CSIRO PUBLISHING
www.publish.csiro.au/journals/ajea
Australian Journal of Experimental Agriculture, 2005, **45**, 831–836

The transition from quantitative trait loci to diagnostic test in cattle and other livestock

W. Barendse

CSIRO Livestock Industries, Queensland Bioscience Precinct, 306 Carmody Road, St Lucia, Qld 4067, Australia.
Email: Bill.Barendse@csiro.au

Abstract. The efficient identification of the genes that influence quantitative traits requires: large sample sizes; the analysis of large numbers of polymorphisms in and around genes or surrogates for these; repeated testing in independent samples; the realisation that the inheritance patterns of quantitative trait loci will show the full range of effects found for genes that affect discrete traits; and choosing the appropriate genetic structure of the sample and the kind of DNA polymorphism for the different stages in the identification of the quantitative trait loci. The choice of trait is critical to the successful production of diagnostic tests. Since this is the most important single factor affecting whether a test will be commercialised, not only due to the economic importance of the trait, but whether there are easy, alternative methods to improve the trait that are cheaper to implement than a DNA test.

Additional keywords: QTL, sheep, pigs, DNA, genomics.

Introduction

The existence of quantitative trait loci (QTL), or Mendelian factors that influence quantitative traits, has been known since the dawn of genetic analysis, but their identification has proved to be more difficult than originally thought. Over the past 80 years techniques to identify them have been developed. Not only has evidence for their pervasive existence been found, but their location to gene maps have become commonplace. A few diagnostic tests for QTL have been reported and some of these have been commercialised. In this overview the steps of efficiently identifying the genetic basis for QTL in livestock are described, including the resources that are needed and some of the pitfalls in the validation of diagnostic tests for QTL. These are discussed under the following 4 sections: challenges to mapping QTL, choosing genes for QTL, resources for QTL and their use, and from QTL to diagnostic test.

Challenges to mapping QTL

Genes affecting quantitative traits are difficult to identify because their effects are small to moderate, many genes affect the trait, and the genomes of birds and mammals (where most of the QTL mapping occurs) are very large, of the order of 3 Gb pairs. Each of these factors makes it difficult to identify the genetic basis of QTL and have consequences that differentiate mapping genes for QTL from other, simpler forms of gene mapping. These consequences lead to specialised uses for standard techniques, and knowing which tool to use at which stage is critical to success in identifying the genetic basis for QTL.

The smaller the mean difference between genotypes, and hence the more their distributions overlap, the more inefficient it is to detect linkage between the trait and a gene map, due to the misclassification of genotypes. Ott (1985) summarises the drop in efficiency, but the rules of thumb are that (i) if the difference between means is ≥3 s.d. then the efficiency is not <50% and (ii) that if the means differ by less than 1 s.d. then the efficiency is never >15%. This means that to have equivalent power to detect a QTL for a trait where the phenotypes of each genotype never overlap, the sample size needs to be doubled for mean differences of 3 s.d. and at least 7 times greater if the mean differences are 1 s.d. Most QTL have small to moderate effects (Morton and Lio 1997), <1.5 s.d., and most would be considered to be part of the polygenic background. Moreover, if the difference between means is <0.5 of a s.d. then the combined distribution will be broad but there will not be a suggestion of bimodality. A segregation analysis will not be able to follow the inheritance of alleles and so assigning genotypes on the basis of the phenotype will be impossible.

The second major factor, the presence of several genes affecting the trait, means that a high trait value will be due to the effects of several genes, not to the effects of the particular gene under investigation. This means, first, that the provisional genotypes cannot be assigned accurately to individuals merely because they have extremely large or small trait values. Second, the combination of genes may have an additive effect on the phenotype or may have an epistatic, dominant or multiplicative effect on the trait. Thus average trait values may not behave in a simple or uniformly predictable fashion, depending upon the interactions that

10.1071/EA05067 0816-1089/05/080831

may occur between genotypes. These interactions occur with major genes, for example the breed background effects for pale soft exudative meat in pigs and double muscling in cattle (e.g. Otsu *et al.* 1991; Grobet *et al.* 1997), where they can be explicitly recognised because the genotypes will show clear differences in phenotype. However, with QTL, because of the small sizes of effect, discrete phenotypic differences will not be found and so these interactions are not explicit, but carefully designed studies can find them (e.g. Wang *et al.* 1999; Luo *et al.* 2001). What may be seen, instead, is the inexplicable failure to replicate a linkage between a QTL and a particular gene or chromosomal segment. Possible explanations for this include different environments, differences in the genetic background, or the QTL was not there in the first place. These explanations highlight the lack of tools to describe the actual processes, and the relative intractability of mapping QTL due to the small sizes of effects.

