DETECTION AND UTILISATION OF QUANTITATIVE TRAIT LOCI IN DAIRY CATTLE

Promotoren:	dr ir E.W. Brascamp hoogleraar in de veefokkerij
	dr D.J. Garrick A.L. Rae Chair in Animal Breeding & Genetics, Massey University, New Zealand
Co-promotor:	dr ir J.A.M. van Arendonk persoonlijk hoogleraar bij de Leerstoelgroep Fokkerij en Genetica

DETECTION AND UTILISATION OF QUANTITATIVE TRAIT LOCI IN DAIRY CATTLE

RICHARD J. SPELMAN

Proefschrift

Ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van de Landbouwuniversiteit Wageningen,
dr. C.M. Karssen,
in het openbaar te verdeligen
op vrijdag 4 september 1998
des namiddags te vier uur in de Aula

ISN 957-913

Detection and utilisation of quantitative trait loci in dairy cattle.
SPELMAN, RICHARD JOHN 1998.

Doctoral thesis, Wageningen Agricultural University, P.O. Box 338, 6700 AH Wageningen, The Netherlands.

ISBN: 90-5485-896-6

Cover: Kathy Morgan Art & Design, Palmerston North, New Zealand.

ABSTRACT

This thesis focuses on the detection of quantitative trait loci in dairy cattle and their utilisation in breeding programmes. Analysis of one bovine chromosome for quantitative trait loci for milk production traits is described. Through stochastic simulation, the effect of incorrect parameter estimates for quantitative trait locus effect and position on genetic response from marker assisted selection is investigated. Also through stochastic simulation the effect of reducing flanking-marker bracket size on genetic response from marker assisted selection is examined. Strategies to confirm the existence and size of quantitative trait loci identified in a genome scan are outlined. Simulation is used to estimate improvements in rate of genetic gain from marker assisted selection for two scenarios, the current situation and a futuristic setting. The general discussion of this thesis addresses the use of significance levels in quantitative trait loci detection, experimental designs to identify further quantitative trait loci in the New Zealand dairy industry, and the current and possible future application of marker assisted selection in dairy breeding programmes.

Dedicated to Nana and Granddad

...do not waste time and energy on things that you can not control, concentrate on things that you can.

...a positive mind is an incredibly powerful entity.

STELLINGEN

1. To ensure near uniform information content over a chromosome, highly informative markers or two markers should be positioned at the end of the chromosome (this thesis).

2. Verification studies should be undertaken before QTL are implemented in marker assisted selection (this thesis).

3. Reproductive technologies need to be used when applying within family marker assisted selection (this thesis).

4. Selective DNA pooling is a very powerful experimental design to detect QTL in dairy cattle (DARVASI and SOLLER 1994, this thesis).

5. Thresholds are not absurd – people who use them foolishly are (LANDER and KRUGLYAK 1996).

6. The finding of TERWILLIGER et al. (1997) that true positive peaks in genome scans are expected to be broader than false positive peaks could be utilised with the height of the peak to decrease the false positive rate.

7. Selection of parents based solely on BLUP estimated breeding values and thus ignoring average co-ancestry of selected animals (equivalently inbreeding), is a short-sighted breeding strategy.

8. The conclusion of DE ROO (1988) that avoiding the mating of relatives only postpones, but does not prevent the increase in inbreeding, has been disproved by CABALLERO et al. (1996), who have shown that the rate of long-term inbreeding can be influenced by controlled mating systems.

9. Highly trained triathletes can improve their performance from a high intensity, low volume taper, lasting between 9 and 15 days (LEE and MACLEOD 1998).

10. Dutch real estate agents should have shares in home decorating businesses.

11. Een Kiwi is een inheemse Nieuw Zeelandse vogel, of een ander woord voor een Nieuw Zeelander, maar geen fruit.

Richard Spelman - Detection and utilisation of quantitative trait loci in dairy cattle.
September 4th 1998, Wageningen.

