Rev. sci. tech. Off. int. Epiz., 2005, **24** (1), 379-391

The use of marker-assisted selection in animal breeding and biotechnology

J.L. Williams

Division of Genetics and Genomics, Roslin Institute (Edinburgh), Roslin, Midlothian, Scotland, EH25 9PS, United Kingdom

Summary

Improvement of livestock has focused on the selective breeding of individuals with superior phenotypes. With the development of increasingly advanced statistical methods that maximise selection for genetic gain, this simple approach has been extremely successful in increasing the quantity of agricultural output. However, information now available on the organisation and functioning of the genome could be used in breeding programmes to improve a range of traits. Many traits are under the control of several genetic loci, each of which contribute to the variation in the trait and hence are called quantitative trait loci (QTL). While genetic markers for QTL that are linked to the trait gene could be used to choose animals for selective breeding programmes, the most effective markers are the functional mutations within the trait genes. Strategies to identify markers for traits and the application of these markers are described by reference to examples of loci that control a range of different traits.

Keywords

Genetic markers – Genome mapping – Linkage – Quantitative trait loci – Selective breeding.

Introduction

Traditionally, selective breeding in livestock was aimed at improving the genetics of local populations of animals to allow them to survive and thrive in the prevailing environment, thus providing food and a source of traction for the indigenous communities. This selective breeding led to the development of animals with characteristic phenotypes that could be classified as distinct breeds. In 1993 there were 783 cattle breeds, 863 sheep breeds, and 263 pig breeds worldwide; although, over recent years the number of breeds in each species has been rapidly declining (23). The diversity of phenotypes displayed by the various breeds of livestock is controlled by an equally broad genetic diversity, which provides the opportunity for the selection of animals with superior performance in specific desirable traits, such as growth rate, composition of products (e.g. milk, meat, egg, etc.), fertility, survival in different environments, and resistance to disease. Technological advances, like the use of artificial insemination (AI) and embryo transfer in some species and the application of statistical methods to maximise selection for genetic gain, have been applied in the developed world. These advances have resulted in dramatic improvements in simple production traits that can be readily measured. Consequently, in countries where the economic environment supports high input agriculture, there has been a dramatic increase in the level of productivity from the selective improvement of livestock.

Genetic selection

Potential problems

Improvements in one trait achieved by selective breeding are often associated with losses in other traits. Whereas the primary factors that promote natural selection are reproductive success and resistance to disease challenge, artificial selection in domestic species has, for the most

part, ignored health traits in favour of improving productivity traits. Nevertheless, productivity is dependent on good health, and in order to maintain the health status of domesticated species management strategies have become increasingly dependent on veterinary interventions and the prophylactic use of antibiotics. Recently there have been numerous health problems in man associated with the transfer of diseases from livestock (zoonoses), such as infection of the human population with *E. coli* and *Salmonella* through contact with farm animals and contaminated food products. In Europe, the suggestion that a variant of Creutzfeldt-Jakob disease in man arose from the bovine spongiform encephalopathy epidemic in cattle in the United Kingdom has heightened concerns over livestock management. The occurrence of avian influenza in Asia is an ongoing cause of concern with respect to human health worldwide. With an increased awareness of zoonoses there is public pressure for animal health to have a higher priority in livestock production systems. Additionally, the increasing demand by consumers for naturally produced products and the breakdown in effective antibiotic treatment, as a result of drug resistance in pathogens, provides a strong case for breeding animals with better natural resistance to disease.

There are many diverse pressures and concerns associated with livestock production. For example, consumers demand particular qualities in their food, such as a specific fat composition, while still insisting on the lowest possible price. For the producer, past selection choices have resulted in a decrease in fertility (e.g. in dairy cattle and broiler poultry) that is threatening the viability of production. Dairy farmers are also faced with increasing lameness in their herds, while poultry breeders have to cope with birds that have brittle bones. In addition, the inadvertent selection for genetic defects linked to desirable production characteristics is a potential risk. Moreover, intensive selection can lead to a narrowing of the genetic diversity in a species, which reduces the genetic variation available for future selection, and also potentially concentrates genetic defects.

Thus, while selection of animals based on phenotypic qualities has been extremely successful in increasing the quantity of produce and reducing production costs, a limited number of production traits have been improved. In order to respond to public demand and to develop a sustainable industry, it will be necessary to address the potential problems associated with traditional selection approaches by fully exploiting the new technologies available for the selection of genetically superior animals.

Opportunities for applying molecular genetics

If simple phenotype guided selection is used in isolation there are inevitably some selection choices that will have conflicting outcomes, considering the diverse range of traits that are important at different levels of the production chain. For example, alleles of a particular gene may be beneficial for one trait, but have negative effects on another. In most cases, different genes will be involved in controlling different traits. However, when the genes controlling different traits are located very close together on a chromosome it may appear that only one locus controls both traits because alleles at linked loci will generally be inherited together. Even with very closely linked genes, the alleles that are found together in the progeny of a particular individual will change because of recombination between them (Fig. 1). Knowing the alleles at particular genetic loci will enable the identification of individuals that carry the beneficial alleles and allow for direct selection of genetically superior animals at several loci simultaneously. Therefore, in theory at least, a strategy to select for improved performance in a number of traits could be developed using genetic markers, even when at the phenotypic level the traits may seem to be in conflict.

The major challenge that faces molecular geneticists is to identify markers for genes that control the phenotypic variation in the target traits. Two types of marker can be considered. First, markers that are sufficiently close to the trait gene on the chromosome such that, in most cases, alleles at the marker and the trait gene are inherited together. This type of marker is called a linked marker. At the population level alleles at linked markers cannot be used to predict the phenotype until the association between alleles at the marker and alleles at the trait-gene is known (called 'phase'). To determine phase, inheritance of the marker and trait gene has to be studied in a family. However, information on phase is only valid within that family and may change in subsequent generations through recombination. The second type of marker is a functional

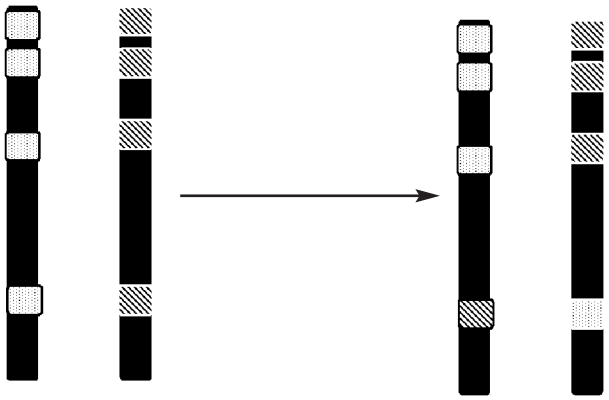


Fig. 1
Recombination between loci is more likely to occur between loci that are far apart than those that are close together

polymorphism in the gene that controls variation in the trait. These markers are called 'direct' markers. Once the functional polymorphism is known it is possible to predict the effect of particular alleles in all animals in a population, without first having to determine the phase. Therefore, 'direct' markers are more useful than 'linked' markers for predicting the phenotypic variation of target traits within a population (7).

