

## Genomic selection - Revolutionary breeding practice in Domestic animals

Manjit Panigrahi, Subhashree Parida

Indian Veterinary Research Institute,  
Izatnagar - 243122 Dist. Bareilly, UP, India

Corresponding author: Manjit Panigrahi, email: manjit707@gmail.com

Received: 01-12-2011, Accepted: 24-12-2011, Published Online: 06-05-2012

doi: 10.5455/vetworld.2012.433-436

### Abstract

With more and more Single Nucleotide Polymorphisms (SNPs) being identified throughout the genome, some of those SNPs will be found to be located within candidate genes, allowing the researchers to use the candidate gene approach on a genome-wide scale. SNPs have some drawbacks but when compared to other markers they are more efficient and SNP consortium is growing to meet the requirements of genome-wide scans. Genomic selection should be able to at least double the rate of genetic gain in the dairy industry but the incorporation of genomic information into the breeding programs must be carefully considered. One needs to have around 2000 genotypes means a large reference population (population with both phenotype and genotype recorded) to achieve meaningful increases in accuracy. The available information, selection objectives, production circumstances and benefit/cost analysis must be evaluated in order to decide whether or not the population is suitable for GS implementation, and which would be the most convenient way, if any, for its implementation. Animal breeders will need to lead the way on the integration of genomic and phenotypic data into a new era of genome-enabled animal improvement and management.

Keywords: Breeding practice, Domestic Animals, DNA, Genome

### To cite this article:

Panigrahi M, Parida S (2012) Genomic selection - Revolutionary breeding practice in Domestic animals, *Vet World*, 5(7): 433-436, doi: 10.5455/vetworld.2012.433-436

### Introduction

In the past century DNA has moved from being a mysterious molecule only of interest to a handful of specialists to being the heart of a technology that is transforming many aspects of the way we all live. The study of human genetics has recently undergone a dramatic transition with the completion of both the sequencing of the human genome and the mapping of human haplotypes. In February 2001, two largely independent draft versions of the human genome were published. Both studies estimated that there are 30,000 to 40,000 genes in the human genome, roughly one-third the numbers of previous estimates. After human (2001) and Mouse, Whole Genome Sequence of the chicken (2004), Dog (2005), Bovine (2006), Horse (2007) and Pig (2009) are entirely available in the public domain. However, we still have to make guesses at the actual number of genes and the work assigned to them because not all of the genome sequence is properly annotated. Humans differ in genetic makeup by only 0.1%, but that small part of the genome contains the key differences that can

determine a person's susceptibility to disease or any adverse reaction to drug. Since the completion of the human genome project, there has been a huge effort in establishing maps of genetic markers, especially of the most abundant source of DNA variation, the single nucleotide polymorphism (SNP). A subsequent major quest is to identify the variation in the human genome causally involved in the genetic etiology of complex diseases and drug response. Studies to explore which genes are involved in a trait include essentially two major approaches.

### Candidate Gene approach

In the hypothesis based approaches the involvement of candidate genes and pathways is being investigated by testing for genetic association. A candidate gene is a gene, located in a chromosome region suspected of being involved in the expression of a trait such as a disease, whose protein product suggests that it could be the gene in question. A candidate gene can also be identified by linkage analysis to a region of the genome. In this approach, genotyping is targeted to

functional and positional candidate genes [1]. Functional candidates are genes that have been shown or are suspected to have a functional role in the phenotype of interest, for example, lipid lowering by a drug. Allelic variants of such genes may be causal for the observed natural trait variation. In this case, DNA polymorphisms located within the candidate gene or physically close to it will be associated with trait variation. To select the most promising candidates from a large number of functional candidate genes, gene sequences are tested for linkage to QTL (quantitative trait loci) for the trait of interest by molecular mapping, thereby identifying positional candidates [2]. The candidate gene approach mostly depends on and takes advantage of the knowledge available in the literature and databases on physiology, biochemistry and molecular genetics of a trait of interest. Association studies with candidate genes have been widely used for the study of complex diseases. However, this approach has been criticized because of non-replication of results and limits on its ability to include all possible causative genes and polymorphisms.

#### Genomic Selection

Alternatively, recently explorative approaches are applied which include genome-wide scans that first localize the gene(s) and subsequently identify the causal genetic variation at that locus. In this approach, populations of individuals related by descent are genotyped with DNA-based markers covering the whole genome, ideally positioned at regular intervals on the physical map. The genome-wide approach essentially requires a sequenced genome and a large number of markers and is therefore expensive and statistically complex [3]. Out of the different kinds of molecular markers, i.e. restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPDs), amplified fragment length polymorphisms (AFLPs), microsatellites and single nucleotide polymorphisms (SNPs), for the near future, SNPs are singled out as promising markers. SNPs (pronounced 'snips') are single base-pair variations in DNA. It is the most common source of genetic variation in the human genome, accounting for about 90% of sequence differences [4].

