SECTION III

Marker-assisted selection in livestock – case studies

CHAPTER 10

Strategies, limitations and opportunities for marker-assisted selection in livestock

Jack C.M. Dekkers and Julius H.J. van der Werf

SUMMARY

This chapter reviews the principles, opportunities and limitations for detection of quantitative trait loci (QTL) in livestock and for their use in genetic improvement programmes. Alternate strategies for QTL detection are discussed, as are methods for inclusion of marker and QTL information in genetic evaluation. Practical issues regarding implementation of marker-assisted selection (MAS) for selection in breed crosses and for selection within breeds are described, along with likely routes towards achieving that goal. Opportunities and challenges are also discussed for the use of molecular information for genetic improvement of livestock in developing countries.

INTRODUCTION

Since the 1970s, the discovery of technology that enables identification and genotyping of large numbers of genetic markers, and research that demonstrated how this technology could be used to identify genomic regions that control variation in quantitative traits and how the resulting QTL could be used to enhance selection, have raised high expectations for the application of gene- (GAS) or marker-assisted selection (MAS) in livestock. Yet, to date, the application of GAS or MAS in livestock has been limited (see e.g. review by Dekkers, 2004 and the case study chapters that follow). However, recent further advances in technology, combined with a substantial reduction in the cost of genotyping, have stimulated renewed interest in the large-scale application of MAS in livestock.

Successful application of MAS in breeding programmes requires advances in the following five areas:

- *Gene mapping*: identification and mapping of genes and genetic polymorphisms.
- *Marker genotyping*: genotyping of large numbers of individuals for large numbers of markers at a reasonable cost for both QTL detection and routine application for MAS.
- *QTL detection*: detection and estimation of associations of identified genes and genetic markers with economic traits.
- *Genetic evaluation*: integration of phenotypic and genotypic data in statistical methods to estimate breeding values of individuals in a breeding population.
- *MAS*: development of breeding strategies and programmes for the use of molecular genetic information in selection and mating programmes.

This chapter outlines the main strategies for the application of MAS in livestock and identifies and discusses the limitations and opportunities for successful MAS in commercial breeding programmes. It concludes by discussing limitations and opportunities for applying MAS in developing countries.

MARKERS AND LINKAGE DISEQUILIBRIUM

Over the past decades, a substantial number of alternate types of genetic markers have become available to study the genetic architecture of traits and for their use in MAS, including restriction fragment length polymorphisms (RFLPs), microsatellites, amplified fragment length polymorphisms (AFLPs) and single nucleotide polymorphisms (SNPs). Detailed information on these markers can be found elsewhere in this publication. Although alternate marker types have their own advantages and disadvantages, depending on their abundance in the genome, degree of polymorphism, and ease and cost of genotyping, what is crucial for their use for both QTL detection and MAS is the extent of linkage disequilibrium (LD) that they have in the population with loci that contribute to genetic variation for the trait. Linkage disequilibrium relates to dependence of alleles at different loci and is central to both QTL detection and MAS. Thus, a thorough understanding of LD and of the factors that affect the presence and extent of LD in populations is essential for a discussion of both QTL detection and MAS.

Linkage disequilibrium

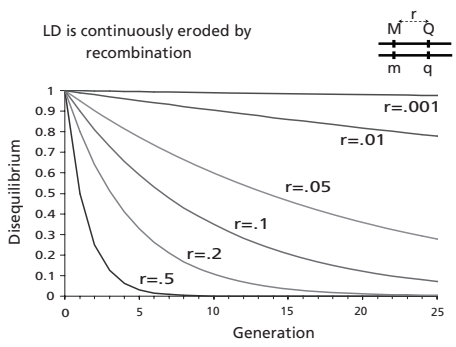
Consider a marker locus with alleles M and m and a QTL with alleles Q and q that is on the same chromosome as the marker, i.e. the marker and the QTL are linked. An individual that is heterozygous for both loci would have genotype $MmQq$. Alleles at the two loci are arranged in *haplotypes* on the two chromosomes of a homologous

pair that each individual carries. An individual with genotype *MmQq* could have the following two haplotypes: *MQ*/*mq*, where the / separates the two homologous chromosomes. Alternatively, it could carry the haplotypes *Mq*/*mQ*. This alternative arrangement of linked alleles on homologous chromosomes is referred to as the marker-QTL *linkage phase*. The arrangement of alleles in haplotypes is important because progeny inherit one of the two haplotypes that a parent carries, barring recombination.

The presence of linkage equilibrium (LE) or disequilibrium relates to the relative frequencies of alternative haplotypes in the population. In a population that is in linkage *equilibrium*, alleles at two loci are randomly assorted into haplotypes. In other words, chromosomes or haplotypes that carry marker allele *M* are no more likely to carry QTL allele *Q* than chromosomes that carry marker allele *m*. In technical terms, the frequency of the *MQ* haplotypes is equal to the product of the population allele frequency of *M* and the frequency of *Q*. Thus, if a marker and QTL are in linkage *equilibrium*, there is no value in knowing an individual's marker genotype because it provides no information on QTL genotype. If the marker and QTL are in linkage *dis*equilibrium, however, there will be a difference in the probability of carrying *Q* between chromosomes that carry *M* and *m* marker alleles and, therefore, a difference in mean phenotype between marker genotypes would also be expected.

The main factors that create LD in a population are mutation, selection, drift (inbreeding), and migration or crossing. See Goddard and Meuwissen (2005) for further background on these topics. The main factor that breaks down LD is recombination, which can rearrange haplotypes that

FIGURE 1
Break-up of LD over generations

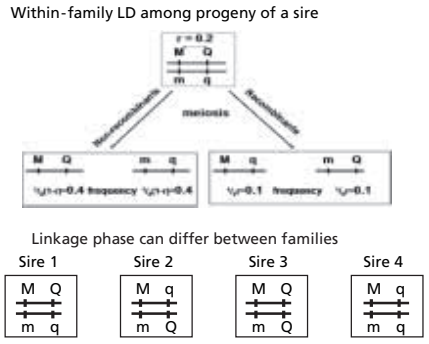


exist within a parent in every generation. Figure 1 shows the effect of recombination (r) on the decay of LD over generations. The rate of decay depends on the rate of recombination between the loci. For tightly linked loci, any LD that has been created will persist over many generations but, for loosely linked loci (r > 0.1), LD will decline rapidly over generations.

Population-wide versus within-family LD

Although a marker and a linked QTL may be in LE across the population, LD will always exist *within* a family, even between loosely linked loci. Consider a double heterozygous sire with haplotypes *MQ*/*mq* (Figure 2). The genotype of this sire is identical to that of an F_1 cross between inbred lines. This sire will produce four types of gametes: non-recombinants *MQ* and *mq* and recombinants *Mq* and *mQ*. As non-recombinants will have higher frequency, depending on the recombination rate between the marker and QTL, this sire will produce gametes that will be in LD. Furthermore, this LD will extend over a larger distance (Figure 1), because it has undergone only one generation of recombination. This specific type of LD,

FIGURE 2
Within-family marker-QTL LD



however, only exists within this family; progeny from another sire, e.g. an *Mq*/*mQ* sire, will also show LD, but the LD is in the opposite direction because of the different marker-QTL linkage phase in the sire (Figure 2). On the other hand, *MQ*/*mQ* and *Mq*/*mq* sire families will not be in LD because the QTL does not segregate in these families. When pooled across families these four types of LD will cancel each other out, resulting in linkage *equilibrium* across the population. Nevertheless, the within-family LD can be used to detect QTL and for MAS provided the differences in linkage phase are taken into account, as will be demonstrated later.

QTL DETECTION AND TYPES OF MARKERS FOR MAS

Application of molecular genetics for genetic improvement relies on the ability to genotype individuals for specific genetic loci. For these purposes, three types of observable genetic loci can be distinguished, as described by Dekkers, 2004:

- direct markers: loci for which the functional polymorphism can be genotyped;
- LD-markers: loci in population-wide LD with the functional mutation;
- LE-markers: loci in population-wide linkage equilibrium with the functional mutation but which can be used for QTL detection and MAS based on within-family LD.

For these alternate types of markers, different strategies are appropriate to detect QTL in livestock populations. These are summarized in Table 1 and will be described in more detail. Strategies for QTL detection in livestock differ from those used in plants because of the lack of inbred lines.

QTL detection using LD markers within crosses

Crossing two breeds that differ in allele and, therefore, haplotype frequencies, creates extensive LD in the crossbred population. This LD extends over large distances

TABLE 1
Summary of strategies for QTL detection in livestock

Type of population	Within crosses		Outbred population			
	F2/Backcross	Advanced intercross	Half- or full-sib families	Extended pedigree	Non-pedigreed population sample	
Type of markers	LD markers		LE markers		LD markers	
Genome coverage	Genome-wide		Genome-wide		Candidate gene regions	Genome-wide
Marker density	Sparse	Denser	Sparse	More dense	Few loci	Dense
Type of LD used	Population-wide LD		Within-family LD		Population-wide LD	
Number of generations of recombination used for mapping	1	>1	1	>1	>>1	
Extent of LD around QTL	Long	Smaller	Long	Smaller	Small	
Map resolution	Poor	Better	Poor	Better	High	

because it has undergone only one generation of recombination in the F_2 (Figure 1). Thus, although these markers may be in LE with QTL within the parental breeds, they will be in partial LD with the QTL in the crossbred population if the marker and QTL differ in frequency between the breeds. This population-wide LD enables detection of QTL that differ between the parental breeds based on a genome scan with only a limited number of markers spread over the genome (~ every 15 to 20 cM). This approach has formed the basis for the extensive use of F_2 or backcrosses between breeds or lines for QTL detection, in particular in pigs, poultry and beef cattle (see Andersson, 2001 for a review). The extensive LD enables detection of QTL that are some distance from the markers but also limits the accuracy (map resolution) with which the position of the QTL can be determined.

More extensive population-wide LD is also expected to exist in synthetic lines, i.e. lines that were created from a cross in recent history. These can be set up on an experimental basis through advanced intercross lines (Darvasi and Soller, 1995) or be available as commercial breeding lines. Depending on the number of generations since the cross, the extent of LD will have eroded over generations and will, therefore, span shorter distances than in F_2 populations (Figure 1). This will require a more dense marker map to scan the genome with equivalent power as in an F_2 but will enable more precise positioning of the QTL.

QTL detection using LE markers in outbred populations

As linkage phases between the marker and QTL can differ from family to family, use of within-family LD for QTL detection requires QTL effects to be fitted on a within-family basis, rather than across the population. Similar to F_2 or backcrosses, the extent of within-family LD is extensive and, thus, genome-wide coverage is provided by a limited number of markers but significant markers may be some distance from the QTL, resulting in poor map resolution. Thus, LE markers can be readily detected on a genome-wide basis using large half-sib families, requiring only sparse marker maps (~15 to 20 cM spacing). Many examples of successful applications of this methodology for detection of QTL regions are available in the literature, in particular for dairy cattle, utilizing the large paternal half-sib structures that are available through extensive use of artificial insemination (see Weller, Chapter 12).

QTL detection using LE markers can also be applied to extended pedigrees by modelling the co-segregation of markers and QTL (Fernando and Grossman, 1989). These approaches use statistical models that are described further in the section on genetic evaluation using LE markers. Depending on the number of generations with phenotypes and marker genotypes that are included in the analysis, map resolution will be better than with analysis of half-sib families because multiple rounds of recombination are included in the data set.

QTL detection using LD markers in outbred populations

The amount and extent of LD that exists in the populations that are used for genetic improvement are the net result of all forces that create and break down LD and are, therefore, the result of the breeding and selection history of each population, along with random sampling. On this basis, populations that have been closed for many generations are expected to be in linkage *equilibrium*, except for closely linked loci.

Thus, in those populations, only markers that are tightly linked to QTL may show an association with phenotype (Figure 1), and even then there is no guarantee because of the chance effects of random sampling.

There are two strategies to find markers that are in population-wide LD with QTL (see Table 1):

- evaluating markers that are in, or close to, genes that are thought to be associated with the trait of interest (candidate genes);
- a genome scan using a high-density marker map, with a marker every 0.5 to 2 cM.

The success of both approaches obviously depends on the extent of LD in the population. Studies in human populations have generally found that LD extends over less than 1 cM. Thus, many markers are needed to obtain sufficient marker coverage in human populations to enable detection of QTL based on population-wide LD. Opportunities to utilize population-wide LD to detect QTL in livestock populations may be considerably greater because of the effects of selection and inbreeding. Indeed, Farnir *et al.* (2000) identified substantial LD in the Dutch Holstein population, which extended over 5 cM. Similar results have been observed in other livestock species (e.g. in poultry, Heifetz *et al.*, 2005). The presence of extensive LD in livestock populations is advantageous for QTL detection, but disadvantageous for identifying the causative mutations of these QTL; with extensive LD, markers that are some distance from the causative mutation can show an association with phenotype.

The candidate gene approach utilizes knowledge from species that are rich in genome information (e.g. human, mouse), effects of mutations in other species, previously identified QTL regions, and/or knowledge of the physiological basis of traits, to identify genes that are thought to play a role in the physiology of the trait. Following mapping and identification of polymorphisms within the gene, associations of genotype at the candidate gene with phenotype can be estimated (Rothschild and Plastow, 1999).

Whereas the candidate gene approach focuses on LD within chosen regions of the genome, recent advances in genome technology have enabled sequencing of entire genomes, including of several livestock species; the genomes of the chicken and cattle have been sequenced and public sequencing of the genome of the pig is under way. In addition, sequencing has been used to identify large numbers of positions in the genome that include SNPs, i.e. DNA base positions that show variation. For example, in the chicken, over 2.8 million SNPs were identified by comparing the sequence of the Red Jungle Fowl with that of three domesticated breeds (International Chicken Polymorphism Map Consortium, 2004). This, combined with reducing costs of genotyping, now enables detection of QTL using LD-mapping with high-density marker maps.

QTL detection using combined LD and linkage analysis in outbred populations

As markers may not be in complete LD with the QTL, both population-wide associations of markers with QTL and co-segregation of markers and QTL within families can be used to detect QTL. Using these combined properties of being both LD and LE markers, methods have been developed to combine LD and linkage information. These methods are further explored under genetic evaluation models in what follows.

INCORPORATING MARKER INFORMATION IN GENETIC EVALUATION PROGRAMMES

The value of genotypic information for predicting the genetic merit of animals is dependent on the predictive ability of the marker genotypes. The three types of molecular loci described previously differ not only in methods of detection but also in methods of their incorporation in genetic evaluation procedures. Whereas direct and, to a lesser degree, LD markers, allow selection on genotype across the population, use of LE markers must allow for different linkage phases between markers and QTL from family to family, i.e. LE markers are family specific and family specific information must be derived. As discussed later in this chapter, this makes LE markers a lot less attractive for use in breeding programmes. In this section, the different types of models that have been proposed for genetic evaluation based on marker information are described and this is followed by a brief description of some practical issues regarding implementation of such methods and the likely routes towards achieving that goal.

Modelling QTL effects in genetic evaluation

By using QTL information in genetic evaluation, in principle, part of the assumed polygenic variation is substituted by a separate effect due to a genetic polymorphism at a known locus. This has the immediate effect of having a much better handle on the Mendelian sampling process, as phenotypic co-variance can be evaluated based on specific genetic similarity rather than on an average relationship. For example, on average two full sibs share 50 percent of their alleles, but at a specific locus it is now possible to know whether these full sibs carry exactly the same complete genotype (both paternal and maternal alleles are in common), or actually have a completely different genotype. The actual degree of similarity of full sibs at a QTL can thus vary between 0 and 1. This additional information helps to better evaluate the genetic merit due to specific QTL, and to better predict offspring that do not yet have phenotypic measurements.

A number of different approaches have been described to accommodate marker information in genetic evaluation. Roughly, these methods can be distinguished through their modelling of the QTL effect and through the type of genetic marker information used. The QTL effect can be modelled as random or fixed, while the molecular information comes from LE, LD or direct markers.

With a fixed QTL model, regression on genotype probabilities would be used in genetic evaluation to account for the effect of QTL polymorphisms. In the simplest additive QTL model, suitable for estimating breeding values, simple regressions could be included on the probability of carrying the favourable mutation. Regression can be on known genotypes (class variables), or probabilities can be derived for ungenotyped animals in a general complex pedigree (Kinghorn, 1999). A fixed QTL model is sensible if few alleles are known to be segregating, and where dominance and/or epistasis are important. The model also assumes effects being the same across families. The effects of various genotypes could be fitted separately, giving power to account for dominance and epistasis in case of multiple QTL. For selection purposes, a fixed QTL effect, if additive, would be added to the polygenic estimated breeding values (EBVs), similar to breed effects in across-breed evaluations. The advantage of a fixed QTL model is the limited number of effects that need to be fitted.

Alternatively, QTL effects could be modelled as random effects, with each individual having a different QTL effect. Co-variances are based on the probability of QTL alleles being identical by descent rather than on numerator relationships as in the usual animal model with polygenic effects. With full knowledge about segregation, this would effectively fit all founder alleles as different effects. The random QTL model was first described by Fernando and Grossman (1989), where for each animal both the paternal and the maternal allele were fitted. Without loss of information, these effects can be collapsed into one genotypic effect for each animal (Pong-Wong *et al.,* 2001). The random QTL model makes no assumptions about number of alleles at a QTL and it automatically accommodates possible interaction effects of QTL with genetic background (families or lines). Therefore, the random QTL model is less reliant on assumptions about homogeneity of QTL effects. The random QTL model is a natural extension to the usual mixed model and seems therefore a logical way to incorporate genotype information into an overall genetic evaluation system. These models result in EBVs for QTL effects along with a polygenic EBV. The total EBV is the simple sum of these estimates. One of the main computational limitations of this method, however, is the large number of equations that must be solved, which increases by two per animal for each QTL that is fitted. Thus, the number of QTL regions that can be incorporated is limited.

Genetic evaluation using direct markers

When the genotype of an actual functional mutation is available, no pedigree information is needed to predict the genotypic effect, as QTL genotypes are measured directly. When there is only a small number of alleles, the number of specific genotypes is limited. In genetic evaluation, it would seem appropriate to treat the genotype effect as a fixed effect, i.e. the assumption is that genotype differences are the same in different families and herds or flocks. Such assumptions might be reasonable for a bi-allelic QTL model in a relatively homogeneous population. Alternatively, random QTL models could be used with different effects for different founder alleles, or even QTL by environment interactions. In both fixed and random QTL models, genotype probabilities can be derived for individuals with missing genotypes.

Genetic evaluation using LE markers

When the genotype test is not for the gene itself, but for a linked marker, QTL probabilities derived from marker genotypes will be affected by the recombination rate between marker and QTL and by the extent of LD between the QTL and marker across the population. If LD between the QTL and a linked marker only exists within families, marker effects or, at a minimum, marker-QTL linkage phase must be determined separately for each family. This requires marker genotypes and phenotypes on family members. If linkage between the marker and QTL is loose, phenotypic records must be from close relatives of the selection candidate because associations will erode quickly through recombination. With progeny data, marker-QTL effects or linkage phases can be determined based on simple statistical tests that contrast the mean phenotype of progeny that inherited alternate marker alleles from the common parent. A more comprehensive approach is based on Fernando and Grossman's (1989) random QTL model, where marker information from complex pedigrees can be used

to derive co-variances between QTL effects, yielding best linear unbiased prediction (BLUP) of breeding value for both polygenic and QTL effects. Random effects of paternal and maternal QTL alleles are added to the standard animal model with random polygenic breeding values. The variance-co-variance structure of the random QTL effects, also known as the gametic relationship matrix (GRM), is based on probabilities of identity by descent (IBD), and is now derived from co-segregation of markers and QTL within a family. Probabilities of IBD derived from pedigree and marker data link QTL allele effects that are expected to be equal or similar, therefore using data from relatives to estimate an individual's QTL effects. For example, if two paternal half-sibs i and j have inherited the same paternal allele for markers that flank the QTL (with recombination rate r), they are likely IBD for the paternal QTL allele and the correlation between the effects of their paternal QTL alleles will be $(1\text{-}r)^2$. The method is appealing, but computationally demanding for large-scale evaluations, especially when not all animals are genotyped and complex procedures must be applied to derive IBD probabilities.

Genetic evaluation using LD markers

Most QTL projects have moved towards fine mapping where the final result is a marker or marker haplotype in LD with the QTL, if not the direct mutation. A haplotype of marker alleles close enough to the putative QTL is likely to be in LD with QTL alleles. Such a marker test provides information about QTL genotype across families, and is in a sense not very different from a direct marker. The most convenient way to include genotypic information from marker haplotypes in genetic evaluation systems is through the random QTL model. In their original paper, Fernando and Grossman (1989) derived IBD from genotype data on single markers and recombination rates between marker and QTL. However, the random QTL model is more versatile, and co-variances based on IBD probabilities can also use information beyond pedigree, based on LD. The latter can be derived from marker or haplotype similarity, e.g. based on a number of marker genotypes surrounding a putative QTL. Meuwissen and Goddard (2001) proposed using both linkage and LD information to derive IBD-based co-variances (termed LDL analysis). Lee and van der Werf (2005) showed that with denser markers, the value of linkage information, and therefore pedigree, reduces. Hence, when QTL positions become more accurately defined, genetic information from close markers (within a few cM) can be used increasingly to derive LD-based IBD probabilities, thereby defining co-variances between random QTL effects without the need for a family structure or information through pedigree.

