

UTILIZATION OF MOLECULAR TECHNOLOGY IN THE GENETIC IMPROVEMENT OF LIVESTOCK

A. Sabry

SUMMARY

This paper presents a general overview of the principles of using molecular technology in the genetic improvement of livestock, with special emphasis on dairy cattle. The developments in molecular biology and statistics as well as the paralleled improvement of computation speed have allowed possibility of mapping and identifying areas on the genome that give raise to phenotypic quantitative variation. These genomic areas are coined as Quantitative Trait Loci (QTL). Application of molecular technology in the genetic improvement of livestock involves (1) developing of DNA markers, (2) construction of genetic map, (3) QTL mapping, (4) QTL fine mapping, (5) QTL characterization and cloning, (6) QTL expression, (7) marker assisted selection (MAS). Strategies and limitation of such application are discussed; moreover, examples of trait loci for which the causative gene and mutation have been identified are also reviewed. The mapped QTLs introduce a powerful tool in genetic improvement of livestock through marker assisted selection.

INTRODUCTION

The discovery of DNA markers (i.e. a small DNA segment with easily detectable sequence polymorphism in different individuals within the species) starts a new era of DNA technology in biology. That is, now a given phenotype can be traced back to its causative genetic factor(s). In agriculture these DNA markers open new possibilities to identify Quantitative Trait Loci (QTL i.e. chromosomal segments identified by statistical analysis of complex traits) affecting low-heritability traits such as disease resistance / susceptibility and reproduction traits.

Animal breeders are keen on utilizing QTL in breeding programs. Information about QTL might be used to improve the efficiency of breeding programs and to give an efficient tools for more rapid genetic improvement, providing breeding companies with a competitive asset. Finally open the 'black box' which allow more realistic models for phenotypic variations, response to selection and evolutionary process.

One approach to identify QTL is the genome-wide QTL scan. That is a process of symmetrically scanning the entire genome of a mapping population (e.g. crosses between inbred lines, full-sib or half-sib families) with regularly spaced DNA markers whose exact location is known. Guided with phenotypic recodes, one can identify a genetic regions associated or 'in linkage' with the phenotype of interest, by observing the affected individuals share certain marker allele located in those regions more frequently than expected by chance. The objective of this work is to present a general overview of the principles of genome-wide QTL scan. The major steps for genome-wide QTL scan are: (1) developing of DNA markers, (2) construction of genetic map (3) QTL mapping (4) QTL fine mapping.

1. DNA Markers

At the molecular level, chromosomes are composed of DNA and proteins. The DNA carries the genetic information while protein components provide enzymatic and structure functions important for replication, recombination and segregation of the chromosome.

The main part of the chromosome contains a mixture of coding sequences (exons), intragenic (introns) and intergenic non-coding sequences, regulatory sequences, tandem repeats (mainly minisatellites and microsatellites) and other dispersed repeats.

DNA markers are often separated into two groups: Type I loci and Type II loci.

Type I loci functional genes often only slightly polymorphic and are not considered immediately useful for QTL maps. However, functional genes are evolutionary conserved across species and therefore, essential for comparative mapping.

Type II loci highly polymorphic species-specific DNA markers that are usually not evolutionary conserved across species.

RFLPs based on nucleotide changes that occurs as a results of creation or abolition of a restriction endonuclease recognition site, Then DNA acquires or loses the ability to be cleaved by a particular restriction endonuclease. Because of diallelic nature of RFLP markers they are not useful for QTL mapping in animal species individuals will be heterozygous.

Microsatellites Repeatitive DNA with repeats ranging from 1–6 bp. Microsatellites are highly polymorphic, abundant and often found in non-coding regions of the genes. Microsatellites are The most commonly used marker in QTL mapping.

SNPs are defined as a base pair location in which the frequency of the most of the most common base pair is less than 99%. Unlike Microsatellites which is usually multiple alleles SNPs are diallelic. SNPs are prevalent and provide more potential markers near or in any locus of interest than other types of polymorphic markers. Some SNPs are located in in coding regions and directly affect the protein function. SNPs are more stably inherited than microsatellites. SNPs are more suitable for than microsatellites for high throughput genetic analysis, using DNA microarray technology.

A comprehensive review of DNA markers can be found in Liu, 1998, Montaldo and Meza-herrera, 1998, Cunningham and Meghan, 2001 or Beuzen *et al.*, 2001 DNA markers are classified into Type I loci and Type II loci.

2. Linkage map

The convention has been to divide genome mapping methods into two categories that is, genetic (linkage) map and physical maps.

Genetic (linkage) mapping is based on the use of genetic techniques to construct maps showing the positions of genes and other sequence features on a genome. Genetic (linkage) map is a linear arrangement of a group of genes and markers. A linkage map is based on homologous recombination during meiosis. Alleles of linked markers travel together during meiosis process. Linkage map is constructed by

following the recombination events in genetically defined population. In classical genetics studies linkage maps are constructed by conducting a crossbreeding experiment such as intercrossing or backcrossing to follow the recombination events on the bases of alleles passed from parents to offspring.