Now the question arises, why are SNPs used, as opposed to the more polymorphic microsatellite loci? The reason is subtle, namely mutation rates. SNPs have very low mutation rates and are closely spaced, and hence any decay in association between a SNP and linked QTL is entirely due to recombination. Such is

not the case for STRs, which have high mutation rates (often around 1/1000 to 1/250 per generation). Genome-scans with microsatellites will result ~350 markers with one marker every 5-10cM (1cM ~ 1000 bases) whereas with SNPs 3000 – 10000 markers can be obtained with one SNP in every 5Kb.

Genome-Wide Approaches can complement the candidate gene approach by providing relatively unbiased survey of the genome to identify previously unrecognized candidate genes. Genome wide scan is a very extensive and elaborate study method for determining the effects of various allelic variants occurring throughout the genome and the drug response in a disease condition [5]. This method involves identification of all the allelic variants in the entire human genome and the creation of a SNP map. The first major GWA (genome wide association) study in humans was published in Nature in February 2007 by *Sladek* and his coworkers [6] in a study searching for type II diabetes variants. The work was mainly carried out in the Genome Quebec centre at McGill University although it included collaboration with scientists of Imperial College, London and other research institutions. The group tested 3,92,935 single-nucleotide polymorphisms and identified several associations. Ironically, the developments in molecular genetics over the last 20 years or so have acted as catalysts to bring human and animal complex and quantitative trait genetics closer together. The development of molecular markers, first RFLPs, then microsatellites and now SNPs, have increasingly made the dissection of complex and quantitative traits feasible, into the effects of individual genes and networks of genes in both animals and humans. In human population studies and subsequently in the major livestock species, attention has been switched to association-based studies harnessing the linkage disequilibrium across the population that can be found with dense marker panels (upto 1 million SNPs in humans compared to around 60 thousand currently used in cattle and pigs).

#### Application of Genomic selection in Animal breeding

Animal breeding, one can say, is at the crossroads. A range of new tools and technologies is becoming available and this may change the methods used and the organization of breeding systems. In the last few years, it has been witnessed that animal breeding and genetics evolves and expands from substantiation of basic population genetics to development of quantitative genetics leading to today's relatively sophisticated

genetic prediction systems. Genomic selection, a variant of marker assisted selection (MAS) has become one of the main areas of interest throughout the world. It is well known that most economically important traits for livestock are quantitative traits and in traditional breeding, breeders improve these traits by recording phenotypes and pedigrees. But, for traits with low heritability, sex-limited traits, longevity traits, and traits that are difficult or expensive to measure (such as carcass traits); traditional selection methods have been of little use or are inefficient. With the advance of molecular genetics in the past 20 years, genetic markers have been used in animal breeding programs. These new technologies can change both the rates and systems of genetic change, by increasing accuracy, intensity or timeliness of selection. Unlike traditional MAS which uses only the few markers that are thought to have large effects on the underlying trait of interest to predict breeding values, Genomic selection (GS) simultaneously uses high density (HD) markers (particularly SNPs) covering the whole genome to predict the breeding values of each genotyped individual [7]. A breeding value predicted in this way is termed the genomic estimated breeding value (GEBV). In GS, it is assumed that each gene or QTL (quantitative trait locus) affecting the trait of interest is in linkage disequilibrium (LD) with at least one of the markers. Theoretically, all genetic variance can be tracked by markers if the marker density is high enough. Hence, GS has shown to overcome the drawback of traditional MAS and predict breeding values more accurately. Such experiments had used reference populations of between 650 and 4,500 progeny-tested Holstein-Friesian bulls, genotyped for approximately 50,000 genome-wide markers. Reliabilities of GEBV for young bulls lacking progeny testing results, in the reference population were found to be between 20 and 67%.

To calculate GEBV, a prediction equation based on the SNP is first derived. A variety of methods has been suggested for the calculation of GEBV, ranging from BLUP, Bayesian methods such as BayesB [8], and machine learning techniques. These methods differ in their assumptions about the underlying genetic model: BLUP assumes the infinitesimal model, i.e. large numbers of genes each with small effects are scattered along the chromosomes; the machine learning technique assumes that there are a limited number of genes, and thus also a limited number of SNPs, worthwhile to be fitted; and BayesB is in between, i.e. it assumes few genes with large effects and many genes with small effect.

The application of GS in dairy cattle has been reported in many countries, including USA, Canada, Australia, Norway, New Zealand, Netherland, Denmark, Germany and Ireland. The traits currently being evaluated using genomic information include all the target traits that are included in the traditional breeding programs. A reasonable explanation is that the bovine DGAT1 gene located on chromosome 14 explains a considerable proportion of the genetic variance for fat percentage. This would result in an obvious deviation from the infinitesimal model for this trait which might favor BayesB. Different types of chips based on the use of SNPs are also available for livestock and can be used to achieve different objectives. The bovine 50 K SNP chip has become the standard tool for breeding industries in dairy cattle, and a higher density chip 800 K SNP is also available to screen for more genes and for a deeper implementation of genomic selection. A smaller and less expensive 3 K SNP chip is also available now to screen large populations [9].