Lee and van der Werf (2006) have shown that LD information results in a very dense GRM. Genetic evaluation, which is usually based on mixed model equations that are relatively sparse, is currently not feasible computationally for the LDL method for a large number of individuals and alternative models are needed. One approach is to model population-wide LD by simply including the marker genotype or haplotype as a fixed effect in the animal model evaluation, as suggested by Fernando (2004). An advantage of modelling population-wide LD effects as fixed rather than random is that fewer assumptions about population history are needed. A disadvantage is that estimates are not "BLUPed", i.e. regressed towards a mean depending on

the amount of information that is available to estimate their effects. This will be important if some of the genotype or haplotype effects cannot be estimated with substantial accuracy because the number of individuals with that genotype or haplotype is limited. Haplotype effects could also be fitted as random, but more development is needed in this area.

Whole genome approach for genetic evaluation using high-density LD markers

With more and more QTL being discovered, the polygenic component will slowly be replaced by multiple QTL effects, the inheritance of each being followed by marker brackets or more generally by information on haplotypes. Nejati-Javaremi, Smith and Gibson (1997) presented the concept of the total allelic relationship, where the co-variance between two individuals was derived from allelic identity by descent, or by state (based on molecular marker information), with each location weighted by the variance explained by that region. This approach contrasts with the average relationships derived from pedigree that are used in the numerator relationship matrix. Nejati-Javaremi, Smith and Gibson (1997) showed that using total allelic relationship resulted in a higher selection response than pedigree based relationships, because it more accurately accounts for the variation in the additive genetic relationships between individuals. Therefore, the gain of following inheritance at specific genome locations contributes to more accurate genetic evaluation, and is able to deal more specifically with within and between loci interactions and with specific modes of inheritance at different QTL.

When large-scale marker genotyping becomes cheap and available to breeders at low cost, this approach could even be used for non-detected QTL and genetic evaluation could be based on a "whole genome approach" (Meuwissen, Hayes and Goddard, 2001). In this approach, marker haplotypes are fitted as independent random effects for each, e.g. 1 cM region of the genome. In the work by Meuwissen, Hayes and Goddard (2001), variances associated with each haplotype were either assumed to be equal for each chromosomal region or estimated from the data using Bayesian procedures with alternate prior distributions. In essence, this procedure estimates breeding values for each haplotype, and EBVs of individuals are computed by simply summing EBVs for the haplotypes that they contain.

Using this procedure, Meuwissen, Hayes and Goddard (2001) demonstrated through simulation, that for populations with an effective population size of 100 and a spacing of 1 or 2 cM between informative markers across the genome, sufficient LD was present to predict genetic values with substantial accuracy for several generations based on associations of marker haplotypes with phenotype on as few as 500 individuals. It should be noted that, in the approach proposed by these authors, no polygenic effect is included since all regions of the genome are included in the model. It may, however, be useful to include a polygenic effect because LD between markers and QTL will not be complete for all regions. In addition, this model assumes that haplotype effects are independent within and across regions. Incorporating IBD probabilities to model co-variances between haplotypes within a region as in Meuwissen and Goddard (2000), and by incorporating co-variances between adjacent regions caused by LD between regions, could lead to further improvements but would also lead to increasing computational demands.

In general, for the purpose of increased genetic change of economically important quantitative traits, and in the context of well recorded and efficient breeding programmes, there is no need to have knowledge of functional mutations since nearby markers will have a high predictive value about genetic merit. Moreover, the benefit from the extra investment and time spent on finding functional mutations might be superseded by the genetic change that can be made in the breeding programme in the meantime.

Implementation of marker-assisted genetic evaluation

It is important to note that, for most of the gene marker tests currently on the market, integration with existing systems for genetic evaluation is not obvious. This is because the gene testing is either for a Mendelian characteristic, or it predicts phenotypic differences for traits that are not the same as those in current genetic evaluation. Moreover, breeders would not only be interested in more accurate EBVs based on gene markers, but they would also want to know the actual QTL genotypes for their breeding animals. This information on individual genotype will become less relevant if more gene tests become available and if testing becomes cheaper and more widespread. This might still take some years. Thus, as gene marker testing is gradually introduced, it is more likely to create additional selection criteria to consider and it will take some time before QTL information is seamlessly and optimally integrated in existing genetic evaluation programmes. In particular, if genetic evaluation is based on information from many different breeding units, such as in cattle or sheep, genotyping information will initially be available for only a small proportion of the breeding animals, possibly not justifying a total overhaul of the system for genetic evaluation. Simple ad hoc procedures where QTL effects are estimated and presented separately as additional effects are initially a more likely route to implementation.

Solutions for fixed QTL genotype effects, along with genotype probabilities as outputs of genetic evaluation, might be interesting to breeders and, compared with random QTL effects, may be more likely to be presented and used separately from polygenic EBVs. This would also be the case for genotypic information on Mendelian characters, where there is no polygenic component.

INCORPORATING MAS IN SELECTION PROGRAMMES

Molecular information can be used to enhance both the processes of integrating superior qualities of different breeds and within-breed selection. These strategies are further described below.

Between-breed selection

Crossing breeds results in extensive LD, which can be capitalized upon using MAS in a number of ways. If a large proportion of breed differences in the trait(s) of interest are due to a small number of genes, gene introgression strategies can be used. If a larger number of genes is involved, MAS within a synthetic line is the preferred method of improvement.

Marker-assisted introgression

Introgression of the desirable allele at a target gene from a donor to a recipient breed is accomplished by multiple backcrosses to the recipient, followed by one or more generations of intercrossing. The aim of the backcross generations is to produce

individuals that carry one copy of the donor QTL allele but that are similar to the recipient breed for the rest of the genome. The aim of the intercrossing phase is to fix the donor allele at the QTL. Marker information can enhance the effectiveness of the backcrossing phase of gene introgression strategies by: (i) identifying carriers of the target gene(s) (foreground selection); and (ii) enhancing recovery of the recipient genetic background (background selection). The effectiveness of the intercrossing phase can also be enhanced through foreground selection on the target gene(s). If the target gene cannot be genotyped directly, carrier individuals can be identified based on markers that flank the QTL at <10 cM, because of the extensive LD in crosses. The markers must have breed-specific alleles in order to identify line origin. For the introgression of multiple target genes, gene pyramiding strategies can be used during the backcrossing phase to reduce the number of individuals required (Hospital and Charcosset, 1997; Koudandé *et al.,* 2000). For background selection, markers are used that are spread over the genome at <20 cM intervals, such that most genes that affect the trait will be within 10 cM from a marker. Combining foreground and background selection, selection will be for the donor breed segment around the target locus but for recipient breed segments in the rest of the genome. Foreground selection will result in selection for both the target locus and for donor breed loci that are linked to this locus, some of which could have an unfavourable effect on performance. To reduce this so-called linkage drag around the target locus, in the molecular score used for background selection greater emphasis can be given to markers that are in the neighbourhood of the target locus (apart from the flanking markers, which are used in foreground selection).

Most studies have considered marker-assisted introgression (MAI) of single QTL (e.g. Hospital and Charcosset, 1997) but often several QTL must be introgressed simultaneously. Koudandé *et al.* (2000) showed that large populations are needed to obtain sufficient individuals that are heterozygous for all QTL in the backcrossing phase. This would make MAI not feasible in livestock breeding programmes. In many cases, however, immediate fixation of introgressed QTL alleles may not be required. Instead, the objective of the backcrossing phase can be to enrich the recipient breed with the favourable donor QTL alleles at sufficiently high frequency for selection following backcrossing. The effectiveness of such strategies was demonstrated by Chaiwong *et al.* (2002).

Marker-assisted improvement of synthetic lines

In MAI studies it is usually assumed that the aim is to recover the recipient breed genotype, except for the donor QTL. An alternative objective could be to aim simply for individuals with highest merit. Selection would then be for QTL genotype as well as EBV, estimated across breeds or lines. This EBV selection would replace background selection, as recovery of the recipient genotype is achieved through selection on genetic merit rather than through selecting for breed of origin. This strategy would be more competitive if the original breeds overlap in merit, and indeed, as was shown by Dominik *et al.* (2007), background selection based on anonymous markers would be less profitable.

Strategies for using markers to select within a hybrid population were first proposed by Lande and Thompson (1990). These strategies capitalize on population-wide LD that initially exists in crosses

between lines or breeds. Thus, marker-QTL associations identified in the F_2 generation can be selected for several generations, until the QTL or markers are fixed or the disequilibrium disappears. Zhang and Smith (1992) evaluated the use of markers in such a situation with selection on BLUP EBV. Although both studies considered the ideal situation of a cross with inbred lines, there will be opportunities to utilize a limited number of markers to select for favourable QTL regions that are detected in crosses between breeds, thereby enhancing the development of superior synthetics. Pyasatian, Fernando and Dekkers (2006) investigated use of the whole genome approach of Meuwissen, Hayes and Goddard (2001) for MAS in a cross by including all markers as random effects in the model for genetic evaluation. They showed that this resulted in substantially greater responses to selection than selection on identified QTL regions only. Due to the much greater LD, whole genome selection in a cross can be accomplished with a much smaller number of markers compared with the number required for whole genome selection in an outbred population.

Within-breed selection

The procedures described previously for incorporating markers in genetic evaluation result in estimates of breeding values associated for QTL, together with estimates of polygenic breeding values. Alternatively, if molecular data are not incorporated into genetic evaluations, as will be the case for more ad hoc approaches and for gene tests for Mendelian characteristics, separate selection criteria will be available that capture the molecular information. The following three selection strategies can then be distinguished (Dekkers, 2004):

- select on the QTL information alone;
- tandem selection, with selection on QTL followed by selection on polygenic EBV;
- selection on the sum of the QTL and polygenic EBV.

Selection on QTL or marker information alone ignores information that is available on all other genes (polygenes) that affect the trait and is expected to result in the lowest response to selection unless all genes that affect the trait are included in the QTL EBV. This strategy does not, however, require additional phenotypes other than those that are needed to estimate marker effects, and can be attractive when phenotype is difficult or expensive to record (e.g. disease traits, meat quality, etc.). Selection on the sum of the QTL and polygenic EBV is expected to result in maximum response in the short term, but may be suboptimal in the longer term because of losses in polygenic response (Gibson, 1994). Indexes of QTL and polygenic EBV can be derived that maximize longer-term response (Dekkers and van Arendonk, 1998) or a combination of short- and longer-term responses (Dekkers and Chakraborty, 2001). However, if selection is on multiple QTL and emphasis is on maximizing shorter-term response, selection on the sum of QTL and polygenic EBV is expected to be close to optimal. Optimizing selection on a number of EBVs, indexes and genotypes, while also considering inbreeding rate and other practical considerations is not a trivial task. Kinghorn, Meszaros and Vagg (2002) have proposed a mate selection approach that could be used to handle such problems, and it can be expected that with more widespread use of genotypic information for a larger number of regions, specific knowledge about individual QTL becomes less interesting and will simply contribute to prediction of whole EBV or whole genotype.

Meuwissen and Goddard (1996) published a simulation study that looked at the main characteristics determining efficiency of MAS using LE markers. They found that MAS could improve the rate of genetic improvement up to 64 percent by selecting on the sum of QTL and polygenic EBV. Their work also demonstrated that MAS is mainly useful for traits where phenotypic measurement is less valuable because of: (i) low heritability; (ii) sex-limited expression; (iii) availability only after sexual maturity; and (iv) necessity to sacrifice the animal (e.g. slaughter traits). Selection of animals based on (most probable) QTL genotype will allow earlier and more accurate selection, increasing the short- and medium-term selection response.

Most simulation studies have assumed complete marker genotype information but in practice only a limited number of individuals will be genotyped. However, in an advanced breeding programme with complete information on phenotype and pedigree information, marker and QTL genotype probabilities could be derived for un-genotyped animals and genotyping strategies could be optimized to achieve a high value for the investments made. Marshall, Henshall and van der Werf (2002) looked at strategies to minimize genotyping cost in a sheep breeding programme. Close to maximal gain could be achieved when genotyping was undertaken only for high ranking males and animals whose marker genotype probability could not be derived with enough certainty based on information on relatives. Marshall, van der Werf and Henshall (2004) also looked at progeny testing of sires to determine family-specific marker-QTL phase within a breeding nucleus. Again, testing of a limited number of males provided a lot of information about phase for several generations of breeding animals, as progeny tested sires have relationships with descendants. However, in breeding programmes for more extensive production systems (beef, sheep), pedigree recording is often incomplete and only a small proportion of animals are genotyped. Moreover, these genotyped animals are not necessarily the key breeding animals. The utility of linked markers will be even more limited if pedigree relationships cannot be used to resolve genotype probabilities and marker-QTL phase of un-genotyped individuals.

A second point of caution is that many studies on MAS have taken a single-trait approach and shown that genetic markers could have a large impact on responses for traits that are difficult to improve by phenotypic selection. However, within the context of a multitrait breeding objective, the overall impact of such markers on the breeding goal may be less because a greater response for one trait often appears at the expense of another. For example, genetic markers for carcass traits improve the ability to select (i.e. earlier, with higher accuracy) for such traits, but selection emphasis for other traits is reduced. Therefore, the overall effect of MAS on the breeding programme will generally be much smaller than predicted for single trait MAS-favourable cases. The main effects of MAS would be to shift the selection response in favour of the marked traits, rather than achieving much additional overall response. Hence, while it will be easier to select for carcass and disease resistance, further improvement for these traits will be at the expense of genetic change for production traits (growth, milk).

The impact of MAS on the rate of genetic gain may be limited in conventional breeding programmes (ranging up to perhaps 10 percent extra gain) unless

the variation in profitability is dominated by traits that are hard to measure. However, new technologies often lead to other breeding programme designs being closer to optimal. Genotypic information has extra value in the case of early selection and where within-family variance can be exploited, which is particularly the case in programmes where reproductive technologies are used. Reproductive technologies usually lead to early selection and more emphasis on between-family selection. DNA marker technology and reproductive technologies are therefore highly synergistic and complementary (van der Werf and Marshall, 2005) and gene markers have much more value in such programmes. Gene marker information is also clearly valuable in introgression programmes, as demonstrated by simulation (Chaiwong *et al.*, 2002; Dominik *et al.*, 2006) as well as in practice (Nimbkar, Pardeshi and Ghalsasi, 2005). Yet, although these examples are favourable to the value of gene marker information, the added value of MAS still relies heavily on a high degree of trait and pedigree recording.

OPPORTUNITIES FOR MAS IN DEVELOPING COUNTRIES

Complete phenotypic and pedigree information is often only available in intensive breeding units. Therefore, in the context of low input production systems, some questions can be raised concerning the validity and practicality of the simulation studies described above, and it would be more difficult to realize the value of marker information. It would be harder and more expensive to determine the linkage phase in the case of using linked markers. Moreover, even if the genetic marker were a direct or LD marker, its effect on phenotype would have to be estimated for the population and the environment in which it is used. This would require phenotypes and genotypes on a sample of a rather homogeneous population to avoid spurious associations that could result from unknown population stratification. Therefore, a gene marker for a QTL is likely to be most successful in an environment with intensive pedigree and performance recording. Nevertheless, in low input environments, direct and LD markers will be more useful than LE markers because the latter require routine recording of phenotypes and genotypes to estimate QTL effects within families.

In addition to MAS within local breeds, several other strategies for breed improvement could be pursued in developing countries, including gene introgression and MAS within synthetic breeds. This would be most advantageous for introducing specific disease resistance alleles into breeds with improved production characteristics to make them more tolerant to the environments encountered in developing countries. Gene introgression is, however, a long and expensive process and only worthwhile for genes with large effects. MAS within synthetic breeds, e.g. a cross between local and improved temperate climate breeds, can allow development of a breed that is based on the best of both breeds (e.g. Zhang and Smith, 1992). Because of the extensive LD within the cross, a limited number of markers would be needed. Care should, however, be taken to avoid the impact of genotype x environment interactions if MAS is implemented in a more controlled environment.

REFERENCES

Andersson, L. 2001. Genetic dissection of phenotypic diversity in farm animals. *Nature Revs. Genet*. 2: 130–138.

Chaiwong, N., Dekkers, J.C.M., Fernando, R.L. & Rothschild, M.F. 2002. Introgressing multiple QTL in backcross breeding programs of limited size. *Proc. 7th Wld. Congr. Genet. Appl. Livest. Prodn*. Electronic Communication No. 22: 08. Montpellier, France.

Darvasi, A. & Soller, M. 1995. Advanced intercross lines: an experimental population for fine genetic mapping. *Genetics* 141: 1199–1207.

Dekkers, J.C.M. & van Arendonk, J.A.M. 1998. Optimizing selection for quantitative traits with information on an identified locus in outbred populations, *Genet. Res*. 71: 257–275.

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

Dekkers J.C.M. & Chakraborty, R. 2001. Potential gain from optimizing multi-generation selection on an identified quantitative trait locus. *J. Anim. Sci.* 79: 2975–2990.

Dominik, S., Henshall, J., O'Grady, J. & Marshall, K.J. 2007. Factors influencing the efficiency of a marker assisted introgression program in merino sheep. *Genet. Sel. Evol*. In press.

Farnir, F., Coppieters, W., Arranz, J.-J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marcq, F., Moreau, L., Mni, M., Nezer, C., Simon, P., Vanmanshoven, P., Wagenaar, D. & Georges, M. 2000. Extensive genome-wide linkage disequilibrium in cattle. *Genome Res*. 10: 220–227.

Fernando, R.L. 2004. Incorporating molecular markers into genetic evaluation. Session G6.1. *Proc. 55th Meeting of the European Association of Animal Production*. 5–9 September 2004, Bled, Slovenia.

Fernando, R.L. & Grossman, M. 1989. Marker-assisted selection using best linear unbiased prediction. *Genet. Sel. Evol*. 21: 467–477.

Gibson, J.P. 1994. Short-term gain at the expense of long-term response with selection of identified loci, *Proc. 5th Wld. Cong. Genet. Appl. Livest. Prodn.* CD-ROM Communication No. 21: 201–204. University of Guelph, Canada.

Goddard, M.E. & Meuwissen, T.H.E. 2005. The use of linkage disequilibrium to map quantitative trait loci. *Austr. J. Exp. Agric*. 45: 837–845.

Heifetz, E.M., Fulton, J. E., O'Sullivan, N., Zhao, H., Dekkers, J.C.M. & Soller, M. 2005. Extent and consistency across generations of linkage disequilibrium in commercial layer chicken breeding populations. *Genetics* 171: 1173–1181.

Hospital, F. & Charcosset, A. 1997. Marker-assisted introgression of quantitative trait loci. *Genetics* 147: 1469–1485.

International Chicken Polymorphism Map Consortium. 2004. A genetic variation map for chicken with 2.8 million single nucleotide polymorphisms. *Nature* 432: 717–722.

Kinghorn, B.P. 1999. Use of segregation analysis to reduce genotyping costs. *J. Anim. Breed. Genet.* 116: 175–180.

Kinghorn, B.P., Meszaros, S.A. & Vagg, R.D. 2002. Dynamic tactical decision systems for animal breeding, *Proc. 7th Wld. Congr. Genet. Appl. Livest. Prodn*. Communication No. 23-07. Montpellier, France.

Koudandé, O.D., Iraqi, F., Thomson, P.C., Teale, A.J. & van Arendonk, J.A.M. 2000. Strategies to optimize marker-assisted introgression of multiple unlinked QTL. *Mammal. Genome* 11: 145–150.

Lande, R. & Thompson, R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124: 743–56.

Lee, S.H. & van der Werf, J.H.J. 2005. The role of pedigree information in combined linkage disequilibrium and linkage mapping of quantitative trait loci in a general complex pedigree. *Genetics* 169: 455–466.

Lee, S.H. & van der Werf, J.H.J. 2006. An efficient variance component approach implementing an average information REML suitable for combined LD and linkage mapping with a general complex pedigree. *Genet. Sel. Evol*. 38: 25–43.

Marshall, K., Henshall, J. & van der Werf, J.H.J. 2002. Response from marker-assisted selection when various proportions of animals are marker typed: a multiple trait simulation study relevant to the sheep meat industry. *Anim. Sci.* 74: 223–232.

Marshall, K.J., van der Werf J.H.J. & Henshall, J. 2004. Exploring major gene - marker phase typing strategies in marker assisted selection schemes. *Anim. Sci.* 78: 213–227.

Meuwissen, T.H.E. & Goddard, M.E. 1996. The use of marker haplotypes in animal breeding schemes. *Genet. Sel. Evol*. 28: 161–176.

Meuwissen, T.H.E. & Goddard, M.E. 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics* 155: 421–430.

Meuwissen, T.H.E. & Goddard, M.E. 2001. Prediction of identity by descent probabilities from marker haplotypes. *Genet. Sel. Evol*. 33: 605–634.

Meuwissen, T.H.E., Hayes, B. & Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.

Nejati-Javaremi, A., Smith, C. & Gibson, J.P. 1997. Effect of total allelic relationship on accuracy and response to selection. *J. Anim. Sci*. 75: 1738–1745.

Nimbkar, C., Pardeshi, V. & Ghalsasi, P. 2005. Evaluation of the utility of the FecB gene to improve the productivity of Deccani sheep in Maharashtra, India. pp. 145–154. *In* H.P.S. Makkar & G.J. Viljoen, eds. *Applications of gene-based technologies for improving animal production and health in developing countries.* Netherlands, Springer.

Pong-Wong, R., George, A.W., Woolliams, J.A. & Haley, C.S. 2001. A simple and rapid method for calculating identity-by-descent matrices using multiple markers. *Genet. Sel. Evol*. 33: 453–471.

