Introduction:
Selection based on phenotypic and pedigree information has been tremendously effective in improving the dairy animals for milk production, milk components and reproductive traits for more than 40 years. With the advent of application of Marker assisted selection (MAS) in dairy cattle we could preselect young candidate bulls prior to progeny testing, thus increasing the selection differentials, shortening generation interval and increase the rate of genetic gain. Once a quantitative trait locus (QTL) is identified, it is necessary to identify families in the breeding population which are segregating that QTL. If a QTL has been fine mapped with respect to closely linked markers that are in linkage disequilibrium with the QTL, the associations between specific marker haplotypes and QTL alleles should hold across populations and need not be re-established for individual family. Selection of such QTL can be undertaken throughout the population rather than only specific families, thereby simplifying the process of MAS. The availability of these genetic maps of cattle has allowed the whole genome to be evaluated for QTL with major effect.

Genome-wide association study (GWAS):
This is a DNA based marker technology wherein it is possible to identify genome regions (QTLs) under lying complex traits. Quantitative trait loci method of genetic evaluation has greater potential for selection than depending on phenotype and pedigree (Khatkar et al., 2004). It is a process for inspection and screening of detectable common genetic variants (single-nucleotide polymorphisms) in individuals to identify the variant(s) associated with the trait under study. GWAS is based on the premise that a causal variant is located on a haplotype, and therefore a marker allele in Linkage Disequilibrium with the causal variant should show an association with a trait of interest (Zhao et al., 2007).

Basic applications of GWAS in molecular animal breeding:
- Has the power to detect causal variants with modest effects.
- Has the power to define narrower genomic regions harbouiring causal variants.
- It is accepted as the primary approach for gene finding and has high success in identifying disease genes.
- It has more power to detect QTL and provides more precise estimates of QTL locations compared to linkage studies.
- To associate a trait with a region in the genome, in order to map the clinically and/or economically important QTLs.

GWAS in livestock:
The GWAS has been successfully extended towards animal science, for detecting the differentially expressed genes as well as identification of the central key gene(s) underlying the trait(s) of interest. It encompasses disease tolerance/susceptibility, production vis-a-vis reproduction traits and growth traits, as well. The first whole genome scans in dairy cattle were initiated by Georges et al. (1995). Since that time several QTL and candidates genes for milk production, reproduction, functional, and conformation traits have been described for several Bos taurus autosomes and most of these regions have been mapped in multiple studies (Ashwell et al., 2004; Schnabel et al., 2005).

The association studies are conducted by studying the genetic polymorphic of many loci across the genome and associating it with various phenotypic traits.

Experimental strategies to study the genetic polymorphism:

1. **Linkage studies:** These rely on genetic map knowledge, searching quantitative trait loci (QTL) by using family information and comparing segregation patterns of genetic marker and the traits being analyzed. For instance, milk production traits in dairy cattle was estimated to be controlled by 150 QTL (Hayes et al., 2005), but there are even more QTLs because the power to detect these QTL was not 100% (Goddard and Hayes, 2007).

2. **Candidate gene approaches**

   a. **QTL based candidate gene sequencing**
   At present, direct nucleotide sequencing is one of the high throughput methods for mutation detection, and is the most accurate method to determine the exact nature of a polymorphism. Thus it has only limited utility when the polymorphism is present in a minor fraction of the total DNA (in case of pooled samples of DNA) because of low sensitivity.

   b. **Microsatellites marker analysis:**
   The microsatellites (or SSRs, Simple Sequence Repeats) have proven to be very useful for the purpose of unveiling genetic diversity in animal as well as plant species (Ashwell et al., 2004; Schnabel et al., 2005). Microsatellite markers have been commonly used for genetic mapping, linkage analysis and to trace inheritance patterns. It is assumed that extensive polymorphisms observed with respect to microsatellites results in causing variation in the phenotype. Hence this concept of microsatellite variation is used in evolutionary studies and measures the amount of gene flow. Microsatellite markers were selected at approximately 20-cM intervals from published bovine maps (Ashwell et al., 2004) for detection of QTL affecting milk production, health and reproductive traits in Holstein cattle.

   c. **Single Nucleotide Polymorphisms (SNPs)**
   A SNP (single nucleotide polymorphism) is just a single base change in a DNA sequence, with a usual alternative of two possible nucleotides at a given position. For such a base position with sequence alternatives in genomic DNA to be considered as an SNP, it is considered that the least frequent allele should have a frequency of 1% or greater. The alignment of multiple sequence fragments of the samples taken into study and representing the same region on the genome will allow for the discovery of sequence variants.