India is blessed with huge livestock population. Apart from varied climatic conditions, it has a vast genetic diversity which exists as a result of 3.5-4.5 billion years of evolution. Domesticated cattle are the descendants of groups of races of *Bos primigenius*, known as 'urus' or 'auroch' and thought to be domesticated around 6000 B.C. India, the goldmine and vast reservoir of germplasm of indigenous cattle accounts for 15% of world's cattle population. It has approximately 19.9 crores of cattle, comprising of 16.6 crores indigenous cattle and 3.3 crores crossbred cattle. It is also the largest producer of milk in the world. This enormous milk yield can be attributed to both indigenous and crossbred cattle population of the country. In fact, in this country indigenous cattle are able to perform better under harsh climatic conditions, are resistant to diseases and able to convert coarse and dry fodder into valuable animal products.

However, about 75-80% of total indigenous animals are nondescript and hence remain undocumented. This is in sharp contrast to developed countries where each animal is known for its pedigree. Livestock development in India is mostly characterized by policy inadequacies, lack of political will and vision, adoption of programmes without any long-term impact and wasteful government expenditure on subsidies. The almost total absence of effective genetic improvement and dissemination programmes is a fundamental constraint to the development of the livestock sector in India.

Keeping all these factors in consideration

recently one Indo-Denmark Workshop on Genomic Selection in Cattle and Buffaloes was held at National Agricultural Science Centre Complex, New Delhi, India from 11<sup>th</sup> to 12<sup>th</sup> April 2011. The purpose of the workshop was to exchange knowledge and enhance cooperation between India and Denmark, focusing on the potential and feasibility of genomic selection applied to animal breeding.

Recommendations of the meeting:

- Traditional selection tools need to be combined with biotechnological tools to have higher rates of genetic improvement. Phenotypic and SNP data have to be combined for selection of animals using multivariate analysis for genomic selection.
- Quality data is a key for the success of genomic selection. More numbers of sire families need to be developed besides improvement in the accuracy of performance of recording systems.
- Highly-skilled manpower in the area of bioinformatics, database development and management need to be developed under the exchange programme.
- Basic studies to understand Genotype and Environment interactions and adaptive traits at genomic level need to be undertaken.
- It would be necessary to evaluate performance of genomically tested exotic bulls under Indian conditions.
- Genomic selection can be initiated in India by utilizing reference population of Friesian crosses and indigenous cattle available in various Institutional herds.
- Collaborative programme on genomic analysis and capacity building needs to be undertaken.
- Breeding strategies for genomic selection of cattle in organized herds need to be developed and tested.
- Scientific exchange and joint collaborative research and training programmes need to be conducted in the areas of Bioinformatics, association studies and Genome Wide Selection.

#### Conclusion

With more and more SNPs being identified throughout the genome, some of those SNPs will be found to be located within candidate genes, allowing the researchers to use the candidate gene approach on a genome-wide scale. SNPs have some drawbacks but when compared to other markers they are more

efficient and SNP consortium is growing to meet the requirements of genome-wide scans. Genomic selection should be able to at least double the rate of genetic gain in the dairy industry [10] but the incorporation of genomic information into the breeding programs must be carefully considered. One needs to have around 2000 genotypes means a large reference population (population with both phenotype and genotype recorded) to achieve meaningful increases in accuracy. The available information, selection objectives, production circumstances and benefit/cost analysis must be evaluated in order to decide whether or not the population is suitable for GS implementation, and which would be the most convenient way, if any, for its implementation. Animal breeders will need to lead the way on the integration of genomic and phenotypic data into a new era of genome-enabled animal improvement and management.

#### References

1. Pflieger S., Lefebvre V., and Causse M. (2001). The candidate gene approach in plant genetics: A review. *Mol. Breed.*, 7: 275–291.
2. Pajerowska K.M., Parker J.E., and Gebhardt C. (2005). Potato homologues of *Arabidopsis thaliana* genes functional in defense signaling - Identification, genetic mapping and molecular cloning. *Mol. Plant. Microbe Interact.*, 18:1107–1119.
3. Hirschhorn J.N., and Daly K.J. (2005) Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.*, 6:95–108.
4. Collins F.S., Patrinos A., Jordan E., Chakravarti A., Gesteland R., Walters L. (1998) New Goals for the U.S. Human Genome Project. *Science*, 282:682–689.
5. Zhang Z., Zhang Q., Ding X.D. (2011). Advances in genomic selection in domestic animals. *Animal Genetics*, 56 (25): 2655-2663.
6. Sladek R., Rocheleau G., Rung J. (2007). A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*, 445 (7130): 881–5.
7. Goddard M.E., Hayes B.J. (2007). Genomic selection. *J. Anim. Breed. Genet.*, 124: 323–330.
8. Meuwissen T.H.E., Hayes B.J., Goddard M.E. (2001) Prediction of total genetic value using genome wide dense marker maps. *Genetics*, 157: 1819–1829.
9. Eggen A. “Different chips for different population segments” in Proceedings of Interbull Meeting, Riga, UK, 2010, <http://www.interbull.org/>
10. Hayes B.J., Bowman P.J., Chamberlain A.C., Verbyla K., Goddard M.E. (2009). Accuracy of genomic breeding values in multi-breed dairy cattle population. *Genet Sel Evol.*, 41- 51.

\* \* \* \* \*