Pyasatian, N., Fernando, R.L. & Dekkers, J.C.M. 2006. Genomic selection for composite line development using low density marker maps. In *Proc. 8th Wrld. Congr. Genet. Appl. Livest. Prodn.*, Paper 22–65. Belo Horizonte, Brazil.

Rothschild, M.F. & Plastow, G.S. 1999. Advances in pig genomics and industry applications. *AgBiotechNet.* 1: 1–7.

van der Werf, J.H.J. & Marshall, K. 2005. Combining gene-based methods and reproductive technologies to enhance genetic improvement of livestock in developing countries, pp. 131–144. *In* H.P.S Makkar & G.J. Viljoen, eds. *Applications of gene-based technologies for improving animal production and health in developing countries*. Netherlands, Springer.

Zhang, W. & Smith, C. 1992. Computer simulation of marker-assisted selection utilizing linkage disequilibrium. *Theor. Appl. Genet*. 83: 813–820.

CHAPTER 11

Marker-assisted selection in poultry

Dirk-Jan de Koning and Paul M. Hocking

SUMMARY

Among livestock species, chicken has the most extensive genomics toolbox available for detection of quantitative trait loci (QTL) and marker-assisted selection (MAS). The uptake of MAS is therefore not limited by technical resources but mostly by the priorities and financial constraints of the few remaining poultry breeding companies. With the cost of genotyping decreasing rapidly, an increase in the use of direct trait- single nucleotide polymorphism (SNP)-associations in MAS can be predicted.

CURRENT STATUS OF CHICKEN BREEDING PROGRAMMES

Poultry production has been the fastest growing livestock industry over the last decades especially in middle- and low-income countries (Taha, 2003). In 2001, poultry production accounted for 70 million tonnes of poultry meat and 47 million tonnes of eggs (Arthur and Albers, 2003). Among poultry, chicken account for 85 percent of meat production and 96 percent of egg production (Bilgili, 2001; Arthur and Albers, 2003; Taha, 2003). While chickens have been domesticated and selected for thousands of years, modern poultry breeding started during the 1950s. One of the most notable features is the diversification between chickens bred for meat production (broilers) and those bred for table egg production (layers). This is a result of the negative genetic correlation in chicken between growth and reproductive traits. Within breeds, there is a separation into male and female lines that are crossed to produce commercial hybrids. In broilers, male lines are selected for growth and carcass quality whereas in female lines less emphasis is placed on growth and more on reproductive traits such as egg production and hatchability. In table egg-laying chickens, male lines are selected for high egg production and high egg weight whereas in female lines selection may emphasize rate of lay with less attention to egg size. In both broiler and layer lines the primary selection goal is the improvement of feed efficiency and economic gain.

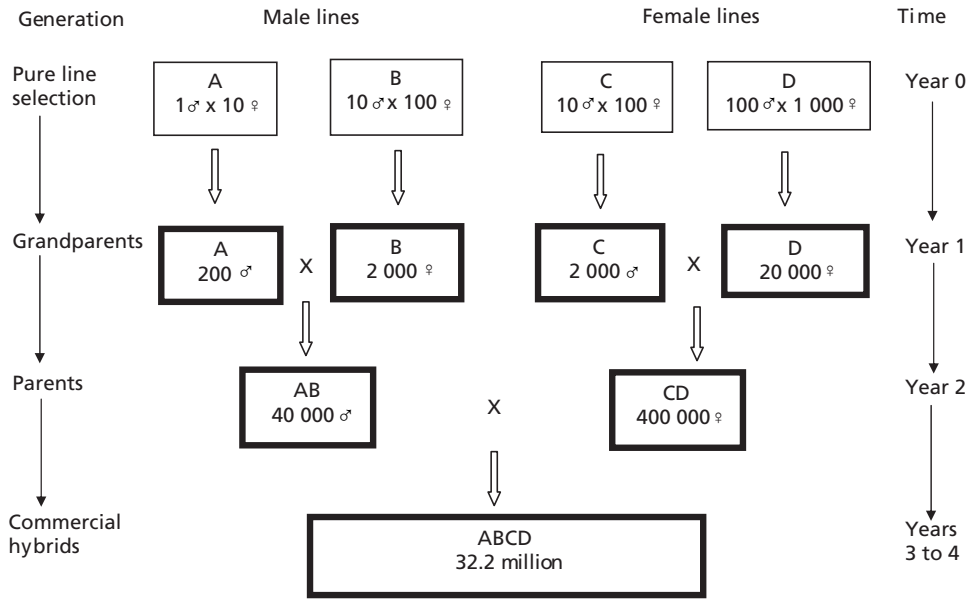
Significant heterosis for fitness traits in poultry is well established and all commercial poultry (chickens, turkeys and ducks) are hybrids that are produced in a selection and multiplication pyramid that is illustrated in Figure 1. Crossing male and female lines maximizes heterosis at the grandparent and parent levels of the hierarchy, and allows traits that have been genetically improved in different lines to be combined in the commercial birds. The power of this structure to deliver large economic gains in chickens is a result of their high reproductive rate and short generation interval and is clearly illustrated by this example of an egg-laying improvement programme. Even greater numerical efficiency is possible in broilers: a single pen containing ten females and one male at the nucleus level might produce 150 great-grandparents after selection (line D of Figure 1); these will produce 50 female offspring each or 7 500 grandparents in a year and these grandparents will generate 375 000 female parent stock during the succeeding year. These hybrid parent females will each produce over 130 male and female offspring and generate nearly 50 million commercial broilers or 70 000 tonnes of meat. The figure illustrates the rapidity with which genetic improvement at the nucleus level can be disseminated to commercial flocks and the fact that relatively few pure-line birds are needed to produce very large numbers of commercial layers.

The existence of this breeding structure results in rapid transmission of genetic change to commercial flocks (about four years), including traits that might be improved by MAS. Conversely, undesirable genetic change can also be disseminated very quickly to a very large number of birds. In practice, far more birds are kept at the nucleus level than shown in Figure 1 where the numbers presented are purely for illustrative purposes.

STATUS OF FUNCTIONAL GENOMICS IN CHICKEN

Among the various livestock species, chicken has the most comprehensive genomic tool-

FIGURE 1
Multiplication pyramid for a four-strain commercial hybrid layer



The boxes in bold are the multiplication (crossing) phase. The top line of the boxes indicates the lowest level of the pure-line (nucleus) selection stage. The numbers in the boxes refer to the minimal numbers of birds in line A, and corresponding numbers in lines B, C and D, that are required to generate hybrid laying hens at the commercial level and the numbers that result from this process. The time scale is indicated on the right hand side of the figure assuming a generation interval of one year.

Adapted from Bowman, 1974.

box. The chicken genome consists of 39 pairs of chromosomes: eight cytologically distinct macrochromosomes, the sex chromosomes Z and W and 30 pairs of cytologically indistinguishable microchromosomes. Linkage maps were developed initially using three separate mapping populations (Bumstead and Palyga, 1992; Crittenden *et al.*, 1993; Groenen *et al.*, 1998) that were later merged to provide a consensus map with 1 889 markers (Groenen *et al.*, 2000). A good overview of the consensus linkage map and the cytogenetic map can be found in the *First Report of Chicken Genes and Chromosomes 2000* and its successor in 2005 (Schmid *et al.*, 2000, 2005). All chicken maps can be viewed at www.thearkdb.org.

More recently, the chicken genome became the first livestock genome to be sequenced with a six-fold coverage (six full genome equivalents) (Hillier *et al.*, 2004). The chicken genome sequence can be browsed via a number of Web sites, which are summarized at www.chicken-genome.org/resources/databases.html. The genome sequence effort was accompanied by partial sequencing of three distinct poultry breeds (a broiler, a layer and a Chinese Silky), to identify SNPs between and among these and the reference sequence of the Red Jungle Fowl. This resulted in an SNP map consisting of about 2.8 million SNPs (Wong *et al.*, 2004). The chicken polymorphism database (ChickVD) can be browsed at: http://chicken.genomics.org.cn/index.jsp

(Wang *et al.*, 2005). The SNP map will facilitate the development of genome-wide SNP assays, containing between 5 000 and 20 000 SNPs per assay.

For the study of gene expression, there are various complementary DNA (cDNA) microarrays available, varying from targeted arrays (immune, neuroendocrine, embryo) to whole genome generic arrays. Recently, a whole-genome Affymetrix chip was developed in collaboration with the chicken genomics community (www.affymetrix.com and www.chicken-genome.org/resources/affymetrix-faq1.htm). Altogether, this provides a very comprehensive toolbox to study the functional genomics of chicken, whether this be an individual gene or the entire genome.

CURRENT UPTAKE OF MAS IN CHICKEN

Implementation of MAS requires knowledge of marker-trait associations based on QTL and candidate gene studies, and ideally from studies of the underlying genetic mechanisms. There have been a large number of QTL studies in chicken covering a wide range of traits including growth, meat quality, egg production, disease resistance (both infectious diseases and production diseases) and behaviour. These studies have recently been reviewed (Hocking, 2005). A total of 27 papers reported 114 genome-wide significant QTL from experimental crosses largely involving White Leghorn and broiler lines. A summary of the QTL that have been detected is presented in Table 1. While the abundance of QTL would indicate ample opportunity for MAS in chicken, it must be noted that nearly all studies were carried out in experimental crosses and hence the results do not reflect QTL within selected populations. However, these results do provide a good starting point to search for QTL within commercial populations, as demonstrated for growth and carcass traits where many published QTL also explained variation within a broiler dam line (de Koning *et al.*, 2003; de Koning *et al.*, 2004). To the authors' knowledge, there are no other QTL studies within commercial lines of poultry in the public domain. Of the QTL from experimental crosses, only a small number has been followed up by fine mapping analyses and the responsible gene mutation has only been described for some disease resistance QTL (Liu *et al.*, 2001a, b; Liu *et al.*, 2003).

A good example of how QTL mapping combined with functional studies can identify functional variants is for Marek's disease. Marek's disease (MD) is an infec-

TABLE 1
Quantitative traits and chromosomal locations in experimental chicken crosses

Trait	Chromosome (number of QTL)	Total QTL	Number of papers
Behaviour/fear	1,2(3),3,4(2),7,10,27, E22	11	5
Body fat	1(2),3,5,7(2),15,28	8	3
Body weight	1(7),2(4),3(4),4(5),5,8(2),11,12,13(2),27(3)Z(2)	32	9
Carcass quality	1(2),2,3,4(2).5(2),6(2),7(3),8(2),9,13(2),27,Z(2)	21	1
Disease resistance	1(4),2(2),3(2),4(2),5(5),6(2),7,8,14,18,27,Z	23	10
Egg number	8,Z(2)	3	1
Egg quality	2,11,Z	3	2
Egg weight	1,2,3,4(3),14,23,Z	9	3
Feed intake	1,4	2	2
Sexual maturity	Z(2)	2	2

Source: Hocking, 2005.

tious viral disease caused by a member of the herpes virus family and costs the poultry industry about US$1 000 million per annum. An F_2 cross between resistant and susceptible lines was challenged experimentally and genotyped, providing the data for a QTL analysis that resulted in a total of seven QTL for susceptibility to MD (Vallejo *et al.*, 1998; Yonash *et al.*, 1999). Subsequently, the founder lines of the F_2 cross were used for a micro-array study to identify genes that were differentially expressed between the two lines following artificial infection. Fifteen of these genes were mapped onto the chicken genome and two of them mapped to a QTL region for resistance to MD (Liu *et al.*, 2001a). At the same time, protein interaction studies between a viral protein (SORF2) and a chicken splenic cDNA library revealed an interaction with the chicken growth hormone (GH) (Liu *et al.*, 2001b). This led to the detection of a polymorphism in the GH gene that was associated with differences in the number of tumours between the susceptible and the resistant line (Liu *et al.*, 2001b). GH coincided with a QTL for resistance and was differentially expressed between founder lines (Liu *et al.*, 2001a).

Alongside the various genome scans for QTL, a large number of candidate gene studies have been carried out. The majority of studies summarized in Table 2 have been conducted on White Leghorn strains and have utilized restriction fragment length polymorphisms (RFLPs), SNPs or single strand conformation polymorphisms (SSCPs). These techniques require both that the gene is known and that the experimenter is able to sequence part of the gene to detect polymorphisms that distinguish the experimental lines.

Candidate gene studies have been used in two ways. First, candidate genes may be used merely as a marker for a trait (typically disease) based on prior knowledge and, second, and much less often, to search for the mutation within a gene that is associated with phenotypic variation in a trait. Currently, potential (candidate) genes for a QTL may be obtained from a knowledge of physiology (Dunn *et al.*, 2004) or comparative linkage maps (i.e. locating genes that are in the location of the QTL based on common areas of the gene-rich genomes of different species, usually human and mouse). There are likely to be many more of the second type of candidate gene studies as information from large-scale gene expression and proteomic experiments begin to suggest novel gene candidates for traits of commercial and biological importance. It should also be noted that there is good evidence that genetic variation is not limited to genomic DNA: associations between polymorphisms in mitochondrial genes and MD resistance, body weight and egg shell quality were reported by Li *et al.* (1998a, b).

Despite great enthusiasm for breeding companies to be involved in functional genomics research in poultry, there are very few applications of MAS in commercial poultry breeding. One existing example is the use of blood group markers to improve resistance to MD where selection of haplotypes B^{21} and B^{12} based on conventional serological tests has been widely used (McKay, 1998). In discussions with the industry it is clear that most interest is in QTL or candidate genes for resistance to diseases like MD or ascites, a genetic condition associated with pulmonary hypertension, leading to mortality in fast growing birds. There is also considerable interest among breeders of layer lines for egg quality, especially egg shell quality because of its importance for food

TABLE 2
Association of candidate genes with quantitative traits in poultry

Trait	Chromosomes[1]	Gene symbols	References
Age at first egg	1,2,3	GH, NPY, ODC	Feng *et al.,* 1997; Dunn *et al.,* 2004; Parsanejad *et al.,* 2004
Disease resistance (*E. coli*)	16	MHC1, MHC4, TAP2	Yonash *et al.,* 1999
Disease resistance (MD[2])	1,NK	GH, LY6E	Kuhnlein *et al.,* 1997; Liu *et al.,* 2001a,b and 2003
Disease resistance (Sal[3])	4,6,7,16,19, 1,17,NK	TNC, PSAP, NRAMP1[4], MHC1,CASP1, IAP1, TLR4, TLR5	Hu *et al.,* 1997; Lamont *et al.,* 2002; Leveque *et al.* 2003; Liu and Lamont, 2003; Iqbal *et al.,* 2005
Double yolked eggs	10	GNRHR	Dunn *et al.,* 2004
Egg production	Z,1,20	GHR, GH, PEPCK	Feng *et al.,* 1997; Kuhnlein *et al.,* 1997; Parsanejad *et al.,* 2003
Egg weight	1	IGF1	Nagaraja *et al.,* 2000
Eggshell quality	1,3,20	IGF1, ODC, PEPCK	Nagaraja *et al.,* 2000; Parsanejad *et al.,* 2003, 2004
Body fat	1,1,5,Z	GH, IGF1, TGFβ3, GHR	Feng et al., 1998; Fotouhi *et al.,* 1993; Li *et al.,* 2003; Zhou *et al.,* 2005
Feed efficiency	3,20	ODC, PEPCK	Parsanejad et al., 2003 and 2004
Body weight/carcass quality	1,3,5,Z,1,1,1	IGF1, ODC, TGFβ3, GHR, APOA2, PIT1	Feng *et al.,* 1998; Li *et al.,* 2003; Jiang *et al.,* 2004; Parsanejad *et al.,* 2004; Li *et al.,* 2005; Zhou *et al.,* 2005
Organ weight (spleen)	3,5,32	TGFβ2, TGFβ3, TGFβ4[5]	Li *et al.,* 2003
Skeletal traits	1,3,5,32	IGF1, TGFβ2, TGFβ3, TGFβ4[5]	Li *et al.,* 2003; Zhou *et al.,* 2005

[1] NK = gene has not yet been assigned to a chromosome.
[2] Marek's Disease.
[3] Salmonellosis.
[4] Now known as Slc11a1.
[5] TGFβ4 in the paper is now known to be TGFβ1.

safety. For production traits such as growth and egg numbers, breeders make sufficient progress using traditional selection methods, and they expect little improvement from MAS for such traits unless markers can be used to increase the accuracy of selection. Nonetheless, among breeders of broiler stock there is interest in markers for traits that are difficult to measure such as feed efficiency and meat quality in addition to disease resistance.

POTENTIAL FOR MAS IN CHICKEN

The technical aspects and potential implications of implementing MAS in livestock are discussed in Chapter 10 and Dekkers (2004), and van der Beek and van Arendonk (1996) evaluated the technical aspects of MAS in poultry breeding. A review of the potential of MAS in poultry is provided by Muir (2003) but this includes many of the technical issues that are common across livestock species. This chapter therefore focuses on poultry-specific issues, and readers are referred to Chapter 10 or Muir (2003) for a more comprehensive overview of applications and limitations of MAS.

Muir (2003) identified two cases where MAS could increase the selection intensity in poultry breeding: (i) traits that are measured later in life or are costly to measure (such as egg production and feed efficiency for broiler breeders); and (ii) selection within full-sib families for sex-limited traits (e.g. male chicks for egg production). Accuracy of selection can also be improved via MAS when selecting between full-sib families for sex-limited traits and traits that cannot be measured directly on one or both

sexes and/or have a low heritability (e.g. egg production, disease resistance, carcass quality and welfare traits).

Limiting factors for application of MAS (Muir, 2003) include biological factors (reproductive capacity) and many theoretical considerations related to the effectiveness of MAS (e.g. diverting selection pressure from polygenes to a single marked gene), which are generally applicable to MAS in livestock (Dekkers, 2004; Chapter 10). One of the concerns of Muir (2003) is the expected lack of major QTL for traits that have been under selection for many generations (following simulation results). However, recent QTL studies within commercial lines of pigs (Evans *et al.,* 2003; Nagamine *et al.,* 2003) and poultry (de Koning *et al.,* 2003, 2004) have demonstrated that many sizeable QTL are still segregating in commercial populations despite decades of selection.

There is strong academic interest in chicken genomics outside agriculture from, among others, developmental biologists and evolutionary geneticists, and this has contributed greatly to the development of the current functional genomics toolbox available for chicken. Among livestock species, chickens are best placed to pioneer new approaches where QTL studies are complemented by gene expression studies (Liu *et al.,* 2001a) or where they become fully integrated within "genetical genomics" (de Koning, Carlborg and Haley, 2005; de Koning and Haley, 2005).

If poultry breeders decide to embrace MAS, one of the main questions is whether they are prepared to re-structure their breeding programmes around MAS or implement these around their current breeding strategies. Adopting the terminology of Dekkers (2004), there are three levels of MAS: gene-assisted selection (GAS) where the functional mutation and its effects are known; linkage disequilibrium MAS (LD-MAS) where a marker (or marker haplotypes) is in population-wide disequilibrium with a QTL; and linkage equilibrium MAS (LE-MAS) where markers are in Hardy-Weinberg equilibrium with the QTL at the population level, but linkage disequilibrium exists within families. A fourth type of MAS that was recently proposed is "genome-wide MAS" (GW-MAS), where dense markers (i.e. SNPs) across the genome are used to predict the genetic merit of an individual without targeting any individual QTL or measuring (expensive) phenotypes on every generation (Meuwissen, Hayes, and Goddard, 2001). Integrating current evaluations with MAS is most straightforward for GAS and LD-MAS because the QTL effect can be included in routine evaluations as a fixed effect (Chapter 10). LE-MAS, on the other hand, requires extensive genotyping and fairly complicated statistical procedures (Wang, Fernando and Grossman, 1998), while GW-MAS reduces the genome to a "black-box" but does not require selection of QTL using arbitrary thresholds. Furthermore, the dense marker information required for GW-MAS may dispense with often faulty pedigree records because all pedigree information is encoded in the genome-wide genotypes.

In terms of quantitative genetic theory, there are ongoing developments in the tools required to detect and evaluate QTL in arbitrary pedigrees, moving away from strictly additive-dominance models to epistasis and parent-of-origin effects (Liu, Jansen and Lin, 2002; Shete and Amos, 2002). At the same time, the technology to analyse more than 10 000 SNPs in a single assay is available, and a cost of as little as US$0.02 per genotype is likely for chicken

SNPs in the near future. However, the fine mapping and characterization of identified QTL remain costly and time-consuming processes and are often restricted to the most promising QTL, resulting in hundreds of QTL that will never make it past the stage of mapping to a 30 cM confidence interval.

While current research and developments in poultry functional genomics are relevant to all four possible applications of MAS to livestock, poultry breeders need to decide at what level they want to exploit molecular information and for which traits.

The emerging picture is that breeders are more comfortable with known gene mutations as this provides an easy route to implementation as well as knowledge about the underlying biology. Furthermore, there is concern that the marker-trait linkage will break down over a relatively few generations of selection in large commercial flocks. While candidate gene studies would provide the quickest route to implementation, fine mapping and characterization of QTL (e.g. using expression studies) may reveal gene variants that are not obvious candidate genes for quantitative traits.

POTENTIAL FOR MAS IN POULTRY IN DEVELOPING COUNTRIES

Owing to the relatively low value of single animals, the high reproductive rate in poultry and good portability of eggs or day-old hatchlings, the concentration of resources is very high in the poultry breeding industry and all poultry breeding is privately owned. Fifty years ago there were many primary breeders in each and every industrialized country, but not so long ago there were only 20 breeding companies worldwide. Today, three groups of primary breeders dominate the international layer market. Equally, in the chicken meat industry, there are four major players in broiler breeding worldwide (Flock and Preisinger, 2002). The concentration process is probably now complete, and the present players are sufficient to meet the global supply for 700 000 million eggs as final products. A similar trend is expected in the pig industry, where international breeding companies of hybrid products are increasing their market share (Preisinger, 2004). For large-scale farming of broilers and layers in developing countries there are additional challenges with regard to heat stress and potential disease pressure. With increasing poultry production in developing countries, breeding companies may give priority to using breeding and molecular tools to address these additional challenges. While chickens are very efficient in converting grain into valuable meat and egg protein, and smallholder chicken production can be valuable for sustaining the livelihoods of farmers in the developing world, this type of poultry production would require robust dual-purpose (meat and egg) birds, rather than specialized broiler and layer lines. It is unlikely that the commercial breeders will develop such lines but there may be scope for national or international research organizations to do so. Any MAS would have to be done at the institutional level where the line is developed and would necessitate prior knowledge of trait-marker associations at the farm level. The implementation of whole genome SNP approaches to farm level recording might facilitate progress in this area but the challenges, both practical and theoretical, are formidable.

CONCLUDING REMARKS

Among livestock species, chicken have by far the most comprehensive genomic toolbox. However, uptake of MAS will

depend strongly on whether the industry wishes to supplement its current selection programme with a known gene variant or whether it is prepared to restructure breeding programmes around MAS. Compared with, for instance, the dairy cattle industry, the poultry breeding community may be slower to embrace emerging complex approaches to MAS. This is somewhat surprising because the closed structure of the poultry breeding pyramid offers much better protection of intellectual property than the dairy cattle industry where semen from highest ranking bulls is available for all. On the other hand, the fact that blood groups have been used to select for resistance to MD suggests that poultry breeders have some experience and skills in this type of selection. Poultry breeding companies contribute significantly to poultry genomics research but may not be fully convinced about the economic feasibility of MAS. To implement MAS successfully, a company must tackle the problems of identifying the traits to select and their economic significance, the lack of current knowledge of the genes or markers associated with these traits, and their association with other economic selection criteria. The current "toolbox" provides the means to answer some of these questions but there are obvious concerns about human and capital resources and the potential loss of gains in other traits in a competitive market. Coupled with these reservations must be the very evident success of current breeding programmes in achieving many desirable commercial goals.

ACKNOWLEDGEMENTS

The Roslin Institute is supported by a core strategic grant from the Biotechnology and Biological Sciences Research Council (BBSRC). The authors gratefully acknowledge information on breeding objectives from representatives of the poultry breeding industry.

REFERENCES

Arthur, J.A. & Albers, G.A.A. 2003. Industrial perspective on problems and issues associated with poultry breeding. *In* W. M. Muir & S. E. Aggrey, eds. *Poultry genetics, breeding and biotechnology,* pp. 1–12. Wallingford, UK, CABI Publishing.

Bilgili, S.F. 2001. Poultry products and processing in the international market place. *Proc. Internat. Anim. Agric. & Food Sci. Conf*. Indianapolis, IN, USA (available at www.fass.org/fass01/pdfs/Bilgili.pdf).

Bowman, J.C. 1974. *An introduction to animal breeding*. The Institute of Biology's Studies in Biology No 46. London, Edward Arnold.

Bumstead, N. & Palyga, J. 1992. A preliminary linkage map of the chicken genome. *Genomics* 13: 690–697.

Crittenden, L.B., Provencher, L., Sanatangello, L., Levin, I., Abplanalp, H., Briles, R.W., Briles, W.E. & Dodgson, J.B. 1993. Characterization of a Red Jungle Fowl by White Leghorn backcross reference population for molecular mapping of the chicken genome. *Poultry Sci.* 72: 334–348.

de Koning, D.J., Carlborg, O. & Haley. C.S. 2005. The genetic dissection of immune response using gene-expression studies and genome mapping. *Vet. Immun. & Immunopath.* 105: 343–352.

de Koning, D.J. & Haley, C.S. 2005. Genetical genomics in humans and model organisms. *Trends Genetics* 21: 377–381.

de Koning, D.J., Haley, C.S., Windsor, D., Hocking, P.M., Griffin, H., Morris, A., Vincent, J. & Burt, D.W. 2004. Segregation of QTL for production traits in commercial meat-type chickens. *Genetic. Res.* 83: 211–220.

de Koning, D.J., Windsor, D., Hocking, P.M., Burt, D.W., Law, A., Haley, C.S., Morris, A., Vincent, J. & Griffin, H. 2003. Quantitative trait locus detection in commercial broiler lines using candidate regions. *J. Anim. Sci.* 81:1158–1165.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sc.* 82 E-Suppl: E313–E328.

Dunn, I.C., Miao, Y.W., Morris, A., Romanov, M.N., Wilson, P.W. & Waddington, D. 2004. A study of association between genetic markers in candidate genes and reproductive traits in one generation of a commercial broiler breeder hen population. *Heredity* 92: 128–134.

Evans, G.J., Giuffra, E., Sanchez, A., Kerje, S., Davalos, G., Vidal, O., Illan, S., Noguera, J.L., Varona, L., Velander, I., Southwood, O.I., de Koning, D.J., Haley, C.S., Plastow, G.S. & Andersson, L. 2003. Identification of quantitative trait loci for production traits in commercial pig populations. *Genetics* 164: 621–627.

Feng, X.P., Kuhnlein, U., Aggrey, S.E., Gavora, J.S. & Zadworny, D. 1997. Trait association of genetic markers in the growth hormone and the growth hormone receptor gene in a White Leghorn strain. *Poultry Science* 76: 1770–1775.

Feng, X.P., Kuhnlein, U., Fairfull, R.W., Aggrey, S.E., Yao, J. & Zadworny, D. 1998. A genetic marker in the growth hormone receptor gene associated with body weight in chickens. *J. Heredity* 89: 355–359.

Flock, D.K. & Preisinger, R. 2002. Breeding plans for poultry with emphasis on sustainability. *7th World Congress on Genetics Applied to Livestock Production*, 19–23 August, 2002, Montpellier, France, Session 24: Sustainable breeding plans in developed countries. Communication 24-02.

Fotouhi, N., Karatzas, C.N., Kuhnlein, U. & Zadworny, D. 1993. Identification of growth-hormone DNA polymorphisms which respond to divergent selection for abdominal fat-content in chickens. *Theor. Appl. Genet.* 85: 931–936.

Groenen, M.A., Cheng, H.H., Bumstead, N., Benkel, B.F., Briles, W.E., Burke, T., Burt, D.W., Crittenden, L.B., Dodgson, J., Hillel, J., Lamont, S., de Leon, A.P., Soller, M., Takahashi, H. & Vignal, A. 2000. A consensus linkage map of the chicken genome. *Genome Research* 10: 137–147.

Groenen, M.A., Crooijmans, R.P., Veenendaal, A., Cheng, H.H., Siwek, M. & van der Poel, J.J. 1998. A comprehensive microsatellite linkage map of the chicken genome. *Genomics* 49: 265–274.

Hillier, L.W., Miller, W., Birney, E., Warren, W., Hardison, R.C., Ponting, C.P., Bork, P., Burt, D.W., Groenen, M.A., Delany, M.E., Dodgson, .J.B., Chinwalla, A.T., Cliften, P.F., Clifton, S.W., Delehaunty, K.D., Fronick, C., Fulton, R.S., Graves, T.A., Kremitzki, C., Layman, D., Magrini, V., McPherson, J.D., Miner, T.L., Minx, P., Nash, W.E., Nhan, M.N., Nelson, J.O., Oddy, L.G., Pohl, C.S., Randall-Maher, J., Smith, S.M., Wallis, J.W., Yang, S.P., Romanov, M.N., Rondelli, C.M., Paton, B., Smith, J., Morrice, D., Daniels, L., Tempest, H.G., Robertson, L., Masabanda, J.S., Griffin, D.K., Vignal, A., Fillon, V., Jacobbson, L., Kerje, S., Andersson, L., Crooijmans, R.P., Aerts, J., van der Poel, J.J., Ellegren, H.. Caldwell, R.B., Hubbard, S.J., Grafham, D.V., Kierzek, A.M., McLaren, S.R., Overton, I.M., Arakawa, H., Beattie, K.J., Bezzubov, Y., Boardman, P.E., Bonfield, J.K., Croning, M.D., Davies, R.M., Francis, M.D., Humphray, S.J., Scott, C.E., Taylor, R.G., Tickle, C., Brown, W.R., Rogers, J., Buerstedde, J.M., Wilson, S.A., Stubbs, L., Ovcharenko, I., Gordon, L., Lucas, S., Miller, M.M., Inoko, H., Shiina, T., Kaufman, J., Salomonsen, J., Skjoedt, K., Wong, G.K., Wang, J., Liu, B., Wang, J., Yu, J., Yang, H., Nefedov, M., Koriabine, M., Dejong, P.J., Goodstadt, L., Webber, C., Dickens, N.J., Letunic, I., Suyama, M., Torrents, D., Von, M.C., Zdobnov, E.M., Makova, K., Nekrutenko, A.,

Elnitski, L., Eswara, P., King, D.C., Yang, S., Tyekucheva, S., Radakrishnan, A., Harris, R.S., Chiaromonte, F., Taylor, J., He, J., Rijnkels, M., Griffiths-Jones, S., Ureta-Vidal A., Hoffman, M.M., Severin, J., Searle, S.M., Law, A.S., Speed, D., Waddington, D., Cheng, Z., Tuzun, E., Eichler, E., Bao, Z., Flicek, P., Shteynberg, D.D., Brent, M.R., Bye, J.M., Huckle, E.J., Chatterji, S., Dewey, C., Pachter, L., Kouranov, A., Mourelatos, Z., Hatzigeorgiou, A.G., Paterson, A.H., Ivarie, R., Brandstrom, M., Axelsson, E., Backstrom, N., Berlin, S.,Webster, M.T., Pourquie, O., Reymond, A., Ucla, C., Antonarakis, S.E., Long, M.J.J., Emerson, J.J Betran, E., Dupanloup, I., Kaessmann, H., Hinrichs, A.S., Bejerano, G., Furey, T.S., Harte, R.A., Raney, B., Siepel, A., Kent, W.J., Haussler, D., Eyras, E., Castelo, R., Abril, J.F., Castellano, S., Camara, F., Parra, G., Guigo, R., Bourque, G., Tesler, G., Pevzner, P.A., Smit, A., Fulton, L.A., Mardis, E.R. & Wilson, R.K. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432: 695–716.

Hocking, P.M. 2005. Review of QTL results in chicken. *Wld. Poultry Sci. J.* 61: 215–226.

Hu, J.X., Bumstead, N., Barrow, P., Sebastein, G., Olien, L., Morgan, K. & Malo, D. 1997. Resistance to salmonellosis in the chicken is linked to NRAMP1 and TNC. *Genome Res.* 7: 693–704.

Iqbal, M., Philbin, V.J., Withanage, G.S.K., Wigley, P., Beal, R.K., Goodchild, M.J., Barrow, P., McConnell, I., Maskell, D.J., Young, J., Bumstead, N., Boyd, Y. & Smith A.L. 2005. Identification and functional characterization of chicken Toll-like receptor 5 reveals a fundamental role in the biology of infection with *Salmonella enterica* serovar *typhimurium*. *Infect. & Immunity* 73: 2344–2350.

Jiang, R., Li, J.,Qu, Li, H. & Yang, N. 2004. A new single nucleotide polymorphism in the chicken pituitary- specific transcription factor (POULF1) gene associated with growth rate. *Anim. Genet.* 35: 344–346.

Kuhnlein, U., Ni, L., Weigend, S., Gavora, J.S., Fairfull, W. & Zadworny, D. 1997. DNA polymorphisms in the chicken growth hormone gene: response to selection for disease resistance and association with egg production. *Anim. Genet.* 28: 116–123.

Lamont, S.J., Kaiser, M.G. & Liu, W. 2002. Candidate genes for resistance to *Salmonella enteritidis* colonization in chickens as detected in a novel genetic cross. *Vet. Immun. & Immunopath.* 87: 423–428.

Leveque, G., Forgetta, V., Morroll, S., Smith, A.L., Bumstead, N., Barrow, P., Loredo-Osti, J.C. , Morgan, K. & Malo, D. 2003. Allelic variation in TLR4 is linked to susceptibility to *Salmonella enterica* serovar *typhimurium* infection in chickens. *Infect. & Immunity* 71: 1116–1124.

Li, S., Aggrey, S.E., Zadworny, D., Fairfull, W. & Kuhnlein, U. 1998a. Evidence for a genetic variation in the mitochondrial genome affecting traits in White Leghorn chickens. *J. Hered.* 89: 222–226.

Li, S., Zadworny, D., Aggrey, S.E. & Kuhnlein, U. 1998b. Mitochondrial pepck: A highly polymorphic gene with alleles co- selected with Marek's disease resistance in chickens. *Anim. Genet.* 29: 395–397.

Li, H., Deeb, N., Zhou, H., Mitchell, A.D., Ashwell, C.M. & Lamont, S.J. 2003. Chicken quantitative trait loci for growth and body composition associated with transforming growth factor-beta genes. *Poult. Sci.* 82: 347–356.

Li, H., Deeb, N., Zhou, H., Ashwell, C.M. & Lamont, S.J. 2005. Chicken quantitative trait loci for growth and body composition associated with the very low density apolipoprotein-II gene. *Poult. Sci.* 84: 697–703.

Liu, W. & Lamont, S.J. 2003. Candidate gene approach: Potential association of caspase-1, inhibitor of apoptosis protein-1, and Prosaposin gene polymorphisms with response to *Salmonella enteritidis* challenge or vaccination in young chicks. *Anim. Biotech.* 14: 61–76.

Liu, H.C., Cheng, H.H., Tirunagaru, V., Sofer, L. & Burnside, J. 2001a. A strategy to identify positional candidate genes conferring Marek's disease resistance by integrating DNA microarrays and genetic mapping. *Anim. Genet.* 32: 351–359.

Liu, H.C., Kung, H.J., Fulton, J.E., Morgan, R.W. & Cheng. H.H. 2001b. Growth hormone interacts with the Marek's disease virus SORF2 protein and is associated with disease resistance in chicken. *Proc. Natl. Acad. Sci. USA* 98: 9203–9208.

Liu, H.C., Niikura, M., Fulton, J.E. & Cheng. H.H. 2003. Identification of chicken lymphocyte antigen 6 complex, locus E (LY6E, alias SCA2) as a putative Marek's disease resistance gene via a virus-host protein interaction screen. *Cytogenet. & Genome Res.* 102: 304–308.

Liu, Y., Jansen, G.B. &. Lin, C.Y. 2002. The covariance between relatives conditional on genetic markers. *Genet. Select. Evoln.* 34: 657–678.

McKay, J.C. 1998. A poultry breeder's approach to avian neoplasia. *Avian Pathology* 27: S74–S77.

Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.

Muir, W.M. 2003. Incorporating molecular information in breeding programmes: applications and limitations. *In* W.M. Muir & S.E. Aggrey, eds. *Poultry Genetics, Breeding and Biotechnology*, pp. 549–562. Wallingford, UK, CABI Publishing.

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: 629–635.

Nagaraja, S.C., Aggrey, S.E., Yao, J., Zadworny, D., Fairfull R.W. & Kuhnlein, U. 2000. Trait association of a genetic marker near the IGF-I gene in egg-laying chickens. J. *Heredity* 91: 150–156.

Parsanejad, R., Torkamanzehi, A., Zadworny, D. & Kuhnlein, U. 2003. Alleles of cytosolic phosphoenolpyruvate carboxykinase (PEPCK): trait association and interaction with mitochondrial PEPCK in a strain of white leghorn chickens. *Poultry Sci.* 82: 1708–1715.

Parsanejad, R., Praslickova, D., Zadworny, D. & Kuhnlein, U. 2004. Ornithine decarboxylase: haplotype structure and trait associations in White Leghorn chickens. *Poultry Sci.* 83: 1518–1523.

Preisinger, R. 2004. Internationale Tendenzen der Tierzüchtung und die Rolle der Zuchtunternehmen. *Züchtungskunde* 76(6): 395–402.

Schmid, M., Nanda, I., Guttenbach, M., Steinlein, C., Hoehn, M., Schartl, M., Haaf, T., Weigend, S., Fries, R., Buerstedde, J.M., Wimmers, K., Burt, D.W., Smith, J., A'Hara, S., Law, A., Griffin, D.K., Bumstead, N., Kaufman, J., Thomson, P.A., Burke, T., Groenen, M.A., Crooijmans, R.P., Vignal, A., Fillon, V., Morisson, M., Pitel, F.,Tixier-Boichard, M.T., Ladjali-Mohammedi, K., Hillel, J., Maki-Tanila, A., Cheng, H.H., Delany, M.E., Burnside, J. & Mizuno, S. 2000. First report on chicken genes and chromosomes 2000. *Cytogenet. & Cell Genet.* 90: 169–218.

Schmid, M., Nanda, I., Hoehn, H., Schartl, M., Haaf, T., Buerstedde, J.M., Arakawa, H., Caldwell, R.B., Weigend, S., Burt, D.W., Smith, J., Griffin, D.K., Masabanda, J.S., Groenen, M.A., Crooijmans, R.P., Vignal, A., Fillon, V., Morisson, M., Pitel, F., Vignoles, M., Garrigues, A., Gellin, J., Rodionov, A.V., Galkina, S.A., Lukina, N.A., Ben-Ari, G., Blum, S., Hillel, J., Twito, T., Lavi, U., David, L., Feldman, M.W., Delany, M.E., Conley, C.A., Fowler, V.M., Hedges, S.B., Godbout, R., Katyal, S., Smith, C., Hudson, Q., Sinclair, A. & Mizuno, S. 2005. Second report on chicken genes and chromosomes 2005. *Cytogen.& Genome Res.* 109: 415–479.

Shete, S. & Amos, C.I. 2002. Testing for genetic linkage in families by a variance components approach in the presence of genomic imprinting. *Am. J. Human Genet.* 70: 751–757.

Taha, F.A. 2003. *The poultry sector in middle income countries and its feed requirements: the case of Egypt.* Outlook Report No. WRS03-02, pp. 3–14. Washington, DC, Economic Research Service, USDA.

Vallejo, R.L., Bacon, L.D., Liu, H.C., Witter, R.L., Groenen, M.A., Hillel, J. & Cheng, H.H. 1998. Genetic mapping of quantitative trait loci affecting susceptibility to Marek's disease virus induced tumors in F2 intercross chickens. *Genetics* 148: 349–360.

van der Beek, S. &. van Arendonk, J.A.M. 1996. Marker-assisted selection in an outbred poultry nucleus. *Anim. Sc.* 62: 171–180.

Wang, T., Fernando, R.L. & Grossman, M. 1998. Genetic evaluation by best linear unbiased prediction using marker and trait information in a multibreed population. *Genetics* 148: 507–515.

Wang, J., He, X., Ruan, J., Dai, M., Chen, J., Zhang, Y., Hu, Y., Ye, C., Li, S., Cong, L., Fang, L., Liu, B., Li, S., Wang, J., Burt, D.W., Wong, G.K., Yu, J., Yang, H. & Wang, J. 2005. ChickVD: a sequence variation database for the chicken genome. *Nucl. Acid Res.* 33: D438–D441.

Wong, G.K., Liu, B., Wang, J., Zhang, Y., Yang, X., Zhang, Z., Meng, Q., Zhou, J., Li, D., Zhang, J., Ni, P., Li, S., Ran, L., Li, H., Zhang, J., Li, R., Li, S., Zheng, H., Lin, W., Li, G., Wang, X., Zhao, W., Li, J., Ye, C., Dai, M., Ruane, J., Zhou, Y., Li, Y., He, X., Zhang, Y., Wang, J., Huang, X., Tong, W., Chen, J., Ye, J., Chen, C., Wei, N., Li, G., Dong, L., Lan, F., Sun, Y., Zhang, Z., Yang, Z., Yu, Y., Huang, Y., He, D., Xi, Y., Wei, D., Qi, Q., Li, W., Shi, J., Wang, M., Xie, F., Wang, J., Zhang, X., Wang, P., Zhao, Y., Li, N., Yang, N., Dong, W., Hu, S., Zeng, C., Zheng, W., Hao, B., Hillier, L.W., Yang, S.P., Warren, W.C., Wilson, R.K., Brandstrom, M., Ellegren, H., Crooijmans, R.P., van der Poel, J.J., Bovenhuis, H., Groenen, M.A., Ovcharenko, I., Gordon, L., Stubbs, L., Lucas, S., Glavina, T., Aerts, A., Kaiser, P., Rothwell, L., Young, J.R., Rogers, S., Walker, B.A., van, H.A., Kaufman, J., Bumstead, N., Lamont, S.J., Zhou, H., Hocking, P.M., Morrice, D., de Koning, D.J., Law, A., Bartley, N., Burt, D.W., Hunt, H., Cheng, H.H., Gunnarsson, U., Wahlberg, P., Andersson, L., Kindlund, E., Tammi, M.T., Andersson, B., Webber, C., Ponting, C.P., Overton, I.M., Boardman, P.E., Tang, H., Hubbard, S.J., Wilson, S.A., Yu, J., Wang, J. & Yang, H. 2004. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* 432: 717–722.

Yonash, N., Bacon, L.D. Witter, R.L. & Cheng. H.H. 1999. High resolution mapping and identification of new quantitative trait loci (QTL) affecting susceptibility to Marek's disease. *Anim. Genet.* 30: 126–135.

Zhou, H., Mitchell, A.D., McMurtry, J.P., Ashwell, C.M. & Lamont, S.J. 2005. Insulin-like growth factor-1 gene polymorphism associations with growth, body composition, skeleton integrity, and metabolic traits in chickens. *Poult. Sci.* 84: 212–219.

CHAPTER 12

Marker-assisted selection in dairy cattle

Joel Ira Weller

SUMMARY

Considering the long generation interval, the high value of each individual, the very limited female fertility and the fact that nearly all economic traits are expressed only in females, it would seem that cattle should be a nearly ideal species for application of marker-assisted selection (MAS). As genetic gains are cumulative and eternal, application of new technologies that increase rates of genetic gain can be profitable even if the nominal annual costs are several times the value of the nominal additional annual genetic gain. Complete genome scans for quantitative trait loci (QTL) based on the granddaughter design have been completed for most commercial dairy cattle populations, and significant across-study effects for economic traits have been found on chromosomes 1, 3, 6, 9, 10, 14 and 20. Quantitative trait loci associated with trypanotolerance have been detected in a cross between the African N'Dama and the Boran breeds as the first step in the introgression of these genes into breeds susceptible to trypanosomosis. In dairy cattle, the actual DNA polymorphism has been determined twice, for QTL on BTA 6 and BTA 14. In both cases the polymorphism caused a non-conservative amino acid change, and both QTL chiefly affect fat and protein concentration. Most theoretical studies have estimated the expected gains that can be obtained by MAS to be in the range of a 5 to 20 percent increase in the rates of genetic gain obtained by traditional selection programmes. Applied MAS programmes have commenced for French and German Holsteins. In both programmes genetic evaluations including QTL effects are computed by variants of marker-assisted best linear unbiased prediction (MA-BLUP).

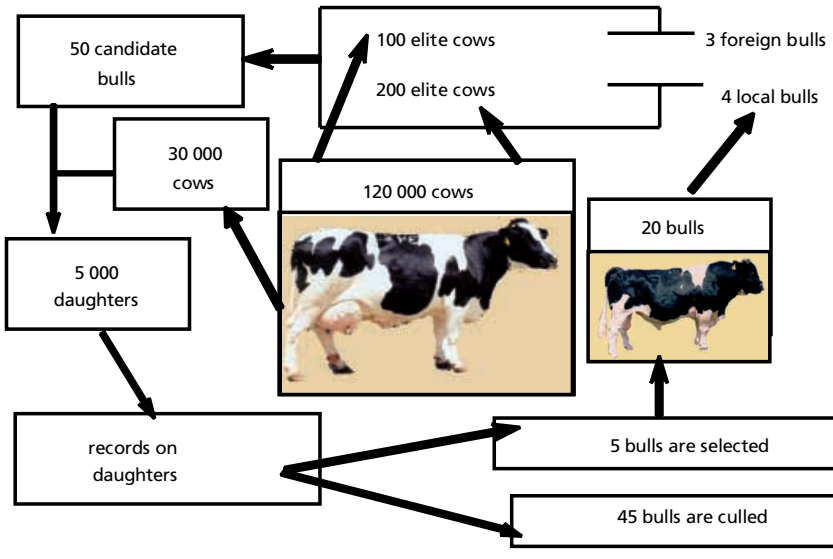
INTRODUCTION

Compared with other agricultural species, dairy cattle are unique in terms of the value of each animal, their long generation interval and the very limited fertility of females. Thus unlike plant and poultry breeding, most dairy cattle breeding programmes are based on selection within the commercial population. Similarly, detection of quantitative trait loci (QTL) and marker-assisted selection (MAS) programmes are generally based on analysis of existing populations. The specific requirements of dairy cattle breeding have led to the generation of very large data banks in most developed countries, which are available for analysis. In this chapter, dairy cattle breeding programmes in the developed and developing countries are reviewed and compared. The important issues in the application of MAS are then outlined. These include economic considerations based on phenotypic selection, the current status of cattle marker maps, methods to detect QTL and to estimate QTL effects and location suitable for dairy cattle, the current state of QTL detection in dairy cattle, methods to incorporate information from genetic markers in genetic evaluation systems, methods to identify the actual polymorphisms responsible for observed QTL and description of the reported results, methods and theory for MAS in dairy cattle, the current status of MAS and, finally, the future prospects for MAS in dairy cattle.

DAIRY CATTLE BREEDING PROGRAMMES IN DEVELOPED COUNTRIES

In most developed countries, dairy cattle breeding programmes are based on the "progeny test" (PT) design. The PT is the design of choice for moderate to large dairy cattle populations, including the United States Holsteins, which include over ten million animals. An example of the Israeli PT design is given in Figure 1.

FIGURE 1
The Israeli Holstein breeding programme



This population consists of approximately 120 000 cows of which 90 percent are milk recorded. Approximately 20 bulls are used for general service. Each year about 300 elite cows are selected as bull dams. These are mated to the two to four best local bulls and an equal number of foreign bulls to produce approximately 50 bull calves for progeny testing. At the age of one year, the bull calves reach sexual maturity, and approximately 1 000 semen samples are collected from each young bull. These bulls are mated to approximately 30 000 first parity cows to produce about 5 000 daughters, or 100 daughters per young bull. Gestation length for cattle is nine months. Thus the young bulls are approximately two years old when their daughters are born, and are close to four when their daughters calve and begin their first lactation. At the completion of their daughters' first lactations, most of the young bulls are culled. Only four to five are returned to general service, and a similar number of the old proven sires are culled. By this time the selected bulls are approximately five years old.

Various studies have shown that rates of genetic gain by a PT scheme are about 0.1 to 0.2 genetic standard deviations of the selection index per year (Nicholas and Smith, 1983; Israel and Weller, 2000). The PT was devised to take advantage of the nearly unlimited fertility of males. However, compared with breeding schemes for other species, the PT has several major weaknesses. First, for a PT system to be effective, the population must include at least several tens of thousands of animals with recording on production traits and paternity. Inaccurate recording can significantly reduce rates of genetic gain (Israel and Weller, 2000). Second, generation intervals, especially along the sire-to-dam and sire-to-sire paths, are much longer than the biological requirements. The increase in generation interval reduces genetic gain per year. As artificial insemination (AI) institutes generally pay a premium price for male calves of elite cows, these cows are often given preferential treatment in order to increase their genetic evaluations (Powell and Norman, 1988). The small number of bulls actually used for general service, and the even smaller number of bulls used as bull sires, tends to reduce the effective population size, which increases inbreeding and decreases genetic variance in the population. The effective population size of the United States Holstein population with ten million cows has been estimated at about 100 (Farnir *et al.*, 2000). Finally, there is virtually no selection along the dam-to-dam path. Generally, 70-80 percent of healthy female calves produced are used as replacements.

Various studies have suggested that selection intensities along the dam-to-dam path could be increased by application of multiple ovulation and embryo transfer (MOET) and sexed semen. Costs of both technologies are still prohibitively high to be applied to the entire population, as shown below. To overcome this problem for MOET, Nicholas and Smith (1983) proposed a "nucleus" breeding scheme. In nucleus schemes, the selection population consists of several hundred individuals, and bulls are not progeny tested. Instead, bulls are selected based on the genetic evaluations of their dams and sisters, which shortens the generation interval on the sire-to-dam and sire-to-sire paths, but reduces the reliabilities of the genetic evaluations. Dams of bulls and cows are selected based chiefly on their own production records, and MOET is applied to increase the number of progeny per dam. As the selection population consists of only several hundred individuals, MOET costs

are manageable if costs are spread over the entire national dairy industry. Rates of genetic gain within the nucleus are thus higher than can be obtained by a national PT design. This gain is transferred to the general population through the use of bulls from the nucleus population. In addition to the greater overall rate of genetic gain, the nucleus scheme has the advantage that it is necessary to collect data on a much smaller population, which should reduce costs and increase accuracy. The disadvantages of MOET are that overall costs and rates of increase of inbreeding will be greater unless steps are taken to reduce inbreeding. However, these steps will also slightly decrease rates of genetic gain. In practice, no country has replaced its standard PT scheme with a nucleus breeding programme.

DAIRY CATTLE BREEDING IN DEVELOPING COUNTRIES

The genus *Bos* includes five to seven species, of which *Bos taurus* and *Bos indicus* are the most widespread and economically important. *B. taurus* is the main dairy cattle species, and is found generally in temperate climates. Several tropical and subtropical cattle breeds are the result of crosses between B. *taurus* and B. *indicus*, which interbreed freely. In the tropics, cows need at least some degree of tolerance to environmental stress due to poor nutrition, heat and disease challenge to sustain relatively high production levels (Cunningham, 1989). Tropical breeds are adapted to these stresses but have low milk yield, whereas more productive temperate breeds cannot withstand the harsh tropical conditions, to the point of not being able to sustain their numbers (de Vaccaro, 1990). Furthermore, most tropical countries are developing countries, which lack systematic large-scale milk and pedigree recording.

A number of studies have been conducted on crosses between imported and local breeds in the tropics. Generally, the F_1 *B. taurus* x *B. indicus* crosses are economically superior to either of the purebred strains (FAO, 1987). The heterosis effect of the F_1 cross is due to genes for disease resistance from the local parent, and genes for milk production from the imported strain (Smith, 1988; Cunningham, 1989). However, this heterosis is lost in future generations if the F_1 is backcrossed to either parental strain. Madalena (1993) presented an F_1 continuous replacement scheme to capitalize on its superiority. Recently, Kosgey, Kahi and van Arendonk (2005) proposed a closed adult nucleus MOET scheme to increase milk production in tropical crossbred cattle.

ECONOMIC CONSIDERATIONS IN APPLYING MAS TO DAIRY CATTLE

For any new technique to be economically viable, overall gains must be greater than overall costs. This also applies to using MAS within a dairy cattle breeding programme. However, unlike investment in new equipment, genetic gains never "wear out", i.e. breeding is unique in that genetic gains are cumulative and eternal. Thus, as shown by Weller (1994, 2001) investments in MAS or other techniques that enhance breeding programmes are economically viable even if "nominal" costs are greater than "nominal" gains.

For example, consider an ongoing breeding programme with a constant rate of genetic gain per year. Assume that the annual rate of genetic gain has a nominal economical value of V. The cumulative discounted returns to year T, R_v, will be a function of the nominal annual returns, the discount rate, d, the profit horizon, T, and the number of years from the beginning of the programme until

first returns are realized, t. R_v is computed as follows (Hill, 1971):

$$R_v = V\frac{r_d^t - r_d^{T+1}}{(1-r_d)^2} - \frac{(T-t+1)r_d^{T+1}}{1-r_d} \quad \{1\}$$

where r_d= 1/(1+d). For example, with d = 0.08, T = 20 years, and t = 5 years, R_v = 32.58V. That is, the cumulative returns are equal to nearly 33 times the nominal annual returns. For an infinite profit horizon, Equation {1} reduces to:

$$R_v = \frac{Vr^t}{(1-r_d)^2} = \frac{V}{d^2(1+d)^{t-2}} \quad \{2\}$$

and R_v = 124.04V.

The value of nominal annual genetic gain will now be compared with the annual costs of a breeding programme, assuming a fixed nominal cost per year. Costs, unlike genetic gain, only have an effect in the year they occur. Assuming that annual costs are equal during the length of the breeding programme, and that first costs occur in the year after the base year, C_T, the net present value of the total costs of the breeding programme is computed as follows:

$$C_T = \frac{C_c r_d (1-r_d^T)}{1-r_d} \quad \{3\}$$

where C_c = annual costs of the breeding programme. Using the same values for T and d, C_T = 9.82C_c. Thus, with a profit horizon of 20 years, cumulative profit is positive if V > 0.31C_c. For an infinite profit horizon, C_T = 12.5C_c, and profit will be positive if V > 0.1C_c.

Therefore, a breeding programme can be profitable even if the nominal annual costs are several times the value of the nominal annual genetic gain. For example, consider the United States of America dairy cattle population, which consists of about ten million cows. Annual genetic gain is about 100 kg milk per year. The value of a 1 kg gain in milk production has been estimated at US$0.1 (Weller, 1994). Thus, the nominal annual value of a 10 percent increase in the rate of genetic gain (10 kg per year) is:

V = (10 kg per cow per year)
(US$0.1 per kg)(10 000 000 cows) = {4}
US$10 000 000 per year

The cumulative value with a profit horizon of 20 years and an 8 percent discount rate would be US$326 million, and break-even annual costs for a technology that increases annual genetic gain by 10 percent are US$32 million per year. Thus, it would be profitable to spend quite a lot for a relatively small genetic gain.

The value of genetic gain to a specific breeding enterprise will generally be less than the gain to the general economy. This is because most of the gains obtained by breeding will be passed on to the consumers. Brascamp, van Arendonk and Groen (1993) considered the economic value of MAS based on changes in returns from semen sales for a breeding organization operating in a competitive market. In this case, a breeding firm that adopts a MAS programme can increase its returns either by increasing its market share or increasing the mean price of a semen dose. Although the value of genetic gain will be less, relatively small changes in genetic merit can result in large changes in market share.

CURRENT STATUS OF MARKER MAPS IN CATTLE

Cattle have 29 pairs of autosomes and one pair of sex chromosomes. All the autosomes are acrocentric, and map units are

scored from the centromere. Chromosomes are denoted with the prefix "BTA" (*B. taurus*). Similar to other mammals, the bovine DNA includes 3×10^9 base pairs (bp), and the map length is approximately 3 000 cM. The human genome is estimated to encode 20 000–25 000 protein-coding genes (International Human Genome Sequencing Consortium, 2004), and it can be assumed that the number of genes in other mammals, including cattle, should be quite similar. Thus, a single map unit, on average, includes approximately eight genes and one million bp.

As in other animal species, microsatellites are still the marker of choice for map construction due to their prevalence and high polymorphism. Although single nucleotide polymorphisms (SNPs) are much more prevalent, genetic maps based on SNPs are still in the future. More than 50 000 SNPs have been identified in humans, but only several thousand have been validated in cattle (www.afns.ualberta.ca/Hosted/Bovine%20Genomics/), and rates of polymorphism are generally unknown. With the completion of the six-fold coverage of the bovine genome by the Bovine Genome Sequencing Project at Baylor College of Medicine (www.hgsc.bcm.tmc.edu/projects/bovine/) many more SNPs will be identified.

Several genetic maps are available on the internet. The United States Meat Animal Research Center (MARC) (www.marc.usda.gov/) includes thousands of markers, chiefly microsatellites. The ArkDB database system, hosted at Roslin Institute, includes data from several published maps (www.thearkdb.org/). The Commonwealth Scientific and Industrial Research Organization (CSIRO) livestock industries cattle genome marker map is built upon data provided by the University of Sydney's comparative location database (www.livestockgenomics.csiro.au/perl/gbrowse.cgi/cattlemap/). This map combined all publicly-available maps into a single integrated map that currently includes 9 400 markers.

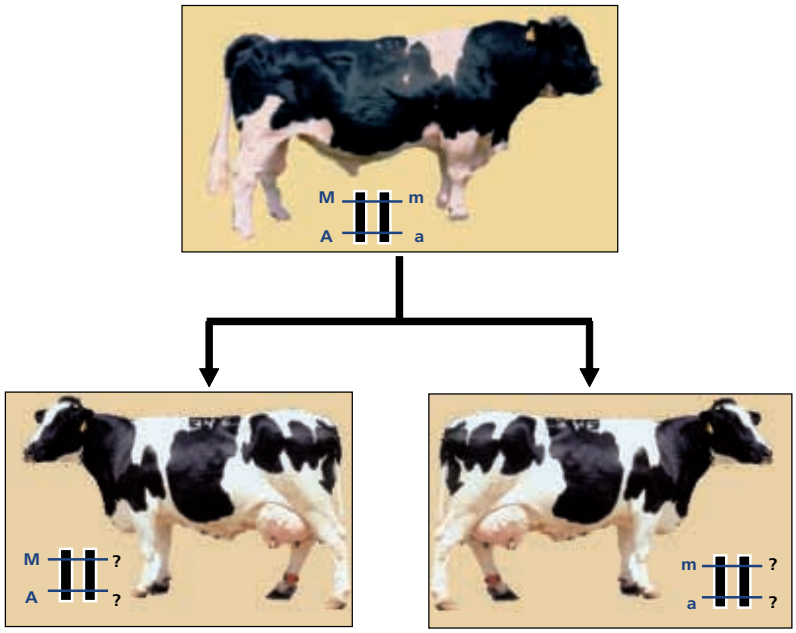
METHODS OF QTL DETECTION SUITABLE FOR COMMERCIAL DAIRY CATTLE POPULATIONS

Detection of QTL requires generation of linkage disequilibrium (LD) between the genetic markers and QTL. In plants, this is generally accomplished by crosses between inbred lines but, for the reasons noted in the introduction, this is not a viable option for dairy cattle in developed countries, in which all analyses must be based on analysis of the existing population. Detection of QTL in developing countries is considered below. For advanced commercial populations, the "daughter" and "granddaughter" designs, which make use of the existence of large half-sib families, are most appropriate for QTL analysis (Weller, Kashi and Soller, 1990). These designs are presented in Figures 2 and 3.

Both designs are similar to the backcross design for crosses between inbred lines in that only the alleles of one parent are followed in the progeny. Thus, similar to the backcross design, dominance cannot be estimated. These designs differ from crosses between inbred lines in that several families are analysed in which the linkage phase between QTL and genetic markers may differ. In addition, any specific QTL will be heterozygous in only a fraction of the families included in the analysis. Thus, QTL effects must be estimated within families, and these designs are therefore less powerful per individual genotyped than designs based on crosses between inbred lines.

The granddaughter design has the advantage of greater statistical power per

FIGURE 2
The daughter design



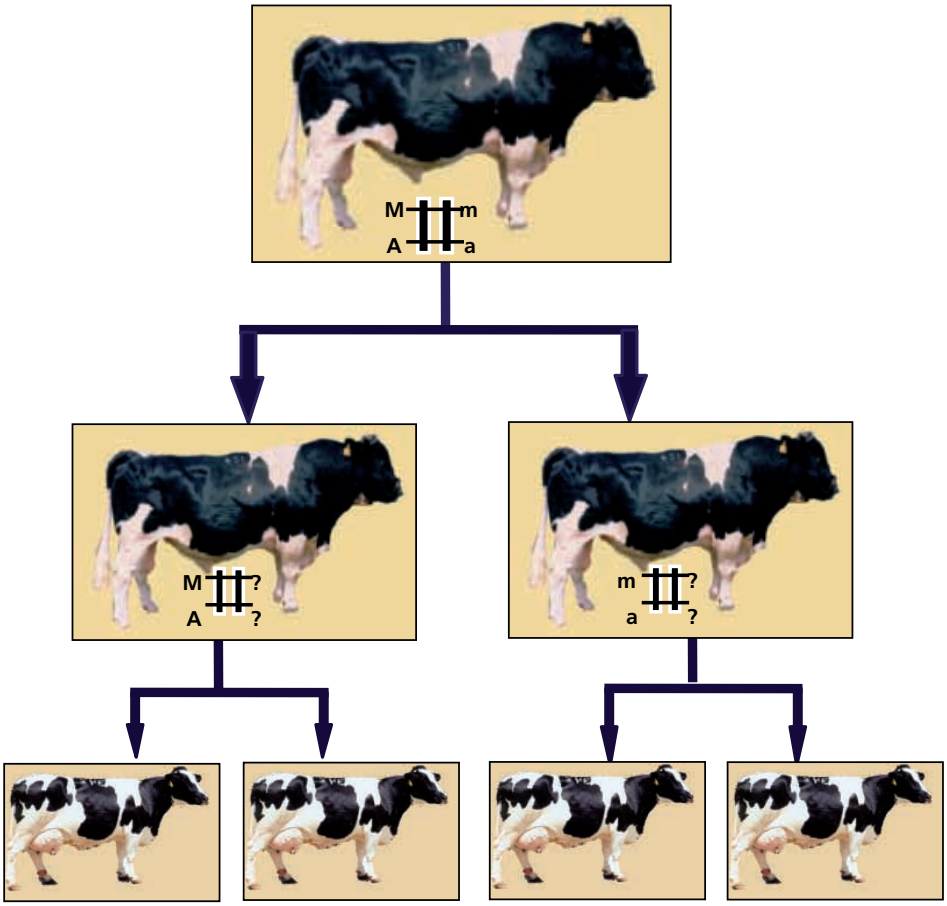
Only a single family is shown, although in practice several families will be analysed jointly. The sire is assumed to be heterozygous for a QTL and a linked genetic marker. The two alleles of the marker locus are denoted "M" and "m", and the two alleles of the QTL are denoted "A" and "a". Alleles of maternal origin are denoted by question marks.

individual genotyped. As each genotype is associated with multiple phenotypic records, the power per individual genotyped in the granddaughter design can be four-fold the power of the daughter design (Weller, Kashi and Soller, 1990). The disadvantage of this design is that the appropriate data structure (hundreds of progeny tested bulls, sons of a limited number of sires) is found only in the largest dairy cattle populations. Both daughter and granddaughter designs are less powerful per individual genotyped than designs based on analysis of inbred lines. Furthermore, the half-sib designs have the disadvantage that progeny with the same genotype as the sire are uninformative, because the progeny could have received either paternal allele.

Additional experimental designs have also been proposed. Coppieters *et al.* (1999) proposed the "great-granddaughter design". One of the disadvantages of the granddaughter design is that the number of progeny-tested sons of most sires is too low to obtain reasonable power to detect QTL of moderate effects. Coppieters *et al.* (1999) proposed that power can be increased by also genotyping progeny-tested grandsons of the grandsire. Inclusion of the grandsons is complicated by the fact that there is another generation of meiosis between the grandsire and his grandson.

A significant drawback of all the designs considered above is that they give no indication of the number of QTL alleles segregating in the population or their rela-

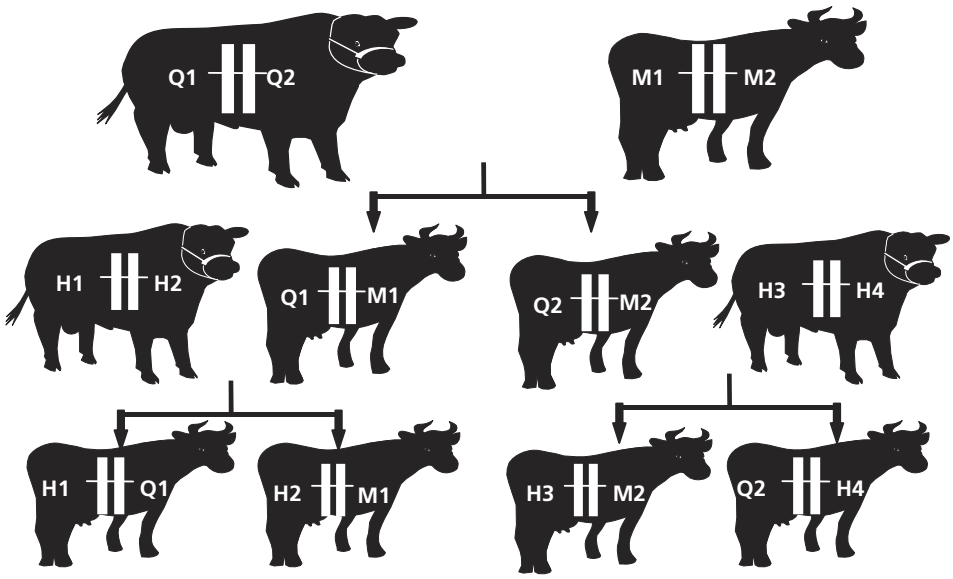
FIGURE 3
The granddaughter design



The grandsire is assumed to be heterozygous for a QTL and a linked genetic marker. As in Figure 2, only a single family is shown. The two alleles of the marker locus are denote "M" and "m", and the two alleles of the QTL are denoted "A" and "a". Alleles of maternal origin are denoted by question marks. Genotypes are not listed for the granddaughters because they were not genotyped.

tive frequencies. To answer this question, Weller *et al*. (2002) proposed the "modified granddaughter design" presented in Figure 4. Assume that a segregating QTL for a trait of interest has been detected and mapped to a short chromosomal segment using either a daughter or a granddaughter design. Consider the maternal granddaughters of a grandsire with a significant contrast between his two paternal alleles. This grandsire will be denoted the "heterozygous grandsire". Each maternal granddaughter will receive one allele from her sire, who is assumed to be unrelated to the heterozygous grandsire, and one allele from her dam, who is a daughter of the heterozygous grandsire. Of these granddaughters, one-quarter should receive the grandpaternal QTL allele with the positive effect, one-quarter should receive the

FIGURE 4
The modified granddaughter design



Only alleles for the QTL are shown. Alleles originating in the heterozygous grandsire are termed "Q1" and Q2". Alleles originating in the grand-dams are termed "M1" and "M2". Alleles originating in the sires are termed "H1", "H2", "H3" and "H4".

negative grandpaternal QTL allele, and half should receive neither grandpaternal allele. In the third case, the granddaughter received one of the QTL alleles of her grand-dam, the mate of the heterozygous grandsire. These grand-dams can be considered a random sample of the general population with respect to the allelic distribution of the QTL. All genetic and environmental effects not linked to the chromosomal segment in question are assumed to be randomly distributed among the granddaughters, or are included in the analysis model. Thus, unlike the daughter or granddaughter designs, it is possible to compare the effects of the two grandpaternal alleles with the mean QTL population effect.

Assuming that the QTL is "functionally biallelic" (i.e. there are only two alleles with differential expression relative to the quantitative trait), and that allele origin can be determined in the granddaughters, the relative frequencies of the two QTL alleles in the population can be determined by comparing the mean values of the three groups of granddaughters for the quantitative trait. Using the modified granddaughter design it is also possible to estimate the number of alleles segregating in the population, and to determine if the same alleles are segregating in different cattle populations. Weller *et al.* (2002) estimated the frequency of the QTL allele that increases fat and protein concentration on BTA6 in the Israeli Holstein population as 0.69 and 0.63, relative to fat and protein percent, by the modified granddaughter design. This corresponded closely to the frequency of 0.69 estimated for the Y581 allele of the ABCG2 gene for cows born during the same time period (Cohen-Zinder *et al.*, 2005).

METHODS TO ESTIMATE QTL EFFECTS AND LOCATION IN DAIRY CATTLE

If a significant effect on a quantitative trait is associated with a genetic marker, the difference between the means of marker genotype classes will be a biased estimate of the QTL effect due to recombination between the QTL and the genetic marker. Weller (1986) first demonstrated that maximum likelihood (ML) methodology could be used to obtain estimates of QTL location and effect unbiased by recombination, while Lander and Botstein (1989) proposed interval mapping, based on ML for a QTL bracketed between two markers. Haley and Knott (1992) and Martinez and Curnow (1992) proposed an interval mapping method based on non-linear regression, which was easier to apply than ML. Their methods are not directly applicable to half-sib designs because, as noted previously, linkage relationships between the QTL and the genetic markers will be different across families, and in some families the common ancestor will be homozygous for the QTL. Furthermore, if multiple QTL alleles are segregating in the population, or if the observed effect is due to several tightly linked QTL, the magnitude of the effect will also differ across families.

A method suitable for interval mapping that accounts for these problems has been developed by Knott, Elsen and Haley (1996) and has been applied to nearly all daughter and granddaughter design analyses. Their method is a modification of the non-linear regression method, and assumes a single QTL location for all families, but estimates a separate QTL effect for each family. This method has the advantage that, unlike ML, it can readily deal with missing and uninformative genotypes for some markers. Mackinnon and Weller (1995) proposed an ML method to estimate both QTL location and effect for half-sib designs under the assumption that only two QTL alleles are segregating in the population. Using this method it is also possible to estimate QTL genotype of the common parent of each family. However, these determinations are accurate only for relatively large QTL. The method of Mackinnon and Weller (1995) is more difficult to apply than the method of Knott, Elsen and Haley (1996), and has not come into general usage.

Lander and Botstein (1989) proposed the LOD-score (logarithm of the odds to the base 10) drop-off method to estimate confidence intervals for QTL location, but several studies have shown that this seriously underestimate the actual value (e.g. Darvasi *et al.*, 1993). The non-parametric bootstrap method (Visscher, Thompson and Haley, 1996) was found to be more accurate, but tends to overestimate confidence intervals. Bennewitz, Reinsch and Kalm (2003) proposed improvements to the bootstrap method that result in shorter but still unbiased confidence intervals.

Most studies to detect QTL in dairy cattle have considered many markers and multiple traits. In some studies nearly the entire genome was analysed, which raises a serious problem with respect to the appropriate threshold to declare significance. If normal point-wise significance levels of 5 or 1 percent are used, many marker-trait combinations will show "significance" by chance. While this is a problem for all QTL genome scans, it is even more severe for dairy cattle in which multiple half-sib families are analysed, in addition to multiple markers and traits. Several solutions to this problem have been proposed, none of which is completely satisfactory. The only solution to deal adequately with both multiple traits and families in addition to multiple markers is the false discovery rate (Weller *et al.*, 1998).

The QTL effects derived from either daughter or granddaughter by ML or non-linear regression will still be biased for several reasons. First, the usual assumptions of interval mapping, a single QTL segregating within the marker interval and no QTL in adjacent intervals, often do not reflect reality. Second, the dependant variable is generally an "adjusted" record, either daughter yield deviations (DYD; VanRaden and Wiggans, 1991) or genetic evaluations. Israel and Weller (1998) demonstrated that QTL effects derived from analysis of either genetic evaluations, yield deviations or DYD will be underestimated. In addition to this downward bias, there are two sources of upward bias for QTL effects. First, the direction of the effects is generally arbitrary, and therefore absolute values are retained and all effects are >0. Third, only the effects deemed "significant" are retained, and this is a selected sample (Georges *et al.,* 1995). Bayesian analysis methods that account for bias of QTL effect due to selection have recently been developed by Weller, Schlezinger and Ron (2005).

CURRENT STATUS OF QTL DETECTION IN DAIRY CATTLE

Genome scans by the granddaughter design have been completed for Holsteins from Canada (Nadesalingam, Plante and Gibson, 2001), the Netherlands (Spelman *et al.,* 1996; Schrooten *et al.,* 2000), France (Bennewitz, *et al.,* 2003a; Boichard *et al.,* 2003), Germany (Bennewitz, *et al.,* 2003a; Kuhn *et al.,* 2003a), New Zealand (Spelman *et al.,* 1999), and the United States (Georges *et al.,* 1995; Ashwell *et al.,* 1996, 1997, 1998a, 1998b, 2004; Ashwell, Van Tassell and Sonstegard, 2001; Zhang *et al.,* 1998; Ashwell and Van Tassell, 1999; Heyen *et al.,* 1999); Finnish Ayrshires (Vilkki *et al.,* 1997; Viitala *et al.,* 2003; Schulman *et al.,* 2004); French Normande and Montbeliarde cattle (Boichard *et al.,* 2003); Norwegian cattle in Norway (Klungland *et al.,* 2001; Olsen *et al.,* 2002); and Swedish Red and White (SRB) (Holmberg and Andersson-Eklund, 2004). Daughter design analyses have been performed for Israeli Holsteins (Mosig *et al.,* 2001; Ron *et al.,* 2004). Most studies have considered the five economic milk production traits: milk, fat and protein production, and fat and protein concentration, although a number of studies have also considered somatic cell score (SCS), female fertility, herd life, calving traits, health traits, temperament and conformation traits. The SCS is a log base 2 function of the concentration of somatic cells, and has been shown to be a useful indicator of udder health. Results are summarized in Table 1.

Results for milk, fat and protein production, fat and protein concentration, and SCS from most of the studies listed above are summarized at www.vetsci.usyd.edu.au/reprogen/QTL_Map/. Results from these traits, and many others including meat production, are summarized at http://bovineqtl.tamu.edu. Significant effects were found on all 29 autosomes, but most effects were found only in single studies and have not been repeated. Khatkar *et al.* (2004) performed a meta-analysis, combining data from most of these studies, and found significant across-study effects on chromosomes 1, 3, 6, 9, 10, 14 and 20.

METHODS OF INCORPORATING INFORMATION FROM GENETIC MARKERS IN GENETIC EVALUATION SYSTEMS

Heritabilities of most economic traits in dairy cattle are low to moderate. Genetic evaluation of dairy cattle is complicated by confounding between genetic and

TABLE 1
Summary of dairy cattle genome scans

Experimental design	Breed	Country	Traits analysed	References
Granddaughter	Ayrshire	Finland	Milk production[1]	Vilkki *et al.,* 1997; de Koning *et al.,* 2001; Viitala *et al.,* 2003
			SCS[2], mastitis, other treatments	Schulman *et al.,* 2004
	Jersey	New Zealand	Conformation	Spelman, Garrick and van Arendonk, 1999
	Holstein	Canadian	Milk production	Plante *et al.,* 2001
		France	Milk production	Boichard *et al.,* 2003
		Germany	Milk production	Thomsen *et al.,* 2001
			Functional	Kuhn *et al.,* 2003
			Conformation, temperament, milking speed	Hiendleder *et al.,* 2003
		Netherlands	conformation, SCS, fertility, calving, milking speed, gestation, birth weight, temperament	Schrooten *et al.,* 2000
		New Zealand	Conformation	Spelman, Garrick and van Arendonk, 1999
		USA	Milk production	Ashwell *et al.,* 1998b; Ashwell and Tassell 1999; Ashwell *et al.,* 1997, 2004; Ashwell, Van Tassell and Sonstegard, 2001;Georges *et al.,* 1995; Heyen *et al.,* 1999; Zhang *et al.,* 1998
			SCS	Ashwell *et al.,* 1996, 1997, 1998b; Ashwell and Van Tassell, 1999; Heyen *et al.,* 1999
			Herdlife	Heyen *et al.,* 1999
			Conformation	Ashwell *et al.,* 1998a, 1998b; Ashwell and Van Tassell, 1999
			Fertility	Ashwell *et al.,* 2004
	Montbeliarde	France	Milk production	Boichard *et al.,* 2003
	Normande	France	Milk production	Boichard *et al.,* 2003
	Norwegian	Norway	Milk production	Olsen *et al.,* 2002
			SCS, mastitis	Klungland *et al.,* 2001
	Swedish	Sweden	SCS, mastitis, other diseases	Holmberg and Andersson-Eklund, 2004
Daughter	Holstein	Israel	Milk production, SCS, fertility, herdlife	Ron *et al.,* 2004
			% protein	Mosig *et al.,* 2001

[1] Milk, fat, and protein production, and fat and protein concentration.
[2] Somatic cell concentration

environmental factors. Cows are scattered over many different herds with different management levels, and distribution of sires across herds is not random or orthogonal. Furthermore, cows generally produce multiple lactations that are correlated. In order to account for the limited heritability, and co-variances among relatives, genetic effects are generally assumed to be random, while most environmental effects are assumed to be fixed. Thus, genetic evaluation is performed by the mixed model using best linear unbiased prediction (BLUP) methodology (Henderson, 1984).

Beginning in the late 1980s, the model of choice for genetic evaluation for milk production traits was the individual animal model, in which a genetic effect is computed

for each animal, including animals that did not have production records (Westall and van Vleck, 1987). Genetic evaluations for these animals are derived via the numerator relationship matrix, which is included in the model. In addition, a "permanent environmental" effect is computed for each animal with records to account for similarities among multiple records of the same cow that are not due to additive genetic effects. As noted previously, analysis of QTL effects has generally been based on analysis of genetic evaluations or DYD, which are the adjusted means of the daughter records of a bull but which, unlike genetic evaluations, are not regressed. However, the statistical properties of DYD are not well understood, and QTL effects derived from analysis of DYD are still biased (Israel and Weller, 1998). Theoretically, it should be possible to derive unbiased QTL estimates if these effects are incorporated into a genetic evaluation scheme based on analysis of the actual records, such as the animal model. In practice, the inclusion of QTL effects into genetic evaluation models is complicated by three main factors:

- actual QTL location is unknown, and there is only partial linkage between genetic markers and QTL;
- linkage phase between genetic markers and QTL differs among individuals, and is generally unknown;
- only a small fraction of the population is genotyped.

An analysis including only genotyped individuals is not a viable option as it will generally not be possible to derive accurate fixed effects, such as herd-year-seasons, from this sample.

Fernando and Grossman (1989) proposed modifying the individual animal model described above to a "gametic" model that assumes the two QTL alleles of each individual are random effects sampled from a distribution with a known variance. They developed a method to estimate breeding values for all individuals in a population, including QTL effects via linkage to genetic markers, provided that all animals are genotyped and the heritability and recombination frequency between the QTL and the genetic marker are known. This model is suitable for any population structure and can also incorporate non-linked polygenic effects and other "nuisance" effects such as herd or block. The basic model assumes only a single record per individual, but can be adapted readily to a situation of multiple records per animal. This method is also denoted "marker-assisted BLUP" or "MA-BLUP".

Each individual with unknown ancestors is assumed to have two unique alleles for the QTL, which are "sampled" from an infinite population of alleles. For animals that are not genotyped, the probability of receiving either allele from either parent will be equal. However, if both the parent and progeny are genotyped for a linked genetic marker, then the probability of receiving a specific parental allele for a QTL linked to the genetic marker will be a function of the progeny marker genotype and recombination frequency. Based on these probabilities, Fernando and Grossman (1989) demonstrated how a variance-co-variance matrix could be constructed for the QTL gametic effects. They further described a simple algorithm to invert this matrix analogous to Henderson's method for inverting the numerator relationship matrix. This method has been extended to handle multiple markers and traits (Goddard, 1992). Cantet and Smith (1991) demonstrated that the number of equations could be significantly reduced by analysis of the reduced animal model.

The disadvantages of this model are that it assumes that both recombination frequency and the variance due to the QTL are known *a priori*. Studies on simulated data have demonstrated that although restricted maximum likelihood methodology can be used to estimate these parameters, they are completely confounded for a single marker locus (van Arendonk *et al.,* 1994). Methods to estimate the variance contributed by QTL with multiple markers were developed by Grignola, Hoeschele and Tier (1996). Furthermore, as each individual with unknown parents is assumed to have two unique alleles, the prediction error variances of the effects for any individual will be quite large and, therefore, not very informative. Finally, the assumption of a normal distribution of possible QTL allele effects may not be realistic.

Israel and Weller (1998) proposed an alternative method that assumes that only two QTL alleles are segregating in the population, and that either a daughter or granddaughter design has been applied to determine QTL genotypes of the family ancestors. The QTL effect is then included in the complete animal model analysis as a fixed effect. For individuals that are not genotyped, probabilities of receiving either allele are included as regression constants. These probabilities can be readily computed for the entire population using the segregation analysis method of Kerr and Kinghorn (1996). Israel and Weller (1998) assumed complete linkage between the QTL and a single marker. Israel and Weller (2002) extended the method to QTL analysis based on flanking marker, using the method of Whittaker, Thompsom and Visscher (1996) to estimate QTL effects and location from the regression estimates of flanking markers. This method has been tested extensively on simulated populations, and was able to derive unbiased estimates of QTL effect and location. It has also been applied to actual data from the Israeli Holstein population for a segregating QTL on chromosome 14 that affected milk production traits (Weller *et al.,* 2003). However, in this case the QTL effect was underestimated. Further research is required to determine the reason for this discrepancy.

METHODS FOR QTL DETECTION AND MAS IN DEVELOPING COUNTRIES

As noted previously, dairy cattle breeding in tropical and subtropical countries is generally based on crossbreeding between high production breeds adapted to temperate climates, and tropical strains which are adapted to the local environment, including resistance to local diseases. In other animal species, synthetic strains have been produced by selecting those individuals that retain the positive characteristics from both strains. For example, the Assaf sheep breed was produced from a cross between the Middle East Awassi breed and the East Friesian breed (www.sheep101.info/breedsA.html). In dairy cattle, the problem of an appropriate strategy for future generations has not been adequately solved, for reasons considered previously. If the economically important genes were identified, then the time and effort required for production of the desired synthetic strains could be reduced.

Visscher, Haley and Thompson (1996) considered the situation in which the recipient strain is an outbred population in an ongoing selection programme, and the introgressed genes are QTL. Markers flanking the QTL will be required in order to select backcross progeny that received the donor QTL allele. As there will be uncertainty with respect to the QTL location,

the flanking markers must be sufficiently close to the QTL so that it will be possible to determine with relative certainty that the QTL is in fact located between the flanking markers. Although marker-assisted introgression does decrease the number of generations required to obtain fixation of the desired allele, it increases two key cost elements. First, with traditional introgression, half of the progeny will carry the donor allele for the introgressed gene, and all of these can be used as parents in the next generation. However, if only a small fraction of the progeny is selected based on genetic markers, then many more individuals must be produced each generation. Second, genotyping costs for a large number of markers at each generation will also be significant.

Crosses between cattle breeds can also be used for QTL detection and they have been used in developing countries. In most plant species, the parental lines are completely inbred, and there will be complete LD in the F_2 or backcross generation. However, cattle are outbreeders and in crosses between breeds there will therefore only be partial LD between segregating QTL and linked genetic markers. Song, Soller and Genizi (1999) proposed the full-sib intercross line (FSIL) design for QTL detection and mapping for crosses between strains of outcrossing species. They assumed that the two parental strains differ in allelic frequencies, but were not at fixation for alternative QTL alleles.

For given statistical power, the FSIL design requires only slightly more individuals than an F_2 design derived from an inbred line cross, but six- to ten-fold fewer than a half-sib or full-sib design. In addition, as the population is maintained by continued intercrossing, DNA samples and phenotypic information can be accumulated across generations. Continued intercrossing in future generations also leads to map expansion, and thus to increased mapping accuracy in the later generations. An FSIL can therefore be used for fine mapping of QTL and this is considered below in detail.

Although these methods have not as yet been applied to detect QTL related to milk production, they have been applied to QTL for disease resistance. Trypanosomosis (sleeping sickness) is a major constraint on livestock productivity in sub-Saharan Africa. Hanotte *et al.* (2003) mapped QTL affecting trypanotolerance in a cross between the "tolerant" N'Dama breed and the susceptible Boran breed. Putative QTL affecting 16 traits associated with disease susceptibility were mapped tentatively to 18 autosomes. Excluding chromosomes with ambiguous effects, the allele associated with resistance was derived from the N'Dama strain for nine QTL and from the Boran strain for five QTL. These results are consistent with many plant crossbreeding experiments in which the strain with overall phenotypic inferiority for the quantitative trait nevertheless harbours QTL alleles that are superior to the alleles present in the phenotypically superior strain (e.g. Weller, Soller and Brody, 1988).

FROM QTL TO QTN – THEORY

As noted by Darvasi and Soller (1997), with a saturated genetic map, the resolving power for QTL will be a function of the experimental design, number of individuals genotyped and QTL effect. Weller and Soller (2004) computed that the 95 percent confidence interval (CI) in percent recombination for half-sib designs, including the daughter and granddaughter designs, was $3073/d^2N$, where d is the QTL substitution effect in units of the standard deviation, and

N is the sample size. In the case of a granddaughter design, the units for the standard deviation will be either units of the bulls' DYD or genetic evaluations. For example, if d is 0.5 and N is 400, the CI will be 31 percent recombination, or approximately 35 cM. Thus, except for the largest QTL, CIs will generally include several tens of cM. Considering that each cattle cM includes ~8 genes and one million bp, detection of the actual polymorphism responsible for the observed QTL effects (the quantitative trait nucleotide, QTN) appears at first glance to be a "mission impossible".

Various strategies have been proposed to reduce the CI based on multiple crosses, but most are not applicable to dairy cattle (e.g. Darvasi, 1998). Meuwissen and Goddard (2000) proposed that CI for QTL location could be reduced to individual cM by application of LD mapping. If a QTL polymorphism is due to a relatively recent mutation or to a relatively recent introduction from another population, then it should be possible to detect population-wide LD between the QTL and closely linked genetic markers. The closer the marker to the QTL, the greater will be the extent of LD. They developed a method to estimate QTL location and CI based on LD between a QTL and a series of closely linked markers. The CI can be further reduced by combining linkage and LD mapping (Meuwissen *et al.,* 2002), and by a multitrait analysis (Meuwissen and Goddard, 2004). However, unless the QTL effect is very large, the CI will still extend over several cM.

In order to determine the actual gene responsible for the QTL, most studies have used the "candidate gene" approach, i.e. to determine a likely candidate among the genes within the CI, based on known gene function, or specific gene expression in the organ of interest. Examples are given in the following section. However, even if a polymorphism is detected in the candidate gene and the polymorphism has a major LD effect on the QTL, how does one prove that this polymorphism is not merely in LD with the actual QTN?

Mackay (2001) proposed two alternatives for proof positive that a candidate polymorphism is in fact the QTN, namely, co-segregation of intragenic recombinant genotypes in a candidate gene with the QTL phenotype, and functional complementation where the trait phenotype is "rescued" in a transgenic organism. Neither of these is applicable to QTL in dairy cattle. In this case, Mackay (2001) postulated that the only option to achieve the standard of rigorous proof for identification of a gene underlying a QTL in commercial animal populations is to collect "multiple pieces of evidence, no single one of which is convincing, but which together consistently point to a candidate gene". Evidence can be provided by concordance of polymorphism with deduced QTL genotype, quantitative differences of gene expression in physiologically relevant organs, SNP capable of encoding a non-conservative amino acid change, protein differences in cows with contrasting genotypes for the QTN, orthologous QTL in other species (genes that are derived from a common ancestral gene) and alteration of gene protein in bovine cell lines by "short interfering RNA" (siRNA) technology. (The siRNA molecules bind with proteins to form a unit called the "RNA-induced silencing complex" that suppresses the expression of the gene to which it corresponds in the viral genome, silencing the gene from which the siRNA is derived.)

For dairy cattle, to date, the most compelling evidence is "concordance", i.e. that

the deduced QTL genotypes of a sample of individuals correspond completely to their genotypes for the putative QTN. All individuals heterozygous for the QTL should be heterozygous for the putative QTN, with the same QTN allele associated with the same QTL allele in all individuals, and all individuals homozygous for the QTL should also be homozygous for the QTN. Theoretically, the sample of individuals analysed should be large enough to reject statistically the hypothesis that concordance was obtained by chance. However, in dairy cattle, the only individuals for which QTL genotype can be derived with any level of reliability are sires that have been analysed by either a daughter or granddaughter design, and the number of these individuals will always be limited. Furthermore, there is at present no accepted theory to compute concordance probabilities by chance, considering that any polymorphism very close to the QTN will display significant LD. Several studies have addressed the problem (Cohen-Zinder *et al.*, 2005; Schnabel *et al.*, 2005). The case for identification of the QTN is clearly more compelling if concordance is obtained in two different populations.

FROM QTL TO QTN – RESULTS

To date, the QTN has been determined in two cases in dairy cattle, on BTA 6 and BTA 14. In both cases the QTL chiefly affected fat and protein concentration and the QTL effect was large enough that the confidence interval for QTL location was <10 cM. A QTL on BTA 14 near the centromere that chiefly affected fat quantity and both fat and protein concentration in both the United States and Israeli Holstein populations was first detected by Ron *et al.* (1998), and further studies were able to map the QTL to a region of approximately 10 cM (Coppieters *et al.*, 1999). In 2002, two studies independently showed that a mis-sense mutation, causing replacement of a lysine residue with alanine in exon VIII of the gene acylCoA:diacyglycerol acyltransferase (DGAT1), is the QTN (Grisart *et al.*, 2002; Winter *et al.*, 2002). Discovery was aided by the fact that DGAT1 was an obvious physiological candidate. In addition to mapping to the putative QTL region, DGAT1 encodes a microsomal enzyme that catalyses the final step of triglyceride synthesis and mice lacking both copies of DGAT1 are completely devoid of milk secretion. Complete concordance between this polymorphism and the QTL was found in three different dairy breeds.

The QTL near the middle of BTA 6 affecting protein concentration was first detected by Georges *et al.* (1995) in the United States Holstein population. This QTL was then detected in several other Holstein populations, including Finnish Ayrshire cattle (Velmala *et al.*, 1999) and Norwegian cattle (Olsen *et al.*, 2002). Ron *et al.* (2001) reduced the CI to 4 cM centred on microsatellite BM143. Olsen *et al.* (2002) used physical mapping and combined linkage and LD mapping to determine that this QTL is located within a 420 000 bp region between the genes ABCG2 and LAP3.

In 2005, two research groups claimed to have found the QTN in two different genes. Schnabel *et al.* (2005) claimed that the QTN is located in a poly-A sequence in the promoter region of the osteopontin gene, also denoted SPP1, while Cohen-Zinder *et al*. (2005) claimed that the QTN is a mis-sense mutation in exon 14 of the ABCG2 gene. Both studies based their claim on gene function and concordance of bulls with known genotypes. Both genes are dif-

ferentially expressed in the mammary gland during lactation, as compared with the liver. Furthermore, anti-sense SPP1 transgenic mice displayed abnormal mammary gland differentiation and milk secretion (Nemir *et al.,* 2000).

Schnabel *et al.* (2005) found concordance based on four heterozygous and four homozygous sires for the United States Holstein population, as determined by a granddaughter design, while Cohen-Zinder *et al.* (2005) found concordance for three heterozygous and 15 homozygous sires from both the United States and Israeli Holstein populations. Cohen-Zinder *et al.* (2005) also analysed the site proposed by Schnabel *et al.* (2005), and found that this site was hyper-variable in that at least four single nucleotide changes were found within the 20 bp region centred on the poly-A sequence. Eight of nine Israeli sires analysed by the daughter design were heterozygous for at least one of these polymorphisms.

Many studies have found a QTL affecting all five milk production traits and SCS near the middle of BTA 20. Blott *et al.* (2003) claimed that a mis-sense mutation in the bovine growth hormone receptor was responsible for the QTL affecting milk yield and composition on BTA 20, but did not find concordance for the bulls heterozygous for the QTL. Thus, this polymorphism may be responsible for only part of the observed effect on BTA 20, or may be a physiologically neutral mutation in LD with the QTN.

For both the QTL on BTA 6 and 14, the polymorphisms analysed apparently do not account for the entire effect observed in these chromosomal regions (Bennewitz *et al.,* 2004a; Kuhn *et al.,* 2004; Cohen-Zinder *et al.,* 2005). The effect associated with the mis-sense mutation in ABCG2 explains the entire effect observed on milk yield and fat and protein concentration, but does not explain the effects associated with fat and protein yield. It is likely that in the near future additional QTN will be resolved. As noted, the meta-analysis (Khatkar *et al.,* 2004) found significant effects on BTA 1, 3, 9 and 10, in addition to the effects described on BTA 6, 14 and 20.

METHODS AND THEORY FOR MAS IN DAIRY CATTLE

Considering the long generation interval, the high value of each individual, the very limited female fertility and the fact that nearly all economic traits are expressed only in females, it would seem that dairy cattle should be a nearly ideal species for application of MAS. However, most theoretical studies have been rather pessimistic with respect to the expected gains that can be obtained by MAS. As noted by Weller (2001), MAS can potentially increase annual genetic gain by increasing the accuracy of evaluation, increasing the selection intensity and decreasing the generation interval.

The following dairy cattle breeding schemes that incorporate MAS have been proposed:

- a standard PT system, with information from genetic markers being used to increase the accuracy of sire evaluations in addition to phenotypic information from daughter records (Meuwissen and van Arendonk, 1992);
- a MOET nucleus breeding scheme in which marker information is used to select sires for service in the MOET population, in addition to phenotypic information on half-sisters (Meuwissen and van Arendonk, 1992);
- PT schemes in which information on genetic markers is used to preselect young sires for entrance into the PT (Kashi,

Hallerman and Soller, 1990; Mackinnon and Georges, 1998);
- selection of bull sires without a PT, based on half-sib records and genetic markers (Spelman, Garrick and van Arendonk, 1999);
- selection of sires in a half-sib scheme, based on half-sib records and genetic markers (Spelman, Garrick and van Arendonk, 1999);
- use of genetic markers to reduce errors in parentage determination (Israel and Weller, 2000).

Meuwissen and van Arendonk (1992) found that inclusion of marker information to increase the accuracy of sire evaluations increased the rate of genetic gain by only 5 percent when the markers explained 25 percent of the genetic variance. This result is not surprising considering that the accuracy of sire evaluations based on a PT of 50 to 100 daughters is already quite high. In "open" and "closed" nucleus breeding schemes, rates of genetic gain were increased by 26 and 22 percent, respectively. The advantage of MAS in this case is greater, because young sires are not progeny tested, and their reliabilities based only on half-sib information are much lower.

Mackinnon and Georges (1998) proposed "top-down" and "bottom-up" strategies to apply the third scheme listed above, preselection of young sires prior to PT. In the "top-down" strategy, QTL genotypes are determined for the elite sires used as bull sires by a granddaughter design. If a dense marker map is available, it will then be possible to determine which QTL allele is passed to each son. Elite bulls from among these sons are then selected as bull sires for the next generation. If the original sire was heterozygous for a QTL, it can be determined which of his sons received the favourable allele. Sons of these sires are then genotyped and selected based on whether they received the favourable grandpaternal QTL alleles. It is assumed that the dams of the candidate sires are also genotyped, and that these cows will be progeny of the sires evaluated by a granddaughter design. Thus, grandpaternal alleles inherited via the candidates' dams can also be traced. A disadvantage of this scheme is that only the grandpaternal alleles are followed. Some of the sons of the original sires that were evaluated by a granddaughter design will also have received the favourable QTL allele from their dams, but not via the genotyped grandsires. However, young sires will be selected based only on the grandpaternal haplotypes.

In the "bottom-up" scheme, QTL genotypes of elite sires are determined by a daughter design. These sires are then used as bull sires. The candidate bulls are then pre-selected for those QTL heterozygous in their sires, based on which paternal haplotype they received. As the QTL phase is evaluated on the sires of the bull calves (the candidates for selection), no selection pressure is "wasted" as in the "top-down" scheme. In addition, this design can be applied to a much smaller population, because only several hundred daughters are required to evaluate each bull sire. On the negative side, more daughters than sons must be genotyped to determine QTL genotype. Mackinnon and Georges (1998) assumed that in either scheme it will not be necessary to increase mean generation interval above that of a traditional PT programme, although this will probably not be the case (Weller, 2001).

Kashi, Hallerman and Soller (1990), Mackinnon and Georges (1998), and Israel and Weller (2004) all addressed the problem that QTL determination will be subject to error. Deciding that a specific sire is homozygous for the QTL when in fact

the sire is heterozygous will be denoted the "type I" error. Deciding that the QTL is heterozygous in a specific sire, while the sire is in reality homozygous will be denoted the "type II" error. In the first case, segregating QTL will be missed while, in the second case, selection for the positive QTL allele will be applied to no advantage. All three studies found that genetic gains will be maximized with a relatively large proportion of type I errors, between 5 and 20 percent. This is due to the fact that as type I error increases, type II error decreases, and more real effects will be detected and applied in selection. A third type of error is theoretically possible, i.e. determining correctly that the ancestor is heterozygous for the QTL, but incorrect determination of QTL phase relative to the genetic markers. However, Israel and Weller (2004) showed by simulation that this error never occurred even when the type I error rate was set at 20 percent.

Spelman, Garrick and van Arendonk (1999) considered three different breeding schemes by deterministic simulation:

- a standard PT with the inclusion of QTL data;
- the same scheme except that young bulls without PT could also be used as bull sires based on QTL information;
- a scheme in which young sires could be used as both bull sires and cow sires in the general population, based on QTL information.

It was assumed that only bulls were genotyped but that, once genotyped, the information on QTL genotype and effect was known without error. It was then possible to conduct a completely deterministic analysis. They varied the fraction of the genetic variance controlled by known QTL from zero to 100 percent. Even without MAS, a slight gain was obtained by allowing young sires to be used as bull sires, and a genetic gain of 9 percent was obtained if young sires with superior evaluations were also used directly as both sires of sires and in general service. As noted previously, the genetic gain was limited where MAS was used only to increase the accuracy of young bull evaluations for a standard PT scheme because the accuracy of the bull evaluations was already high. Thus, even if all the genetic variance was accounted for by QTL, the genetic gain was less than 25 percent. However, if young sires are selected for general service based on known QTL, the rate of genetic progress can be doubled. The maximum rate of genetic gain that can be obtained in the third scheme, the "all bulls" scheme, was 2.2 times the rate of genetic gain in a standard PT. Theoretically, with half of the genetic variance due to known QTL, the rate of genetic gain obtained was greater than that possible with nucleus breeding schemes.

The final scheme, with use of genetic markers to reduce parentage errors, is the most certain to produce gains, as it does not rely on QTL genotype determination, which may be erroneous. Weller *et al.* (2004) genotyped 6 040 Israeli Holstein cows from 181 Kibbutz herds for 104 microsatellites. The frequency of rejected paternity was 11.7 percent, and most errors were due to inseminator mistakes. Most advanced breeding schemes already use genetic markers to confirm parentage of young sires. Israel and Weller (2002) found by simulations that if the parentage of bull dams and the test daughters of young sires are also verified, genetic gain increased by 4.3 percent compared with a breeding programme with 10 percent incorrect paternity. This scheme is economically justified if genotyping costs per individual are no more than US$15.

CURRENT STATUS OF MAS IN DAIRY CATTLE

Two ongoing MAS programmes in dairy cattle have been reported to date, in French and German Holsteins (Boichard *et al.,* 2002, 2006; Bennewitz *et al.,* 2004b). Currently in the German programme, markers on three chromosomes are used. The MA-BLUP evaluations (Fernando and Grossman, 1989) are computed at the VIT-computing centre in Verden, and are distributed to Holstein breeders who can use these evaluations for selection of bull dams and preselection of sires for progeny testing. The MA-BLUP algorithm only includes equations for bulls and bull dams, and the dependent variable is the bull's DYD (Bennewitz *et al.*, 2003b). Linkage equilibrium throughout the population is assumed. To close the gap between the grandsire families analysed in the German granddaughter design and the current generation of bulls, 3 600 bulls were genotyped in 2002. As then, about 800 bulls have been evaluated each year (N. Reinsch, personal communication). Only bulls and bull dams are genotyped as tissue samples are already collected for paternity testing. Thus additional costs due to MAS are low and even a very modest genetic gain can be economically justified. This scheme is similar to the "top-down" scheme of Mackinnon and Georges (1998) in that evaluation of the sons is used to determine which grandsires are heterozygous for the QTL and their linkage phase. This information is then used to select grandsons based on which haplotype was passed from their sires. It differs from the scheme of Mackinnon and Georges (1998) in that the grandsons are preselected for PT based on MA-BLUP evaluations, which include general pedigree information in addition to genotypes.

The French MAS programme includes elements of both the "top-down" and "bottom-up" MAS designs. Similar to the German programme, genetic evaluations including marker information were computed by a variant of MA-BLUP, and only genotyped animals and non-genotyped connecting ancestors were included in the algorithm. Genotyped females were characterized by their average performance based on pre-corrected records (with the appropriate weight), whereas males were characterized by twice the yield deviation of their non-genotyped daughters. Twelve chromosomal segments, ranging in length from 5 to 30 cM, are analysed. Regions with putative QTL affecting milk production or composition are located on BTA 3, 6, 7, 14, 19, 20 and 26; segments affecting mastitis resistance are located on BTA 10, 15 and 21; and chromosomal segments affecting fertility are located on BTA 1, 7 and 21. Each region was found to affect one to four traits and on average three regions with segregating QTL were found for each trait. Each region is monitored by two to four evenly spaced microsatellites, and each animal included in the MAS programme is genotyped for at least 43 markers. Sires and dams of candidates for selection, all male AI ancestors, up to 60 AI uncles of candidates, and sampling daughters of bull sires and their dams are genotyped. The number of genotyped animals was 8 000 in 2001 and 50 000 in 2006. An additional 10 000 animals are genotyped per year, with equal proportions of candidates for selection and historical animals.

FUTURE PROSPECTIVE FOR MAS IN DAIRY CATTLE

Although the first large experiment in QTL detection in dairy cattle was published in 1961 by Neimann-Sørensen and Robertson, in 1985 it still looked as if

MAS was a long way off for commercial animal populations as there were very few known genetic markers and methodology was rudimentary. In the last 20 years there have been huge advances in both DNA technology and statistical methodology, and it can now be stated with near certainty that the technology is available to detect and map accurately segregating QTL in dairy cattle. Furthermore, although many effects reported in the literature are "false positives", there is a wealth of evidence that several QTL are in fact real as a number of effects have been repeated across numerous experiments, and the actual QTN have been identified for at least two QTL.

The main limitation at this point to detecting and mapping more QTL is the sample sizes available, especially the number of progeny tested bulls per family. To map QTL of smaller magnitude accurately, it will be necessary to combine data across experiments (e.g. Khatkar *et al.,* 2004) or significantly increase sample sizes. This can only be done by genotyping cows, even though power per individual genotyped will be lower.

The fact that only two countries have actually started MAS programmes highlights the current limitations to practical application of MAS. To date, very few segregating QTL with economic impact have been identified in commercial dairy cattle populations. Of the two QTNs that have been detected, each has disadvantages with respect to application in MAS. The allele of DGAT1 that increases fat production and decreases water content in the milk, both desirable, also decreases protein yield, which is undesirable (Weller *et al.,* 2003). The allele of ABCG2 that decreases milk production and increases protein percent is clearly the favourable allele in nearly all current selection indices, but this allele is already at a very high frequency in all major dairy cattle populations (Ron *et al.,* 2006).

In addition to the limitation of definitively identified QTL with economic value, suitable software for genetic evaluation including QTL effects is also a limiting factor. At present, those countries that are applying MAS are using two-step procedures, i.e. a preliminary analysis to compute genetic evaluations based only on pedigree and phenotypic data, and then a second analysis in which the genetic evaluations are "adjusted" for QTL effects. Ideally a single algorithm should be used to derive genetic evaluations for the entire population including the effects of known QTL.

ACKNOWLEDGEMENTS

I thank M. Ron, E. Seroussi and M. Ashwell for their input.

REFERENCES

Ashwell, M.S. & Van Tassell, C.P. 1999. Detection of putative loci affecting milk, health, and type traits in a US Holstein population using 70 microsatellite markers in a genome scan. *J. Dairy Sci.* 82: 2497–2502.

Ashwell, M.S., Van Tassell, C.P. & Sonstegard, T.S. 2001. A genome scan to identify quantitative trait loci affecting economically important traits in a US Holstein population. *J. Dairy Sci.* 84: 2535–2542.

Ashwell, M.S., Rexroad, C.E., Miller, R.H. & VanRaden, P.M. 1996. Mapping economic trait loci for somatic cell score in Holstein cattle using microsatellite markers and selective genotyping. *Anim. Genet*. 27: 235–242.

Ashwell, M.S., Rexroad, C.E., Miller, R.H., VanRaden, P.M. & Da, Y. 1997. Detection of loci affecting milk production and health traits in an elite US Holstein population using microsatellite markers. *Anim. Genet*. 28: 216–222.

Ashwell, M.S., Da, Y., VanRaden, P.M., Rexroad, C.E. & Miller, R.H. 1998a. Detection of putative loci affecting conformational type traits in an elite population of United States Holsteins using microsatellite markers. *J. Dairy Sci*. 81: 1120–1125.

Ashwell, M.S., Da, Y., Van Tassell, C.P., VanRaden, P.M., Miller, R.H. & Rexroad, C.E. 1998b. Detection of putative loci affecting milk production and composition, health, and type traits in a United States Holstein population. *J. Dairy Sci*. 81: 3309–3314.

Ashwell, M.S., Heyen, D.W., Sonstegard, T.S., Van Tassell, C.P., DA, Y., VanRaden, P.M., Ron, M., Weller, J.I. & Lewin, H.A. 2004. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. *J. Dairy Sci*. 87: 468–475.

Bennewitz, J., Reinsch, N. & Kalm, E. 2003. Comparison of several bootstrap methods for bias reduction of QTL effect estimates. *J. Anim. Breed. Genet*. 120: 403–416.

Bennewitz, J., Reinsch, N., Grohs, C., Leveziel, H., Malafosse, A., Thomsen, H., Xu, N.Y., Looft, C., Kuhn, C., Brockmann, G.A., Schwerin, M., Weimann, C., Hiendleder, S., Erhardt, G., Medjugorac, I., Russ, I., Forster., M., Brenig, B., Reinhardt, F., Reents, R., Averdunk, G., Blumel, J., Boichard, D. & Kalm, E. 2003a. Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle. *Genet. Sel. Evol.* 35: 319–38.

Bennewitz, J., Reinsch, N., Thomsen, H., Szyda, J., Reinhart, F., Kuhn, C., Schwerin, M., Erhardt, G., Weimann, C. & Kalm, E. 2003b. Marker assisted selection in German Holstein dairy cattle breeding: outline of the program and marker assisted breeding value estimation. *Ann. Meet. Europ. Ass. Anim. Prod*. 54th Session G1.9. Rome, Italy.

Bennewitz, J., Reinsch, N., Paul, S., Looft, C., Kaupe, B., Weimann, C., Erhardt, G., Thaller, G., Kuhn, C., Schwerin, M., Thomsen, H., Reinhardt, F., Reents, R. & Kalm, E. 2004a. The DGAT1 K232A mutation is not solely responsible for the milk production quantitative trait locus on the bovine chromosome 14. *J. Dairy Sci*. 87: 431–442.

Bennewitz, J., Reinsch, N., Reinhardt, F., Liu, Z. & Kalm, E. 2004b. Top down preselection using marker assisted estimates of breeding values in dairy cattle. *J. Anim. Breed. Genet*. 121: 307–318.

Blott, S., Kim, J., Moisio, S., Schmidt-Kuntzel, A., Cornet, A., Berzi, P., Cambisano, N., Ford, C., Grisart, B., Johnson, D., Karim, L., Simon, P., Snell, R., Spelman, R., Wong, J., Vilkki, J., Georges, M., Farnir, F. & Coppieters, W. 2003. Molecular dissection of a quantitative trait locus: A phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* 163: 253–266.

Boichard, D., Fritz, S., Rossignol, M.N., Boscher, M.Y., Malafosse, A. & Colleau, J.J. 2002. Implementation of marker-assisted selection in French dairy cattle. *Proc. 7th World Cong. Genet. Appl. Livest. Prod.* 33: 19–22. Montpellier, France.

Boichard, D., Fritz, S., Rossignol, M.N., Guillaume, F., Colleau, J.J. & Druet, T. 2006. Implementation of marker-assisted selection: practical lessons from dairy cattle. *Proc. 8th World Cong. Genet. Appl. Livest. Prod.* 22: 11. 13–18 August 2006, Belo Horizonte, MG, Brazil.

Boichard, D., Grohs, C., Bourgeois, F., Cerqueira, F., Faugeras, R., Neau, A., Rupp, R., Amigues, Y., Boscher, M.Y. & Leveziel, H. 2003. Detection of genes influencing economic traits in three French dairy cattle breeds. *Genet. Sel. Evol*. 35: 77–101.

Brascamp, E.W., van Arendonk J.A.M. & Groen A.F. 1993. Economic appraisal of the utilization of genetic markers in dairy cattle breeding. *J. Dairy Sci*. 76: 1204–1213.

Cantet, R.J.C. & Smith, C. 1991. Reduced animal model for marker assisted selection using best linear unbiased prediction. *Genet. Sel. Evol*. 23: 221–233.

Cohen-Zinder, M., Seroussi, E., Larkin, D.M., Loor, J.J., Everts-van der Wind, A., Lee, J.H., Drackley, J.K., Mark R. Band, M.R., Hernandez, A.G., Shani, M., Lewin, H.A., Weller, J.I. & Ron, M. 2005. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Research* 15: 936–944.

Coppieters, W., Kvasz, A., Arranz, J.J., Grisart, B., Riquet, J., Farnir, F. & Georges, M. 1999. The great-grand-daughter design: a simple strategy to increase the power of a grand-daughter design for QTL mapping. *Genetical Research* 74: 189–199.

Cunningham, E.P. 1989. Formulation of breeding plans for dairy and dual-purpose cattle. *Brazil. J. Genet*. 12: 81.

Darvasi, A. 1998. Experimental strategies for the genetic dissection of complex traits in animal models. *Nat. Genet*. 18: 19–24.

Darvasi, A. & Soller, M. 1997. A simple method to calculate resolving power and confidence interval of QTL map location. *Behavior. Genet*. 27: 125–132.

Darvasi, A., Vinreb, A., Minke, V., Weller, J.I. & Soller, M. 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 134: 943–951.

de Koning, D.J., Schulman, N.F., Elo, K., Moisio, S., Kinos, R., Vilkki, J. & Maki-Tanila, A. 2001. Mapping of multiple quantitative trait loci by simple regression in half-sib designs. *J. Anim. Sci*. 79: 616–622.

de Vaccaro, L. 1990. Survival of European dairy breeds and their crosses with Zebus in the tropics. *Anim. Breed. Abstr*. 58: 475–494.

Fernando R.L. & Grossman, M. 1989. Marker assisted selection using best linear unbiased prediction. *Genet. Sel. Evol*. 21: 467–477.

FAO. 1987. Crossbreeding *Bos indicus* and *Bos taurus* for milk production in the tropics, by E.P. Cunningham & O. Syrstad. *Anim. Prod. Health Paper No. 68*. Rome.

Farnir, F., Coppieters, W., Arranz, J.J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marcq., F., Moreau, L., Mni, M., Nezer, C., Simon, P., Vanmanshoven, P., Wagenaar, D. & Georges, M. 2000. Extensive genome-wide linkage disequilibrium in cattle. *Genome Research* 10: 220–227.

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

Goddard, M.E. 1992. A mixed model for analyses of data on multiple genetic markers. *Theor. Appl. Genet.* 83: 878–886.

Grignola, F.E., Hoeschele, I. & Tier, B. 1996. Mapping quantitative trait loci in outcross populations via residual maximum likelihood. I. Methodology. *Genet. Sel. Evol*. 28: 479–490.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C. Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M. & Snell, R. 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research* 12: 222–231.

Haley, C.S. & Knott, S.A. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69: 315–324.

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. Nat. Acad. Sci. USA* 100: 7443–7448.

Henderson, C.R. 1984. *Applications of linear models in animal breeding*. University of Guelph, Ontario, ON, Canada.

Heyen, D.W., Weller J.I., Ron, M., Band, M., Beever, J.E., Feldmesser, E., Da, Y., Wiggans, G.R., VanRaden, P.M. & Lewin, H.A. 1999. A genome scan for QTL influencing milk production and health traits in dairy cattle. *Physiol. Genom*. 1: 165–175.

Hiendleder, S., Thomsen, H., Reinsch, N., Bennewitz, J., Leyhe-Horn, B., Looft, C., Xu, N., Medjugorac, I., Russ, I., Kuhn, C., Brockmann, G.A., Blumel, J., Brenig, B., Reinhardt, F., Reents, R., Averdunk, G., Schwerin, M., Forster, M., Kalm, E. & Erhardt, G. 2003. Mapping of QTL for body conformation and behavior in cattle. *J. Heredity* 94: 496–506.

Hill, W.G. 1971. Investment appraisal for national breeding programmes. *Anim. Prodn*. 13: 37–50.

Holmberg, M. & Andersson-Eklund, L. 2004. Quantitative trait loci affecting health traits in Swedish dairy cattle. *J. Dairy Sci.* 87: 2653–2659.

International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431:931–945.

Israel, C. & Weller, J.I. 1998. Estimation of candidate gene effects in dairy cattle populations. *J. Dairy Sci*. 81: 1653–1662.

Israel, C. & Weller, J.I. 2000. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. *J. Dairy Sci*. 83: 181–187.

Israel, C. & Weller, J.I. 2002. Estimation of quantitative trait loci effects in dairy cattle populations. *J. Dairy Sci*. 85: 1285–1297.

Israel, C. & Weller, J.I. 2004. Effect of type I error threshold on marker-assisted selection in dairy cattle. *Livest. Prod. Sci.* 85: 189–199.

Kashi, Y., Hallerman, E. & Soller, M. 1990. Marker-assisted selection of candidate bulls for progeny testing programmes. *Anim. Prod*. 51: 63–74.

Kerr, R. J. & Kinghorn, B.P. 1996. An efficient algorithm for segregating analysis in large populations. *J. Anim. Breed. Genet*. 113: 457–469.

Khatkar, M.S., Thomson, P.C., Tammen, I. & Raadsma, H.W. 2004. Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet. Sel. Evol*. 36: 163–190.

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

Knott, S.A., Elsen J.M. & Haley, C.S. 1996. Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theor. Appl. Genet*. 93: 71–80.

Kosgey, I. S., Kahi, K.A. & van Arendonk J.A.M. 2005. Evaluation of closed adult nucleus multiple ovulation and embryo transfer and conventional progeny testing breeding schemes for milk production in tropical crossbred cattle. *J. Dairy Sci*. 88: 1582–1594.

Kuhn, C., Bennewitz, J., Reinsch, N., Thomsen, H., Looft, C., Brockmann, G.A., Schwerin, M., Weimann, C., Hiendieder, S., Erhardt, G., Medjugorac, I., Forster, M., Brenig, B., Reinhardt, F., Reents, R., Russ, I., Averdunk, G., Blumel, J. & Kalm, E. 2003. Quantitative trait loci mapping of functional traits in the German Holstein cattle population. *J. Dairy Sci*. 86: 360–368.

Kuhn, C., Thaller, G., Winter, A., Bininda-Emonds, O.R.P., Kaupe, B., Erhardt, G., Bennewitz, J., Schwerin, M. & Fries, R. 2004. Evidence for multiple alleles at the DGAT1 locus better explains a quantitative tip trait locus with major effect on milk fat content in cattle. *Genetics* 167: 1873–1881.

Lander, E.S. & Botstein, D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.

Mackay, T.F. 2001. The genetic architecture of quantitative traits. *Ann. Rev. Genet.* 35: 303–339.

Mackinnon, M.J. & Georges, M.A.J. 1998. Marker-assisted preselection of young dairy sires prior to progeny-testing. *Livest. Prod. Sci.* 54: 229–250.

Mackinnon, M.J. & Weller, J.I. 1995. Methodology and accuracy of estimation of quantitative trait loci parameters in a half-sib design using maximum likelihood. *Genetics* 141: 755–770.

Madalena, F.E. 1993. A simple scheme to utilize heterosis in tropical dairy cattle. *World Anim. Rev.* 74/75: 17–25.

Martinez, O. & Curnow, R.N. 1992. Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theor. Appl. Genet*. 85: 480–488.

Meuwissen, T.H.E. & Goddard, M.E. 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics* 155: 421–430.

Meuwissen, T.H.E. & Goddard, M.E. 2004. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genet. Sel. Evol* 36: 261–279.

Meuwissen, T.H.E. & van Arendonk, J.A.M. 1992. Potential improvements in rate of genetic gain from marker assisted selection in dairy cattle breeding schemes. *J. Dairy Sci.* 75: 1651–1659.

Meuwissen, T.H.E., Karlsen, A., Lien, S., Olsaker, I. & Goddard, M.E. 2002. Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* 16: 373–379.

Mosig, M.O., Lipkin, E., Khutoreskaya, G., Tchourzyna, E., Soller, M. & Friedmann, A. 2001. A whole genome scan for quantitative trait loci affecting milk protein percentage in Israeli-Holstein cattle, by means of selective milk DNA pooling in a daughter design, using an adjusted false discovery rate criterion. *Genetics* 157: 1683–1698.

Nadesalingam, J., Plante, Y. & Gibson, J.P. 2001. Detection of QTL for milk production on chromosomes 1 and 6 of Holstein cattle. *Mammal. Genome* 12: 27–31.

Neimann-Sørensen, A. & Robertson, A. 1961. The association between blood groups and several production characters in three Danish cattle breeds. *Acta. Agr. Scand*. 11: 163–196.

Nemir, M., Bhattacharyya, D., Li, X., Singh, K., Mukherjee, A.B. & Mukherjee, B.B. 2000. Targeted inhibition of osteopontin expression in the mammary gland causes abnormal morphogenesis and lactation deficiency. *J. Biol. Chem*. 275: 969–976.

Nicholas, F.W. & Smith C. 1983. Increased rates of genetic change in dairy cattle by embryo transfer and splitting. *Anim Prod*. 36: 341–353.

Olsen, H.G., Gomez-Raya, L., Vage, D.I., Olsaker, I., Klungland, H., Svendsen, M., Adnoy, T., Sabry, A., Klemetsdal, G., Schulman, N., Kramer, W., Thaller, G., Ronningen, K. & Lien, S.

2002. A genome scan for quantitative trait loci affecting milk production in Norwegian dairy cattle. *J. Dairy Sci.* 85: 3124–3130.

Plante, Y., Gibson, J.P., Nadesalingam, J., Mehrabani-Yeganeh, H., Lefebvre, S., Vandervoort, G. & Jansen, G.B. 2001. Detection of quantitative trait loci affecting milk production traits on 10 chromosomes in Holstein cattle. *J. Dairy Sci*. 84: 1516–1524.

Powell, R.L. & Norman, H.D. 1988. Accuracy of cow indexes according to repeatability, evaluation, herd yield, and registry status. *J. Dairy Sci*. 71: 2232–2240.

Ron, M., Heyen, D.W., Weller, J.I., Band, M., Feldmesser, E., Pasternak, H., Da, Y., Wiggans, G.R., Vanraden, P.M., Ezra, E. & Lewin, H.A. 1998. Detection and analysis of a locus affecting milk concentration in the US and Israeli dairy cattle populations. *Proc. 6th World Cong. Genet. Appl. Livest. Prod*. 26: 422–425. Armidale, Australia.

Ron, M., Kliger D., Feldmesser E., Seroussi E., Ezra E. & Weller J. I. 2001. Multiple quantitative trait locus analysis of bovine chromosome 6 in the Israeli Holstein population by a daughter design. *Genetics* 159:727–735.

Ron, M., Cohen-Zinder, M., Peter, C., Weller, J.I. & Erhardt, G. 2006. ABCG2 polymorphism in *Bos indicus* and *Bos taurus* cattle breeds. *J. Dairy Sci.* **89**: 4921–4923.

Ron, M., Feldmesser, E., Golik, M., Tager-Cohen, I., Kliger, D., Reiss, V., Domochovsky, R., Alus, O., Seroussi, E., Ezra, E. & Weller, J.I. 2004. A complete genome scan of the Israeli Holstein population for quantitative trait loci by a daughter design. *J. Dairy Sci*. 87: 476–490.

Schnabel, R.D., Kim, J.J., Ashwell, M.S., Sonstegard, T.S., Van Tassell, C.P., Connor, E.E. & Taylor, J.F. 2005. Fine-mapping milk production quantitative trait loci on BTA6: Analysis of the bovine osteopontin gene. *Proc. Nat. Acad. Sci. USA* 102: 6896–6901.

Schrooten, C., Bovenhuis, H., Coppieters, W. & van Arendonk, J.A.M. 2000. Whole genome scan to detect quantitative trait loci for conformation and functional traits in dairy cattle. *J. Dairy Sci*. 83: 795–806.

Schulman, N.F., Viitala, S.M., de Koning, D.J., Virta, J., Maki-Tanila, A. & Vilkki, J.H. 2004. Quantitative trait loci for health traits in Finnish Ayrshire cattle. *J. Dairy Sci*. 87: 443–449.

Smith, C. 1988. Genetic improvement of livestock, using nucleus breeding units. *World Anim. Rev*. 65: 2–10.

Song, J-Z., Soller, M. & Genizi, A. 1999. The full-sib intercross line (FSIL): a QTL mapping design for outcrossing species. *Genet. Research* 73: 61–73.

Spelman, R.J., Garrick, D.J. & van Arendonk, J.A.M. 1999. Utilization of genetic variation by marker assisted selection in commercial diary cattle populations. *Livest. Prod. Sci.* 59: 51–60.

Spelman, R.J., Coppieters, W., Karim, L., van Arendonk, J.A.M. & Bovenhuis, H. 1996. Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. *Genetics* 144: 1799–1808.

Spelman, R.J., Huisman, A.E., Singireddy, S.R., Coppieters, W., Arranz, J., Georges, M. & Garrick, D.J. 1999. Quantitative trait loci analysis on 17 nonproduction traits in the New Zealand dairy population. *J. Dairy Sci*. 82: 2514–2516.

Thomsen, H., Reinsch, N., Xu, N., Looft, C., Grupe, S., Kuhn, C., Brockmann, G.A., Schwerin, M., Leyhe-Horn, B., Hiendleder, S., Erhardt, G., Medjugorac, I., Russ, I., Forster, M., Brenig, B., Reinhardt, Kalm, E. 2001. Comparison of estimated breeding values, daughter yield deviations and deregressed proofs within a whole genome scan for QTL. *J. Anim. Breed. Genet*. 118: 357–370.

van Arendonk, J.A.M., Bovehuis, H., van der Beek, S. & Groen, A.F. 1994. Detection and exploitation of markers linked to quantitative traits in farm animals. *Proc. 5th World Cong. Genet. Appl. Livest. Prod*. 21:193–200. Guelph, Canada.

VanRaden, P.M. & Wiggans, G.R. 1991. Derivation, calculation and use of national animal model information. *J. Dairy Sci*. 74: 2737–2746.

Velmala, R.J., Vilkki, H.J., Elo, K.T., de Koning, D.J. & Maki-Tanila, A.V. 1999. A search for quantitative trait loci for milk production traits on chromosome 6 in Finnish Ayrshire cattle. *Anim. Genet*. 30: 136–143.

Viitala, S.M., Schulman, N.F., de Koning, D.J., Elo, K., Kinos, R., Virta, A., Virta, J., Maki-Tanila, A. & Vilkki, H.J. 2003. Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle. *J. Dairy Sci*. 86: 1828–1836.

Vilkki, H.J., de Koning, D.J., Elo, K.T., Velmala, R. & Maki-Tanila, A. 1997. Multiple marker mapping of quantitative trait loci of Finnish dairy cattle by regression, *J. Dairy Sci.* 80: 198–204.

Visscher, P.M., Haley, C.S. & Thompson, R. 1996. Marker-assisted introgression in backcross breeding programs. *Genetics* 144: 1923–1932.

Visscher, P.M., Thompson, R. & Haley, C.S. 1996. Confidence intervals in QTL mapping by bootstrapping. *Genetics* 143: 1013–1020.

Weller, J.I. 1986. Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. *Biometrics* 42: 627–640.

Weller, J.I. 1994. *Economic aspects of animal breeding.* London, Chapman & Hall. 244 pp.

Weller, J.I. 2001. *Quantitative trait loci analysis in animals*. London, CABI Publishing. 287 pp.

Weller, J.I. & Soller, M. 2004. An analytical formula to estimate confidence interval of QTL location with a saturated genetic map as a function of experimental design. *Theor. Appl. Genet.* 109: 1224–1229.

Weller, J. I., Kashi, Y. & Soller, M. 1990. Power of "daughter" and "granddaughter" designs for genetic mapping of quantitative traits in dairy cattle using genetic markers. *J. Dairy Sci*. 73: 2525–2537.

Weller, J.I., Shlezinger, M. & Ron, M. 2005. Correcting for bias in estimation of quantitative trait loci effects. *Genet. Sel. Evol*. 37: 501–522.

Weller, J.I., Soller, M. & Brody, T. 1988. Linkage analysis of quantitative traits in an interspecific cross of tomato (*L. esculentum* x *L. pimpinellifolium*) by means of genetic markers. *Genetics* 118: 329–339.

Weller, J.I., Song, J.-Z., Heyen, D.W., Lewin, H.A. & Ron M. 1998. A new approach to the problem of multiple comparisons in the genetic dissection of complex traits. *Genetics* 150: 1699–1706.

Weller, J.I., Weller, H., Kliger, D. & Ron, M. 2002. Estimation of quantitative trait locus allele frequency via a modified granddaughter design. *Genetics* 162: 841–849.

Weller, J.I., Golik, M., Seroussi, E., Ezra, E. & Ron, M. 2003. Population-wide analysis of a QTL affecting milk-fat production in the Israeli Holstein population. *J. Dairy Sci*. 86: 2219–2227.

Weller, J.I., Feldmesser, E., Golik, M., Tager-Cohen, I., Domochovsky, R., Alus, O., Ezra, E. & Ron, M. 2004. Factors affecting incorrect paternity assignment in the Israeli Holstein population. *J. Dairy Sci.* 87: 2627–2640.

Westell, R.A. & van Vleck, L.D. 1987. Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. *J. Dairy Sci*. 70: 1106–1017.

Whittaker, J.C., Thompsom, R. & Visscher, P.M. 1996. On the mapping of QTL by regression of phenotype on marker-type. *Heredity* 77: 23–32.

Winter, A., Kramer, W., Werner, F.A.O., Kollers, S., Kata, S., Durstewitz, G., Buitkamp, J., Womack, J.E., Thaller, G. & Fries, R. 2002. Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. *Proc. Nat. Acad. Sci. USA* 99: 9300–9305.

Zhang, Q., Boichard, D., Hoeschele, I., Ernst, C., Eggen, A., Murkve, B., Pfister-Genskow, M., Witte, L.A., Grignola, F.E., Uimari, P., Thaller, G. & Bishop, M.D. 1998. Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. *Genetics* 149: 1959–1973.