In general, association studies have to be performed in order to statistically establish that particular alleles are associated with one or more phenotypic traits (Mullen et al 2011). SNP have been used for the detection and localization of QTL for complex traits in dairy cattle (Hayes et al., 2010). For large scale identification of validation and analysis of genotypic variation in cattle, the SNP genome wide data base comprising of set of SNP markers spanning the whole bovine genome can be used. (Bovine Genome Project. http://www.hgsc.bcm.tmc.edu/projects/bovine)

**Methods of SNP analysis**
1. **SNP association:** A method for determining the type of base present at a given SNP locus (Allele discrimination) Plots the frequency on a graph and gives the results like how many are T/T, C/T, b/C/C? It gives the frequency percentage. It also gives the allelic frequency T and C.
2. **Method for reporting the presence of the allele (signal detection). This is done by**
3. Detecting linkage between a single marker and a QTL. QTL detection in a genome:

4. Principle of QTL detection in a genome:
1. Large proportion of QTL’s is likely to be generated by mutations outside the protein coding regions.
2. Number of genes underlying quantitative traits and their distribution effects. The hypothesis that the majority of the genetic variation in quantitative trait is controlled by few genes of large effect. This is supported by a number of cases where the causative mutation underlying the QTL effect has been discovered. Eg. DGAT I gene- responsible for 43% of genetic effect in fat % in HF (Grisart et al., 2002).
3. Detecting linkage between a single marker and a QTL. QTL detected using neutral molecular markers which are randomly scattered throughout the genome. Variation at the markers genotype accounts for the phenotypic variation in the quantitative trait.
4. Two marker QTL linkage mapping. The power of QTL detection is increased by knowing the flanking marker.
5. Interval mapping for inferring location and effect of QTL’s in the genome. When the genome scan is continued the existence of QTL’s at regular interval is not segregating in the population.

Two methods are used to detect this:

a. Maximum likelihood method: This method proceeds by evaluating each of the putative QTL position. This gives the likelihood of odds (LOD score) for each putative QTL position.

b. Regression approach: Phenotype data are regressed on the QTL allele probabilities at each putative location as inferred from flanking markers.

Several statistical approaches have been developed for whole genome scans and QTL mapping projects, including the least square method, based on regression of phenotype on marker genotypes or haplotypes, and random effects models based on identity by descent (IBD) approaches (Kolbehdari et al., 2005). High density SNP marker genotypes have increased the feasibility of QTL detection and mapping using historical population-wide linkage disequilibrium (LD). LD mapping method requires a marker allele to be in LD with the QTL allele across the entire population.

QTL mapping experiments for genome scanning for quantitative trait loci

Family Based designs

Design 1: Crossing of inbred lines or divergent lines

By crossing two inbred lines we can create an F 1 which is in linkage disequilibrium between markers and QTL’s. Random mating of F 1’s results in F 2’s which are in high level of linkage disequilibrium. In case if the original parental lines are not inbred and they belong to different breeds that is they are not homozygous for majority of loci may result in lesser segregation of markers and QTL’s in F 2. For the analysis ML method or regression method is used for both phenotypic and genotypic data.

The power of this design depends on

i) Number of offspring’s studied

ii) Size of the QTL effect

iii) Degree of the QTL dominance

The software, QTL Express (http://qtl.cap.ed.ac.uk) analyses data from F2, half sib and sib pair families to detect QTL.

Design 2: Half sib design

This is done in species where the male’s reproductive capacity is exploited wherein number of half sibs is produced. Marker alleles of the sire are evaluated for linkage to putative QTL’s. Unrelated sires are mated to unrelated dam1, dam 2 and dam 3 etc. and their half sib families are analysed by ML or regression method for

i. Most likely linkage phase of genetic markers for each sire

ii. Probabilities of inheriting each of the two sire gametes are calculated for fixed position along the chromosome, conditional on the marker genotypes of the progeny.

The power of this design depends on

i) Number of sires and half sibs per sire.

ii) At least one sire used should be heterozygous at QTL

iii) Large sire families.

Eg: GWAS studies for QTL’s of milk production (Georges et al., 1995)

Design 3: Complex pedigree designs

To detect a QTL this explains 20% of the genetic variance for a trait. 8000 sib pairs (2 sibs in each pair). Extra power is gained if extended pedigrees can be genotyped. The pedigrees should be more distant in relationship than the parent. Analysis is done by regression method (Georges et al., 1995), ML method (Hoeschele et al., 1997) and Variance component method (Georges et al., 2000).