PREFACE

I am appreciative of the opportunity and support given by Livestock Improvement Corporation to undertake my PhD. Specifically, Harvey Tempero for stimulating my interest in animal breeding, and Brian Wickham for giving me the possibility to further my education.

Thanks to Johan van Arendonk and Dorian Garrick for your support, ideas and supervision over the last three years while I have been at Wageningen Agricultural University and Massey University, respectively. I am very appreciative of the patience and 'wisdom' of Henk Bovenhuis in answering my daily questions. Thanks also to Pim Brascamp for your constructive comments on my manuscripts.

Over the three years studying both in Wageningen and Massey, I have had the opportunity to work and socialise with a multifarious group of people. Firstly, the 'Kiwis', cheers to PC, Nicolas, Pav, Satish, Luis, Alistair, Jose, Reddy, Annor, Hugh, Pauline, Donna. Secondly, the 'Dutchies', bedankt to Marco, Jan-Thijs, Piter, Vera, Bronwyn, Richard, Dirk-Jan, Bart, Jaco, Gerard, Andy, Ab, Martien, Alex, Luc, Egbert, Erwin, Delphin, Jascha, Maria, Carolien, Antonia, Sijne, Tette, Jan, Liesbeth, Pim, Birgitte, Bernardo, Ant, Ada.

Special thanks to Kaye, for your support and interest, and of course thanks to Mum and Dad for your encouragement from an early age in my education.

...one's limits are those that we set ourselves.

TABLE OF CONTENTS

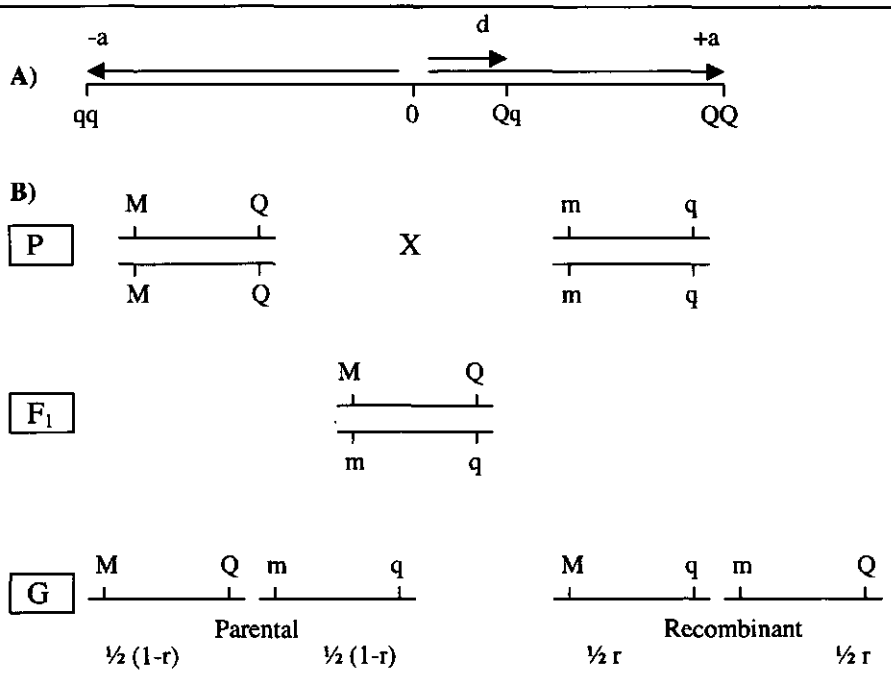
CHAPTER ONE

GENERAL INTRODUCTION

SAX (1923) was the first to demonstrate linkage between a Mendelian marker and a locus that affected a quantitative trait (QTL). SAX's experiment with garden beans (*Phaseolus vulgaris*) involved the crossing of a large-seed "eyed" variety to a small-seed white variety. The parental types differed greatly in bean (seed) weight and also at 3 Mendelian loci that affected bean colour: *P/p* (colour/white), *T/t* (extended colour/"eye") and *Y/y* (mottled/uniform). The F_1 plants were self-fertilised, and the resulting F_2 were weighed and classified according to seed coat pattern. In the F_2 generation, the white beans (*pp*) weighed significantly less than the coloured beans (*P*-); uniform beans (*yy*) weighed less than mottled (*Y*-) and extended beans (*T*-) less than "eyed" (*tt*). Therefore, the marker alleles *P*, *t*, and *Y* were found to be associated with factors contributing to large seed size, even though the *Y* allele came from the small parent. SOLLER (1990) nicely illustrated the phenomenon of linkage between a marker and QTL, for a cross between inbred lines (Figure 1, adapted from SOLLER 1990).

