The Veterinary Journal 2000, **160**, 42–52
doi: 10.1053/tvjl.2000.0468, available online at http://www.idealibrary.com on IDEAL®

Review

Molecular markers and their use in animal breeding

N. D. BEUZEN*, M. J. STEAR† and K. C. CHANG*

**Department of Veterinary Pathology, University of Glasgow, Bearsden Road Glasgow G61 1QH, UK †Department of Veterinary Clinical Studies, University of Glasgow, Bearsden Road, Glasgow G61 1QH, UK*

SUMMARY

The use of DNA markers to define the genetic makeup (genotype) and predict the performance of an animal is a powerful aid to animal breeding. One strategy is known as marker-assisted selection (MAS). MAS facilitates the exploitation of existing genetic diversity in breeding populations and can be used to improve a whole range of desirable traits. DNA markers are, by definition, polymorphic, and the methods used to define DNA markers include restriction fragment length polymorphisms (RFLPs), microsatellites, and single nucleotide polymorphisms (SNPs). Linkage analysis, association analysis and analysis of gene function can be used to determine which polymorphisms are useful markers for desirable traits. Future prospects include the use of high throughput DNA microarray (DNA chip) technology which could revolutionize animal breeding in the next millennium.

KEYWORDS: Polymorphism; molecular markers; marker-assisted selection; quantitative trait loci; candidate genes; animal breeding; quantitative genetics; molecular genetics.

INTRODUCTION

This century has seen the widespread application of statistical methods to the study of genetic inheritance, 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, milk production, wool production and litter size are multifactorial; both genes and the environment influence them. Quantitative genetics has also provided methods to maximize the response to selection (Nicholas, 1996). Unsurprisingly, animal breeding in more advanced countries relies heavily 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 next century is likely to see both quantitative and molecular genetics dominate the theory and practice of animal breeding. Genetic evaluation has usually started by analysing phenotypes to identify genetic influences, whereas molecular genetics often begins with known alleles or DNA sequences and then examines their influence on phenotypes. Eukaryotic genomes show considerable DNA sequence variations (polymorphisms) between species and among individuals within a species. Three types of DNA polymorphisms are particularly well characterized. They are: (1) restriction fragment length polymorphisms (RELPs), (2) microsatellites, and (3) single nucleotide polymorphisms (SNPs). This review focuses on these three types of polymorphisms and their potential use in animal breeding.

Correspondence to: K. C. Chang, Tel.: +44 (0) 141 330 4123; Fax.: +44 (0) 141 330 5602; E-mail: k.chang@vet.gla.ac.uk

1090-0233/00/010042 + 11 $35.00/0

DEFINITION OF DNA POLYMORPHISMS

Restriction fragment length polymorphisms

Nucleotide changes occur in all eukaryotic genomes. If the change results in the creation, or abolition, of a restriction endonuclease recognition site, then the DNA sequence acquires or loses the ability to be cleaved by a particular restriction endonuclease. If a recognition site is absent then digestion with the relevant restriction enzyme will generate a long fragment. If a recognition site is present then digestion with the relevant restriction enzyme will generate two shorter fragments. If the recognition site is in only one of two parental alleles, digestion will produce two different electrophoretic patterns: a long fragment and two shorter fragments (Fig. 1). This is the basis of restriction fragment length polymorphism (RFLP), which was among the first techniques to be used for typing DNA polymorphisms. While RFLP mapping is a powerful and extensively used technique, its gel-based approach is inconvenient for high throughput screening. Additionally, most mutations do not result in the abolition or creation of restriction endonuclease sites, making such mutations impossible to detect by RFLP analysis.

Microsatellites

Microsatellites are two- to six-nucleotide repeats, interspersed throughout the genome. Microsatellites are highly polymorphic and abundant, often found in non-coding regions of genes. The most common dinucleotide motif in mammals is $(CA)_n$, where n is the number of repeats (Rohrer *et al.*, 1994). In avian species, the frequency of $(CA)_{\geq 10}$ is estimated at once every 140 to 180 kb, and that of $(CA)_{\geq 14}$ is one every 350 to 450 kb (Primmer *et al.*, 1997). 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. One study put the mutation rate as high as $1/10^3$ to $1/10^4$ per locus per generation (Small *et al.*, 1998). Slippage of DNA polymerase

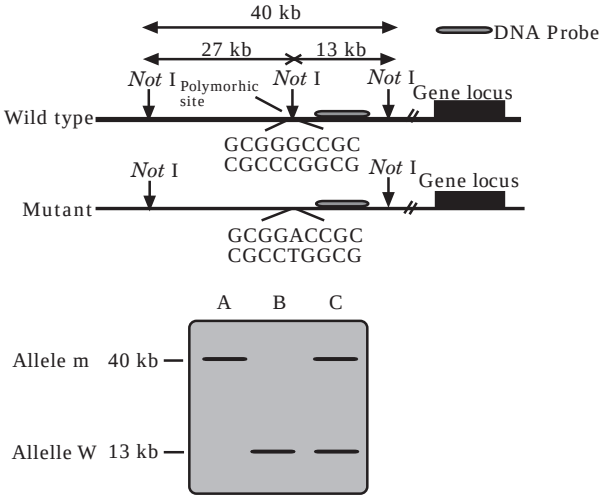
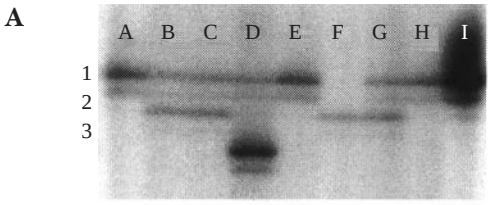


Fig. 1. Restriction fragment length polymorphism. (upper panel) Two alleles (W and m) of one gene presented as presence or absence of *Not* I restriction site, due to a G→A point mutation. (lower panel) Genomic DNA of three individuals (A, B, and C) were digested with *Not* I, separated by gel electrophoresis, transferred onto a nylon membrane, hybridized with an appropriate radioactive DNA probe and subjected to autoradiography. Lane A, homozygous for allele m. Lane B, homozygous for allele W. Lane C, heterozygous for alleles m and W.