A further complication is that the mechanisms of genetic control differ between traits. The variation seen in some traits is directly controlled by a single gene (monogenic traits), which may have a limited number of alleles. In the simplest situation a gene will have two alleles: one allele will be associated with one phenotype and another allele with a different phenotype (e.g. black versus brown coat colour in cattle: the brown coat color occurs as a result of a mutation in the melanocyte hormone receptor gene, which results in the creation of a different allele that alters its function). However, the traits that are important in livestock production are generally more complex and have a very large range of variation in the observed phenotype. Growth rate and milk yield are examples of two traits that exhibit a continuous phenotypic variation. Such traits are called quantitative traits. The variation in quantitative traits is controlled by several genetic loci (called quantitative trait loci [QTL]), each of which is responsible for a small amount of the overall variation (1).

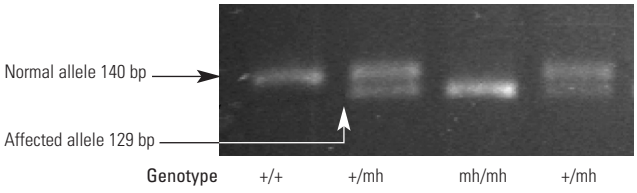
Genetic markers

Genetic markers can take a number of forms and in the simplest definition are: an observable genetically controlled variation that follows a mendelian pattern of inheritance. An example of such a marker is coat colour, which was used in early selection programmes for establishing various breeds of livestock, e.g. the white face in Hereford cattle and the white belt in Belted Galloway cattle and Saddleback pigs. More recently, protein polymorphisms, particularly blood groups, have been used to verify pedigrees in several species, including man, horses, cattle, and dogs. Such protein markers are generally impractical for use as genomic markers as they are relatively infrequent and, in some cases, the protein is expressed at low levels, making detection difficult, or is only found in specific tissues.

The earliest form of deoxyribonucleic acid (DNA) marker used to construct the first true genomic maps was the restriction fragment length polymorphism (RFLP). Restriction fragment length polymorphism analysis uses bacterial restriction enzymes to bind and cut DNA molecules at highly specific recognition sequences that are typically four to six base pairs (bp) long. There are a large number of restriction enzymes, each of which has a different specific recognition site. In theory, variations at restriction enzyme recognition sites in the genome could be identified by digesting genomic DNA with a restriction enzyme and observing the pattern of fragments produced by gel electrophoresis. However, due to the large number of recognition sites within the genome, too many fragments are produced for simultaneous examination. Therefore, the fragments produced by a genomic digest are separated by gel electrophoresis and transferred onto a solid matrix. Polymorphisms within a particular gene are then revealed by hybridisation with a radioactively labelled gene-specific probe. This approach allows each locus to be studied separately and is a powerful way to examine variations at a particular point in a given gene. The RFLP technique has been used to screen for carriers of genetic defects, e.g. bovine leukocyte adhesion deficiency (BLAD) in Holstein cattle. On occasion, it is desirable to screen several loci simultaneously, e.g. to produce a 'genetic fingerprint" to identify specific individuals. Several techniques are available to reveal RFLPs at a tractable number of loci at the same time. While not commonly employed at present, a method for the detection of RFLPs used probes for repeated sequences that occur within the genome. Mini-satellite repeated sequences are typically 20 bp to 50 bp in length, are often repeated many times at a particular locus, and may occur at 10 to 100 different sites in the genome. Variations in the number of repeats present at a particular locus give rise to a large number of alleles. These variable number tandem repeat (VNTR) markers have been used to identify relationships between individuals in wild populations, to verify pedigrees in many species, and in genetic mapping studies (10, 18). However, analysis of a single locus or multiple loci by this method required the use of radioactive probes and was slow and cumbersome.

The polymerase chain reaction

The development of the polymerase chain reaction (PCR) technique has revolutionised molecular genetics. This technique uses two short single strand DNA primers (typically 20 bp to 25 bp long) to initiate DNA replication at a specific point on the DNA molecule. A thermostable DNA polymerase is then used to copy the DNA by extending the primers and synthesising complementary strands of DNA. Repeatedly denaturing the DNA, reannealing the primers, and copying the DNA will exponentially amplify the target sequence between the primers. Following the amplification process, sufficient DNA is available for analysis directly by electrophoresis. The PCR-amplified DNA can be digested with a restriction enzyme and visualised by gel electrophoresis to determine if the PCR fragments have been cleaved. Polymerase chain reaction and PCR-RFLP are frequently used in diagnostic testing to determine the genotype at a known genetic mutation (Fig. 2).



The normal allele, denoted '+', is seen as a 140 bp fragment. The affected allele, denoted 'mh', has an 11 bp deletion and is seen as a 129 bp allele. Genotypes are given below the image as +/+ homozygous wild type, +/mh carrier, and mh/mh homozygous affected.

Fig. 2
Genotype of a genetic mutation in the myostatin gene determined by polymerase chain reaction (31)

Microsatellite markers

Microsatellite loci typically contain five to 20 copies of a short sequence motif that is between 2 bp and 4 bp in length and is repeated in tandem. The number of repeat units varies between individuals resulting in a large number of alleles for a given locus. The relatively large number of alleles at microsatellite loci and their amenability to PCR amplification make them excellent markers for use in genetic studies.