In 1961, NEIMANN-SORENSON and ROBERTSON undertook one of the first attempts to identify associations between blood groups and milk production traits in three Danish cattle breeds. They did not detect any significant associations between the blood group genes and genes that had measurable effects on the milk production traits. However, the small number of blood groups and serum protein polymorphisms limited that study and other studies in subsequent years that used these polymorphisms as markers.

Restriction fragment length polymorphisms (RFLPs) DNA markers, developed in the 1970's, removed the limitation on number of markers with the potential of over 100,000 RFLP loci. Two limitations that RFLP markers had were the time required for genotyping and more importantly their low heterozygosity. Marker development continued and minisatellites and microsatellites were discovered. Today microsatellites have become the marker of preference, which may be replaced by single nucleotide polymorphisms in the near future. There are many microsatellite loci distributed evenly over the genome, and each locus usually has many alleles and therefore is highly informative. With the identification of this new class of marker(s) and the development of the polymerase chain reaction, it has enabled highly



C)

QTL genotypes	Marker genotypes MM	mm
QQ	$(1-r)^2$	r^2
Qq	$2(1-r)r$	$2(1-r)r$
qq	r^2	$(1-r)^2$

D) $\overline{MM} - \overline{mm} = 2(1-2r)a$

FIGURE 1: Linkage between a marker locus and QTL in the F_2 generation of a cross between inbred lines. M and m, are the alleles at the marker locus; Q and q, are the alleles at the QTL; a and d, are the main effect and dominance effect at the QTL; and r is the proportion of recombination between the marker locus and QTL. **A)** The QTL gene effects are shown for the two homozygotes and the heterozygote QTL classes. **B)** The genotypes of the parents indicate that M and Q are linked in one parental inbred line, and m and q are linked in the other parental line. The F_1 progeny are heterozygous for both the marker and QTL loci, with MQ inherited from one parental line and mq from the other. In the gametes (G) formed by the F_1, there are 4 classes: parental gametes where no recombination between marker and QTL loci has occurred, and recombinant gametes where recombination has occurred. The frequency of parental and recombinant F_1 gametes are equal when no linkage between marker and QTL (r = 0.5). **C)** The relative QTL genotype frequencies for the F_2 progeny that have alternative homozygous marker genotypes. **D)** The mean of the MM F_2 progeny ($\overline{MM}$) is the frequency of the QTL genotype multiplied by the gene effects given in A). The mean for the mm F_2 progeny ($\overline{mm}$) can be calculated in the same manner. The difference between the mean values of the alternative homozygous marker genotypes in the F_2 generation ($\overline{MM} - \overline{mm}$), produces an estimate where the effects of a and r are confounded. Interval mapping allows separate estimates of QTL additive effect (a) and recombination rate (r).

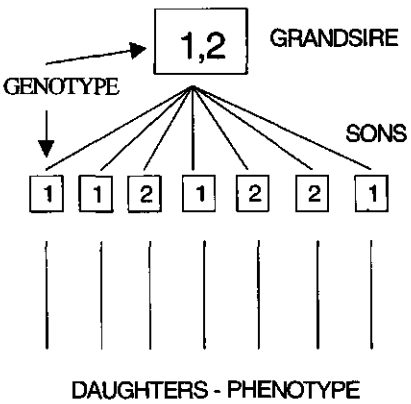
informative marker loci to be identified with lower time requirements than that needed in the late 1970's. SOLLER (1990) stated that mapping the bovine genome with respect to many of the QTL, was a realisable project and a goal that should be seriously considered in setting research priorities for the dairy cattle community in the last decade of the century.