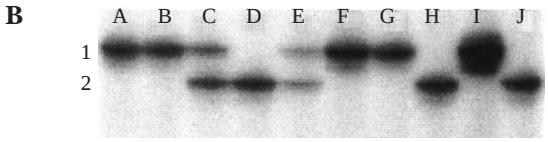


Fig. 2A. Microsatellite allelic length polymorphism of the porcine myosin heavy chain slow/I locus, by PCR using primers as shown in Table Ia. Three alleles (1, 2 and 3) are found in lanes A to I which correspond to the nine individual pigs. The genotypes are (1.1) for pigs A, E, H and I, (2.2) for pig F, (1.2) for pigs B, C, and G, and (1.3) for pig D. Note that additional band seen below each major band is characteristic of *Taq* polymerase in microsatellite amplification.

Fig. 2B. Microsatellite allelic length polymorphism of the porcine myosin heavy chain fast 2a locus, by PCR using primers as shown in Table Ib. Two alleles (1 and 2) are found in lanes A to J which correspond to the 10 individual pigs. The genotypes are (1.1) for pigs A, B, F, G and I, (2.2) for pigs D, H and J, and (1.2) for pigs C and E.

Table Ia
Part of a microsatellite locus found in the porcine slow/I myosin heavy chain. The length $(TG)_n$, where *n* is the number of repeats, varies according to specific allelic form. Upstream (sense) and downstream (antisense) primers (underlined) were derived from either side of the core repeat region (Chang, 1998)

5′—ATGTCTGTATCTGAGCATGTGAGAACTACAATGCCTAGAGAA(**TG**)$_n$TGGGTGGTGGGTGGTGTGTGT
GTTGTGAACAGTGTGTACTGTCTGAATGCACGGTG
AGGGAGGCCACGACCTCACTGCCTAAAAAGCAGAAA
TATCAAGTGGGACACATGTTT-3′

Table Ib
Part of a microsatellite locus found in the porcine fast 2a myosin heavy chain. The length $(TA)_n$, where *n* is the number of repeats, varies according to specific allelic form. Upstream (sense) and downstream (antisense) primers (underlined) were derived from either side of the core repeat region (Chang, 1998)

5′—TGGATTCAAGATTTACTTTCATTATCCTACAAA
TTCCAGCAAATTATATTTGTCAATGAAGGAAAAA
GAGTTGGACATTTAAAAAATATAATTTCAAAAGT
TATAAA(**TA**)$_n$GCTTTACATTTTTACAACCTTCTTGTG
CAAACTTTCTTCACGATGGTAATTATCAAGAACATAC
ATACTGTGTGCCCTTCTTCATCAGGAAATATGTCCT
TACTTCAAGTATAAAGCCA-3′

and mismatch repair during replication appear to be the mechanisms generating diversity of microsatellite length. Microsatellite length variation is easily detected by the polymerase chain reaction (PCR) using unique flanking primer sequences. PCR primers developed for one species can sometimes function in related species (Moore *et al.*, 1991; Rubinsztein *et al.*, 1995).

Microsatellite-derived markers represent a powerful way of mapping genes controlling economic traits. Once a simple repeat region is identified, by sequencing its immediate flanking regions, specific primers can be designed for PCR for genotyping. Two examples of microsatellite genotyping are shown in Tables Ia and b and Figures 2a and b. Traditionally, the size of a microsatellite PCR product is determined by electrophoresis in a 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 an inherent weakness in accurately determining DNA size. Currently, as an alternative to the gel-based approach to size DNA products, there are a number of technological developments based on mass spectrometry. One new trend is the use of MALDI-TOF (Matrix Assisted Laser Desorption Ionisation Time of Flight) mass spectrometry, as a rapid and reliable way of discriminating polymorphic fragments (Braun *et al.*, 1997). The main drawback of MALDI-TOF mass spectrometry is high apparatus cost, limiting its widespread availability. To obtain good spectral reading, specifically prepared ultrapure primers and reagents are required, which in turn increase the cost of analysis (Haff & Smirnov, 1997).

Single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) involve the substitution of one nucleotide for another, or the addition or deletion of one or a few nucleotides. There are four major reasons for the increasing interest in the use of SNPs as markers for genetic analysis. Firstly, they are prevalent and provide more potential markers near or in any locus of interest than other types of polymorphism such as microsatellites. For example, in human genomic DNA there appears to be an SNP approximately every 1000 bases (Landegren *et al.*, 1998). Secondly, some SNPs are located in coding regions and directly affect protein function. These SNPs may be directly responsible for some of the variations among individuals in important traits. Thirdly, SNPs are more stably inherited than microsatellites, making them more suited as long-term selection markers. Finally, SNPs are more suitable than microsatellites for high throughput genetic analysis, using DNA microarray technology (Lipshutz *et al.*, 1999).

