Genetic variation, maps and markers

Peter Fennessy

Natural variation is the basis of evolution. The contributions of Darwin and Mendel last century laid the foundation for the study of genetics and evolution. More than 40 years ago, Watson and Crick revealed the structure of DNA closing their paper with "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material". Subsequently the solution to how the base sequence encodes the amino acids and hence the protein sequence was also found providing the mechanism whereby mutations which altered the DNA sequence could lead to changes in the protein product. Since that time, our understanding of how such mutations affect the genotype of the species has advanced rapidly so that now we are well on the way to a map of the human genome.

Chromosomes, DNA and genes

While most mammalian species have around 3 billion base pairs, the number of chromosomes varies. Thus the sheep has a diploid number (2n) of 54 (26 pairs of autosomes plus the sex chromosomes), while goats and cattle have 60, red deer and wapiti have 68, and the horse has 62. The actual number of genes is not known although the estimates range from 50,000 to over 100,000. However all of the DNA does not code for proteins - in mammals it is estimated that more than 90% may be non-coding DNA but this does not mean it is not important.

Despite the huge differences between mammalian species, there are remarkably conserved groupings of the same genes across species. The basis of genetic variation between species is the organisation of these groups across chromosomes, the inversions of groups or the re-organisation of small segments of the groups which result in the groups being broken up. While estimates vary and will increase as knowledge expands, calculations have indicated that around 100 reshuffles contribute the difference between humans and mice. Of course, there are also differences in the actual DNA sequence but the basic gene product produced across the various mammalian species is the same. Thus while the basic organisation of the genome is similar across the mammals, it is the differences which are important.

Within a species, it is largely variation within DNA sequences which is the cause of genetic differences between individuals. However substantial reorganisational changes can occur without markedly affecting the viability of the animal or its fertility. Good examples would be Robertsonian translocations where two separate chromosomes become joined. In cattle the combination of chromosomes 1 and 29 is a well recognised mutation.

Genes and genetic variation

Variation in the DNA sequence of particular genes (including the non-coding portion) can influence the phenotype of the animal. While the obvious effect may well be on the protein sequence, there may also be subtle effects on the timing of gene expression during development or at later times of life. For example, several mutations in sheep such as *Booroola, Inverdale* and *Callipyge* are likely due to effect on the timing of gene expression in the developing foetus.

These three genes are all examples of major genes affecting productive (or product quality) traits in sheep. In fact, the reason that these mutations have been described is that their effects on the phenotype of the carrier animal and/or its progeny are profound. Their locations on particular chromosomes have also been defined although the actual genes responsible for the phenotypes have not been located. However, the genes responsible for several recognised mutations in domestic animals have been discovered. Examples include: the malignant hyperthermia/porcine stress syndrome (which results in pale soft exudative meat and also increases lean meat yield) and which is due to a mutation of the ryanodine receptor gene; bovine leucocyte adhesion deficiency (BLAD) which is due to a mutation in the beta-2 integrin gene; DUMPS or deficiency of uridine monophosphate synthase which is due to a nonsense mutation in the gene in cattle. In the homozygous case, the animal dies at around day 40 of gestation. Consequently the mutation has important effects on fertility due to abortion of homozygotes; the heterozygote cows also have very high levels of orotic acid in their milk as the UMPS deficiency reduces the conversion of orotic acid to uridine monophosphate.

Genotype and phenotype

The *Booroola, Inverdale* and *Callipyge* genes are all interesting for other reasons and indicate some of the complexity of genetics and also highlight aspects of the interaction between the genotype and the environment which result in phenotypes which may not be quite as expected.

Booroola (FecB) is located on chromosome 6 and is an example of simple autosomal inheritance. Each copy of the gene increases ovulation rate by about 1.6 but the effect on litter size is different. A single copy will likely increase litter size by about one lamb whereas two copies will increase it by about 1.6 lambs. Thus the (intrauterine) environment modifies the expression of the genotype. Similarly in terms of lambs weaned, the effects of the gene are reduced due to the higher mortality of multiple-born lambs associated with lower birth weight.

Inverdale (Fec X^I) is located on the X chromosome. Thus all daughters of a carrier ram will inherit the gene which in the heterozygous state increases ovulation rate and litter size. However homozygous females have streak gonads and are sterile. The gene is believed to be active early in the developing embryo/foetus as the effects of the gene can be observed in the foetus. As well there is the perplexing question of X-inactivation to consider where one copy of the X chromosome in every cell of a female mammal is randomly inactivated. However the process is not absolute in that parts of the X are not inactivated and as well inactivation may be relaxed at stages during development.