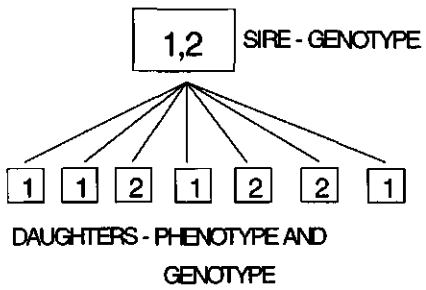
The increase in the number of DNA markers and development of methods for genotyping, enabled large-scale projects to be undertaken to identify marker-QTL associations. Two designs to identify marker-QTL associations in dairy cattle were investigated by WELLER *et al.* (1990); i.e. the granddaughter design (Figure 2a) and the daughter design (Figure 2b). In the granddaughter design a sire and his progeny-tested sons are genotyped and phenotypic records are collected on the granddaughters of the sire, to calculate daughter group means for the sons. Segregation of heterozygous marker loci is followed for the two alleles from the sire to each of the sons. Significant differences in daughter group means, between the two groups of sons inheriting alternate sire alleles at a marker locus, indicates the presence of a linked QTL. This is the same as the situation outlined in Figure 1, but in the granddaughter design the grandsire is preferably heterozygous at marker and QTL loci and marker information from the dam is ignored. The daughter design is similar to the granddaughter design but daughters of the sire are genotyped and phenotyped. The granddaughter design has 3-4 times more statistical power than the daughter design, for the same number of genotype assessments, due to greater accuracy of measuring genetic differences in performance from daughter group means than from single observations on lactating cows (WELLER 1990).

When marker-QTL associations have been identified and located to chromosomal segments, the marked QTL can be utilised in breeding schemes by marker assisted selection. Favourable theoretical genetic (SOLLER 1978; KASHI *et al.* 1990; MEUWISSEN and VAN ARENDONK 1992) and economic (BRASCAMP *et al.* 1993) responses to marker assisted selection have been reported for dairy cattle breeding schemes.

FIGURE 2a: Granddaughter design: Grandsire is heterozygous for a marker and the two alleles (1,2) are traced to the grandsire's sons.

FIGURE 2b: Daughter design: Sire is heterozygous for a marker and the two alleles (1,2) are traced to the sire's daughters.





AIM

The aim of this thesis is to identify and resolve aspects in QTL identification and their successful utilisation by marker assisted selection in dairy cattle. The thesis examines issues of QTL detection and the sensitivity of marker assisted selection to differing levels of precision in estimation of QTL location and variance. The thesis also studies different marker assisted selection strategies for dairy cattle breeding programmes.

OUTLINE

Livestock Improvement Corporation (New Zealand) and Holland Genetics (The Netherlands) established a QTL experiment involving granddaughter and daughter designs. The objective of the experiment was to identify chromosomal regions that affect milk production traits and utilise these regions in marker assisted selection breeding programmes.

In Chapter 2, bovine chromosome *six* was analysed for associations between nine microsatellite markers and five milk production traits; fat yield, protein yield, milk yield, fat percentage, and protein percentage, in the Dutch Holstein-Friesian population. Estimates for QTL effect or variance and location, such as those

estimated for chromosome *six*, will be used in marker assisted selection. Ninety-five percent confidence intervals for QTL location from an experiment of this size are some 50 centiMorgans (cM), and QTL effect estimates are likely to be overestimated. In Chapter 3, we investigate the effect of inaccurate estimation of QTL variance and location on marker assisted selection genetic response, through simulation of an adult multiple ovulation and embryo transfer nucleus breeding scheme. Marker assisted selection genetic response was calculated for QTL variation being over-estimated by a factor of 2 and 3, and QTL location being in error by 5, 10 and 15 cM. In addition, marker assisted selection genetic response for a non-existent QTL that was estimated to explain 5% or 10% of the phenotypic variance was investigated.

