GENETIC DISSECTION OF QUANTITATIVE TRAITS

H. Roshdy

Department of Animal Production, Faculty of Agriculture, Cairo University

INTRODUCTION

The last century has seen the widespread application of statistical methods to study and understand the genetic architecture of quantitative traits, especially in farm animals. The resultant science of quantitative genetics has been an impressive intellectual achievement. It has shown that many important traits, such as growth rate, lean meat conversion efficiency, milk production, disease resistance, wool production, and litter size are multifactorial or quantitative traits; which are influenced by both genetic (often more than one gene) and non-genetic factors (environment). Genes that affect quantitative traits are known as Quantitative Trait Loci (QTL). Quantitative genetics has also provided methods to maximize the response to selection. Unsurprisingly, animal breeding in more advanced countries relies heavily nowadays on the principles of quantitative genetics. So far, despite impressive achievements in the understanding of gene structure and expression, molecular genetics has made little direct contribution to animal breeding.

The current century is likely to see both quantitative and molecular genetics dominate the theory and practice of animal breeding. Genetic evaluation has begun by analysing phenotypes to identify genetic effects, whereas molecular genetics often starts with some known alleles or DNA sequences and then examines their influence on phenotypes. Eukaryotic genomes show considerable DNA sequence variations (polymorphisms) between different species and among individuals within a given species.

Identification of Quantitative Trait Loci:

Genetic dissection of QTL involves the following phases: 1) development of genetic markers, 2) construction of genetic maps, 3) collection of phenotypic data and genotyping individuals, 4) locating QTL and estimating its effect, and 5) using associations detected between markers and quantitative traits in breeding programmes.

1. Development of Genetic Markers:

From the time of Mendel until the 1980s the only genetic markers available to geneticists were either simple phenotype markers, such as eye colour in *Drosophila*, or protein polymorphisms, like the haemoglobin blood groups. As a result of using these markers, some quite detailed genetic linkage maps were developed for model species such as mice and *Drosophila*, but there were great limitations for linkage map construction in livestock species.

The advent of technologies to study variation at the DNA level, in particular the Polymerase Chain Reaction (PCR) technique, suddenly removed the obstacle of marker availability so that polymorphic genetic markers in all species could be

developed and linkage mapping projects could also be planned. Such genetic markers gave the ability to track the inheritance of the linked segments of the genome in suitable pedigrees. A significant association between the inheritance of a particular marker allele and a measured quantitative trait provides evidence that a QTL is linked to the marker in question. By following a reasonable set of genetic markers spread approximately across the genome in an appropriate pedigree, it is possible to identify all the major QTL influencing genetic variation in a trait in that population. This approach was first used in plants to determine QTL controlling the difference between genetically divergent lines (Paterson *et al.*, 1988). The first similar study in livestock identified genes controlling differences between the wild boar and commercial pigs was carried out by Andersson *et al.* (1994). This was followed by the identification of QTL within the Holstein cattle population (Georges *et al.*, 1995) and now by results of many other studies, both within populations and in crosses between populations.

The ability of undertake large-scale QTL studies followed the development of molecular markers. Nowadays, molecular marker technology is moving on rapidly and providing new markers for the study of genetic variation. Restriction Fragment Length Polymorphisms (RFLPs) were the predominant DNA marker initially, but many other types of markers, such as Random Amplified Polymorphic DNA (RAPDs), Amplified Fragment Lenght Polymorphisms (AFLPs), Single Stranded Conformation Polymorphisms (SSCPs), Variable Number of Tandem Repeats (VNTR, which include both Minisatellites and Microsatellites), and Single Nucleotide Polymorphisms (SNPs) have been developed and have or are being used. The current maps of farm animals and humans are mainly dominated by microsatellite markers. DNA microarray and chip based technologies may in the near future provide promising tools for rapid genotyping of large numbers of individuals for several thousand or more polymorphisms.