The large size of avian and mammalian genomes is both a help and a hindrance. In QTL mapping studies, the location of the QTL is usually defined to a region of about 30 cM or about 30 Mb (e.g. Paterson *et al.* 1988; Andersson *et al.* 1994; Georges *et al.* 1995; Davis *et al.* 1998; Stone *et al.* 1999; Beh *et al.* 2002; Hanotte *et al.* 2003). The large size and multiplicity of genes, estimated to be between 20000 and 25000 for humans (International Human Genome Sequencing Consortium 2004), means that in any region of the genome containing a QTL, there will be many plausible alternatives that need systematic exclusion. For such a chunk of a mammalian genome one would expect on average about 2–300 genes including the QTL (International Human Genome Sequencing Consortium 2001, 2004; Mouse Genome Sequencing Consortium 2002). Like any gene mapping, the comprehensive mapping of QTL means exploring the entire genetic material. For any trait, there are usually 0 or 1 major genes segregating at any time, so the entire genome might require screening to find them, and could require scanning a large part of the genome before a linkage is found. For QTL, there are usually many genes affecting any trait. Experiments over several decades indicate that we should expect that there will be a QTL on every chromosome and 5–10 industrially useful QTL or 1 every 2–3 chromosomes in mammals (e.g. Paterson *et al.* 1988; Andersson *et al.* 1994; Georges *et al.* 1995; Davis *et al.* 1998; Stone *et al.* 1999; Beh *et al.* 2002; Hanotte *et al.* 2003; Nagamine *et al.* 2003). Indeed, every region of 20 cM should contain a QTL, if the *Drosophila* experiments are applicable to mammals (e.g. Breese and Mather 1957; Shrimpton and Robertson 1988). Simulations show that the expectations of size and distribution of QTL depend upon the genetic models that are conceptualised, and that the genetic locations of QTL of large effect may be illusory (Franklin and Mayo 1996). Experiments in cattle show (i) that there are usually 2–3 large QTL per experiment (c.f. above and reviews e.g. Dominik 2005; Kuhn *et al.* 2005; Purvis and Franklin 2005), (ii) that the QTL are on different chromosomes from one study to the next, although several studies have confirmed the location of QTL to the same genetic region (c.f. above), and (iii) that there are background and environmental effects on the detection of QTL, or that there are regional or country differences in the detection of QTL (e.g. Paterson *et al.* 1991). Moreover, since the sample sizes are generally of the order of 2–500 animals, there is not the statistical power to detect all the QTL that are segregating in a particular experiment, which would act to exacerbate the lack of consistency in finding QTL from one experiment to another.

Choosing genes for QTL

The 2 ways to choose the genes to test are (i) based upon their known function (biochemistry and physiology) or tissue distribution, also known as the candidate gene approach, and (ii) based upon their location in the genome, or the positional cloning strategy although the most efficient strategy is the combination of these called the positional candidate approach. The use of positional candidates has come to dominate gene mapping since Collins (1995) showed the rising power of this approach and there is a large literature of population associations between positional candidate genes and traits. Obviously, a candidate gene for a trait will not always be seen as such by all groups working in the field, since it is a commonplace that individuals will have different scientific backgrounds. So a balance needs to be struck between the extremes of using only genes that are positional candidates and testing all possible genes in the interval.

Resources for QTL mapping and their use

Two resources are essential for QTL mapping, the first is mapped polymorphisms and the second is DNA from individuals that have been measured for a large number of traits of interest. In the past, due to limitations in the number of accurately mapped polymorphisms or a lack of detailed sequence around a gene, libraries of large insert clones, such as bacterial artificial chromosomes (BAC) were considered essential. These are still useful, but in cattle, for example, with a genome sequence, they are not essential. The use of model organisms is also possible, and is often used in human–mouse studies. However, there are no examples in livestock where a QTL fully described in a mouse led to the identification of a QTL in cattle. Mouse models did lead to the myostatin double muscling mutation (Grobet *et al.* 1997) but that is a major gene affecting a quantitative trait.