Fine mapping molecular tools can improve the precision of QTL location, and in the future may lead to the identification of the gene itself. The impact on marker assisted selection genetic response from getting closer to the QTL, and thus having smaller flanking QTL-marker bracket sizes, is investigated in Chapter 4. In addition, the genetic response from having two QTL identified on the same and different chromosome(s) is investigated. Both of these aspects are investigated with the same stochastic model used in Chapter 3.

Results from QTL experiments can not be readily implemented into breeding schemes through marker assisted selection. This is due to uncertainty about whether the QTL identified in the experiments are real or statistical artefacts, and whether the QTL are segregating in the current breeding population. In Chapter 5, methods are outlined that can be used to confirm QTL results. These methods include the establishment of another experiment or combining the results from different experiments.

Marker assisted selection schemes that utilise QTL information to pre-select progeny test bulls on a within-family basis are the most practical application of QTL results in the short-term. This is due to technical difficulties with ungenotyped animals in across-family marker assisted selection schemes that use BLUP procedures incorporating marker information. Two within-family marker assisted selection schemes were evaluated genetically and economically in Chapter 6, using stochastic

simulation for the New Zealand dairy breeding scheme. The importance of female reproductive performance to the genetic and economic response for marker assisted selection was also investigated.

Molecular technology has progressed dramatically in the last five years, and is expected to continue to improve at least at the same rate over the next five years. Therefore today's restrictions may be non-existent in the near future. In Chapter 7, marker assisted selection schemes are investigated where a large proportion of the genetic variation is identified, and the loci are known, or are in linkage disequilibrium with a marker, thus enabling across-family marker assisted selection.

In the general discussion, firstly, the setting of critical values is discussed and outlined with different methods. Secondly, experimental designs to identify and mark further genetic variance in the New Zealand dairy industry, and their experimental power, are outlined. The major factors that have contributed to a wide variety of simulated genetic responses from marker assisted selection are detailed, and the implementation of marker assisted selection in the New Zealand dairy industry by Livestock Improvement Corporation is described. Finally some thoughts on the application of MAS in the future are given.

LITERATURE CITED

BRASCAMP, E. W., J.A.M. VAN ARENDONK, and A. F. GROEN, 1993 Economic appraisal of the utilisation of genetic markers in dairy cattle breeding. J. Dairy Sci. **76:** 1204-1213.

KASHI, Y., E. HALLERMAN, and M. SOLLER, 1990 Marker assisted selection of candidate bulls for progeny testing programmes. Anim. Prod. **51:** 63-74.

MEUWISSEN, T.H.E., and J.A.M. VAN ARENDONK, 1992 Potential improvements in rate of genetic gain from marker assisted selection in dairy cattle breeding schemes. J. Dairy Sci. **75:** 1651-1659.

NEIMANN-SORENSON, A., and A. ROBERTSON, 1961 The association between blood groups and several production characteristics in three Danish cattle breeds. Acta. Agric. Scan. **11:** 163-196.

SAX, K., 1923 The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics **8:** 552-560.

SOLLER, M., 1978 The use of loci associated with quantitative effects in dairy cattle improvement. Anim. Prod. **27:** 133-139.

SOLLER, M., 1990 Genetic mapping of the bovine genome using deoxyribonucleic acid-level markers to identify loci affecting quantitative traits of economic importance. J. Dairy Sci. **73:** 2628-2646.

WELLER, J. I., Y. KASHI, and M. SOLLER, 1990 Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J. Dairy Sci. **73:** 2525-2537.

CHAPTER TWO

QUANTITATIVE TRAIT LOCI ANALYSIS FOR FIVE MILK PRODUCTION TRAITS ON CHROMOSOME *SIX* IN THE DUTCH HOLSTEIN-FRIESIAN POPULATION

RICHARD J. SPELMAN, WOUTER COPPIETERS, LATIFA KARIM,

JOHAN A.M. VAN ARENDONK and HENK BOVENHUIS

PUBLISHED IN GENETICS (1996) **144:** 1799-1808.