Single nucleotide polymorphisms and genome sequencing

Genetic variations fall into two classes: insertions or deletions of DNA sequences (indels) or changes to the nucleotide sequence (often affecting individual bases). Single nucleotide polymorphisms (SNPs) are much more frequent than indels and occur at high frequency in both non-coding regions and coding regions of the genome. Current estimates from genome sequencing projects indicate that SNPs occur every 200 bp on average. Single nucleotide polymorphisms within coding regions may have no effect on the protein coded by the gene (silent polymorphisms) or may result in a change in a single amino acid in the protein sequence. The latter are most likely to be the functional polymorphisms that are responsible for the phenotypic variation in traits. However, in some cases, the functional polymorphism responsible for variations in a trait may occur in intergenic regions (DNA sequences located between genes), e.g. insulin-like growth factor 2 (IGF2 – see the section entitled 'Carcass composition in pigs'). An advantage of using single nucleotide polymorphism markers is that they can be detected by methods other than electrophoresis: a method that is considered to be slow and difficult to automate. Following the discovery of many thousands of SNPs by the human sequencing project (17, 22), automated assays (e.g. using fluorescence or mass-spectroscopy) have been developed to genotype SNPs. It is now possible to rapidly genotype hundreds to thousands of individuals for several thousand SNP markers within a few hours.

A project to sequence the bovine genome was initiated in 2003, and the first draft of the genome sequence, which had on average three times coverage of the genome, was made publicly available in November 2004. A project to sequence the pig genome is currently being planned. Like the human genome-sequencing project, the bovine project has identified large numbers of SNPs within the genome. The immediate objective is to characterise and confirm 10,000 of these SNPs, while the ultimate goal is to produce a validated set of 300,000 SNPs.

Identifying markers for specific traits

Identifying genes that control particular traits can be approached in a number of ways. For simple monogenic traits it may be possible to postulate which gene(s) control the observed variation through studying the physiology of the trait and identifying the biochemical pathways that are involved. This information can be coupled with patterns of expression in various tissues, which can then be used to facilitate cloning of the gene involved. In some cases, knowledge of the gene that controls a similar phenotype in another species may suggest a potential analogous candidate gene that could be considered in the species of interest. The candidate genes are then studied to determine if the polymorphisms within the gene can account for the observed variation in the trait. This approach clearly requires good prior knowledge of the trait and the underlying physiology, or relevant information regarding the trait from other species. For more complex traits, several genes are likely to contribute to the observed variation. Even with a good knowledge of the physiology of such a trait other genes may be involved that are not obviously part of the biochemical pathways known to contribute to the variation in the trait. Therefore, for complex traits it may be better to make no prior assumptions regarding the physiology of the trait or the possible candidate genes that control the trait. Instead, a genetic mapping approach can be used.

Genes controlling particular traits are mapped by using markers to track inheritance of chromosomal regions in families in which the trait is segregated, and correlating this information with measurements on the individuals quantifying the trait. This approach will localise the major genes that control particular traits within broad chromosomal regions, which is the starting point for identifying the genes themselves. In practice, identification of the trait genes is achieved using a combination of genetic mapping, to localise the QTL region on a chromosome, and candidate gene or positional cloning approaches, to identify the trait gene within the QTL region. Following the initial low resolution mapping of a QTL, fine mapping

to improve the accuracy of the localisation is performed using additional individuals and markers. Better localisation of the trait gene can also be achieved by identifying the regions of the genome controlling the trait that have been inherited by different branches of a family. Over several generations recombination will reduce the amount of the ancestral genome inherited with the trait gene. Therefore, identifying regions of the genome associated with the trait that are identical by descent (IBD) can refine the location of a QTL to a small chromosomal region (1). Once the QTL region has been fine mapped, and in the absence of the identification of a candidate gene that is responsible for controlling the trait, it is then necessary to clone and sequence the region to provide information on the genes and variations present. Comparing the DNA sequence between individuals that show different phenotypes may then identify the specific genetic difference that controls the variation in the trait.

Quantitative trait loci mapping

Markers for quantitative trait loci mapping

Over the last decade considerable effort has been put into the construction of genetic and physical maps of the genomes of livestock species, e.g. pigs (2, 29) and cattle (3, 19). In the first instance these maps were composed predominantly of anonymous, microsatellite markers, but more recently, genes and expressed sequence tags (ESTs) have been added. The bovine genetic map now contains over 3,800 markers (16). These genetic maps have been used to select markers that are distributed across whole genome. These markers are then used in QTL mapping studies to track the inheritance of specific regions of chromosomes through generations of families. Microsatellite markers are commonly used in these studies because they usually have several alleles and hence the parental origin of a particular marker can usually be determined to track the inheritance of specific regions of chromosomes through generations of families. Microsatellite markers are therefore particularly appropriate for linkage mapping.

Populations used to map quantitative trait loci

In addition to a requirement for genetic markers, in order to map QTL it is necessary to have families that have been measured for the traits of interest and in which the trait of interest is segregating. Unfortunately, the range of traits that are routinely recorded in commercial livestock populations is very limited, and in most cases it is only simple traits, such as growth rate and milk yield that are recorded. An additional consideration is that traits are frequently sex specific. Nevertheless, such commercial populations can be used to map the QTL that control some simple traits. In the case of the dairy industry test bulls are mated to high performance cows and the production traits are measured in the female progeny. Bulls with high performing daughters are then used to produce semen that is sold for AI. By measuring the desirable production traits in a large number of female progeny, the genetic contribution of the bull (breeding value) can be determined with high accuracy. The sons of these 'elite' bulls are in turn progeny tested and the best amongst them used for breeding. Weller *et al.* (36) suggested that tracking chromosomal inheritance from the elite sire to his sons and correlating this information with the predicted breeding value of the son, based on the performance of his daughters, would be an efficient way to map the genes that control milk production traits. This 'grand-daughter' design approach was applied by Georges *et al.* (11) in the US Holstein population to map QTL involved in milk yield. Five QTL for dairy associated traits were identified, many of which have been confirmed independently in subsequent studies in other populations (21, 37).

Methods used to select bulls in the beef industry are different from those used in the dairy industry. A large number of breeds are used for beef production; however, these breeds have not been under as focused selection pressure as the dairy breeds (the Holstein breed in particular). In a number of countries, a few beef production traits, such as growth, fat, and conformation, are systematically recorded and used to select superior bulls for breeding. This data provides limited opportunities for mapping QTL for simple beef production traits. The majority of information on QTL that control beef associated traits comes from specifically bred resource herds that are kept under standardised management, which provides the opportunity to record traits that are difficult or impossible to track in commercial herds. The Meat Animal Research Center in Nebraska used a resource herd of cattle created by crossbreeding between beef breeds to map QTL for several beef associated traits, including growth rate, muscle mass, and fat meat texture (33).

