Genomic Selection

a. Once marker effects are estimated they can be used for a few generations

b. Selection is possible on novel traits, which require expensive phenol typing

c. New breeding strategies can be implemented

d. Increases genetic gain

– By increasing accuracy of selection

– By reducing the generation interval

e. Selects animals before they are of productive and/or reproductive age

f. Reduces/eliminates the need for progeny testing

g. Lowers the rate of inbreeding per generation

Limitations of Genomic Selection

a. Genotyping is still costly

b. Some species have no dense marker maps yet

c. When generation intervals are already low genetic gain due to Genomic selection will be less

d. In large litters accuracy can be gained from information on sibs. This yields lesser advantage to GS in pigs/poultry

e. New method, not fully proven and tested

f. Need to genotype a sufficiently large set of animals for accurate marker estimates

g. For traits of lower heritability, more records are needed

h. Marker estimates must be estimated in population that they will be used in

i. Across breed accuracy is low

j. If generation intervals are shortened substantially then annual inbreeding rates could be higher

Global Scenario of Genomic Selection

- **USA & Canada (N.A.) collaboration** - GEBVs obtained by USDA in collaboration with Canada for Holstein bulls have been released in public every year since 2008. A project at Guelph with 820 bulls was carried out with increased reliabilities of 8%, 5%, 18% and 8% for protein yield, fat yield, somatic cell count and conformation respectively for Genomic selection. (Pryce and Daetwyler, 2011)
- **New Zealand (LIC)** - LIC had the foresight to store DNA from every sire that was progeny tested since 1980. This enabled LIC to genotype sires that were the best, and the worst too, of their progeny test cohort and thus evaluate markers across the genetic range.

The degree of accuracy of GEBVs was measured by their correlation with Progeny test breeding values and was found to be ranging from 0.45 to 0.60 for production traits in HF breed. (Hayes *et al,* 2009)

- **Netherlands (CRV)** - CRV launches InSire bulls – designated to GS selected bulls, since 2008, for Holstein and Jersey breeds. Increases in the reliability using GEBVs were also reported by CRV for Genomic selection as 17%, 14% and 11% for protein production, overall conformation and somatic cell count respectively. (Hayes *et al,* 2009)
- **Australia (ADHIS & co.)** - ADHIS produced genomic based breeding values for bulls in September and December, 2010. Some 2,381 Holstein bulls were included in the September 2010 analysis (2,193 reference bulls and 188 young bulls). In this group of 188 young bulls with almost no daughter performance data, an improvement in reliability across all key traits was evident. Improvement in reliability in Australian breeding values (ABVs) from Genomic selection as compared to those from parent average increased from 21% to 53%, 14% to 42%, 8% to 36% and 12% to 46% for production traits, overall type, fertility and survival respectively. (Hayes *et al,* 2009)
- **Denmark & Sweden (Viking Genetics)** – Viking Genetics got the first genomic indexes of Holstein, Jersey and Red Breeds during the latter part of 2007. Today they have genomic indexes every two months for purposes of purchase of new bulls to their breeding programme. They assign the label of Gen-VikPLUS to their sires with best GEBVs.(Pryce and Daetwyler, 2012).

Genomic Selection in India

Extensive validation of association between genotypes and phenotypes is needed. GS cannot be used if accurate performance records are not available. Only genotyping without phenotyping and efficient data analysis will be wasteful expenditure. Additionally, policy makers need to be convinced of the potential of effective breeding programmes and potential of genomics. One Indo-Denmark Workshop on Genomic Selection in Cattle and Buffalo was held at National Agricultural Science Centre Complex, New Delhi, India from 11th to 12th April 2011. They suggested some recommendation as under for enhancing Genomic Selection in animal breeding. (Panigrahiand Parida, 2011).

1. Traditional selection tool to be combined with biotechnological tools for higher genetic gain per year.
2. Quality data and more no. of sire family arerequiring as per system.
3. Highly skilled manpower is needed for respective areas.

4. Basic studies to understand the interaction of genotype and environment and adaptive traits at genomic level is needed.