Most samples are still limited in size, and although QTL are being discovered in samples with 1–400 individuals, the lack of repeatability in some of them is of concern (e.g. Buchanan *et al.* 2002; Barendse *et al.* 2005). The large sizes (minimum of 1–2000) are required for 2 reasons. First, the effect size is small. To have a threshold of $\alpha = 0.0001$

(Lander and Kruglyak 1995) and a probability of detecting a QTL >99%, requires several thousand individuals, if the differences between means is about 0.1 of a s.d. (Soller *et al.* 1976; Weller 2001). Second, in most QTL experiments a DNA marker or gene map region is tested against many or all of the phenotypes, and a large number of DNA markers are used. This leads to multiple testing, and despite the debate on the best way to address multiple testing (e.g. Perneger 1998), will lead to altered thresholds. Therefore, larger samples are required to demonstrate a statistically significant result. Moreover, it is not sufficient merely to have a single resource for QTL mapping. Separate resources will be required, first to confirm the association that has been demonstrated and second to determine how widely distributed the QTL is among various breeds and populations.

Various shortcuts have been proposed to discover a DNA marker associated with a phenotype of interest, to cut down on the enormous size of a mammalian genome. These vary from random, broadscale anonymous methods to more targeted ones as a genome becomes better known. All of these methods require some form of pooling and selection of extremes, either of DNA samples or DNA markers. For example, it is well known that animals that are phenotypically divergent contribute most to the mapping of a trait, so sampling the top and bottom 10% will reduce the genotyping by 80% (Darvasi 1997). By pooling the extremes, this could lead to genotyping as few as 6 samples plus controls (Collins *et al.* 2000; Mohlke *et al.* 2002). Random generation of DNA markers is simple and running them in parallel easy [e.g. amplified fragment length polymorphism (AFLP)] (Vos *et al.* 1995). Once a DNA marker is found in such systems it is then difficult to identify single gene polymorphisms. A more advanced alternative occurs when the DNA markers are mapped. Then fewer of them need to be chosen since they can be selected to give even coverage of the genetic material (Toda *et al.* 2003). Obviously, selecting the genes on the basis of their functions will reduce the number of polymorphisms that need to be used and this should apply to QTL, although many major physical mutations have resulted from changes in genes not originally thought to be involved in the nuts and bolts of a phenotype (e.g. Riordan *et al.* 1989; Wilson *et al.* 2001). DNA polymorphisms have also been identified after pooling animals of extreme phenotype and then identifying all of the DNA sequences that differ between the pools (e.g. Lisitsyn *et al.* 1994). The general practice so far in mammalian QTL mapping is not to use pooling strategies but to use brute force by attempting to find faster methods to genotype larger samples at lower cost (e.g. Hardenbol *et al.* 2003; genotyping platforms such as Affymetrix, ParAllele, Illumina, Sequenom, ABI SNPlex and Beckman SNPStream).

The detection limit for QTL appears to differ by type of analysis, species, sample structure and evolutionary history of the group being studied. Some sample structures and analytical methods will detect QTL at ≥30 cM (about 30 Mb pairs) while others will not detect QTL unless the distance is <10 kb pairs (Thompson 1997). Multipoint linkage analysis in large pedigrees, using microsatellites, will usually detect associations at a great genetic distance. Resolution can be improved by reducing family size until each family is represented by 1 offspring, analysing smaller groups of markers until only a single is used, moving from microsatellites to haplotypes of single nucleotide polymorphism (SNP) and finally to single SNP. All these measures will reduce, stepwise, the ability to detect distant QTL, or more positively, will increase the resolution of mapping until QTL can be located to particular parts of genes rather than merely to a cluster of genes (Edwards 1997; Thompson 1997; Barendse 2003). Careful choices have to be made to meld together an appropriate population structure to the DNA markers that are available so that associations between DNA markers and traits can be detected more quickly. After which, the QTL region is refined by using different populations and different sets of markers, until genes that affect the QTL or predict the QTL genotypes are identified.

At the highest resolution, not all polymorphisms in a particular gene will show associations to the trait. This has 2 consequences. First, it is possible to come very close to identifying the mutations causing QTL purely by genetic mapping. Second, if DNA markers are chosen at random and used on population structures that enhance high resolution QTL mapping, then a large amount of DNA polymorphisms need to be used. Although this will be less than the 800 000 recommended for humans (Kruglyak 1999; Carlson *et al.* 2003) due to the high level of linkage disequilibrium in cattle and other livestock (Farnir *et al.* 2000; Tenesa *et al.* 2003; McRae *et al.* 2002; Nsengimana *et al.* 2004). Taking a whole genome approach will be necessary to map genes for female fertility where the heritability is low and identifying crosses for interval mapping is essentially impossible. However, at present such large numbers of SNP are not yet feasible for cattle mapping so a staged process of using different sample structures and different types of polymorphisms should be used to progressively home in on the causative gene. Other livestock species lag further behind due to the lack of a genome sequence or resources of tens to hundreds of thousands of SNP.