Microsatellites:

Microsatellite markers, also called short tandem repeats (STR), were described definitly by Weber and May (1989). They are the markers responsible for the recent expansion in genetic linkage maps of farm animals for several reasons: they exhibit a high degree of polymorphism, abundant in the genome, spread regularity throughout the genome, codominant markers, require only small amounts of template DNA for their amplification and are relatively easy to find and characterize (Crawford *et al.*, 2000). In addition, it is assumed they are neutral to selection, the observed genetic diversity being the consequence of two forces only: genetic drift and mutation. Microsatellites are comparatively easy to automate with the possibility of multiplex amplification of up to five loci in a single PCR reaction and of multiple loadings of up to fifteen loci per lane in some highly optimised gel systems.

At the heart of any microsatellite is a simple sequence, either di-, tri- or tetranucleotide, that is repeated between 10 and 50 times. The most common dinucleotide motif in mammals is $(CA)_n$, where *n* is the number of repeats. The mutation rate of microsatellites is thought to be high and there are often large numbers of alleles that vary in size at a single locus. The variation between alleles of the microsatellites is due to the variation in the number of simple sequence repeats. Slippage of DNA polymerase and mismatch repair during replication process appear to be the mechanisms generating diversity of microsatellite length. Microsatellite

length variation is easily detected by PCR using unique flanking primer sequences. One of the disadvantages of the microsatellites is that as they are anonymous DNA fragments, they can only amplify DNA of the most approximate species. PCR primers developed for one species can sometimes function in related species (Montgomery, 2000).

Microsatellite markers represent a powerful way of mapping genes controlling economically important traits. Once a simple repeat region is identified, by sequencing its immediate flanking regions, specific primers can be designed for PCR for genotyping purposes.

Traditionally, the size of microsatellite PCR product is determined by electrophoresis in denaturing polyacrylamide gel. One of the two primers used in the PCR is usually labelled with a fluorescent or radioactive tag. This method of detection generally works well but suffers from weakness in determining DNA size. Currently, as an alternative to gel-based approach to size DNA products, there is a number of technological developments based on mass spectrometry.

2. Construction of Linkage Maps:

Genetic maps are sets of loci arranged in order and separated by distances (in units depending on the type of map), with each set corresponding to a chromosome pair. Locating genes of interest by scanning the genome in a linkage study is best conducted using a set of markers spaced throughout the genome. These markers may be chosen from a marker map. Once linkage has been established, it is needed to refine the location. Maps of genes or expressed sequences are useful in selecting markers for other uses, such as population studies, tracking the inheritance of DNA segments, and parentage identification.

According to Mendelian laws of segregation, loci located on different chromosomes segregate independently. While loci on the same chromosome may show evidence of cosegregation, i.e., alleles passed to progeny are often of the same grandparent. The proportion of times that alleles are not of the same grandparental origin is termed the "recombination fraction" (θ). This statistical measure of the distance between a pair of loci ranges from 0 for tightly linked loci to 0.5 for loci that far apart or located on different chromosomes (Ott, 1991). The knowledge of which alleles of a parent came from which grandparent is called the "phase".

The measure of distance between loci must be additive. Unfortunately, recombination fraction is not an additive measure. This problem can be solved with the aid of mapping functions which transform the observed θ to a scale that is additive. Linkage packages commonly use either the Haldane or Kosambi map functions. Haldane map function assumes that there is no interference between loci, while Kosambi function considers interference as high over short distances, but decreasing with distance. The distance unit expressed by a map function is a Morgan. In general, distances are specified in centimorgan units (cM), where 1 Morgan is equal to 100 cM and 1 cM is approximately 10^6 bases (Ott, 1991).

The past few years have demonstrated that QTL are present and can be mapped in livestock populations. As QTL mapping results accumulate over the next few years, special attention will turn to utilising and attempting to understand and clone QTL. The latter objectives will require higher resolution QTL mapping and hence the use of new approaches to QTL mapping, combined with high dense marker maps. Molecular technology seems set to deliver the necessary marker density, which latter

may allow us to improve our ability to map QTL and estimate their effects and position. The genetic maps of farm animals are not as dense as their murine or human counterpart. Maps of several livestock species, including cattle, sheep and pigs are available through various web sites (http://www.ri.bbsrc.ac.uk; http://locus.jouy.inra.fr; and http://sol.marc.usda.gov).