Callipyge is located on chromosome 18 and is responsible for a very substantial increase in muscle development in the hind leg muscles and the longissimus dorsi (eye muscle) in the sheep. It is an example of a strange form of gene expression known as imprinting. Genomic imprinting is the phenomenon whereby the extent to which an allele is expressed depends on whether it came from the mother or the father. This occurs at a number of locations around the genome. In this case, a lamb receiving the callipyge gene from its sire will express the gene whereas a lamb receiving it from its mother will not and will display the normal phenotype. However with this gene there is an added complication such that the homozygous lamb receiving one copy from each of its parents does not express either allele of the gene and is of normal phenotype.

Gene mapping

Gene mapping is simply defining the location of genes on the map of the individual chromosomes. Thus several thousand genes have now been mapped in some species but in the case of the domestic species, the numbers are in the hundreds. However, maps are of various precision and while the location of some genes is known very precisely, the location of others is much more approximate. There are two basic types of gene maps, namely linkage maps and physical maps. A linkage map is a mathematical map where the location of genes is defined in terms of their distance from one another. This is calculated as a function of the proportion of occasions in an individual when particular alleles of the two genes in question are inherited together (ie from the one parent). In contrast a physical map is where a gene is located on a chromosome using a visualisation technique (eg fluorescent label) and examined microscopically. Since gene sequences are relatively conserved across species, the sequence from one species (eg mouse) can often be used to detect the gene in another species (eg sheep).

Over the last few years AgResearch has constructed simple maps of the sheep and deer genomes where known genes have been located and the relationships between genes defined. The deer map is particularly interesting as it used the extraordinary (Père David's x red deer) x red deer hybrid as the basis. Père David's (PD) deer and red deer are evolutionarily widely separated but both male and female hybrids are fertile. In fact they are the widest mammalian hybrid where both sexes are fertile. The parents are so wide apart that around 90% of the genes so far investigated exhibit fixed differences between the two species. Thus using the backcross offspring of a F1 hybrid sire and red dams, it is possible to develop a relatively accurate linkage map with relatively few progeny. Thus in any one animal we can ascertain which parts of chromosomes are inherited together and which parts originate from the PD.

Genetic link age between markers and traits

The ability to define the inheritance outlined above is the basis of identifying genetic markers for traits. Identification of markers (ie a piece of DNA which is inherited along with a particular trait of interest) for productive traits is the objective of much current research. In the case of the PD deer hybrids, AgResearch is defining the phenotype of the quarter PD animals and then mathematically investigating the relationship between the inheritance of the PD marker and the variation in the trait under question. For example, by doing this one part of a single chromosome has been found to have a marked effect

on gestation length. In fact, animals which got a particular part of this chromosome (which presumably contains the gene or genes which are involved) from the PD grandparent have a markedly longer gestation then those which have all red deer genes in this part of the chromosome. Consequently this genome scanning procedure is a very powerful technique and shows the way to the use of markers in identifying deer carrying particularly valuable genes in the future.

However, markers do not need to be known genes and now the vast majority of markers are actually microsatellites. These are non-coding pieces of DNA within the genome of mammals. They are actually highly variable (due to a moderate mutation rate) but are inherited in a Mendelian manner. Consequently as with the known genes in the PD hybrid case, their inheritance and the parental origin can be defined. Unlike the Deermap, which is based largely on known genes, the Sheepmap has in excess of 600 markers, the vast majority of which are microsatellites. As with the deer, the markers on this map are being used to define the relationship between the inheritance of the markers and traits of importance. For example AgResearch has a number of experiments underway where special crosses have been made, the phenotypes are being measured very carefully, the genotypes (ie markers) defined and the mathematical relationships (ie linkage) examined. They are listed in Table 1. The parental cross is then mated to either one of the strains in the cross (backcross) or to an unrelated population (outcross).