ABSTRACT

Twenty Dutch Holstein-Friesian families, with a total of 715 sires, were evaluated in a granddaughter experiment design for marker-QTL associations. Five traits–milk, fat and protein yield and fat and protein percent–were analysed. Across-family analysis was undertaken using multimarker regression principles. One and two QTL models were fitted. Critical values for the test statistic were calculated empirically by permuting the data. Individual trait distributions of permuted test statistics differed and thus distributions had to be calculated for each trait. Experimentwise critical values, which account for evaluating marker-QTL associations on all 29 autosomal bovine chromosomes and for five traits, were calculated. A QTL for protein percent was identified in one and two QTL models and was significant at the 1 and 2% level, respectively. Extending the multimarker regression approach to an analysis including two QTL was limited by families not being informative at all markers, which resulted in singularity. Below average heterozygosity for the first and last marker lowered information content for the first and last marker bracket. Highly informative markers at the ends of the mapped chromosome would overcome the decrease in information content in the first and last marker bracket and singularity for the two QTL model.

KEY WORDS: Dairy cattle, Quantitative trait loci, Chromosome *six*, Granddaughter design, Regression analysis

INTRODUCTION

Use of deoxyribonucleic acid (DNA) markers to search for loci that affect quantitative traits, known as quantitative trait loci (QTL), has become widespread in recent times. Identifying marker-QTL associations in farm animals may be undertaken in various experimental settings including the so-called 'daughter' or 'granddaughter' designs (GELDERMANN 1975; WELLER *et al.* 1990; VAN DER BEEK *et al.* 1995). Analytical techniques have been developed to identify marker-QTL associations (*e.g.*, WELLER 1986; LANDER and BOTSTEIN 1989; HALEY and KNOTT 1992).

KNOTT *et al.* (1994) developed a multimarker regression method to determine position and effect of QTL. The multimarker technique was demonstrated on simulated data for a half-sib population (KNOTT *et al.* 1994). The issue of calculating appropriate critical values that account for repeated testing has been addressed (*e.g.*, HALEY *et al.* 1994; JANSEN 1993; CHURCHILL and DOERGE 1994). CHURCHILL and DOERGE (1994) developed an empirical method based on the concept of the permutation test and illustrated the method on real data sets derived from F_2 and recombinant inbred plant populations and simulated data from a backcross design.

GEORGES *et al.* (1995) reported five chromosomes that gave evidence (LOD score 3) for the presence of a QTL controlling milk yield in the American Holstein population. Chromosome *six* was one of the five chromosomes identified. The QTL on chromosome *six* increased milk yield but not fat or protein yield and as a result protein and fat percent decreased. BOVENHUIS and WELLER (1994) reported an effect for fat percent that was linked to the casein locus, which is also found on chromosome *six*.

The objective of this study is to demonstrate the application of KNOTT *et al.*'s (1994) multimarker approach and CHURCHILL and DOERGE'S (1994) empirical method of calculating critical values to outbred dairy population data generated from a granddaughter design. Both methods are extended; KNOTT *et al.*'s (1994) multimarker approach to a two-QTL model and CHURCHILL and DOERGE'S (1994) permutation test to accommodate multiple traits in the calculation of critical values. The application of these methods is demonstrated for chromosome *six*.

MATERIALS AND METHOD

Experiment structure: Twenty Holstein-Friesian families from the Netherlands in a granddaughter experiment design were evaluated for marker-QTL associations. Average number of sons per grandsire is 36 with a range of 12–140 (Table 1). To avoid selection bias and its influence on detecting QTL (as described by MACKINNON and GEORGES 1992) selected sons were scrutinised by date of progeny testing within each grandsire family. When there was DNA (semen samples) for only some sons that were progeny tested during a given period, information from this group

of selected sons was not used. The grandsire family for the selected sons was not necessarily removed as there were time periods when all of the progeny tested sons had semen samples retained. Some 80 selected sons were not analysed (selected sons are not in Table 1). If a son was not informative at any of the markers he was still retained in the analysis as he contributed to calculation of the fixed effect of the grandsire (Equation 1).