5. Evaluate gnomically tested exotic bulls under Indian condition.

6. Genomic Selection in India can be initiated by utilising reference population of Friesian crosses and indigenous cattle available in various institutional farms.

7. Collaborative programme on genomic analysis and capacity building is required.

8. Breeding strategies for genomic selection of cattle in organized herd need to be developed and tested.

9. Scientific exchange and joint collaborative research and training programme need to be conducted in the areas of Bioinformatics, association studies and Genome Wide Selection.

Possibilities in India

Large Non-Descript animals need to be graded up with High Genetic Merit Sires of known breeds and establishment of well-organized breeding network, performance recording system and reference population. In India initially globally available 50k SNP chip may be tried in indigenous breeds and meanwhile chip for indigenous animal can be developed. INAPH recording system of NDDB is coming up which can be used as performance records for Genomic Selection in indigenous breeds.

Conclusions

Indian dairy industry is uniquely suited for Genomic Selection. In terms of dairy cattle population there is waste genetic diversity in our country. Thus faster genetic progress is possible with higher accuracy and shorter generation interval. There are several countries implementing Genomic Selection in their breeding programmes. There are Hybrid systems merging Classic and Genomic Selection arising. Usefulness of Genomic Selection depends on the population structure, availability of dense marker maps and large reference population.

Future Prospectus

- Genotyping of more SNP to get clearer ‘picture’ of genetic variation
- Genotyping and get records for more animals
- Refinement and development of new estimation methods
- Accurate performance recording and reliable record keeping as well as precise genotyping initiatives will be required for developing nations like India

References

[1] Boichard, D. (2010). State of the Art of Genomics for Selection. UMR1313 *Animal Genetics and integrative biology*, F-78350 Jouy en Josas, France.

[2] Calus, M.P.L., Meuwissen, T.H.E., De Roos, A.P.W. and Veerkamp, R.F. (2008). Accuracy of Genomic Selection using different methods to define Haplotypes. *Genetics,***178(1)**: 553-561.

[3] Dekkers, J. (2009). Implementation of Genomic Selection in Egg Layer Chickens. *Joint ADSA-CSAS-ASAS Annual Meeting,* No. 34843.

[4] Hastbacka, J., De la Chapelle, A., Kaitila, I., Sistonen, P. and Waever, A. (1992). Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat. Genet.,***2**: 204-211.

[5] Hayes, B.J. and Goddard, M.E. (2001). The distribution of effects of data mining of genes affecting quantitative traits in livestock. *Genet. Sel. Evol.,***33**:209-229.

[6] Hayes, B.J., Visscher, P.M. and Goddard, M.E. (2009). Increased accuracy of artificial selection by using the realised relationship matrix. *Genetics Research,***91**: 47-60.

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

[8] Panigrahi, M. and Parida, S., (2012). Genomic selection – Revolutionary breeding practice in Domestic animals. *Vet. World,***5**(7):433-436.

[9] Pryce, J.E. and Daetwyler, H.D. (2011). Designing dairy cattle breeding schemes under genomic selection: a review of international research. *Animal Production Science,***52**(1): 33-41.

[10] Saatchi M., Miraei-Ashtiani, S.R, Nejati-Javaremi, A., Moradi-Shahrebabak M. and Mehrabani-Yeganeh H. (2010). The impact of information quantity and strength of relationship between training set and validation set on accuracy of genomic estimated breeding values. *Afr J Biotechnol,***9**:438-442

[11] Schaeffer, L.R. (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics,***123**: 218-223.

[12] Van der Werf, J.H.J. (2011). Potential benefit of genomic selection in sheep. Proc. Assoc. *Advmt. Anim. Breed. Genet.,***18**: 38-41.

[13] Van Eenennaam, A.L., Van Der Werf, J.H.H., and Goddard, M.E. (2011). The value of using DNA markers for beef bull selection in the seed stock sector. *Journal of Anim. Sci.,***89**: 307-320.

[14] VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science,***91**: 4414–4423.

[15] VanRaden, P.M. (2010). Genomic evaluation with many more genotypes and phenotypes. 9WCGALP, Leipzig, Germany.