SNPs are generally biallelic systems, which means that there are usually only two alleles in a population. As a consequence, the information content per SNP marker is lower than multiallelic microsatellite markers. Five SNP markers provide similar information to one microsatellite marker, meaning that about 2000 SNPs will be required to cover the equivalent of a 10 cM microsatellite map (http://www.pebio.com/ab/apply/dr/dralbl.html). The number of specific SNPs on the genetic map needs to be greater than the densest microsatellite map currently available. Therefore,

high throughput technologies are needed to screen large numbers of SNPs.

There are a number of methods to detect SNPs. The traditional gel-based approach uses standard molecular techniques, such as sequencing, PCR, restriction digests and various forms of gel electrophoresis, i.e. denaturing gradient gel electrophoresis [DGGE], single-strand conformation polymorphism [SSCP] and cleavase fragment length polymorphism [CFLP] (Parsons & Heflich, 1997; Kwok & Chen, 1998; Nataraj *et al.*, 1999). An elegant technique, cleaved amplified polymorphic sequence [CAPS], devised by Michaels and Amasino (1998), uses mismatch PCR primers to create a new restriction site on one of the alleles. Other methods include the PCR-based TaqMan assay (Landegren *et al.*, 1998), the high-performance liquid chromatography-based WAVE DNA fragment analysis system (Kuklin *et al.*, 1998), and MALDI-TOF mass spectrometry, including the use of peptide nucleic acid probes (Ross *et al.*, 1997). One mass spectrometry approach uses a specific primer to extend into the SNP site using di-deoxynucleotides. The SNP is defined by the particular di-deoxynucleotide incorporated (Fig. 3; Li *et al.*, 1999). All these techniques require prior knowledge of the sequence of the polymorphic site.

One way to detect polymorphisms with a few nucleotide mismatches, is the random amplified polymorphic DNA (RAPD) assay (Cushwa & Medrano, 1996). This assay uses short oligonucleotide primers of arbitrary sequence to amplify discrete regions of the genome. Typically, only one 10-mer oligonucleotide of a particular arbitrary sequence is used in each PCR. RAPD polymorphisms are detected as the presence or absence of

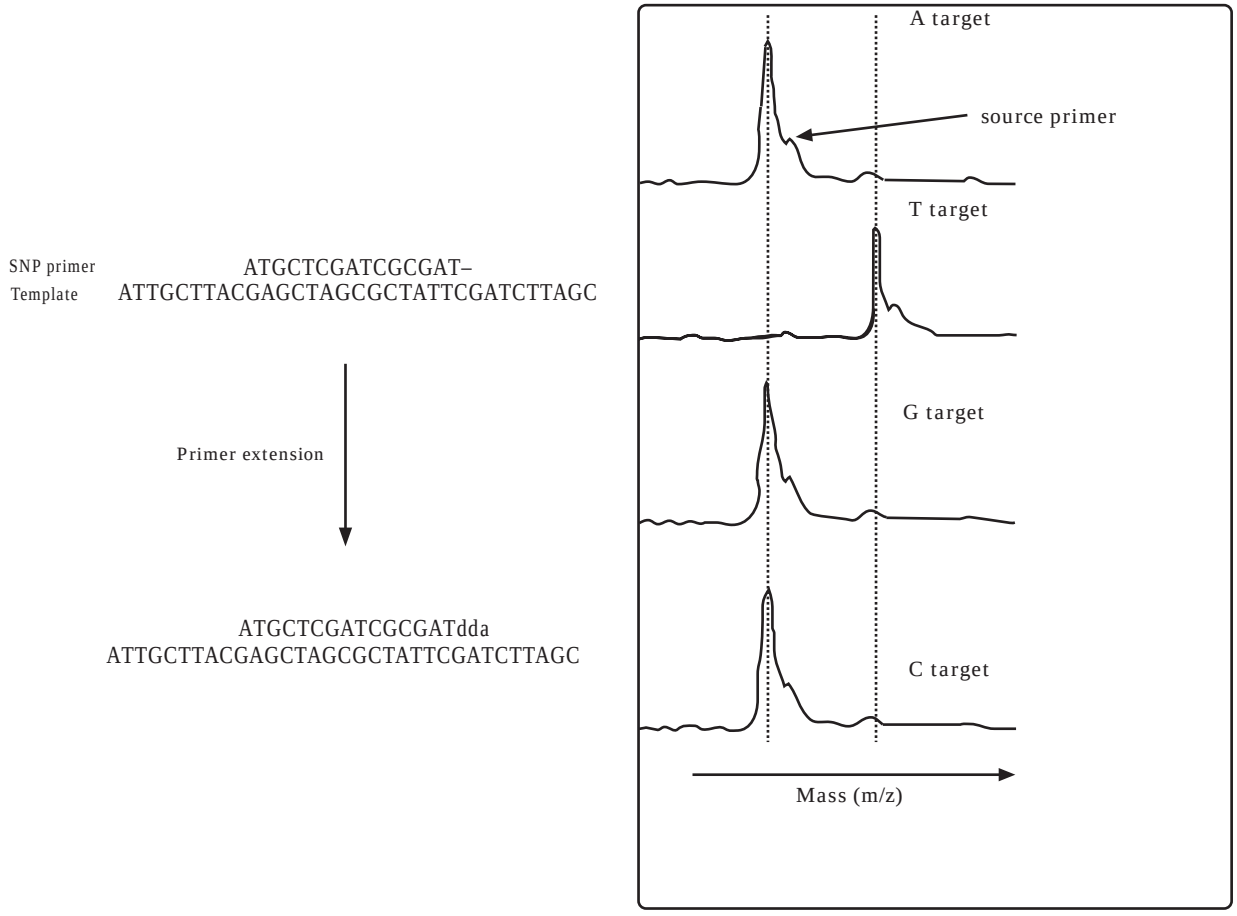


Fig. 3. MALDI-TOF method of SNP detection (Haff & Smirnov, 1997). After annealing of primer and template, primer extension is performed in four different di-deoxynucleotides (ddATP, ddTTP, ddGTP and ddGTP) reactions. Extension products are analysed by MALDI-TOF mass spectrometry. In this case, only ddATP can be incorporated, resulting in a corresponding mass increase.