As a result of extensive use of artificial insemination in animal breeding programs, the structure of some farm animal populations, such as cattle, is based on large numbers of half-sib families. Recombination events can be followed by genotyping parents and their offspring for markers alleles passed on from parent(s) to offspring. In dairy cattle, for example, half-sib sire groups are much larger than the number of offspring per dam. In order to construct a male genetic map, it is necessary to genotype the sires and their half-sib sons.

Cattle have of the order of 3800 markers mapped (<http://www.marc.usda.gov/genome/cattle/cattle.html>) with other livestock species not far behind. This gives a density of markers of approximately one every 1 cM on average.

Physical mapping uses molecular biology technique to examine DNA molecules directly in order to construct maps showing the positions of sequence features including genes.

3. Mapping Quantitative Trait Loci (QTL)

The aim of QTL mapping is to identify chromosomal regions, which contain genes that affect quantitative traits (traits affected by many genes and environmental factors). In order to do so, genetic markers with known positions are identified and associated with phenotypic records.

3.1. Experimental design

The most efficient designs for QTL detection is crosses of inbred lines, because the linkage disequilibrium between marker and QTL is maximized in F1. Therefore, these design are often used in plants but in large animal species such as pig and cattle it is not possible to obtain the inbred lines.

Instead crosses of outbred lines can be used to identify QTL that explain between line differences. The efficiency of this design depends on the difference between the two lines in marker allele frequency, and difference in genetic effects between predominant QTL genotypes of the two lines.

However, to identify QTL within a line designs must take advantage of the structure that exists in the commercial populations. In cattle proven bulls have many male offspring, which in turn are tested by having many female offspring, which are scored phenotypically for traits that are of interest in the breeding scheme. In this structure two designs, daughter design and grand daughter design, Weller et al. (1990) can be used for QTL detection. In the daughter design tested sires and their daughters are typed for genetic markers. Differences in the phenotypic group mean of daughters inheriting alternative alleles from the heterozygote sire, indicates that there is a QTL in the region. In the grand daughter design the proven bulls and their sons are tested for the genetic markers, while daughters of the sons contribute with their phenotypic records to produce a daughter yield deviation (dyd) record for each son. Here, differences in the dyd group means of sons inheriting alternative marker alleles from the heterozygote proven bull indicate there is a QTL in the region. The grand

daughter design is often used in QTL scans, because less genotypes (about a third) are necessary to obtain the same statistical power compared to the daughter design.

In the granddaughter design, genetic relationships between bulls in different families are largely ignored. Since many relationships exist between families this is bound to affect the analysis. It is therefore desirable to include direct relationships between sires and through the dams. This design could be called the complex grand daughter design.

3.2. Statistical methods

Hoeschele *et al.* (1997) have provided an excellent review of the alternative approaches to QTL analysis for outbred populations.

Least squares methods are the simplest methods used in QTL detection. The model assumes QTL genotype effects to be fixed, and regresses probability of the possible genotypes given marker information on the phenotypic data.

In the Maximum Likelihood method variance component models assume random QTL effects. The covariance structure between the QTL effects are modeled using an identity by descent (IBD) matrix, containing probabilities that each combination of alleles are identical by descent given the observed marker genotypes. Polygenic variation not explained by the QTL effects can be modeled as random effects. This model is appropriate for simple and complex designs.

Bayesian linkage analysis is very flexible in terms of the complexity of the QTL model, number of QTL, interaction between QTL, polygenic variation, robust distributions for error variance, and is appropriate for simple or complex designs.

4. QTL Finmapping

The resolution of mapped QTL using conventional linkage based methods is usually low, perhaps 20 cM or more, depending on the number of meioses available (recombinant individuals provide the information about the QTL location) and the magnitude of the QTL effect. Even though for closely linked markers, the number of the recombination events between adjacent markers within these two or three generations is limited and thus dense map will not adding much information, simply because the confidence interval of QTL location will still too large to move to the next step of marker assisted selection and even to identifying the actual genes (Weller, 2001). However, linkage disequilibrium provides an efficient tool that can pinpoint gene location down to less than 1 cM.

Linkage disequilibrium is defined as non-random association of alleles at different loci into gametes. Linkage disequilibrium can exist between alleles of both syntenic and nonsyntenic loci. With linkage disequilibrium, alleles at marker locus and an associated trait locus are non-randomly associated across the population and not just only within families. A number of factors can generate linkage disequilibrium, and several of them are can operate simultaneously:

- (i) population expanding from small number of founders.
- (ii) linkage disequilibrium arises from population admixture when allelic frequencies differ at two loci across subpopulations.
- (iii) stabilizing and directional selection will result in negative linkage disequilibrium as a result of Bulmer effect.
- (iv) positive linkage disequilibrium will occur between characters under

disruptive selection.