A popular experimental design that is used to increase both the genetic and phenotypic variations available for study within a herd is to cross breeds with widely divergent phenotypes, e.g. crossing a dairy with a beef breed. At the Roslin Institute a Holstein × Charolais resource herd is being used to study traits that are important for beef and milk production and also for efficiency, health and fertility. By studying a wide range of traits in the same individuals it will be possible to identify pleiotropic interactions between the traits, i.e. to identify if selection for one trait will have a positive or negative effect on a different trait. Crossing of diverse breeds has also been employed in several QTL studies in pigs. Populations of pigs have been produced by crossing wild boars to several domestic breeds, as well as by crossing the Chinese Meishan to European production breeds, such as the Large White and

Landrace. These diverse crossbred resource populations have been successfully used to identify QTL for carcass and fertility traits (26, 28).

Surprising findings from crosses between animals with diverse phenotypes

Food preferences vary across the world, e.g. in Europe consumers prefer meat that is low in fat, whereas in Japan meat with a high fat content is sought. The selection of stock suited to particular markets has resulted in the production of breeds with widely divergent characteristics. Cattle breeds, such as the Belgian Blue, Charolais, and Limousin, are widely used for beef production in Europe because they are fast growing and produce lean carcasses. In contrast, the Wagu cattle breed is used to produce beef in Japan because it has exceptionally high levels of intra-muscular fat. A similar situation exists in the pig industry. Breeds used extensively in Europe have been selected for lean growth, whereas in the People's Republic of China the Meishan breed is favoured because it has a high fat content. A two-generation crossbred population established from Large White and Meishan pigs has been used to explore the genetic control of lean versus fat carcass growth. The F2 animals were genotyped with markers located throughout the genome and slaughtered to determine muscle and fat content. A QTL analysis correlating production data with the marker information revealed several QTLs that were associated with carcass fat. However, one QTL, on chromosome 7, had a particularly large effect, accounting for about a 30% difference in back fat thickness (8). The most surprising finding, however, was that the 'thin allele' originated from the Meishan breed, which phenotypically is very fat (A. Archibald, personal communication). This demonstrated that the beneficial allele for a particular QTL might not all be fixed in a population that shows the desirable characteristics for the trait.

Mapping trait genes in livestock

The localisation of QTL provides markers linked to the trait genes that could be used in breeding programmes to improve the selection for a particular trait. However, to be applied in this way it is first necessary to determine the phase of the alleles at the markers and trait gene. Therefore, although markers linked to QTL can be used for marker-assisted selection, identifying the underlying genes that control the variation provides more powerful markers. A number of different approaches have been used to identify the genes that control the genetic variation for a wide range of target traits, some of which are discussed below.

Mastitis

Mastitis (an infection of the udder that can be caused by several different pathogens) is the most common disease in dairy cattle. Mastitis infection is both a welfare and an economic problem, the consequences of which include decreased milk yield, increased veterinary treatments, and, in many cases, death (it is the most common cause of death in dairy cattle). The incidence of mastitis is not generally recorded, except in Scandinavia where individual records of disease are kept for every cow as part of a national health-recording programme. Linkage mapping studies using the Scandinavian data collected from half-sibling cows from the same sire have identified an association between putative QTL on chromosomes 3, 4, 14, and 27 and the incidence of mastitis (20). More commonly, somatic cell scores (SCS) are used as an indicator of the incidence of mastitis. These measurements show a correlation ranging between 0.37 and 0.97. The SCS is a measure of the inflammatory response occurring in the udder, possibly following invasion of a pathogen and, thus, does not necessarily indicate clinical or sub-clinical infection; however, SCS are routinely recorded in commercial dairy herds, whereas mastitis is not. Somatic cell scores have been used as an indicator of mastitis in several QTL studies, but a direct link between the QTL for SCS, and resistance and susceptibility to mastitis has not been demonstrated.

For poorly defined traits, such as mastitis resistance, it is difficult to select candidate genes that may control the trait because a large number of different pathogens and physiological mechanisms contribute to the observed variation. One possible candidate locus is the major histo-compatibility complex (MHC) locus. The MHC region includes genes that code for cell surface proteins involved in binding peptides (e.g. peptides from pathogens) and presenting them correctly to cells of the immune system so that an appropriate immune response can be initiated. Thus, the MHC is a good candidate locus when considering variation in the immune response or resistance to disease challenge. There are several MHC genes within the MHC loci that code for sub-units of Class I and Class II cell surface molecules. The Class II molecules are hetero-dimers composed of an α and a β sub-unit. The DRβ gene, which has a number of different alleles, is one of the several MHC genes. The different alleles of the MHC gene differ in the active part of the molecule that is involved in peptide binding and presentation. A number of studies have shown associations between the alleles at the MHC DRβ locus and SCS (30). This suggests that the DRβ alleles carried by an individual may be associated with differences in susceptibility or resistance to mastitis.

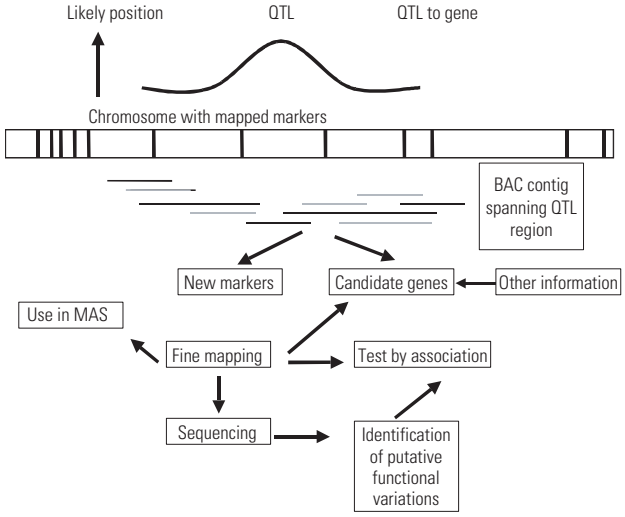
Double muscling

The genomic mapping approach has identified many QTLs for production traits in livestock, but few trait genes have been identified from the chromosomal location of the QTLs found by the mapping studies. One gene that has been identified by this approach is associated with 'double muscling' in cattle. Several European beef breeds are

segregating for this phenotype, which is characterised by muscular hypertrophy and reduced intra-muscular fat (25). Double muscling is also associated with dystocia and, consequently, calving problems, with a high rate of caesarean sections in breeds that express this trait. The most extreme form of double muscling is seen in the Belgian Blue breed in which the trait behaves as if it is controlled by a single major gene.