particular fragments between individual animals after gel electrophoresis. The absence of a particular fragment is the result of sequence differences in one or both of the priming sites which prevents oligonucleotide annealing and subsequent polymerization. A typical arbitrary oligonucleotide will generate up to a dozen or so discrete DNA bands.

The application of DNA microarray or 'DNA-on-Chip' technology to detect SNPs is a potentially powerful tool for high throughput DNA screening (Landegren *et al.*, 1998). Up to 40 000 sequences on a single slide can be screened at one time (Lipshutz *et al.*, 1999). In one experiment, 2.3 Mb of genomic DNA were successfully scanned to identify up to 3241 candidate SNPs, using more than 100 tiling microarrays (Wang *et al.*, 1998). A tiling microarray consists of overlapping sets of four 25-mer-oligonucleotide probes, derived from cDNA sequences, varying by one nucleotide in the central position. A single base substitution will result in a perfect match with one of the four-anchored oligonucleotides (Wang *et al.*, 1998). Although tiling arrays work for base substitution, mutations such as deletion and insertions will not be specifically recognized unless corresponding probes are incorporated in the array.

There are a number of major technical difficulties to overcome before DNA microarray technology can be widely applied to large-scale genotyping of animals (Hacia, 1999). They include the presence of secondary structures within the target and the difficulty in optimizing hybridization conditions over the whole array, due to differences in annealing temperature among oligonucleotides. The preparation of labelled genomic DNA fragments with sufficiently high specific activity for hybridization to immobilized oligonucleotides can also be problematic, especially if a large number of loci are to be screened simultaneously. The key challenge, however, is to identify useful SNPs that can predict the breeding value of an animal. A consortium of 10 pharmaceutical companies and The Wellcome Trust recently announced an initiative to produce a human SNP map. Whilst detecting the presence of a known SNP is relatively straightforward, the main challenge in this ambitious venture must be the identification of new SNPs at frequent intervals throughout the human genome (Taillon-Miller *et al.*, 1999). In farm animals, it is unlikely that SNP mapping will ever be conducted on this scale in the foreseeable future. Nevertheless, DNA microarrays and SNP development are highly relevant for future livestock selection. In the same way that microsatellites have replaced RFLPs, SNPs are expected to supersede microsatellites as the method of choice for detecting DNA polymorphisms.

DETERMINING IF DNA POLYMORPHISMS ARE USEFUL MARKERS

The optimal strategy to determine if a DNA polymorphism is a useful marker for a trait depends upon the nature of inheritance of that trait. There are a wide variety of traits with a genetic component but the vast majority can be explained by one of six patterns of inheritance. They are autosomal dominant, autosomal recessive, sex-linked dominant, sex-linked recessive, incomplete penetrance of a single gene and multifactorial. The first four categories are collectively described as Mendelian traits. Their major interest for animal breeders is that they are responsible for a large number of different diseases (Nicholas, 1996). They are due to mutations in specific genes, either in the coding or in the regulatory regions.

Mendelian traits

Autosomal dominant diseases affect approximately equal numbers of males and females, which distinguishes them from sex-linked diseases. Breeding an affected sire with an affected dam can produce unaffected offspring, which distinguishes them from recessive disorders. If the pedigree of an affected individual is traced back, the disease is present in every generation until the original mutation. Dominant diseases are not normally common in veterinary species because breeders do not usually breed from obviously diseased animals. But some disorders, such as Huntingdon's chorea in humans, develop well after puberty. Other diseases (of bone for example) may not be apparent to the breeder. Multiple exostosis is an example of an autosomal dominant disease in horses. This disease is characterized by benign growths projecting from surface of the bones (Nicholas, 1996). Another example of an autosomal dominant disease of horses is hyperkalaemic periodic paralysis. This is a disease particularly of quarter horses in the USA. Affected individuals have more muscle and breeders inadvertently increased the frequency of this disease by breeding from muscular horses. The biochemical defect is known and is a mutation in the sodium channel in muscle. This reduces the

Table II

Genotypic effects on meat composition traits for the loci of MyHC slow/I and fast 2a based on 220 pigs. The table shows the estimate mean for each genotype and each trait, together with the estimated standard error (SE) of the mean. The statistical significance of the difference between the mean of each genotype is assessed by the probability that the difference could have arisen by chance (column *P*). *P* values of less than 0.05 show that the genotype has a statistically significant effect on the trait in question. Values of $P < 0.01$ and $P < 0.001$ are judged highly statistically significant and very highly statistically significant, respectively. For instance, for the MyHC slow/I locus, the difference between genotypes for % protein in muscle are very highly significant