Several QTL mapping projects are underway in all major livestock species. In general, the planned progression of these projects is the identification of a marker allele associated with a trait of interest, followed by further refinement of the gene position through fine mapping and then eventual gene identification. To date, only a limited number of genes controlling quantitative traits have been identified (for example, Grobet *et al.*, 1997; Grisart *et al.*, 2002). This can largely be attributed to the low resolution of linkage mapping in relation to that needed for positional cloning, and the difficulty of proving a candidate gene is actually the gene of interest (Marshall *et al.*, 2004).

Two main approaches to identify genes for quantitative traits are commonly applied. They are: (1) genome-wide scans and (2) candidate genes. Genome scans using anonymous molecular markers have proved to be effective as a means of detecting QTL. A genome scan can provide a virtual guarantee that a QTL with a given effect will be detected if an adequate study is performed. Genome-wide scans look for associations between the inheritance of a trait and the inheritance of a number of polymorphisms across genome. In principle, if a sufficiently large number of individuals are phenotyped for a quantitative trait, and genotyped for a set of polymorphic markers on all chromosomes it should be possible to map all QTL.

Taking in account these advantages of the genome scan approach, however, it has its drawbacks. In particular, it is relatively costly and time comsuming to genotype several hundred individuals with markers covering the whole genome. In addition, the resolution of mapped QTL is usually low; perhaps 20 cM or more, depending upon the size of the study and the magnitude of the QTL effect. Because mapping resolution is low, it is difficult to move from mapping a QTL to identifying the actual gene. Genome scans are also often performed in a population from a diverse cross that may not be of immediate commercial relevance. These problems partly explain why an alternative strategy for mapping QTL has been adopted by some researchers: the candidate gene approach.

Prior to the availability of microsatellite markers, a candidate gene approach was used to identify QTL for milk production traits. Allelic variants of blood groups, β-lactoglobulin, κ-casein and β-casein, as well as other genes were shown to be associated with chromosome substitution effects for many milk traits (Grisart *et al.*, 2002). If the trait is well understood, there may be one or more genes that are strongly suspected of contributing to genetic variation in the trait. These are candidate genes. Under these circumstances association analysis or linkage analysis is adequate to determine if the candidate genes are QTL. Perhaps the best known example of a QTL that was identified by a candidate gene approach is the porcine oestrogen receptor, which is associated with variation in letter size (Montgomery, 2000). Methods based on candidate genes are not appropriate for poorly understood traits. Even for well-understood traits there may be some advantages in genome-wide scans because they may reveal previously unsuspected loci.

There are two main flavours of the candidate gene approach: comparative and physiological. The comparative approach takes, as a starting point, loci where

polymorphisms are known to have a phenotypic effect in one species and explores them as candidates for similar variation in other species. An excellent example is the identification of the myostatin locus as a cause of double muscling in cattle (Grobet *et al*., 1997) and the identification of the loci controlling colour and patterning in livestock, where genes that were originally identified in the mouse have been found to have similar effects in livestock. The physiological candidate gene strategy focuses on polymorphisms in, or close to, genes with functions that are directly related to the trait of interest. An association between a candidate gene polymorphism and the trait of interest is taken as evidence that the gene is involved in the genetic control of the trait in question.

Although the polymorphism used as a marker may not itself directly control variation in the trait, it is generally assumed that the functional polymorphism is sufficiently close to the marker that the two loci are tightly linked and in linkage disequilibrium. The candidate gene approach thus relies on the presence of linkage disequilibrium within a population at the level of the gene. If there is no linkage disequilibrium, no associations are found (unless the polymorphism detected by the marker itself causes variation). If linkage disequilibrium is widespread, then we may detect associations between a candidate gene and variation caused by a locus or loci that are some distance away. The extent of linkage disequilibrium in livestock populations is not yet certain, but forces causing disequilibrium, such as selection, drift in small populations, a founder effect, and population hybridisation, are all common features of livestock breeding (Weller *et al*., 1990). Recombination tends to erode linkage disequilibrium, but the erosion is slow between closely-linked loci. It is also clear that there is no simple monotonic relationship between distance between loci and disequilibrium.

The merits of the candidate gene approach to QTL detection include the ability to directly use the identified associations in breeding programmes and the fact that the gene carrying the putative causative mutaion is directly identified. In contrast, QTL detected by genome scans are not precisely mapped and linkages only hold up within family. The prior probability of detecting an association via the candidate gene approach would seem to be low, because the number of potential candidate genes is large and only a small proportion of these genes are tested (Crawford *et al*., 2000).