(v) if loci are interacting epistatically, allelic combination that favoured by selection will also be at higher allele frequencies than expected.

When a mutation occurs, or a new allele is introduced to the population, all other polymorphic alleles in that population will be initially in a complete LD with the mutant allele. Over time, however, recombination between the mutant allele and the other loci result in decay of linkage disequilibrium and restore linkage equilibrium between the mutant allele at all but the closely linked loci. The length of the genomic fragment (in cM) surrounding the original mutation in which linkage disequilibrium between the QTL and other loci still exists depends on the average amount of the recombination per generation that occur in this segment of the genome and the number of generation passed since mutation and population size. Measuring the LD region in the current population could be a useful tool for QTL fine mapping.

A number of methods were proposed for fine mapping (e.g./ Spielman et al., 1993, Thaller and Hoeschele, 2000, Meuwissen et al., 2002). The rationale of the fine mapping methods was reviewed by Jannink and Walsh (2001) Spielman et al. (1993), Spielman and Ewens (1996) proposed the method of Transmission Disequilibrium test to overcome this problem. The rationale behind this method is that the affected offspring would inherit marker alleles linked to alleles at a locus for production or disease resistance in larger population.

The test attempts to determine the frequency of transmission of the focal allele to the affected progeny. The test employs family trios consisting of both parents and the disease affected progeny. Such that one of the parents must carry a one copy of the focal marker allele that hypothetically linked to the disease susceptibility allele.

Meuwissen and Goddard (2000) devised the method of gene drop to identify an Identity By Descent (IBD) region that flanks the putative QTL. This is achieved by estimating the covariance between the effects of marker haplotypes. The use of the method on real data has resulted in narrowing the search for a twinning QTL in Norwegian cattle down to 3 cM Meuwissen *et al.* (2002). Furthermore the method (as a part of this thesis) was found to be robust towards violation of the assumptions. With the exception of the presence of genetic heterogeneity in mixed population, the method was able to locate the QTL with only a small loss of the accuracy. More recently Lund et al. (2002) extended this method to include multi-traits. Although the method proposed by Meuwissen and Goddard (2000) is the most relevant for dairy cattle, much work still needed to account for the complexity of the trait.

Thaller and Hoeschele (2000) described alternative fine mapping methods, contrast mapping and substitution mapping, that are both based on the current recombinations.

REFERENCES

- Beuzen, N. D., M. J. Stear and K.C. Chang, 2001. Molecular markers and their use in animal breeding. *The veterinary Journal*, 160(42–52). pages 3
- Cunningham, E.P., and C.M. Meghan, 2001. Biological identification system: genetic markers. *Rev. Sci. tech. Of. Int. Epiz.*, 20, 491–499. pages 3

- Hoeschele, I., P. Uimari, F.E. Grignola, Q. Zhang, and K.M. Gage, 1997. Advances in statistical methods to map quantitative trait loci in outbred populations. *Genetics*, 147, 1445–1457. pages 5
- Jannink, J.-L., and J.B. Walsh, 2001. Association Mapping in Plant Population. Page in press of: Kang, M.S. (ed), Symposium on quantitative genetics and plant breeding in the 21st century. CABI, Wallingford, UK. pages 6
- Liu, B.H. 1998. Statistical Genomics, Linkage, Mapping and QTL analysis. CRS Press LLC. pages 3
- Lund, M.S., P. Sørensen, B. Guldbbrandtsen and D.A. Sorensen, 2002. Multi-trait fine mapping of quantitative trait loci using combined linkage disequilibrium and linkage analysis. submitted. pages 7
- Meuwissen, T.H.E. and M. Goddard, 2000. Finemapping of quantitative trait loci using linkage disequilibrium with closely linked markers. *Genetics*, 155, 421–430. pages 7
- Meuwissen, T.H.E., A. Karlsen, S. Lien, I. Olsaker and M. Goddard, 2002. Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics*, 161, 265–372. pages 6, 7
- Montaldo, H.H. and C.A. Meza-herrera, 1998. Use of molecular markers and major genes in the genetic improvement of livestock. *Electronic Journal of Biotechnology*. pages 3
- Spielman, R.S. and W.J. Ewens, 1996. The TDT and other family-based tests for linkage disequilibrium and association. *Am. J. Hum. Genet.*, 59, 983–989. pages 6
- Spielman, R.S., R.E. McGinnis and W.J. Ewens, 1993. Transmission test for linkage disequilibrium :the insulin gene region and insulindependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.*, 52, 506–516. pages 6
- Thaller, G., and I. Hoeschele, 2000. Fine-mapping of quantitative trait loci in half-sib families using current recombinations. *Genet. Res., Camb*, 76, 87–104. pages 6, 7
- Weller, J.I. 2001. Quantitative Trait Loci Analysis in Animals. CABI. pages 5
- Weller, J.I., Y. Kashi and M. Soller, 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait. *J. Dairy Sci.*, 73, 2525–2537. pages 4