Trait	*Mean SE*	*Mean SE*	*Mean SE*	*Mean SE*	*Mean SE*	*P*
MyHC slow/I						
genotype	s1.s1	s1.s2	s2.s2	s1.s3	s3.s3	
no. of animals	133	66	7	7	7	
% moisture in muscle	73.3 0.11	73.9 0.15	73.7 0.44	72.6 0.44	73.5 0.44	0.008**
% protein in muscle	23.6 0.07	23.0 0.10	23.4 0.28	23.8 0.28	23.1 0.28	<0.001***
% fat in muscle	1.53 0.08	1.35 0.11	1.39 0.32	2.35 0.32	1.96 0.32	0.03*
MyHC fast 2a						
genotype	a1.a1	a1.a2	a2.a2			
no. of animals	101	96	23			
% moisture in muscle	73.4 0.18	73.5 0.16	73.4 0.28			0.85
% protein in muscle	23.3 0.12	23.2 0.11	23.6 0.18			0.04*
% fat in muscle	1.84 0.14	1.74 0.12	1.57 0.20			0.31

membrane potential and can cause increased spasms. The defect can be unambiguously identified by amplification of part of the gene by PCR then incubation of the amplified DNA with labelled probes that hybridize to the mutation but not with normal DNA (Rudolph *et al.*, 1992).

Autosomal recessive diseases occur in approximately equal numbers of males and females. Crossing an affected sire with an affected dam gives only affected offspring. The disease can skip generations, when tracing the inheritance back from an affected individual. An important autosomal recessive disease is bovine leucocyte adhesion deficiency (BLAD). Following bacterial infection, dendritic cells and macrophages release interleukin-1 and other cytokines. These cytokines increase the expression of CD54 molecules on endothelial cells that line the walls of local blood vessels. Circulating neutrophils bind to CD54 through their receptor complex that contains CD11 and CD18. The neutrophils move through the blood vessel walls and attack the bacteria by phagocytosis and extracellular killing. Leucocyte adhesion deficiency occurs in calves that have a mutation that prevents the production of effective CD18. Consequently, these individuals have high levels of circulating neutrophils that cannot extravascate. They are highly susceptible to bacterial infections and are usually dead by 12 months of age. The disease is relatively common in Holstein-Friesians because one very popular sire was a carrier for this mutation. There is a PCR-RFLP screening test for the CD18 mutation (Shuster *et al.*, 1993).

Sex-linked dominant diseases are very rare, and few examples, if any, exist in species of veterinary importance. As with all sex-linked diseases, the disease occurs at different frequencies in males and females. In mammals, crossing affected males with unaffected females gives affected daughters and unaffected sons. While crossing unaffected sires with affected dams gives equal ratios of affected males, unaffected males, affected females and unaffected females. Sex-linked recessive diseases are more common and include some haemophilias. They occur at a greater frequency in mammalian males than females. All offspring of two affected parents are affected and as with autosomal recessives, the disease may skip generations.

Incomplete penetrance

Many important defects and diseases run in families but do not conform precisely to any obvious pattern of Mendelian inheritance. The incidence is higher among relatives of affected individuals than among the general population. This may be due to a similar environment, common allelic forms of genes, or a combination of shared genetic and environmental factors. There are two models to explain familial disorders that do not follow Mendelian rules: the incomplete penetrance model and the multifactorial model of liability with a threshold.

Sometimes the inheritance of a single locus trait may show a departure from Mendelian predictions. If, for example, only 98% of homozygous recessives express the trait, penetrance is said to be less than the expected 100% and is labelled incomplete. One example of incomplete penetrance is scrapie. The PrP gene codes for susceptibility to scrapie in sheep; mutations in positions 136, 154 and 171 determine susceptibility (Dawson *et al.*, 1998). However, on some farms sheep do not develop disease, even those that carry the susceptibility alleles. Similarly, Australian sheep carry the same mutations as British sheep but do not develop scrapie; possibly these sheep are not exposed to an infectious agent.

Another example of incomplete penetrance is malignant hyperthermia in pigs. The ryanodine receptor 1 (RYR1) locus codes for the calcium release channel in skeletal muscle. A point mutation in this gene leads to a life-threatening disorder with skeletal muscle rigidity, hypermetabolism and hyperthermia, which is precipitated by stress. The defective gene is associated with increased leanness and appears to have been inadvertently selected by breeders. It can be detected by PCR-RFLP (Brenig & Leeb, 1998). The clinical severity of a carrier pig depends on the number of mutated ryanodine receptor subunits that make up the tetrameric channel molecule, which to an extent accounts for the phenomenon of incomplete penetrance.

For Mendelian traits and traits with incomplete penetrance, the simplest way to determine if a DNA polymorphism can serve as a marker is linkage analysis (Morton, 1955). In essence, families with affected and unaffected individuals are typed for the marker. If the individuals that inherit the trait are more likely to inherit one alleic form of the marker then the trait is assumed to be linked to the marker locus. The linkage distance can be calculated from the proportion of discordant (recombinant) individuals; those that inherit the trait but not the marker or *vice versa*.

Multifactorial traits

Milk production, daily live weight gain, lean meat conversion efficiency, litter size, disease resistance and fleece weight are some examples of multifactorial traits, which are influenced by genetic (often more than one gene) and non-genetic factors. There are several complementary approaches to identify DNA markers for multifactorial traits. They include association analysis, linkage analysis and genome-wide scans, which are a modified form of linkage analysis. Genes that influence quantitative traits are known as quantitative trait loci (QTL).