Steps:

1. For each QTL positions on the chromosome segment calculate the covariance matrix associated with QTL

2. For each position considered in step 1 construct linear model to estimate QTL variance and other parameters to test the presence of QTL.

Resources for QTL detection in dairy cattle

1. Granddaughter design (GDD) (Weller et al., 1990) and Daughter design (DD). For a DD, genotypic information is
recorded for sires and their daughters, with phenotypic observations made on the daughters. For a GDB, the grand-sires and sires are genotyped, and phenotypic observations are made on the grand-daughters (Liu et al., 2010).

3. Number of heterozygous QTL per sire.
4. Within sire segregation variance.
5. Total number of QTL segregating in the population.

**Precision of QTL mapping:**

i. Likelihood ratio test: This is performed at any position covered by markers across the whole genome. The location with the highest likelihood is the most likely putative QTL position.

ii. To improve the mapping precision is to increase the marker density on the chromosome.

iii. Linkage dis-equilibrium mapping.

To increase the power of QTL:

1. Selective genotyping: QTL mapping in which the analysis of linkage between markers and QTL is carried out by genotyping individuals from the high and low phenotypic traits of trait distribution.

2. Selective DNA pooling: Determination of the linkage between a marker and QTL is based on the distribution of alleles among the pooled DNA samples of the extreme high and low phenotypic groups. Two pools to be genotyped to detect within family contrast. Replicate the pools to increase the accuracy.

3. LD mapping: This is mapping of non random association of alleles between two loci. This requires marker alleles to be in LD with a QTL allele across entire population. Association persisted for considerable number of generations so that marker and QTL are closely linked. This is based on the principle that ancestral chromosome segment is conserved and these regions are identical by descent (IBD).

4. Progeny testing:

5. Interval mapping.

6. Simultaneous searching for multiple QTL

7. Study of disease tagged QTL

8. Combined LD-LA mapping (Meuwissen and Goddard 2001)

While investigating LD in livestock, LD was not only highly variable across any particular chromosome but also there was also significant disequilibrium between alleles of loci located on different chromosomes. LD information when combined with linkage information to filter any spurious LD likelihood peaks. This type of QTL mapping is referred to as LD-LA. The analysis proceed by constructing IBD coefficients between the haplotypes of founder animals, and a second matrix describing the transmission of QTL alleles from the founders to later generations of the genotyped animals. A variance component approach can then be used to calculate the likelihood of QTL’s at each putative position along the genome. When LD-LA is performed, both linkage and LD information contribute to the likelihood profile. Any peaks due to LD or linkage alone are filtered from the profile.

**Transmission disequilibrium test (TDT)**

This is a family based linkage disequilibrium test that offers a powerful way to test for linkage between alleles and phenotypes that is either causal (i.e. The marker locus is the trait locus) or due to linkage disequilibrium. When the marker is extremely close to the trait locus itself the association tests such as TDT can be far more powerful than linkage tests. Given a locus with allele ‘a’ and ‘A’ and the assumption that parental alleles assort independently during gametogenesis, this creates a situation in which the ‘a’ versus ‘A’ alleles are randomly assigned to offspring with equal probability for all offspring in the sample (probability “unless there is segregation distortion). Any association between the offspring phenotype and the ‘a’ allele versus the ‘A’ allele must be due to

i. Either the allelic variation directly causes the variation.

ii. The allelic variation indirectly causes variation in the phenotype through an intermediary phenotype.

iii. The marker locus is in LD with a locus causing (either directly or indirectly) variation in the phenotype (Jiang et al., 2010).

**Uses of QTL mapping approaches:**

1. Reveals gene-gene interaction in biochemical, metabolic, regulatory and developmental pathways.

2. Statistical analysis allows identification of QTL or genomic regions which affect the level of expression of particular genes.

3. Combining gene expression data with linkage analysis can accelerate the identification of the causative mutation underlying QTL effects (unravel at least part of gene regulatory networks controlling gene expression).

**Preference of SNPs in genome wide association studies:**

1. Single nucleotide polymorphisms (SNPs) are ubiquitous polymorphic markers and uniformly distributed throughout genome. The SNP detection technique gradually proceeded from single or a group of candidate genes towards whole genome based technique (Zhang et al., 2007).

2. Discovery of novel SNPs has led to the creation of a public repository and freely available data base called dbSNP harbored by the National Center for Biotechnology Information (NCBI; dbSNP) for all species.