TABLE 1: Experimental design and genetic markers used for chromosome *six*.

	Marker										
	1	2	3	4	5	6	7	8	9		
Grandsire	BM1329	BM143	TGLA37	BM4528	BM415	KCAS	BM4311	BP7	BM2320	Total	Sons
A		1			1		1	1		4	13
B		1		1	1	1	1	1		6	40
C		1			1		1	1		4	22
D		1	1		1	1			1	5	12
E					1		1	1		3	16
F		1			1		1	1	1	5	32
G		1			1	1		1		4	42
H	1	1		1			1		1	5	140
I	1	1	1	1	1	1	1	1		8	20
J		1		1		1			1	4	54
K		1	1	1	1	1				5	23
L		1	1	1	1		1	1	1	7	71
M		1		1		1	1	1	1	6	26
N	1	1	1	1	1	1		1	1	8	12
O		1	1		1	1	1	1		6	75
P	1	1			1		1	1	1	6	60
Q		1	1				1			3	15
R		1		1		1	1	1	1	6	14
S				1					1	2	16
T		1	1	1	1		1	1		6	12
Total	4	18	8	11	14	10	14	14	10		
Map (cM)	0	13	20	31	41	52	54	58	95		

The table details the markers for which grandsires are heterozygous (indicated by a 1) and the total for each grandsire, number of sons for each grandsire, number of grandsires heterozygous at each marker, and marker distances based on HALDANE'S (1919) mapping function.

Nine microsatellite markers were positioned and ordered on chromosome *six* with the ANIMAP programs (D. NIELSON and M. GEORGES unpublished) as described by GEORGES *et al.* (1995). The map for chromosome *six* is 95cM long using HALDANE'S (1919) mapping function (Table 1). For one of the nine markers, the position could not be determined unambiguously. The odds for switching marker six (casein locus) and marker seven were only 2.6:1 in favour of the order that was used in the analysis. With the exception of the orientation of marker six and seven and marker three (TGLA37, GEORGES *et al.* 1995) the map in this study corresponds to that of BISHOP *et al.* (1994).

Grandsire heterozygosity was on average 57% for the nine markers. However, there was large variation in the heterozygosity of the grandsires over the nine markers and also heterozygosity level between markers (Table 1). When the grandsire was heterozygous at a marker locus, it was, on average, known with certainty in 65% of cases which marker allele was transmitted from grandsire to son.

Five traits were analysed for marker-QTL effects; milk, fat and protein yield (termed yield traits) and fat and protein percent (termed percentage traits). Daughter yield deviations (DYDs), weighted averages of a sire's daughter's lactation performances expressed as deviations from the population mean (VAN RADEN and WIGGANS 1991) were used as the phenotypic measurement. DYDs for the percentage traits were calculated from the yield traits. DYDs were taken from the September 1995 evaluation conducted by the Royal Dutch Cattle Syndicate.

Power of this design using the method of WELLER *et al.* (1990) was 0.6 for a bi-allelic QTL of size (half the difference in genetic value between homozygotes) 0.2 phenotypic standard deviation with equal allele frequency for a trait with heritability of 0.3 (*e.g.*, yield traits), type I error (comparisonwise) set to 0.05 and no recombination between marker and QTL with fully informative markers. Power was 0.9 for the same criteria but for a trait with heritability of 0.6 (*e.g.*, percentage traits) and a QTL effect of 0.4 phenotypic standard deviation.

One QTL analysis: Analysis was undertaken using multimarker regression principles as developed by KNOTT *et al.* (1994). Basic steps of multimarker regression were determination of the most likely haplotypes of the two grandsire gametes based on genotypes of his sons. The most likely linkage phase was taken and when both phases were equally likely, one was selected at random. The QTL allele of interest was arbitrarily assigned to the linkage phase denoted linkage phase one. The probability of inheriting the chromosomal segment of linkage phase one at any position was calculated for each son based on information from the closest informative flanking markers. DYDs were then regressed on this conditional probability.