If the trait is well understood, there may be one or more genes that are strongly suspected of contributing to variation in the trait. These are candidate genes. Under these circumstances association analysis or linkage analysis is appropriate to determine if the candidate genes are QTL. Association analysis has been widely used in human genetics but it can give false-positive associations if there is heterogeneity in the study population. The statistical methods used in human populations assume that individuals in the study population are unrelated. These methods are not appropriate in farm animals. Kennedy *et al.* (1992) showed how to allow for genetic relationships and also how to account for any other known variables. This procedure combines many features of association and linkage analysis and is more powerful than either. Similar procedures have been used to identify a locus in or around the major histocompatibility complex that has a strong influence on nematode resistance in sheep (Schwaiger *et al.*, 1995; Buitkamp *et al.*, 1996; Stear *et al.*, 1996; Feichtlbauer, Buitkamp & Stear, manuscript in preparation). 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 litter size (Rothschild, 1996).

A candidate gene approach has also been used to identify markers for meat quality in the pig. Although not many porcine muscle genes have been cloned, there is a wealth of information on the biology of skeletal muscles. Different muscles differ intrinsically in meat quality both between individuals and within a carcass. An important part of this variation is attributed to differences in muscle fibre composition. Many metabolic and biochemical characteristics of muscle fibres have been shown to

vary between different fibre types (Essén-Gustavsson, 1993). In an adult mammal, there are four major fibre types characterized by the expression of four postnatal myosin heavy chain gene (MyHC) isoforms (slow/I, fast 2a, fast 2x and fast 2b). The slow/I and fast 2b fibres are slow-oxidative and fast-glycolytic; they represent two extreme metabolic profiles. The fast 2a and fast 2x fibres are intermediate fast oxidative-glycolytic fibres. Two porcine MyHC microsatellites are shown in Table I and Figure 2. The TG dinucleotide repeat in the slow/I MyHC gene is located in intron 1 and the TA repeat is found in the promoter region of the fast 2a MyHC gene. Three alleles (s1, s2 and s3) were detected in the MyHC slow/I locus, and two alleles (a1 and a2) were found in the MyHC fast 2a locus. There is a significant association between the two MyHC microsatellite loci and several muscle composition parameters (Chang, 1998).

Methods based on candidate genes are not appropriate for poorly understood traits. Even for well-understood traits there may be advantages in genome-wide scans because they may reveal previously unsuspected loci. Genome-wide scans look for associations between the inheritance of a trait and the inheritance of a number of polymorphisms across the genome. The crucial feature is that rather than study polymorphisms singly, the inheritance of a pair of adjacent markers is followed. This is known as interval mapping. Even if neither of the pair is associated with the trait, a QTL lying between them should be detected. In principle, if a sufficiently large number of animals are phenotyped for a quantitative trait, and genotyped with a set of polymorphic markers on all chromosomes it should be possible to map the locations of all QTL.

An essential component for an efficient genome-wide scan is a high-density genetic map. Known loci or markers are placed on a genetic map relative to each other, the order being determined by the recombination between them. The units of recombination are centimorgans (cM); one cM is approximately 10^6 bases. In the mouse, a high-density microsatellite map of the genome had been constructed, with a resolution of 0.3 cM (Rhodes *et al.*, 1998). The genetic maps of farm animals are not as dense as their murine or human counterpart. Nonetheless, extensive linkage maps are already available in some species. Maps of several species, including cattle, pigs and sheep are available through the Roslin Institute web site (http://www.ri.bbsrc.ac.uk/).

Several factors influence the chance of detecting a QTL, if one is 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; Haley, 1995). In order to increase the statistical power of detecting QTL, artificial crosses are often made between divergent populations, e.g. Meishan and Large White pig crosses (Spelman & Bovenhuis, 1998). QTL identified in these animals 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 makes 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.

There have, however, been some encouraging results. In one study, two European Wild Boars were crossed to eight Large White sows to generate F_1 pigs which were crossed *inter se* to yield 200 F_2 offspring (Andersson *et al.*, 1994). These pigs were genotyped for 105 polymorphic markers, and growth rates were recorded from birth to 70 kg. After slaughter, fat levels and intestinal length were recorded. In this study, one or more QTL influencing fat levels appeared to map to the proximal end of chromosome 4.

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 (Knott *et al.*, 1998). The location of most QTL, including the gene for fat levels in pigs, is too imprecise for direct positional cloning. Similarly, in another study in pigs, a QTL which accounted for 11% of the variance could only be mapped to an accuracy of ±4 cM which, again, makes positional cloning to identify the causal gene(s) somewhat impractical (Knott & Haley, 1992). Further studies to narrow the region of interest 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.

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 animal populations. These associations are known as linkage disequilibrium. An association between a candidate gene and a trait could be due to the existence of linkage disequilibrium between the candidate locus and the true QTL. Linkage disequilibrium and functional studies in different populations may be necessary to identify the QTL unequivocally.

EXPLOITATION OF DNA POLYMORPHISMS

The simplest way to avoid dominantly-inherited Mendelian diseases is to avoid breeding with affected individuals. DNA tests are not usually required for dominant diseases unless there are difficulties identifying the individuals with the affected gene and the disease. The simplest strategy to avoid recessive Mendelian diseases is to ensure that at least one parent is not a carrier. For relatively common diseases such as BLAD, sire catalogues will list animals that have been tested and are known to be carriers or non-carriers. Where DNA tests do not exist, crossing a sire with a number of known carriers (females that have given birth to diseased individuals) can indicate if the sire is a carrier. Similar considerations apply to genes with incomplete penetrance.

The exploitation of DNA polymorphisms that are associated with multifactorial traits depends whether the polymorphism defines a marker or defines the QTL itself. If the QTL has been identified then exploitation is relatively straightforward and has been extensively discussed (Simm, 1998). The relative advantage of DNA typing will depend upon the cost and how much improvement DNA typing makes over current methods. The improvement will also depend upon the relative contribution of a QTL to the overall genetic variation. Additionally, genes for traits that are difficult to measure, such as meat quality and disease resistance, are likely to be more useful than genes for readily measured traits, such as total milk production in dairy cows or back fat depth in pigs.