The gene responsible for the double muscled phenotype was mapped using a herd of research cattle that were created by crossing double muscled Belgian Blue cattle to a non-carrier breed. The F1 crossbred animals were backcrossed to the double muscled Belgian Blue to produce a second generation cross in which those homozygous for the gene showed double muscling and heterozygotes did not. These animals were genotyped with a panel of about 200 markers located throughout the genome, and the genotypes were correlated with the phenotypes to localise the gene responsible for the double muscling trait. This approach identified a region on bovine chromosome 2 as the most likely location of the double muscling gene (5).

The first approach that was used to identify the double muscling gene was to select and test candidate genes that were located within the region of chromosome 2 that was suspected to contain the gene. However, none of these genes accounted for the occurrence of double muscling at the population level. Therefore, a fine mapping and positional cloning strategy was initiated by constructing an overlapping set of large fragment DNA clones that spanned the region (Fig. 3). As this work was in progress, an unconnected study in mice identified a new member of the transforming growth factor family of genes. The new member, initially called growth differentiation factor 8 (GDF-8), was expressed mainly in skeletal muscle. Transgenic knockout mice (i.e. mice in which the GDF-8 gene was inactivated by genetic manipulation) developed hyper-muscularity that was similar to the double muscling phenotype in cattle (24). The GDF-8 gene product was found to be a negative regulator of muscle growth and was therefore called myostatin. Subsequent studies in Belgian Blue cattle showed that double muscled animals carried an 11 bp deletion within the coding region of the myostatin gene (14). Analysis of other double muscled breeds of cattle showed that they all carried a mutation in the coding region of the myostatin gene, further supporting the conclusion that this gene controlled the double muscling phenotype. To study the phenotype associated with the mutation, the genotype of individuals can be determined by detection of the 11 bp deletion using amplification fragment length polymorphism (Fig. 2). Using this simple test the frequency of the double muscling allele can be increased or decreased by the selection of animals based on their genotype.



MAS: marker-assisted selection

The first step is to construct a physical map of the region from bacterial artificial chromosome (BAC) clones. This physical map can then be used to identify new markers to fine map the region. In some cases, a good candidate gene may be found within the QTL region that can be tested; otherwise it may be necessary to sequence across the region using the BAC clones as the starting point.

Fig. 3
An illustration of the fine mapping and positional cloning strategy that is used to identify a trait gene starting from a quantitative trait loci (QTL) location

Control of milk fat in bovine milk

Studies, such as the grand-daughter design studies described above, that used commercial herds of dairy cattle have identified a number of QTL for milk yield and composition (11). One of these QTL, which mapped to chromosome 14 (6), has been shown to affect the fat content of milk. In order to fine map the QTL region, a contiguous set of bacterial artificial chromosome (BAC) clones was identified using all available bovine markers for chromosome 14 and some ESTs found on human chromosome 8. Parts of this human chromosome have sections that are in common with bovine chromosome 14 (i.e. the chromosomes share regions of conserved synteny). A gene known to affect milk synthesis identified from studies in mice (acyiCoA:diacylglycerol acyltransferase 1 [DGAT1]) has been located on human chromosome 8 (4) within the region of conserved synteny between human chromosome 8 and bovine chromosome 14. This gene was also identified within the BAC contig that spanned the QTL region. The DGAT1 gene was sequenced from bulls that were predicted to carry different alleles at the chromosome 14 QTL. A number of differences were found, including a two base-pair substitution in exon 12 of the gene that resulted in an amino acid change (alanine to lysine) in the mature protein (13). Examination of this polymorphism in the Holstein

cattle population has shown that cows with the lysine allele have considerably increased milk fat compared to those with the alanine allele. Subsequent studies in Holsteins and other breeds have confirmed that the lysine allele is associated with an increased milk fat content, but have also shown that this allele is related to decreased protein percentage and overall milk yield (32, 34). Thus, breeders could change the milk fat content and, to some extent, the protein level in the milk by selecting animals based on their DGAT1 genotype.

Carcass composition in pigs

Resource populations in which founders with distinct phenotypes are crossed have produced a large amount of information on QTL, particularly on loci that affect carcass traits. Fine mapping of a QTL for muscle and fat depth on chromosome 2 in pigs narrowed the likely position of the underlying gene to a region that included the IGF2 gene (27). Sequencing across the IGF2 locus revealed 258 polymorphisms (35). These polymorphisms could be separated into two separate haplotype clusters that were either of European or Chinese origin. Correlation of the haplotype blocks with the trait revealed a single SNP, a G to A transition within intron 3, that appeared to be either the causative mutation or a quantitative trait nucleotide (QTN). This QTN occurs in a region of DNA that is methylated during imprinting (inactivation of one of the parental chromosomes). It is also thought to be part of a binding site for a protein that regulates gene expression. The allele associated with lean meat has been strongly selected for in European pig breeds.

The IGF2 QTN that controls carcass fat is an interesting example of a genetic variation located in a non-coding region of the genome that has a large effect on a production trait. However, this is not the only example. The Callipyge phenotype in sheep, which is associated with increased muscling, has also been mapped to an imprinted, non-coding region of the sheep genome (9).

Eradicating deleterious alleles

The unfortunate consequence of intensive selection is a reduction in genetic diversity, not only at the loci under selection, but also at all loci across the genome. Selective breeding may also inadvertently increase the frequency of deleterious recessive alleles at loci that are linked to those under selection. In dairy cattle a limited number of sires are used as donors for AI. The same sires are also used as the gene pool to produce the next generation of sires. This practice inevitably leads to an increase in inbreeding and an increased possibility that the progeny will receive the regions of the genome that are identical by descent from each of the parents.