Both the comparative and the physiological candidate gene approaches can, and often are, combined with positional information, thereby becoming the second stage of a genome scan. In this case, a scan using anonymous markers points to the position of a locus in the genome and comparative or physiological candidates are explored as putative causative loci. It is clear that, when combined with positional information, the candidate gene approach is very valuable. In fact, this approach underlies much of the success in identifying causative mutaions in livestock species (Grobet *et al*., 1997). Positional information brings two advantages. Firstly, the statistical thresholds utilised in a genome scan are higher than those often used in association analyses, thus the association is on firmer statistical ground. Secondly, the prior probability of linkage is increased substantially (the number of potential candidates is reduced proportional to the resolution of the mapping). Nonetheless, a locus spanning one percent of the genome may still contain approaching 1000 different genes and hence a number of possible candidate genes.

3. Collection of Phenotypic Data:

A large number of experimental designs have been suggested to detect the genes affecting quantitative traits with the aid of genetic markers. All these designs have several elements in common. Biological, economic, genetic, and statistical considerations have been taken in account to choose the experimental design.

A mapping population needs to provide information on the transmission of alleles from grandparents to progeny. Full-sib pedigrees are useful for obtaining linkage information in species such as mice and poultry. On other hand, in some farmed animal species, such as cattle and sheep, large parental half-sib families are common. The granddaughter design (GDD) was initially proposed for identifying genetic markers associated with QTL in commercial dairy cattle populations, where artificial insemination is widely used and large-sized populations are easily available. This design involves genotyping grandsires and sons at molecular marker loci and analyzing phenotypic data from granddaughters. A major advantage of GDD is that it can increase power of QTL detection with an equal amount of genotyping when compared with a half-sib or daughter design (Weller *et al.*, 1990). The GDD also allows existing populations to be used for QTL identification, which is especially important in cattle because the development of multigeneration resource populations specifically for this purpose is time-consuming and costly.

The efficiency of mapping is improved by the use of highly polymorphic markers. Another strategy to increase the informativeness of families is to use breed crosses, since the most prevalent alleles may be different in distinct breeds. Breed crosses are therefore more likely to be heterozygous.

4. Locating QTL and Estimating its Effec :

All methods used to detect QTL affecting quantitative traits, except segregation analysis, rely on linkage disequilibrium between markers and loci. Several factors influence the chance of detecting a QTL, if one is actually present. They include the size of the effect, the frequency of the alleles, the density of the genetic map, the heritability of the trait, the variation among animals, the number of animals studied and the method of analysis (Weller *et al.*, 1990).

The major part of the current studies identifying QTL in livestock species is based on crosses between genetically diverse populations, e.g. dairy and beef cattle breeds, broiler and layer lines of chickens, and Meishan and European pig breeds. Least squares based interval mapping approaches provide a powerful and rubust method of QTL analysis for such data (Andersson *et al*. 1994). A least squares approach can then be used to fit a genetic model to the inferred inheritance patterns, either within family or across the population. Maximum likelihood based analysis of the data is also possible. Both maximum likelihood and least squares approaches give very similar results in the analysis of data from crosses between inbred lines (Weller *et al.*, 1990).

One common assumption made in the analyses of crosses between populations is that QTL are fixed to alternative alleles in the two populations. QTL identified in the animals resulting from those crosses may not be relevant in commercial populations. Indeed, any potential marker must be confirmed in those populations in which selection is to be carried out. Such an exercise make the cost of marker development

laborious and expensive. Although, many markers for production traits have been described in farm animals, few are being used in breeding schemes.

Once a QTL has been provisionally located, the next step is to identify the specific gene that is responsible for the phenotypes of interest. One strategy is positional cloning that was used in identifying mutations in DGAT1 gene with major effect on milk yield and composition in dairy cattle (Grisart *et al*., 2002). However, the location of most QTL is too imprecise for direct positional cloning. Further studies to narrow the region of interest, that carries a QTL, depend upon finding recombinants between the QTL and the marker. This can be demanding in farm animals because of long generation intervals and, as one gets closer to the QTL, increasingly small recombination rates. Possible options include the comparison of different populations or even a comparison of gene maps in different species to identify candidate genes (Grobet *et al*., 1997).