Table 1. Current AgResearch experiments involving genome scans to define linkage between markers and traits

Parental cross	Mating	Traits examined
Romney x Merino	Backcross	Wool quality, footrot
High FEC x Low FEC	Backcross Outcross	Parasite resistance/susceptibility in terms of faecal egg count (FEC)
High FES x Low FES	Backcross Outcross	Resistance/susceptibility to facial eczema
Fat x Lean	Backcross	Carcass composition and meat quality

Over the next few years, these experiments are expected to define markers which can be used in selection of the required animals for breeding. The objective is to identify markers for traits which are commercially valuable but which are not amenable to simple selection based on the animals phenotype and information from their relatives. Such traits include the disease resistance/susceptibility traits which are either difficult or undesirable to measure (eg resistance to footrot, FE, parasites) and traits which are very difficult to measure without upsetting the animal in question (eg meat quality traits such as tenderness and pH).

Applications - DNA profiling

While some genetic marker tests are already being offered commercially, the major use of DNA marker technology at present is in parentage testing. For many years, parentage testing in cattle, horses and other species has been based on genetic variation in protein types (ie protein polymorphisms). However DNA polymorphisms are far more widespread and this is especially the case with the microsatellites. Thus it is relatively straightforward to carry out parentage testing by monitoring the inheritance of particular microsatellites from parent to offspring. The first step in this procedure is to develop a set of microsatellite markers which show considerable variability within the population, are inherited in a Mendelian manner and which are relatively simple to automate in the laboratory. *Genomnz*, an AgResearch Technology Development Unit at Invermay, now offers DNA parentage tests for deer and cattle using DNA profiles (based on up to 13 markers for deer and 11 for cattle). *Genomnz* also offers a hybrid test for hybrids between North American elk and red deer. The output is the Genometer^{TM}, which provides estimates of the likely mean percentage of elk genes in the particular animal (along with 75 and 95% confidence limits of the estimate). The first specific gene test offered by *Genomnz* is that for the presence of the Booroola gene which uses known markers around the gene.

Genetic markers to marker-assisted selection

The objective of the research searching for genetic linkage between markers and traits is to identify markers which can be used to identify carriers of particular genes (alleles). While markers are being used commercially for a number of disease genes, Booroola and the oestrogen receptor (for ovulation rate in pigs) are the first markers available for productive traits. However currently dairy sires are often tested for a number of disease genes including bovine lymphocyte adhesion deficiency (BLAD), deficiency of undine monophosphate synthase (DUMPS), Weaver (in Brown Swiss cows) and citrillinaemia. The major pig breeding companies also usually test for the presence of the ryanodine receptor mutation and in some cases, a variant of the oestrogen receptor gene.

As noted previously, both Weaver and DUMPS are embryonic lethals where homozygotes die *in utero*. However it seems that both are examples of pleitropy which is the term for situations where a gene affects two or more apparently unrelated traits. In both cases, it seems that heterozygous carriers are favoured by artificial selection for milk production as is also the case with the ryanodine receptor mutation, which is associated with higher lean meat yield in pigs.

These examples of disease genes illustrate some of the potential for marker-assisted selection although in all cases, there is probably a strong case to cull carriers hence avoiding the use of heterozygous sires. However there are some cases, such as the ryanodine receptor mutation, where it could be argued that the productive benefits of the mutation are such that it should be maintained in the population because of the associated effects on leaness. Therefore this is one situation where markers can assist with decision-making such as to avoid the mating of heterozygous males and females.

Genetic variation and the size of gene effects

Up to now, this discussion has highlighted the effects of genes which have a major impact on the trait of interest. However, we all know from experience that variation among the members of a population appears to be continuous and that large genes like Booroola do not appear to be very common. The working hypothesis has been that genetic variation around the mean is due to the action of many genes of relatively small effect. This working hypothesis has served the animal industries very well and is the basis of quantitative genetic theory and its application. The recognition of major genes (or genes which have a large effect) affecting performance traits such as reproduction and growth rate is a relatively recent phenomenon.