If the QTL has not been identified but markers exist, these markers will be specific for the families in which they were identified. Nonetheless, they can still be used to improve the efficiency of animal breeding. For example, the top dairy bulls often contribute several male offspring to the next round of genetic evaluation. If the dairy bull is known to be segregating for a valuable QTL allele, offspring that inherit the two corresponding flanking markers are also likely to inherit the QTL. In contrast, offspring that inherit the other non-QTL allele are less likely to inherit the QTL. Therefore, genetic evaluation could be restricted to those offspring that inherit the appropriate flanking markers. A similar scheme is being used in New Zealand (S. Bishop, personal communication).

FUTURE DEVELOPMENTS

Animal genotyping is a powerful aid to animal breeding. The development of genetic markers can be summarized as follows. Polymorphic loci of known chromosomal location are identified, and methods to type these polymorphisms are developed. These polymorphisms are then tested on individuals with relevant phenotypes, and statistical relationships are calculated. The chance of detecting markers depends upon the relative contribution of the marker to the trait, the degree of linkage between the marker and the trait, and the frequency of desirable marker alleles in the population. The heritability of the trait, the amount of variation among animals, the number of animals studied and the method of analysis are also important. Another essential ingredient of marker development is recording accurate and appropriate phenotypic data on the traits of interest. Some parameters are relatively easy to score, such as litter size or live weight gain, which makes selection relatively straightforward. Other traits, like meat quality and disease resistance, are more demanding to assess. DNA markers are likely to be especially valuable for these hard to measure traits.

Microarray technology is likely to revolutionize selective breeding. A comprehensive profile, based on large numbers of polymorphisms could be generated for each animal. This could provide a potent method to predict the future performance and breeding value of any animal. In establishing microarray technology for genotyping, new strategies need to be developed that will allow the rapid identification of new SNPs. Above all, the key to establishing microarray technology in animal farming rests with central government funding bodies and the farming industry. Their long-term political and financial support is critical for the successful

adoption of this emerging technology. With or without microarray technology, however, animal genotyping, based on microsatellite and RFLP markers, and conventional gel-based separation technology, is already well established and will continue to grow in prominence as more useful markers are identified.

ACKNOWLEDGEMENTS

This work was supported by grants from MAFF, BBSRC, SERAD and Cotswold Pig Development Company.

REFERENCES

ANDERSSON, L., HALEY, C. S., ELLEGREN, H., *et al.* (1994). Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* **263**, 1771–4.

BRAUN, A., LITTLE, D. P., REUTER, D., MÜLLER-MYSOK, B. & KÖSTER, H. (1997). Improved analysis of microsatellites using mass spectrometry. *Genomics* **46**, 18–23.

BRENIG, B. & LEEB, T. (1998). Ryanodine receptors and their role in genetic diseases. *International Journal of Molecular Medicine* **2**, 293–300.

BUITKAMP, J., FILMETHER, P., STEAR, M. J. & EPPLEN, J. T. (1996). Class I and class II *Major Histocompatibility Complex* alleles are associated with faecal egg counts following natural, predominantly *Ostertagia circumcincta* infection. *Parasitology Research* **82**: 693–6.

CHANG, K. C. (1998). Use of myosin heavy chain microsatellite markers to predict muscle composition in pigs. Patent: PCT/GB98/01120.

CUSHWA, W. T. & MEDRANO, J.F. (1996). Applications of the random amplified polymorphic DNA (RAPD) assay for genetic analysis of livestock species. *Animal Biotechnology* **7**, 11–31.

DAWSON, M., HOINVILLE, L. J., HOSIE, B. D. & HUNTER, N. (1998). Guidance on the use of prp genotyping as an aid to the control of scrapie. *The Veterinary Record* 623–5.

ESSÉN-GUSTAVSSON, B. (1993). Muscle-fiber characteristics in pigs and relationships to meat-quality parameters – review. In: Pork quality: genetic and metabolic factors (Puolanne, E. and Demeyer, D. I.), pp. 140–59. CAB International, Ltd, Wallingford UK.

HACIA, J. G. (1999). Resequencing and mutational analysis using oligonucleotide microarrays. *Nature Genetics* **21**, (supp. 1), 42–7.

HAFF, L. A. & SMIRNOV, I. P. (1997). Single nucleotide polymorphism identification assays using a thermostable DNA polymerase and a delayed extraction MALDI-TOF mass spectrometry. *Genome Research* **7**, 378–88.

HALEY, C. S. (1995). Livestock QTLs – Bringing home the bacon? *Trends in Genetics* **11**, 488–92.

KENNEDY, B. W., QUINTON, M. & ARENDONK, J. A. M. (1992). Estimation of effects of single genes on quantitative traits. *Journal of Animal Science* **70**, 2000–12.

KNOTT, S. A. & HALEY, C. S. (1992). Aspect of maximum likelihood interval mapping in a F_2 population. *Genetical Research* **60**, 139–51.

KNOTT, S. A., MARKLUND, L., HALEY, C. S., *et al.* (1998). Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and large white pigs. *Genetics* **149**, 1069–80.

KUKLIN, A., MUNSON, K., GJERDE, D., HAEFELE, R. & TAYLOR, P. (1998). Detection of single-nucleotide polymorphisms with the WAVE DNA fragment analysis system. *Genetical Testing* **1**, 201–6.