However, even the demonstration of a statistically significant relationship between a candidate gene and a quantitative trait does not establish beyond doubt that the candidate gene is directly responsible. The presence of one gene can be associated with the presence of other genes; this is especially true for closely-linked genes in farm animals populations. These associations are known as linkage disequilibrium. An association between a candidate gene and a quantitative trait could be due to the existence of linkage disequilibrium between the candidate locus and the true QTL. Therefore, linkage disequilibrium and functional studies in different populations may be necessary to identify the QTL unequivocally.

5. Incorporating marker information in breeding programmes

Genetic progress in dairy cattle has been based on progeny testing of potential sires, which requires approximately five years to conduct the testing. Marker-assisted selection (MAS), using genetic marker information in selection programmes, was proposed as an alternative to progeny testing for improving selection accuracy and reducing the generation interval (Spelman and Bovenhuis, 1998). Identification of QTL for economically important traits may lead to more efficient breeding programmes using MAS, especially for bulls prior to progeny testing (Marshall *et al*., 2004).

To apply MAS efficiently, the QTL should be localized to a relatively short chromosomal segment. The implementation of MAS with a postulated QTL that is falsely detected causes genetic loss compared to a traditional breeding scheme that has no knowledge of the QTL. Spelman and Bovenhuis (1998) suggested that before MAS is implemented, the detected QTL should be verified in further studies and, in addition, accurate estimates of QTL location and effect are required.

REFERENCES

Andersson, L., Haley C.S., Ellegren H., Knott S.A., Johansson M., Andersson K., Andersson-Eklund L., Edfors-Lilja I., Fredholm M., Hansson I., Hakansson J. and Lundstrom, 1994. Genetic mapping of quantitative trait loci for growth and fatness in pigs. Sci., 263: 1771-1774.

Crawford, A.M., K.G. Dodds and J.C. McEwan, 2000. DNA markers, genetic maps and the identification of QTL: General principales. En *"Breeding for Disease Resistance in Farm Animals"*. II ed. Axford R.F.E., Bishop S.C., Nicholas F.W. y

Owen J.B. (2000). CABI Publishing, CAB International. Wallingford, Oxon OX10 8DE, U.K.

Georges, M., D. Nielsen, M. Machinnon, A. Mishra, R. Okimoto, A.T. Pasquino, L.S. Sargeant, A. Sorensen, M.R. Steele, A.T. Zhao and I. Hoeschele, 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genet., 139: 907-920.

Grisart, B., W. Coppieters F. Farnir, 2002. Positional candidate cloning of a QTL in dairy cattle: identification of missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res., 12: 222-231.

Grobet, L., L.J. Royo, D. Poncelet, D. Pirottin, B. Brouwers, J. Riquet, A. Scheoberlein, D. Susana, F. Ménissier, J. Massabanda, R. Fries, R. Hanset and M. Georges, 1997. A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. Nat. Genet., 17: 71-74.

Marshall, K., J.H.J. Van der Werf and J. Henshall, 2004. Exploring major gene-marker phase-typing strategies in marker-assisted schemes. Anim. Sci., 78: 213-227.

Montgomery, G.W., 2000. Genome mapping in ruminants and map location for genes influecing reproduction. Rev. Reprod., 5: 25 37.

Ott, J., 1991. Analysis of Human Genetic Linkage. The Johns Hopkins Univ. Press, Revised ed. Baltimore, Maryland, U.S.A.

Paterson, A.H., E.S. Lander, J.D. Hewitt, 1988. Resolution of quantitative traits into Mendelian factors by using a complete RFLP linkage map. Nature 355: 721-728.

Spelman, R.J.Y, H. Bovenhuis, 1998. Moving from QTL experimental results to the utilization of QTL in breeding programmes. Anim. Genet., 29(2): 77-84.

Weber, J.L. and P.E. May, 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet., 44: 388-396.

Weller, J.I., Y. Kashi, and M. Soller, 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J. Dairy Sci., 73: 2525-2537.