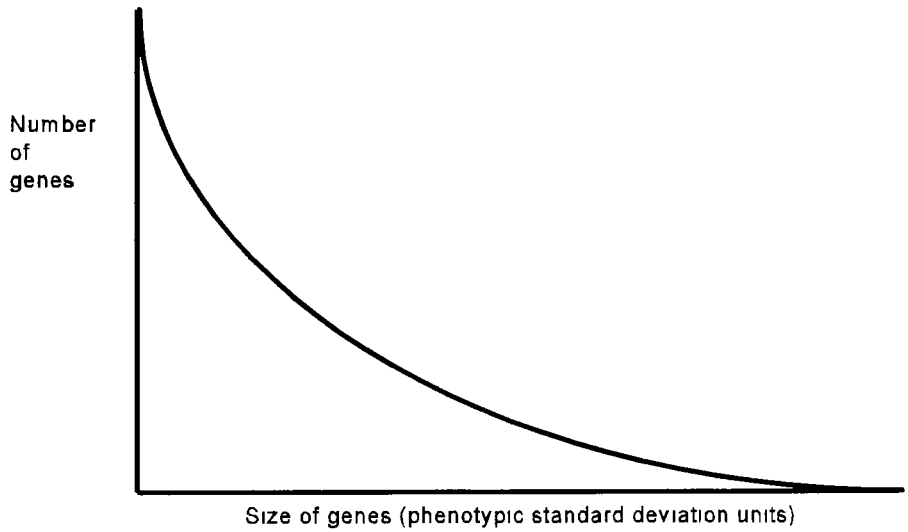

Figure 1 The hypothetical relationship between the number of genes contributing to the variation and a trait and the size of the effect of the individual genes

Figure 1 shows the hypothetical relationship between the number of genes affecting a trait and the relative size of the genes. It clearly shows that we would expect to see very few major genes with smaller effects. Also there are very many genes with smaller effects which simply because they are small would be difficult to detect and manage individually in a breeding programme. It is this recognition which underpins the use of quantitative genetics and mathematically based genetic evaluation systems.

In the last two years the application of new 'gene search' software programmes (eg Findgene, Maggic) has improved our ability to detect the presence of major genes in very well-recorded populations. While these programmes indicate whether or not a major gene is likely to be present and also estimate the size of the gene effect, they cannot tell us anything about its location on the gene map. However they can provide indications of the type of inheritance such as whether the 'gene' is dominant or recessive or whether it is imprinted or not. These new software tools are very powerful and they are now being evaluated in various genetic evaluation schemes. In terms of Figure 1, this type of

software is moving our ability to detect the effect of genes further to the left. The potential for combining this technology with genetic markers is clearly evident.

Genetic markers provide the next opportunity to improve our ability to detect genes and also to manage them in breeding programmes. Firstly, as noted previously, the objective of the research searching for genetic linkage between markers and traits is to identify markers which are inherited along with some of the variation in a trait. That is the marker and the gene contributing to variation in the trait of interest are closely associated on the DNA - that is they are genetically closely linked. Secondly, the intention is to use the genetic marker (ie piece of unique DNA sequence) to then identify carriers of the particular gene. This then gives the breeder the ability to manage the gene.

The detection of genetic linkage between a marker and variation in a trait also effectively enables us to move further to the left in Figure 1, in terms of our ability to detect and manage genes. However at some point, our ability to detect genes accurately using markers and the benefits of using this information will abate - that is, the law of diminishing returns applies. At this point, the quantitative genetic approaches become the only practical way to work with this variation to define the genetic merit of individual animals. The combination of all these techniques, such as markers for major genes, linked markers for smaller effects and normal quantitative genetics approaches into an integrated system is known as marker-assisted selection (MAS).

Marker-assisted selection in practice

Marker-assisted selection (as defined above) involves the incorporation of genetic marker information into the normal mathematically-based systems of genetic evaluation. The gene status of the animal in question is then taken into account in the calculation of the genetic merit or the breeding value of the animal for the particular trait being evaluated. For example in the case of a gene marker related to some part of the variation in parasite resistance or susceptibility, the following factors could be included in the calculations:

- the accuracy of prediction of the presence/absence of the 'gene' or in other words the strength of the genetic linkage between the marker and the trait variation.
- the type of inheritance of the marker/trait (eg dominant, recessive, imprinted).
- the genetic correlation between the 'gene' and other important traits.
- the relative size of the 'gene' effect compared with all other known and unknown genes affecting the trait (ie what proportion of the variation does this 'gene' account for).

Clearly MAS offers considerable promise to the animal industries especially for disease susceptibility/resistance traits and product quality traits which cannot be measured easily. However there is a very large investment required in getting to the point of successfully identifying important markers and then incorporating them into genetic evaluation systems. In addition, to be useful for breeders, the genetic testing and the subsequent analysis must be cost-effective. Naturally this will depend on the breeding structure of the industry and the ability of the breeder to be paid according to the quality of the animals bred. In this respect, the tools are now available for the use of MAS on a limited scale such as with Booroola and callipyge. However it will likely be some years before MAS plays

a major role in identification of genetically superior animals in New Zealand's livestock industries.

Further reading

Nicholas F.W. 1996: Introduction to Veterinary Genetics, Oxford University Press.