KWOK, P. Y. & CHEN, X. (1998). Detection of single nucleotide variations. *Genetic Engineering* **20**, 125–34.

LANDEGREN, U., NILSSON, M. & KWOK, P. Y. (1998). Reading bits of genetic information: Methods for single-nucleotide polymorphism analysis. *Genome Research* **8**, 769–76.

LI, J., BUTLER, J. M., TAN, Y., LIN, H., *et al.* (1999). Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry. *Electrophoresis* **20**, 1258–65.

LIPSHUTZ, R. J., FODOR, S. P. A., GINGERAS, T. R. & LOCKHART, D. J. (1999). High density synthetic oligonucleotide arrays. *Nature Genetics* **21**, (supp. 1), 20–4.

MICHAELS, S. D. & AMASINO, R. M. (1998). A robust method for detecting single-nucleotide changes as polymorphic markers by PCR. *The Plant Journal* **14**, 381–85.

MOORE, S. S., SARGEANT, L. L., KING, T. J., MATTICK, J. S., GEORGES, M. & HETZEL, G. J. (1991). The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics* **10**, 654–60.

MORTON, N. E. (1955). Sequential tests for the presence of linkage. *American Journal of Human Genetics*, **7**, 277–318.

NATARAJ, A. J., OLIVOS-GLANDER, I., KUSUKAWA, N. & HIGHSMITH, W. E. (1999). Single-strand conformation and heteroduplex analysis for gel-based mutation detection. *Electrophoresis* **20**, 1177–85.

NICHOLAS, F. W. (1996). *Introduction to veterinary genetics.* Oxford University Press, New York 64–5.

PARSONS, B. L. & HEFLICH, R. H. (1997). Genotypic selection methods for direct analysis of point mutations. *Mutation Research* **387**, 97–121.

PRIMMER, C. R., RAUDSEPP, T., CHOWDHARY, B. P., MØLLER, A. P. & ELLEGREN, H. (1997). Low frequency of microsatellites in the avian genome. *Genome Research* **7**, 471–82.

RHODES, M., STRAW, R., FERNANDO, S., *et al.* (1998). High resolution microsatellite map of the mouse genome. *Genome Research* **8**, 531–42.

ROHRER, G. A., ALEZANDER, L. J., KEELE, J. W., SMITH, T. P. & BEATTIE, C. W. (1994). A microsatellite linkage map of the porcine genome. *Genetics* **136**, 231–45.

ROSS, P. L., LEE K. & BELGRADER, P. (1997). Discrimination of single-nucleotide polymorphisms in human DNA using peptide nucleic acid probes detected by MALDI-TOF mass spectrometry. *Analytical Chemistry* **69**, 4197–202.

ROTHSCHILD, M. F., JACOBSON, C., VASKE, D., *et al.* (1996). The estrogen receptor locus is associated with major genes influencing litter size in pigs. *Proceedings of the National Academy of Science, USA* **93**, 201–5.

RUBINSZTEIN, D. C., AMOS, W., LEGGO, J. *et al.* (1995). Microsatellite evolution – evidence for directionality and variation in rate between species. *Nature Genetics* **10**, 337–43.

RUDOLPH, J. A., SPIER, S. J., BYRNS, G. ROJAS, C.V., BERNOCO, D. & HOFFMAN, E. P. (1992). Periodic paralysis in quarter horses: a sodium channel mutation disseminated by selective breeding. *Nature Genetics* **2**, 144–47.

SCHWAIGER, F.-W. GOSTOMSKI, D., STEAR, M. J., *et al.* (1995). An ovine *major histocompatibility complex DRB1* allele is associated with low faecal egg counts following natural, predominantly *Ostertagia circumcincta* infection. *International Journal for Parasitology*, **25**, 815–22.

SHUSTER, D. E., KEHRLI, M. E., ACKERMANN, M. R. & GILBERT, R. O. (1993). Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 9225–9.

SIMM, G. (1998). *Genetic improvement of cattle and sheep.* Farming Press, Ipswich.

SMALL, M. P., BEACHAM, T. D. WITHLER, R. E. & NELSON, R. J. (1998). Discriminating coho salmon (Oncorhynchus kisutch) population within the Fraser River, British Columbia, using microsatellite DNA markers. *Molecular Ecology* **7**, 141–55.

SPELMAN, R. J. & BOVENHUIS, H. (1998). Moving from QTL experimental results to the utilization of QTL in breeding programmes. *Animal Genetics* **29**, 77–84.

STEAR, M. J., BAIRDEN, K., BISHOP, S.C., BUITKAMP, J. *et al.* (1996). An ovine lymphocyte antigen is associated with reduced faecal egg counts following natural, predominantly *Ostertagia circumcincta* infection. *International Journal for Parasitology* **26**, 423–8.

TAILLON-MILLER, P., PIERNOT, E. E. & KWOK, P. Y. (1999). Efficient approach to unique single-nucleotide polymorphism discovery. *Genome Research* **9**, 499–505.

WANG, D. G., FAN, J-B., SIAO, C-J., *et al.* (1998). Large scale identification, mapping and genotyping of single nucleotide polymorphism in the human genome. *Science* **280**, 1077–82.

WELLER, J. I., KASHI, Y. & SOLLER, M. (1990). Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *Journal of Dairy Science* **73**, 2525–37.

WILKES, K., JUNG, M., CHEN, Y. & GILDERMANN, H. (1994). Porcine $(GT)_n$ sequences: structure and association with dispersed and tandem repeats. *Genomics* **21**, 63–70.

(Accepted for publication 29 December 1999)