Transition from QTL marker to diagnostic test

There is not much in the literature on the practicalities of commercialisation of DNA marker technology in livestock (Barendse 2002; Rothschild and Plastow 2002) or a ranking of which factors are more important in the commercialisation of the technology. We have been involved in this for 6 years and now some of what is necessary for successful commercialisation is clear to us, apart from interest, finance and a ready market.

These phases, which are not sequential, are: (i) before the work commences the importance of the trait needs to be thoroughly studied and alternatives other than DNA markers for genetic improvement need to be listed; (ii) once a statistically significant allelic association has been described between a DNA marker and a phenotype, the specific SNP association needs to be confirmed; (iii) the size of effect and the economic value should be calculated; (iv) other SNP in the gene which may have better associations or may point to there being several causative mutations in the gene should be tested; (v) the QTL association should be extended to a range of breeds, environments and management systems to determine the general usefulness of the test; and (vi) the test should be configured into an easy, foolproof high-throughput commercial assay.

The second phase, confirmation of the allelic association of the SNP to the trait in an independent population, is important not just for making diagnostic tests but also as a check that the association is real. In genomic studies, where many loci are tested against many traits, it is inevitable that there will be false positives. By testing positives in a second population as specific hypotheses, far less extreme significance thresholds need to be implemented. Importantly, confirmation of a positive provides additional information because the detail can be used to determine how close to the causative mutation the SNP is, and indicates whether the particular SNP can be used for diagnosis. Specifically, if both studies agree that the relationship between the specific genotypes and differences in the phenotype are the same, then the SNP is a candidate as a diagnostic test. Importantly, it indicates that the SNP is at the very least genetically close to the location of SNP that affect the trait. If there is an association, but the specific genotypes do not agree, then there would be greater distance between the SNP and causative mutations and more SNP discovery in the region needs to be done.

The determination of the economic value, the third phase, is dependent upon the size of the allelic substitution, the frequency of the favourable allele, the range of breeds the favourable allele occurs in, negative associations between genotypes and other traits, the existence of other, non-genetic tools to alter the trait, protection and ownership of the intellectual property, and existence of other tests in the market. Integration of all of these factors is not simple and can affect negotiation on the value of the tests, since not all tests will have a favourable aspect for each of these factors. This is an area in which there is no standardisation of approach, possibly because of the complex nature of the inputs.

The fourth phase is to explore other SNP in the genes, to find better predictors of the trait. If random strategies are used to find the SNP then it will only be by chance that those SNP will end up as the SNP used in the commercialised genetic test. While it may take more time to select one gene over another, or to choose the particular part of the gene to look for SNP, rather than use any sequence available, these represent hours of extra bioinformatic analysis rather than weeks or months of laboratory toil lurching from one random SNP to another. Moreover, a comprehensive search of SNP in a gene is probably unnecessary, first because the vast majority of SNP will only show allelic association but will not be causative, second because haplotypes can be combined to give an efficient analysis and adding more SNP to the haplotype in many cases does not materially improve the amount of information (Mira *et al.* 2004) and third because the cost benefit of continuing to try to find SNP once a good predictor has been found, even if the predictor is not causative, will become prohibitive for the marginal improvement in prediction that is afforded. Nevertheless, in some cases, there may be no alternative to finding the causative mutation(s) in a gene, in those situations where haplotypes in different breeds do not give consistent results or where there is the suspicion that there may be more than one causative mutation in a gene.

Much of the data for the fifth phase can be collected as part of the discovery and confirmation of the association if large samples are used in the process and no selection of phenotypes is made to ease the genotyping burden. However, while this might give a preliminary idea of the size of effect, for commercial purposes, industries are so diverse that the inevitable question from producers is 'Does it work in my breed of cattle under my management conditions where I raise my cattle?'. This implies not only a larger variety of samples and environments to be tested, but it also implies clever choices of what samples to use so that generalities for a wide range of producers can be made without exhaustive sample collection.

The final phase, to develop a commercially offered robust high-throughput assay is not always straightforward. Due to constraints of DNA sequence and the way that different kinds of assays work, not all SNP can be assayed using a particular technology. The manufacturers of different integrated genotyping methods (platforms) claim in their bulletins anything from 75 to 95% of SNP can be assayed on their platforms. So several platforms may need to be used since the commercialiser may have radically different platforms available compared with the group doing the research. Clearly, differences in platform do not mean differences in called genotypes. There must be some program where animals of known genotype are exchanged and then genotyped in tandem between the researchers and the commercialisers to ensure that the tests are properly performed.

The relative importance of these phases is becoming clear. Of most importance is the trait itself, whether improvement will provide an economic return to the producer or improve competitiveness of the product compared with alternatives. Although this seems obvious almost all QTL studies published to date include reports of QTL for traits which are

trivially easy to measure, such as growth rate or milk yield for which only a set of scales are necessary; moreover, there appears to be no interest in a DNA marker that will merely improve growth rate. Of the other phases, the next most important is whether the QTL is confirmed, and then whether there is exclusivity of access. The other phases are less important than these. Without confirmation there will be little investment in a marker that might just be a statistical artifact, and the costs of patenting and licensing are large. Serious consideration should be given to whether a particular DNA test warrants patent protection, given that a panel of DNA markers all of which were patented would impose a significant overhead on a DNA testing service. However, a DNA testing service that offered no protected tests would be at the mercy of those that did, since they would have more tests to offer. Of the other factors, it is more important that the allele frequency is favourable than identifying the molecular basis for the QTL. Although the causative mutations will have greater specificity than SNP in allelic association to the causative mutation, if the allele frequency of the favourable allele is too high or too low then testing will not be cost effective. Of lesser importance than these is whether the test is formatted for a high throughput platform, since the initial volume for testing will be low, well within the capacity of restriction fragment length polymorphisms, and alternative platforms can always be used if there is any difficulty with the DNA sequence.

In North America, Europe, Asia, Oceania and Africa much research has been aimed at discovering more QTL and associated tests for industry. This research is being conducted on cattle, pigs, chicken, sheep, horses, goats and other livestock. At present the methods to isolate QTL have matured as have the resources of population samples and DNA polymorphisms. The reagent costs have dropped by 2 orders of magnitude in the last 5 years while throughput of genotyping has increased by 5 orders of magnitude during the same period. Producers and other consumers of genetic tests have become more receptive to their use and the pipeline of getting genetic tests done have become mature. The major challenge, apart from the necessary research to identify the genes affecting QTL and derive diagnostic tests, is to integrate the conventional information on traits with the new genetic information for predicting carcass merit. At present they merely coexist side by side.

Acknowledgments

R. Hawken and I. Purvis provided useful comments and discussion that improved the manuscript. The anonymous reviewers provided some examples that I had inadvertently overlooked, and thereby improved the text.

References

Andersson L, Haley CS, Ellegren H, Knott SA, Johansson M, *et al.* (1994) Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* **263**, 1771–1774.

Barendse W (2002) Development and commercialisation of a genetic marker for marbling of beef in cattle: a case study. In 'Intellectual property rights in animal breeding and genetics'. (Eds MF Rothschild, S Newman) pp. 197–212. (CABI Publishing: Oxford)

Barendse W (2003) DNA markers for marbling. Patent publication WO2004070055.

Barendse W, Bunch RJ, Harrison BE (2005) The leptin C73T missense mutation is not associated with marbling and fatness traits in a large gene mapping experiment in Australian cattle. *Animal Genetics* **36**, 86–88. doi:10.1111/j.1365-2052.2004.01224.x

Beh KJ, Hulme DJ, Callaghan MJ, Leish Z, Lenane I, Windon RG, Maddox JF (2002) A genome scan for quantitative trait loci affecting resistance to *Trichostrongylus colubriformis* in sheep. *Animal Genetics* **33**, 97–106. doi:10.1046/j.1365-2052.2002.00829.x

Breese EL, Mather K (1957) The organisation of polygenic activity within a chromosome in Drosophila I. Hair characters. *Heredity* **11**, 373–395.

Buchanan FC, Fitzsimmons CJ, van Kessel AG, Thue TD, Winkelman-Sim DC, Schmutz SM (2002) Association of a missense mutation in the bovine leptin gene with carcass fat content and leptin mRNA levels. *Genetics, Selection, Evolution.* **34**, 105–116. doi:10.1051/gse:2001006

Carlson CS, Eberle MA, Rieder MJ, Smith JD, Kruglak L, Nickerson DA (2003) Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nature Genetics* **33**, 518–521. doi:10.1038/ng1128

Collins FS (1995) Positional cloning moves from perditional to traditional. *Nature Genetics* **9**, 347–350. doi:10.1038/ng0495-347

Collins HE, Li H, Inda SE, Anderson J, Laiho K, Tuomilehto J, Seldin M (2000) A simple and accurate method for determination of microsatellite total allele content differences between DNA pools. *Human Genetics* **106**, 218–226. doi:10.1007/s004390051031

Darvasi A (1997) The effect of selective genotyping on QTL mapping accuracy. *Mammalian Genome* **8**, 67–68. doi:10.1007/s003359900353

Davis GP, Hetzel DJS, Corbet NJ, Scacheri S, Lowden S, Renaud J, Mayne C, Stevenson R, Moore SS, Byrne K (1998) The mapping of quantitative trait loci for birth weight in a tropical beef herd. In 'Proceedings of the 6th world congress of genetics applied to livestock production. Vol. 26'. pp. 441–444. (WCGALP: Armidale, NSW)

Dominik S (2005) Quantitative trait loci for internal nematode resistance in sheep: a review. *Genetics, Selection, Evolution.* **37**, S83–S96. doi:10.1051/gse:2004027Suppl. 1

Edwards JH (1997) Recessive disease and allelic association. In 'Genetic mapping of disease genes'. (Eds I-H Pawlowitzki, JH Edwards, EA Thompson) pp. 31–57. (Academic Press: San Diego)

Farnir F, Coppieters W, Arranz J-J, Berzi P, Cambisano N, *et al.* (2000) Extensive genome-wide linkage disequilibrium in cattle. *Genome Research* **10**, 220–227. doi:10.1101/gr.10.2.220

Franklin IR, Mayo O (1996) The place of QTL in the basis of quantitative genetics. II. Mapping QTL from a wide cross. In 'Proceedings of the 6th world congress of genetics applied to livestock production. Vol. 26'. pp. 273–276. (WCGALP: Armidale, NSW)

Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, *et al.* (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* **139**, 907–920.

Grobet L, Martin LJR, Poncelet D, Pirottin D, Browers B, *et al.* (1997) A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nature Genetics* **17**, 71–74. doi:10.1038/ng0997-71

Hardenbol P, Baner J, Jain M, Nilsson M, Namsaraev EA, *et al.* (2003) Multiplex genotyping with sequence-tagged molecular inversion probes. *Nature Biotechnology* **21**, 673–678. doi:10.1038/nbt821

Hanotte O, Ronin Y, Agaba M, Nilsson P, Gelhaus A, *et al.* (2003) Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 7443–7448. doi:10.1073/pnas.1232392100

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921. doi:10.1038/35057062

International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945. doi:10.1038/nature03001

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics* **22**, 139–144. doi:10.1038/9642

Kuhn Ch, Leveziel H, Renand G, Goldammer T, Schwerin M, Williams J (2005) Genetic markers for beef quality. In 'Indicators of milk and beef quality'. EAAP publication No. 112. (Eds JF Hocquette, S Gigli) pp. 23–32. (Wageningen Academic Publishers: Wageningen, The Netherlands)

Lander ES, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**, 241–247. doi:10.1038/ng1195-241

Lisitsyn NA, Segre JA, Kusumi K, Lisitsyn NM, Nadeau JH, Frankel WN, Wigler MH, Lander ES (1994) Direct isolation of polymorphic markers linked to a trait by genetically directed representational difference analysis. *Nature Genetics* **6**, 57–63. doi:10.1038/ng0194-57

Luo LJ, Li Z-K, Mei HW, Shu QY, Tabien R, Zhong DB, Ying CS, Stansel JW, Khush GS, Paterson AH (2001) Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. *Genetics* **158**, 1755–1771.

McRae AF, McEwan JC, Dodds KG, Wilson T, Crawford AM, Slate J (2002) Linkage disequilibrium in domestic sheep. *Genetics* **160**, 1113–1122.

Mira MT, Alcais A, Thuc NV, Moraes MO, Di Flumeri C, *et al.* (2004) Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature* **427**, 636–640. doi:10.1038/nature02326

Mohlke KL, Erdos MR, Scott LJ, Fingerlin TE, Jackson AU, Silander K, Hollstein P, Boehnke M, Collins FS (2002) High-throughput screening for evidence of association by using mass spectrometry genotyping on DNA pools. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16928–16933. doi:10.1073/pnas.262661399

Morton NE, Lio P (1997) Oligogenic linkage and map integration. In 'Genetic mapping of disease genes'. (Eds I-H Pawlowitski, JH Edwards, EA Thompson) pp. 17–21. (Academic Press: San Diego)

Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562. doi:10.1038/nature01262

Nagamine Y, Haley CS, Sewalem A, Visscher PM (2003) Quantitative trait loci variation for growth and obesity between and within line of pigs (*Sus scrofa*). *Genetics* **164**, 629–635.

Nsengimana J, Baret P, Haley CS, Visscher PM (2004) Linkage disequilibrium in the domesticated pig. *Genetics* **166**, 1395–1404. doi:10.1534/genetics.166.3.1395

Otsu K, Khanna VK, Archibald AL, Maclennan DH (1991) Cosegregation of porcine malignant hyperthermia and a probable causal mutation in the skeletal-muscle ryanodine receptor gene in backcross families. *Genomics* **11**, 744–750. doi:10.1016/0888-7543(91)90083-Q

Ott J (1985) 'Analysis of human genetic linkage.' (The Johns Hopkins University Press: Baltimore)

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato – comparison across species generations and environments. *Genetics* **127**, 181–197.

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete RFLP linkage map. *Nature* **335**, 721–726. doi:10.1038/335721a0

Perneger TV (1998) What's wrong with Bonferroni adjustments. *British Medical Journal* **316**, 1236–1238.

Purvis IW, Franklin IR (2005) Major genes and QTL influencing wool production and quality: a review. *Genetics, Selection, Evolution* **37**(Suppl. 1), S97–107. doi:10.1051/gse:2004028

Riordan JR, Rommens JM, Kerem BS, Alon N, Rozmahel R, *et al.* (1989) Identification of the cystic-fibrosis gene-cloning and characterization of complementary-DNA. *Science* **245**, 1066–1072.

Rothschild MF, Plastow GS (2002) Development of a genetic marker for litter size in the pig: a case study. In 'Intellectual property rights in animal breeding and genetics'. (Eds MF Rothschild, S Newman) pp. 179–196. (CABI Publishing: Oxford)

Shrimpton AE, Robertson A (1988) The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. I. Allocation of third chromosome sternopleural bristle effects to chromosome sections. *Genetics* **118**, 437–443.

Soller M, Brody T, Genizi A (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theoretical and Applied Genetics* **47**, 35–39. doi:10.1007/BF00277402

Stone RT, Keele JW, Shackelford SD, Kappes SM, Koohmaraie M (1999) A primary screen of the bovine genome for quantitative trait loci affecting carcass and growth traits *Journal of Animal Science* **77**, 1379–1384.

Tenesa A, Knott SA, Ward D, Smith D, Williams JL, Visscher PM (2003) Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. *Journal of Animal Science* **81**, 617–623.

Thompson EA (1997) Conditional gene identity in affected individuals. In 'Genetic mapping of disease genes'. (Eds I-H Pawlowitski, JH Edwards, EA Thompson) pp. 137–146. (Academic Press: San Diego)

Toda T, Momose Y, Murata M, Tamiya G, Yamamoto M, Hattori N, Inoko H (2003) Toward identification of susceptibility genes for sporadic Parkinson's disease. *Journal of Neurology* **250**, iii40–iii43. doi:10.1007/s00415-003-1307-6

Vos P, Hogers R, Bleeker M, Reijans M, Lee T, *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **21**, 4407–4414.

Wang DL, Zhu J, Li ZK, Paterson AH (1999) Mapping QTLs with epistatic effects and QTLxenvironment interactions by mixed linear model approaches. *Theoretical and Applied Genetics* **99**, 1255–1264. doi:10.1007/s001220051331

Weller JI (2001) 'Quantitative trait loci analysis in animals.' (CABI Publishing: New York)

Wilson T, Wu X-Y, Juengel JL, Ross IK, Lumsden JM, *et al.* (2001) Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of Bone Morphogenetic Protein IB Receptor (ALK-6) that is expressed in both oocytes and granulosa cells. *Biology of Reproduction* **64**, 1225–1235.

Received 1 March 2005, accepted 3 June 2005