Bovine leukocyte adhesion deficiency

In the early 1990s a disease of young Holstein calves was observed that was characterised by pneumonia, delayed healing of wounds, and death. This disease followed an autosomal recessive pattern of inheritance and showed a similar pathology to leukocyte adhesion deficiency in man. The defective gene in the human form of the disease encodes the cluster of differentiation antigen 18 (CD18) protein. Subsequent sequencing of this gene from affected calves revealed a mutation within the bovine CD18 gene, and the disease was therefore called bovine leukocyte adhesion deficiency (BLAD). At the time the disease was identified the defective allele had reached a frequency of 15% in bulls used for AI in the United States of America and 8% in the cow population (12). The mutation was traced to Osborndale Ivanhoe: the bull that has had the largest known impact on the Holstein gene pool. Using a PCR RFLP test developed to detect the defect in the CD18 gene, it was possible to identify BLAD carriers and ensure that they were not used for breeding, thus, gradually eliminating this defect from the population.

Using molecular genetic information to understand phenotypic variation

For many decades genetic improvement in livestock has focused on traits that have an impact on the profitability of agricultural enterprises; generally the focus has been on the selection for simple traits that are easy to measure. However, many traits that have an impact on profitability and welfare, such as feed conversion efficiency, fertility, and disease resistance, are difficult to measure and therefore are not routinely recorded. Obtaining the necessary phenotype information to carry out a selective breeding programme would be expensive. Furthermore, using traditional methods, the improvement in one trait is often compromised by simultaneous selection for other traits. Identification of the genes that control desired traits has the potential to enable selection for a trait based on the genotype of animals. In theory, once sufficient knowledge is available, individuals carrying beneficial genes for several traits could be identified using DNA markers and mated to produce progeny that will express many of the desired traits simultaneously. While the scientific community is a long way from acquiring this level of knowledge, research is underway to identify the genes involved in a wide range of traits, as described above. The ultimate goal of molecular genetics is to understand how alleles of the genes controlling various traits interact and control the observed variations in the phenotype.

Introgression of genes between populations

A further potential use of marker technology in breeding is the transfer of genes that control desired phenotypes between breeds. Keeping breeds that have been highly selected for improved production is impossible in some parts of the world, without extensive veterinary input, because of the susceptibility of these 'improved' livestock to endemic diseases. However, indigenous breeds are often resistant to these endemic diseases. If the genes controlling resistance to the specific disease were identified, it would theoretically be possible to transfer them from the indigenous breed into the 'improved breed', thus, producing stock that have an increased production potential and are resistant to endemic disease. The introgression of disease resistance genes into the improved breeds would be achieved initially by crossing the indigenous and improved breeds. The first generation crossbred animals would then be backcrossed to the improved breeds and the animals genotyped for the genes involved in the disease resistance. Animals carrying the favourable alleles would be selected for breeding and backcrossed again to the improved breed. By repeated backcrossing and selective breeding from the animals carrying the favourable disease resistance alleles, it is possible to 'recover' the majority of the genome from the 'improved breed', while maintaining the disease resistance that originated from the indigenous breed.

A QTL study of F2 cattle produced from crossing N'Dama and Boran cattle, which are resistant and susceptible to trypanosomosis, respectively, revealed that 18 QTL regions were associated with resistance to the disease (15). Unfortunately, introgression of such a complex trait from resistant cattle to other breeds could not be considered with marker-assisted selection.

Maintaining genetic diversity

At the phenotypic level, the presence of favourable alleles in an individual is only revealed by measuring the performance of the individual for the trait. For complex traits, animals with above average performance for the trait will have 'good' alleles for several of the genes involved in the trait; although, these may not necessarily be the 'best' alleles. Following the identification of above average individuals, phenotypic selection then relies on continued selection using these 'superior animals'. This strategy inevitably results in a small number of individuals having a large influence on the gene pool, which, in turn, results in losses in genetic diversity, particularly the loss of rare alleles. Knowledge of the genes that control a trait will allow efficient selection for the superior alleles at those genes. However, if additional markers dispersed across the genome are also used to characterise the animals used for breeding, it would be possible to devise strategies to maintain the widest possible genetic diversity, while still making optimum progress in the target trait. Maintaining variation in loci that are not currently under selection is important for sustaining diversity in trait genes that may be selected for in the future.

Future perspectives and conclusions

Knowledge of the genes underling the expression of a trait will allow researchers to search for novel combinations of alleles to make further improvements in desired traits, as the most beneficial alleles and allele combinations may not occur in the production populations. Breeding schemes could then be designed to test novel combinations of alleles at different loci. Some of these novel allele combinations may result in the improvement of particular traits beyond that which would be possible by selecting phenotypically superior animals within a population. It is therefore important to maintain a diverse range of genetic backgrounds to provide sources of variation.

The behaviour of genes (including major genes) that control a trait is likely to be dependent on the genetic background. The myostatin allele found in Belgian Blue cattle is also found in other breeds; however, the phenotype associated with the allele is variable between the breeds. This suggests that there are genes at other loci in the genome that act to modify the phenotypic expression of the major gene. Thus, information is required not only on the major genes that control a trait, but also on the interactions between genes. Further information on gene interactions may be obtained from gene expression studies. Gene expression micro-arrays that allow the expression patterns of many thousands of genes to be assayed simultaneously have now been produced for the majority of livestock species. Compiling information on expression patterns in different tissues and species could reveal co-regulated physiological pathways (i.e. pathways that are regulated by common genes or sets of genes) that are currently not known.

Information on the genes that control commercially important traits is only just emerging from the numerous studies that are underway. For those genes that have been identified, the level of variation within the genes between individuals or populations is not known, nor is the effect of specific variations on phenotypes. As discussed above in the context of double muscling, the effects of even well-characterised variations that are associated with a major phenotype can vary depending on the genetic background. It is therefore premature to start using DNA-based selection widely. However, some DNA tests for specific polymorphisms are being offered commercially, e.g. the

GeneSTAR tests for tenderness (based on variations in the calpastatin gene) and marbling (based on variations in the thyroglobulin gene), and the Ingenity test for fat deposition (based on variations in the leptin gene). These tests can be used by breeders and evaluated in their populations.

Variations in phenotypes may arise from functional differences within the coding regions of genes, which change the structure of the protein that the genes code for, or from differences in the DNA sequences that regulate the expression of the gene. The coding regions of genes can generally be identified, but the role of the non-coding regions is poorly understood. A project is underway to sequence the bovine genome, and sequences are now available for the genomes of several different species. Comparison of the coding and non-coding regions of the genome across species is a further approach that is used to identify functionally significant variations in genes. It will also help to identify regulatory elements in non-coding regions that may account for some of the variation that is seen in particular traits, such as found for fat deposition controlled by the IGF2 gene in pigs or the Callipyge phenotype in sheep. Livestock species are an excellent model for studies on gene function and regulation because of the vast genetic diversity that is present and the extended families that are available to dissect specific regions of the genome by studying recombination throughout generations. What is urgently required are populations that have had phenotypes well recorded for a wide range of traits, so that the wealth of genetic information that is being produced from recent advances in genomics technology can be linked to phenotypic variation and used to explore genome function.

■

Utilisation de la sélection assistée par marqueurs dans le domaine de la biotechnologie et de la reproduction animales

J.L. Williams

Résumé

L'amélioration du bétail est axée sur la reproduction sélective d'individus caractérisés par des phénotypes supérieurs. Grâce au perfectionnement permanent des méthodes statistiques qui optimise la sélection en vue d'un gain génétique, cette méthode simple a permis d'augmenter considérablement la quantité d'animaux produits. Cela étant, les données dont on dispose actuellement sur l'organisation et le fonctionnement du génome pourraient être utilisées dans les programmes d'élevage pour améliorer divers traits. De nombreux caractères sont contrôlés par plusieurs régions du génome, chacune contribuant à la variation d'un caractère ; ces zones sont ainsi appelées loci de caractères quantitatifs (QTL). Si les marqueurs génétiques des QTL qui sont liés au gène responsable d'un caractère peuvent être utilisés pour guider le choix des animaux destinés à des programmes d'élevage sélectif, les marqueurs les plus efficaces sont les mutations fonctionnelles au sein des gènes influençant les caractères. Les stratégies visant à identifier les marqueurs pour la détection des gènes influençant les caractères et l'utilisation de ces marqueurs sont décrites en faisant référence à des exemples de loci qui contrôlent un ensemble de caractères différents.

Mots-clés

Élevage sélectif – Lien – Locus de caractères quantitatifs – Marqueur génétique – Séquençage du génome.

■

La biotecnología y el uso de la selección mediante marcadores en reproducción animal

J.L. Williams

Resumen

Tradicionalmente, los procedimientos de mejora del ganado consistían por lo esencial en seleccionar ejemplares dotados de un fenotipo superior. Gracias a la aparición de métodos estadísticos cada vez más avanzados, que optimizan la selección para obtener las características genéticas deseadas, este sencillo proceder ha sido muy útil para lograr resultados cuantitativamente mejores. Sin embargo, la información que ahora existe sobre la organización y el funcionamiento del genoma podría utilizarse en programas de selección para mejorar una serie de características. Hay numerosos rasgos controlados por varios loci genéticos, denominados loci de rasgos cuantitativos porque cada uno de ellos contribuye a la variación del rasgo en cuestión. Aunque es posible utilizar marcadores genéticos de esos loci (ligados al gen de que se trate) como elemento auxiliar para elegir a los animales destinados a programas de selección, los marcadores más eficaces son las mutaciones funcionales dentro de los propios genes que codifican determinado rasgo. Ofreciendo varios ejemplos de loci que controlan una serie de rasgos distintos, el autor describe métodos para localizar marcadores de rasgos y aplicarlos en la práctica.

Palabras clave

Cartografía genética – Cría selectiva – Ligamiento – Locus de rasgos cuantitativos – Marcador genético.

References

1. Andersson L. & Georges M. (2004). – Domestic animal genomics: deciphering the genetics of complex traits. *Nature Rev. Genet.*, **5**, 202-212.

2. Archibald A.L., Haley C.S., Brown J.F., Couperwhite S., McQueen H.A., Nicholson D., Coppieters W., Van de Weghe A., Stratil A., Wintero A.K. *et al.* (1995). – The PiGMaP consortium linkage map of the pig *(Sus scrofa)*. *Mamm. Genome*, **6** (3), 157-175.

3. Barendse W., Vaiman D., Kemp S.J., Sugimoto Y., Armitage S.M., Williams J.L., *et al.* (1997). – A medium-density genetic linkage map of the bovine genome. *Mamm. Genome*, **8** (1), 21-28.

4. Cases S., Smith S.J., Zheng Y.W., Myers H.M., Lear S.R., Sande E., Novak S., Collins C., Welch C.B., Lusis A.J., Erickson S.K. & Farese R.V. Jr (1998). – Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. *Proc. natl Acad. Sci. USA*, **95** (22), 13018-13023.

5. Charlier C., Coppieters W., Farnir F., Grobet L., Leroy P.L., Michaux C., Mni M., Schwers A., Vanmanshoven P., Hanset R. & Georges M. (1995). – The mh gene causing double-muscling in cattle maps to bovine chromosome 2. *Mammalian Genome*, **6** (11), 788-792.

6. Coppieters W., Riquet J., Arranz J.-J., Berzi P., Cambisano N., Grisart B., Karim B.L., Marcq F., Moreau L., Nezer C., Simon P., Vanmanshoven P., Wagenaar D. & Georges M. (1998). – A QTL with major effect on milk yield and composition maps to bovine chromosome 14. *Mammalian Genome*, **9** (7), 540-544.

7. Dekkers J.C.M. (2004). – Commercial application of marker and gene assisted selection in livestock: strategies and lessons. *J. Anim. Sci.*, **82**, E313-E328.

8. De Koning D.J., Janss L.L.G., Rattink A.P., van Oers P.A.M., de Vries B.J., Groenen M.A.M., van der Poel J.J., de Groot P.N., Brascamp E.W. & van Arendonk J.A.M. (1999). – Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (*Sus scrofa*). *Genetics*, **152** (4), 1679-1690.

9. Freking B.A., Murphy S.K., Wylie A.A., Rhodes S.J., Keele J.W., Leymaster K.A., Jirtle R.L. & Smith T.P.L. (2002). – Identification of the single base change causing the callipyge muscular hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res.*, **12** (10), 1496-1506.

10. Georges M., Lathrop M., Hilbert P., Marcotte A., Schwers A., Swillens S., Vassart G. & Hanset R. (1990). – On the use of DNA fingerprints for linkage studies in cattle. *Genomics*, **6** (3), 461-474.

11. Georges M., Nielsen D., Mackinnon M., Mishra A., Okimoto R., Pasquino A.T., Sargeant L., Sorensen A., Steele M.R., Zhao X., Womack J. & Hoeschele I. (1995). – Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics*, **139** (2), 907-920.

12. Gilbert R.O., Rebhun W.C., Kim C.A., Kehrli M.E. Jr, Shuster D.E. & Ackermann M.R. (1993). – Clinical manifestation of leukocyte adhesion deficiency in cattle: 14 cases (1977-1991). *J. Am. vet. med. Assoc.,* **202** (3), 445-449.

13. Grisart B., Farnir F., Karim L., Cambisano N., Kim J.J., Kvasz A., Mni M., Simon P., Frère J.M., Coppieters W. & Georges M. (2004). – Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. *Proc. natl Acad. Sci. USA*, **101** (8), 2398-2403.

14. Grobet L., Martin L.J.R., Poncelet D., Pirottin D., Brouwers B., Riquet J., Schoeberlein A., Dunner S., Ménissier F., Massabanda J., Fries R., Hanset R. & Georges M. (1997). – A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nature Genet.*, **17** (1), 71-74.

15. Hanotte O., Ronin Y., Agaba M., Nilsson P., Gelhaus A., Horstmann R., Sugimoto Y., Kemp S., Gibson J., Korol A., Soller M. & Teale A. (2003). – Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proc. natl Acad. Sci. USA*, **100** (13), 7443-7448.

16. Ihara N., Takasuga A., Mizoshita K., Takeda H., Sugimoto M., Mizoguchi Y., Hirano T., Itoh T.H.T., Watanabe T., Reed K.M., Snelling W.M., Kappes K.M., Beattie C.W., Bennett G.L. & Sugimoto Y. (2004). – A comprehensive genetic map of the cattle genome based on 3802 microsatellites. *Genet. Res.*, **14** (10A), 1987-1998.

17. International SNP Map Working Group (2001). – A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, **409** (6822), 928-933.

18. Jeffreys A.J., Wilson V. & Thein S.L. (1985). – Hypervariable 'minisatellite' regions in human DNA. *Nature*, **314** (6006), 67-73.

19. Kappes S.S., Keele J.W., Stone R.T., McGraw R.A., Sonstegard T.S., Smith T.P., Lopez-Corrales N.L. & Beattie C.W. (1997). – A second-generation linkage map of the bovine genome. *Genome Res.*, **7** (3), 235-249.

20. Klungland H., Sabry A., Heringstad B., Olsen H.G., Gomez-Raya L., Våge D.I., Olsaker I., Ødegård J., Klemetsdal G., Schulman N., Vilkki J., Ruane J., Aasland M., Rønningen K. & Lien S. (2001). – Quantitative trait loci affecting clinical mastitis and somatic cell count in dairy cattle. *Mamm. Genome*, **12** (11), 837-842.

21. Kühn C.H., Freyer G., Weikard R., Goldammer T. & Schwerin M. (1999). – Detection of QTL for milk production traits in cattle by application of a specifically developed marker map of BTA6. *Anim. Genet.*, **30** (5), 333-340.

22. Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C. *et al.* (2001). – Initial sequencing and analysis of the human genome. *Nature*, **409** (6822), 860-921.

23. Loftus R. & Scherf B. (1993). – World watch list for domestic animal diversity. Food and Agriculture Organization, Rome.

24. McPherron A.C., Lawler A.M. & Lee S.J. (1997). – Regulation of skeletal muscle mass in mice by a new TGF-β superfamily member. *Nature*, **387** (6628), 83-90.

25. Ménissier F. (1982). – General survey of the effect of double muscling on cattle performance. *In* Muscle hypertrophy of genetic origin and its use to improve beef production (J.W.B. King & F. Ménissier, eds). Martinus Nijhoff Publishers, London, 23-53.

26. Nagamine Y., Haley C.S., Sewalem A. & Visscher P.M. (2003). – Quantitative trait loci variation for growth and obesity between and within lines of pigs (*Sus scrofa*). *Genetics*, **164** (2), 629-635.

27. Nezer C., Moreau L., Brouwers B., Coppieters W., Detilleux J., Hanset R., Karim L., Kvasz A., Leroy P. & Georges M. (1999). – An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. *Nature Genet.*, **21** (2), 155-156.

28. Rattink A.P., De Koning D.J., Faivre M., Harlizius B., van Arendonk J.A.M. & Groenen A.M. (2000). – Fine mapping and imprinting analysis for fatness trait QTLs in pigs. *Mamm. Genome*, **11** (8), 656-661.

29. Rohrer G.A., Alexander L.J., Hu Z., Smith T.P., Keel J.W. & Beattie C.W. (1996). – A comprehensive map of the porcine genome. *Genet. Res.*, **6**, 371-391.

30. Sharif S., Mallard B.A., Wilkie B.N., Sargeant J.M., Scott H.M., Dekkers J.C. & Leslie K.E. (1998). – Associations of the bovine major histocompatibility complex DRB3 (BoLA-DRB3) alleles with occurrence of disease and milk somatic cell score in Canadian dairy cattle. *Anim. Genet.*, **29** (3), 185-193.

31. Smith J.A., Lewis A.M., Wiener P. & Williams J.L. (2000). – Genetic variation in the bovine myostatin gene in UK beef cattle: allele frequencies and haplotype analysis in the South Devon. *Anim. Genet.*, **31** (5), 306-309.

32. Spelman R.J., Ford C.A., McElhinney P., Gregory G.C. & Snell R.G. (2002). – Characterization of the DGAT1 gene in the New Zealand Dairy population. *J. Dairy Sci.*, **85** (12), 3514-3517.

33. Stone R.T., Keele J.W., Shackelford S.D., Kappes S.M. & Koohmaraie M. (1999). – A primary screen of the bovine genome for quantitative trait loci affecting carcass and growth traits. *J. Anim. Sci.*, **77** (6), 1379-1384.

34. Thaller G., Kramer W., Winter A., Kaupe B., Erhardt G. & Fries R. (2003). – Effects of DGAT1 variants on milk production traits in German cattle breeds. *J. Anim. Sci.*, **81** (8), 1911-1918.

35. Van Laere S.A., Nguyen M., Braunschweig M., Nezer C., Collette C., Moreau L., Archibald A.L., Haley C.S., Buys N., Tally M., Andersson G., Georges M. & Andersson L. (2003). – A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature*, **425** (6960), 832-836.

36. 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. *J. Dairy Sci.*, **73** (9), 2525-2537.

37. Wiener P., MacLean I., Williams J.L. & Woolliams J.A. (2000). – Testing for the presence of previously identified QTL for milk production traits in new populations. *Anim. Genet.*, **31** (6), 385-395.