**BovineSNP50 v2 DNA Analysis BeadChip**

The BovineSNP50 v2 BeadChip contains 54,609 highly informative SNPs uniformly distributed across the entire genome of major cattle breeds, empowering applications such as genome-wide enabled selection, identification of quantitative trait loci, evaluation of genetic merit of individuals, and comparative genetic studies. This BeadChip was developed by Illumina in collaboration with the USDA-ARS, University of Missouri, and the University of Alberta. More than 24,000 SNP probes target novel SNP loci that were discovered by sequencing three pooled populations of economically important beef and dairy cattle using Illumina’s Genome Analyzer.

- This 24-sample BeadChip represents the highest density genotyping solution for characterizing the genome in dairy and beef cattle.

- The PCR-free, single tube sample preparation significantly reduces labor and potential sample handling errors.

- Laboratory Information Management System (LIMS) and robotic automation are available to accurately and efficiently track samples throughout analysis.

**Different traits studied for QTL mapping**

In dairy cattle, since the seminal work on QTL mapping by Georges et al.(1995), a number of articles have been published concerning detection of QTLs for milk yield (MY), milk protein yield (PY), milk fat yield (FY), milk protein content (%)(PP), milk fat content (%) (FP). So far a total number of 813 QTL underlying milk production traits have been detected via genome scans based on marker-QTL linkage analyses (http://www.animalgenome.org/QTLdb/cattle.html).

Genome wide association studies in Danish Jersey cattle (Mai et al., 2010) for milk production traits revealed 7 QTL’s for milk index in BTA 4 and 5 and 21 QTL’s for fat index and protein index in BTA 4,5,13, 30 and 29. These QTL’s were located in the gene loci of DGAT1, Casein, ARFGAP3, CYP11B1 and CDC like kinased genes. Similar studies in German Holstein cattle (Wang et al., 2012) revealed 4 and 2 QTLs for fat percentage in BTA 14 and 20 respectively where the loci for DGAT1 and GHR genes were reported. Various studies conducted in Dutch Holstein cattle (Bowman et al., 2011, Schopen et al., 2010) revealed QTLs in BTA 14,19 and 29 for fatty acids in milk and the QTL’s for milk protein genes were located in BTA 13, 14, 19 and 26. Studies on Casein index in dairy cattle (Visker et al., 2011) revealed marker locus at BTA 6, 7 and 11. 105 SNP’s associated with amino or more milk production traits was reported in Chinese Holstein cattle (Liu et al., 2010, Jiang et al., 2010). QTL’s affecting milk production and reproductive traits were observed in Chromosome 3, 6, 14, 18 and 29 in Holstein cattle (Ashwell et al., 2004, Schnabel et al., 2005).

Two SNPs associated with BCS, one SNP for Milk production traits
were reported by Mullen et al., (2011) in the studies conducted for growth hormone genes and IL6 genes association with milk production traits. QTL for milk production traits in Norwegian cattle were found in BTA 3, 5, 6, 11, 18 and 20 (Olsen et al., 2011).

Chromosome 22 showed significant associations for calving traits (Johanna et al., 2010). Genome wide studies conducted in Holstein Friesian cattle for susceptibility to Bovine tuberculosis (Emma et al., 2011) revealed 3 SNPs on BTA 22. Studies on calving traits like calving ease and growth related traits in German dairy cattle (Hubert et al., 2010) revealed two SNPs on BTA 14 and 21.

In a study conducted for complex traits like coat colour, milk fat percentage and type traits in Holstein cattle (Hayes et al., 2010) 3 loci for variation in black coat colour were identified. Five SNPs located on BTA 1, and one SNP in each BTA 3, 12, 13, 16, 18 helped to identify genes involved in physiological responses to heat stress (Dikmen et al., 2012).

Challenges in implementing GS programme in India:
In India the challenges are double-Edged. Firstly, the "phenotype gap" due to limited phenome data generation, secondly, lack of integrated DNA bank. This holds true for almost all domestic animal species. The data generation and curation with integrated DNA sample is a herculean task. The predictor and predicted reference population of more than 4000 sample size at least, is required in order to establish the GEBV in each major breed. SNP-Chip based selection of animals is still in its inception. Under Indian circumstances, the indigenous animals should be thoroughly screened for the SNPs. The accuracy of the predicted GEBV is also affected by the type of trait (viz. monogenic or polygenic), and coverage of markers in the model. Whole genome scanning with the help of SNP-Chip will enable the livestock-owner to foretell the genetic worth of their animals at the neo-natal stage. Selection will be based on a prediction equation derived from a reference population that has extensive phenotypic recording and genotypic data. To take maximum advantage of the genomic selection, generation intervals will be shortened as much as reproductive technology will allow. The time has come and we have to gear up before it gets further delayed or never!

REFERENCES: