DESIGNS FOR QTL DETECTION IN LIVESTOCK AND THEIR IMPLICATIONS FOR MAS

D.J de Koning, J.C.M. Dekkers & C.S. Haley
Roslin Institute, Roslin, EH25 9PS, United Kingdom, DJ.deKoning@BBSRC.ac.uk
Iowa State University, Ames, IA, 50011, USA, jdekkers@iastate.edu
Roslin Institute, Roslin, EH25 9PS, United Kingdom, Chris.Haley@BBSRC.ac.uk

Summary
Designs used for detecting QTL in livestock vary from experimental backcross and F_2 populations to half-sib designs that use existing family structures within a commercial population. This contribution discusses how the initial design in which the QTL are detected affects: a) the implementation of MAS and b) the fine mapping of the QTL. The optimal approach depends on the livestock species and the breeding strategies.

Designs for QTL detection in livestock

In comparison to plant species and laboratory animals, genome mapping in livestock faces the following challenges: 1) Inbred lines are not commonly available. 2) Maintenance of experimental populations can be prohibitively expensive. 3) Reproductive capacity and generation interval are often limiting in the choice of experimental design. These factors have to be taken into account in both the design and analysis of QTL experiments. Using a sparse marker map (10-20 cM marker spacing), several designs have been used to detect QTL regions across the genome in livestock.

Experimental crosses. Experimental crosses have been implemented in pigs and poultry because generation intervals are relatively short and the number of offspring is moderate to high. Such crosses have been established between domestic breeds and descendants of their wild progenitors (i.e. Wild boar and commercial Large White pigs or red jungle fowl and White Leghorn chicken) as well as between phenotypically divergent commercial breeds (e.g. egg laying and meat-type chicken). The analysis model is largely the same as that used in the analysis of inbred crosses: marker alleles in the second generation are traced back to their line origin and contrasts for putative QTL are estimated as differences between lines. Because crosses are between outbred lines, the analysis needs to accommodate the fact that founder lines may share alleles at the marker level. Because lines may not be fixed for alternate QTL alleles, these analyses estimate contrasts between average effects of QTL alleles derived from the parental breeds, which only represent estimates of the actual effects if alternative QTL alleles were fixed in the founder lines. This model is straightforward to implement using least squares methodology and has been extended to accommodate more complex genetic models like imprinting, sex-linked and/or sex-specific QTL, and more recently, epistatic QTL interactions. Note that imprinting can be analysed because segregation at the marker loci in the founder lines allows distinction of parental origin of alleles in the heterozygous F_2 animals (an advantage not possible when analysing crosses between inbred lines of model species). Most of these models have been implemented in a user-friendly free analysis package accessible via the web at http://qtl.cap.ed.ac.uk/. Recently, extensions to the line cross model have been proposed where the QTL is modelled to explain variation within lines as well as between lines. Although this provides a more realistic and potentially more powerful approach, computational demands are much higher than for the traditional line-cross approach and, to date, its applications to real data are limited.

Exploiting existing family structures. For sheep and cattle, it is often considered too expensive and time consuming to develop an experimental cross, although some experimental backcross and F_2 populations have been developed. This is only realistically feasible when the cross itself can be developed into an economically viable synthetic or when the phenotype(s) under study has very large economic or welfare relevance. A more common approach is to exploit the large paternal half-sib family structures that occur in species where artificial insemination (AI) and field data recording is widespread, such as in dairy cattle. In these half-sib designs, genotypes are collected on a number of grandsires and their half-sib offspring. Phenotypes are either collected on the half-sib offspring themselves or on a group of progeny from each half-sib. In dairy cattle, where progeny tested bulls have >100 daughters with phenotype records, the 3-generation half-sib design (a granddaughter design) has been successfully implemented in a number of countries. Analyses using least squares generally make no assumptions about number of QTL alleles and estimate a unique QTL effect within every half-sib family. Because half-sib family structures also exist within experimental crosses, these models are sometimes also fitted on line-cross populations to gain further insight into identified QTL by checking which F_1 parents are most likely to be heterozygous for a QTL and allow for different QTL genotypes of the F_1 parents. The same methodology can be extended to accommodate full-sib family structures, provided family sizes are sufficient, but to date this approach has only been used in poultry QTL mapping.

Some analytical methods that were initially developed for analysis of human pedigrees have now also been implemented in the analyses of livestock. These include sib-pair analyses and TDT tests for both qualitative and quantitative traits. A very promising area of QTL analyses is the variance component (VC) model for general pedigrees. In a VC model, the QTL is modelled as a random effect alongside a polygenic component. The VC model accounts for all additive genetic relationships at the QTL by modelling the identical-by-descent (IBD) probabilities between QTL alleles of individuals, which are derived based on observed inheritance of marker alleles. These methods need further development before widespread automated implementation but they have several advantages: 1) they are equally useful for the detection of QTL as for the evaluation of QTL in MAS. 2) They provide genotypic breeding values for every individual in the pedigree. 3) They are applicable across a wide range of species, including humans.

Where a single trait or index is the main focus of interest, a potentially efficient use of genotyping resources can be achieved by selective genotyping, where only the individuals showing a more extreme phenotype within a family or a cross are genotyped. Data collected in this way can be analysed by any of the approaches discussed above, although often it may lead to biased (inflate) estimates of QTL effects (but not to an inflated rate of detection of spurious QTL if robust analytical procedures are used). Selective DNA pooling, where marker allele frequencies in "high" pools are contrasted with those in the "low" pools within families, can reduce the number of genotypings further, but requires more careful design, analysis and consideration of technical aspects if it is not to lead to detection of spurious QTL.

Different types of MAS in livestock

Application of MAS for genetic improvement relies on the level of precision at which a QTL has been identified. For these purposes, we define three levels of refinement:

1) Functional mutations: loci for which the functional polymorphism can be genotyped.

2) LD markers: loci that are in population-wide linkage disequilibrium with the functional mutation.

3) LE markers: loci that are in population-wide linkage equilibrium with the functional mutation.

Gene Assisted Selection (GAS). If the functional mutation underlying a QTL is known, the implementation of this QTL in MAS is straightforward. The effect of the mutation on all breeding goal traits can be quantified and the functional mutation can be included as a systematic effect in the breeding value estimation model. Only the selection candidates themselves need to be genotyped, although genotype probabilities need to be derived for animals that are not genotyped. For applications in other breeds, the effect of a functional mutation will need to be verified, as its effect may be influenced by the background genotype, as has recently been demonstrated for the double muscling locus in beef cattle.

MAS using Linkage Disequilibrium markers (LD-MAS). If a given marker allele or multiple marker haplotype is in population-wide disequilibrium with the functional mutation at a QTL, the implementation of MAS for the given population is nearly as straightforward as in GAS. Genotyping requirements are the same as for GAS, although more markers must be used when working with an LD haplotype. For the longer term, it is important to monitor whether the association between the LD markers and the functional mutation is the same or whether it is weakening because of recombinations between the markers and the QTL. For applications in other breeds or populations, the effect of the haplotype has to be confirmed, which may be possible using an association study of the original haplotype but may require a QTL study for the candidate regions just to confirm that a QTL is segregating. Different effects may be found for the LD markers as a result of different phase with the QTL allele and/or genetic background. A good example is a non-coding polymorphism in the ESR gene in pigs, which has shown increased litter size in several breeds but opposite or no effects in other studies.

MAS using Linkage Equilibrium markers (LE-MAS). For LE-MAS, a detected QTL can be directly implemented in a breeding scheme for the population where it was detected. Because the linkage phase between marker and QTL alleles varies between families, the QTL effect has to be re-assessed continuously and within each family in the MAS process. This requires genotyping for several markers of not only the selection candidates, but also many of their relatives to allow evaluation of the QTL effect. With genotyping becoming cheaper and more advanced statistical methods being developed, LE-MAS could eventually evolve into a system where breeding values are predicted based on genome-wide marker haplotypes equivalent to genome-wide LD-MAS.

Efficiency of MAS. The effectiveness of MAS compared to traditional selection based on estimated breeding values (EBV) depends on the type of traits and the way QTL information is complementing EBV estimation. With few actual livestock MAS results available to date, prediction from simulations vary from 60% additional gain for slaughter traits to an actual loss in response to selection when fixation of a major gene is given too much priority. For short-term gain, GAS can be substantially more efficient than LE-MAS, although LE-MAS could be improved by including prior information on the magnitude of each QTL effect. LD-MAS is expected to obtain only slightly lower responses than GAS.

From QTL detection to GAS/LD-MAS: fine-mapping approaches.
QTL studies in livestock generally detect QTL with confidence intervals >30 cM because of the extensive LD that is created within crosses and exists within two- or three-generation pedigrees. Such regions may contain thousands of genes and do not facilitate LD-MAS, let alone GAS. Strategies to fine-map these QTL are based on the creation or identification of additional recombinations or on exploiting historical recombinations that erode LD, except at short distances. The *de novo* creation of additional recombinations requires a species with reasonable generation intervals while exploiting historical recombinations may require information about the breed history.

Creating additional recombinations: experimental crosses. For experimental crosses, two strategies can be used to increase the number of recombinations around the QTL region.

1) An advanced intercross line (AIL) where the animals from an F_2 are crossed to create an F_3 and so on, until F_n. Every two additional generations of intercrossing is expected to half the confidence interval of QTL position in an AIL compared to an F_2 of the same size. The design is especially useful because, once the population has been created, it can be measured for a number of traits and QTL in different regions of the genome can be mapped to high resolution. The number of generations and sample size required restricts its general use in livestock, although it is being applied in poultry.

2) Repeated backcrossing to one or both founder lines. Animals that are inferred to be segregating for the QTL are crossed back to one of the founder lines. In the offspring, animals with alternative marker haplotypes for the QTL region are selected as parents of the next generation. Phenotype contrasts of marker haplotypes in the offspring show which parents segregate for the QTL and in which area the QTL is most likely to reside. Within segregating parents, offspring with alternative marker haplotypes are then selected to further refine the region. The fact that a relatively limited number of sires and their progeny are needed each generation makes the method practicable in pigs. The number of AIL and BC generations and animals should be chosen to optimise the size of the confidence interval with the risk of 'losing' the power to detect the QTL due to inbreeding and drift, while maintaining a manageable animal and genotyping load.

Exploiting historical recombination: Family designs and general pedigrees. Half-sib families of the same breed can be expected to originate from a limited number of common ancestors. It can be postulated that sires that are heterozygous for a QTL carry the original 'wild-type' allele and a 'mutant' allele that originates from a common founder. Because of population-wide linkage disequilibrium around this mutant allele, heterozygous sires are more likely to share marker alleles around this mutation than expected by chance. Following QTL analysis, the heterozygous sires can be identified and their marker genotypes can be arranged in a high and a low haplotype. With sufficient marker density, the heterozygous sires will show a conserved haplotype around the QTL region, which can be much smaller than the original confidence interval of the QTL. The expected size of the shared haplotype depends (among others) on the number of generations since the founder mutation and effective population size. Further refinement can be made when the same QTL has been detected in other populations or even breeds than the present sample. Methodology that also exploits information coming from the homozygous sires adds further power to the method. This approach has proven successful to move from QTL to functional mutation in both dairy cattle and pigs. A more formal approach based on the same principles has been proposed for general pedigrees using a VC model. The principle is the same as the VC method for QTL detection but the IBD matrix is aggregated by additional IBD probabilities due to population-wide LD that account for genetic relationships with regard to a founder mutation. Because pedigree information is not available to obtain the latter IBD probabilities, they are derived using a coalescent simulation approach, which requires some prior assumptions about population equilibrium, effective population size, and number of generations since mutation.

Selective recombinant genotyping. For this strategy, the high number of recombinations that is needed for fine mapping is acquired by using a very large experimental population. Initially, individuals from the phenotypic tails of the population are typed for markers flanking the confidence interval of a QTL. Subsequently, the animals that are recombinants within the marker interval are typed for additional markers, using genotyping strategies that minimise the number of genotypings required to get full genotypes for the region containing the QTL. Although applicable to F_2, BC, and half-sib designs, the required population size is

at least > 5,000 and more often > 10,000 animals. This restricts the feasibility of this approach in livestock to species where half-sib families of several thousands are not uncommon, such as dairy cattle.

Positional candidate genes. If the QTL maps close to a gene that is expected to play a part in the biology of the trait, this gene is a positional candidate. The population can then be screened for polymorphisms in the gene sequence. The putative effects of these polymorphisms can be tested in an association test or a transmission disequilibrium (TDT) test. Although it requires less genotyping efforts than fine-mapping methods, and several successes have been published, this approach has some drawbacks. 1) The confidence interval of > 30 cM for a QTL contains thousands of genes and several of these could be reasoned to affect the trait. Therefore some degree of fine mapping is desired before attempting to identify positional candidates. 2) Considerable linkage disequilibrium exists per se in an F2 or BC, and, often as a result of drift, selection and/or limited effective population size, in commercial populations. As a result, any positive association that may be found for polymorphisms in a candidate gene may exist for any polymorphism in quite a broad region. This is also observed when the fine-mapping approach has revealed a conserved haplotype in which there is no further variation between carriers of the mutant allele. A significant association for a candidate gene can be used as an LD marker but proving causality might need biological reasoning, expression or other biological studies or ultimately a knockout study in a model species.

From QTL detection to LE-MAS
Given the time and cost required to carry out fine-mapping research, it may be considered to implement LE-MAS immediately after QTL detection. The potential gain of LE-MAS may be smaller than that of LD-MAS or GAS, but if it can be implemented straightaway it can give breeders a genetic and commercial (marketing) advantage. Here we discuss the potential for LE-MAS for the different experimental designs.

MAS within experimental lines. It can be expected that for some QTL that were detected the beneficial allele originates from one line, while for other QTL the beneficial allele was contributed by the second founder line. In this case it may be useful to develop the experimental cross into a commercial synthetic line where markers are used to trace the beneficial alleles. Because of the extensive LD in the cross, this does not require markers that are tightly linked to the QTL. A very specific example of this is marker-assisted introgression (MAI) where one or a few of genes from a donor line are of particular interest to a commercial recipient line. This usually involves repeated backcrossing of the experimental cross to the commercial line where markers are used to select <u>for </u>donor alleles in the QTL region and <u>against</u> donor alleles in other regions. Note that this process offers excellent opportunities for fine mapping of the QTL during the introgression process (see earlier paragraph).

From experimental crosses to commercial populations. The experimental cross may have no potential to be developed into a synthetic because the cross was too divergent or the QTL are expected to segregate within the commercial line as well. The QTL that have been detected explain variation between the founder lines but their segregation within each of the founder lines is unknown. It has been argued that selection should have fixed QTL with big effects between divergent lines a long time ago within the lines. Surprisingly, recent studies in pigs and poultry indicate that regions where QTL have been described in experimental crosses also explain a large amount of variation within commercial lines. In poultry, where meat-type chickens (broilers) have been selected for efficient growth > 50 generations, the same region that explained growth differences between broilers and layers, caused variation in growth

within a broiler line. Therefore, experimental crosses may provide a fast track to MAS in commercial lines. Instead of carrying out expensive and time consuming whole-genome scans within a line, a number of candidate regions identified in experimental crosses can be selected and evaluated within the commercial lines. With a large number of QTL already in the public domain, companies can select regions that are consistent across studies for initial evaluation.

LE-MAS in commercial populations. As pointed out above, implementation of LE-MAS requires large amounts of genotyping and advanced analyses to keep track of the QTL. However, in species such as dairy cattle, data collection is very extensive and DNA samples of selection candidates are taken routinely to verify parentage. Likewise, dairy cattle breeding companies are very experienced in running complex variance component models on extremely large pedigrees. This explains why a number of dairy cattle breeding organisations are the first to have implemented LE-MAS. Further reductions in genotyping costs and development of more advanced statistical techniques might increase the uptake of LE-MAS in livestock species.

Other issues regarding implementation of MAS in livestock

Pleiotropic and non-additive effects of QTL. One important question is why QTL with sizeable effects are still segregating in commercial populations that have been under selection for many decades? A possible explanation is that alleles with unfavourable effects on production traits may have beneficial effect on fitness traits and hence remain at a reasonable frequency in the population. This means that it will be important to evaluate the effects of QTL on fitness, otherwise MAS may result in reduced fitness

Although methods to detect QTL in experimental crosses can accommodate a variety of genetical models, the models used for half-sib families and general pedigrees are usually based on additive QTL. Some QTL may appear additive in the background in which they are detected but give unexpected results when selected in another background as a result of epistasis. It is therefore important to assess any pleiotropic (fitness) and non-additive effects of a QTL before implementing MAS. Again this task is much easier when the functional mutation is known and biological knowledge can be used to infer the consequences of fixing the favourable allele.

Protection of discoveries and investments. One way of protecting your investment is to patent the marker test you are using for MAS. A screen for a functional mutation used in GAS is much more straightforward to patent than a number of anonymous markers that track a QTL region in LE-MAS. Obtaining a patent is another matter that requires a lot of time and money, taking into account that several countries do not acknowledge patents. However, patents can be useful as 'intellectual pocket-money' that can be used to trade intellectual property with other companies ("you can use my test if I can use yours") or as a 'nuisance value' to slow down patents other groups are trying to endorse.

Another issue is how to protect your investment when MAS is implemented and you have animals that are superior to the ones that have been selected traditionally. In the dairy cattle industry, semen of top bulls is available to anybody, including your competitors. In that case, any profit of MAS should be in the short term and/or in savings made on progeny testing schemes. In poultry and pig breeding, breeding companies maintain the selection lines while the 'product' comprises a hybrid animal, or semen/eggs from it, which is unlikely to be useful in the breeding program of a competitor. Therefore any investments using MAS in the selection lines can be protected against competitors to a greater degree than in dairy cattle.

Concluding remarks

It has now been well demonstrated that QTL with considerable effects can be identified in livestock and that with investment, skill and some good fortune, technology can be used to achieve the holy grail of QTL mapping: detection of the functional mutation. Therefore QTL mapping should move on from universities and research centres to applications in the breeding industry. Design of studies should not only focus on power to detect QTL, but also on how this will facilitate implementation of MAS, either via fine-mapping and subsequent LD-MAS or GAS, or via direct application of LE-MAS. Because GAS offers the easiest implementation of MAS and the best characterisation of all QTL effects, research into fine mapping strategies and methodology should stay a main research focus. Analytical tools for general pedigrees must be further developed because of their applications in QTL detection, fine mapping, and MAS. Although the industry is now convinced of the existence of QTL, their utility for breeding programs is yet to be proven.

References
This overview was based on a very large number of publications by many of our colleagues. The following references provide a good starting point for further reading.

Andersson, L. 2001. Genetic dissection of phenotypic diversity in farm animals. *Nature Reviews Genetics* **2:** 130-138. *An overview of mapping designs and studies in livestock*

Dekkers, J.C.M, and F. Hospital, 2002. The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics* **3:** 22-32. *An overview of applications of MAS in plants and livestock.*

Lynch, M. and B. Walsh, 1998. Genetics and analysis of quantitative traits. Sinauer Associates, MA, USA. *All you could wish to know about quantitative traits and their analysis.*