Across family analysis was undertaken by fitting a one QTL model to the data:

$$Y_{ijk} = \mu + gs_i + b_{ik}X_{ijk} + e_{ijk} \qquad [1]$$

where Y_{ijk} is the DYD for the j^{th} son of the i^{th} grandsire at the k^{th} chromosomal position, μ is the overall mean, gs_i is the fixed effect of the i^{th} grandsire, b_{ik} is the regression coefficient for the i^{th} grandsire at the k^{th} chromosomal position, X_{ijk} is the probability of the j^{th} son receiving the chromosomal segment for gamete one from the i^{th} grandsire at the k^{th} position, and e_{ijk} is the random residual.

This model allows multiple QTL alleles. Each grandsire family was constrained to a bi-allelic QTL as only the transmission of marker alleles from the grandsire were considered and grandams contribution ignored. Thus over the twenty families there were, in total, forty possible alleles. For across-family analysis, residual sums of squares (RSS) were summed across families, thus the larger grandsire families contributed to a larger extent to overall RSS. Within grandsire families the number of daughters that each son had varied from those which had only a part progeny test proof to sires that were used extensively as proven sires. Contribution of each sire was weighted according to the number of daughters contributing to the DYD. The weighting factor was based on the variance of the DYD for a son being:

$$\text{Var DYD} = \left[\frac{1+(n-1)\frac{1}{4}h^2}{n}\right]\sigma_p^2 \qquad [2a]$$

where Var DYD is the variance of son's DYD; n is the number of daughters contributing to the DYD; h^2 is the heritability, which was taken as 0.35 for yield traits and 0.75 for percentage traits (average heritabilities from VAN DER WERF and DE BOER 1989); and σ_p^2 is the phenotypic variance.

Assuming equal phenotypic variance for all observations the weighting factor (w) is:

$$w = \left[\frac{1+(n-1)\frac{1}{4}h^2}{n}\right] \qquad [2b]$$

The weighted residual sums of squares is

$$\sum_{j=1}^{n}\frac{1}{wi_j}(y_{ijk} - \mu - gs_i - b_{ik}X_{ijk})^2 \qquad [3a]$$

and the sums of squares explained by the QTL fitted in the model is:

$$\sum_{j=1}^{n}\frac{1}{wi_j}(y_{ijk} - \mu - gs_i)^2 - \sum_{j=1}^{n}\frac{1}{wi_j}(y_{ijk} - \mu - gs_i - b_{ik}X_{ijk})^2 \qquad [3b]$$

where w_{ij} is the weighting factor (equation 2b) for the j^{th} son of the i^{th} grandsire. Equation 3b is equivalent to $R(QTL \mid \mu, g_{si})$ (reduction in residual sums of squares) where QTL represents the QTL fitted (*i.e.* b_{ik} and X_{ijk})

Test statistics were calculated similar to a *F* statistic but were not termed as such because the distribution of the test statistics did not follow a *F* distribution. Test statistics were calculated every centiMorgan over the mapped chromosome.

Critical values: Test statistic critical values were calculated empirically from the permutation method outlined by CHURCHILL and DOERGE (1994). In brief, the permutation test was undertaken by repeatedly randomly shuffling the phenotypic data (DYDs with their weighting factors) within each family and calculating test statistics for each shuffle. The conditional probabilities (X_{ijk}'s) that the DYDs are regressed on were not shuffled. Critical values were calculated from the distribution of test statistics. Comparisonwise, chromosomewise and experimentwise critical values were calculated. Comparisonwise values were calculated each centiMorgan and provided critical values for that point but did not account for repeated testing over the genome or for the five different traits. Chromosomewise values accounted for the multiple, dependent, testing on chromosome *six* and the five traits analysed. The experimentwise critical values account for evaluation of marker-QTL associations on 29 autosomal bovine chromosomes and also the five traits being analysed.

Two-QTL analysis: A two-QTL model was fitted to the data by extending the multimarker regression one-QTL model. The two-QTL model is:

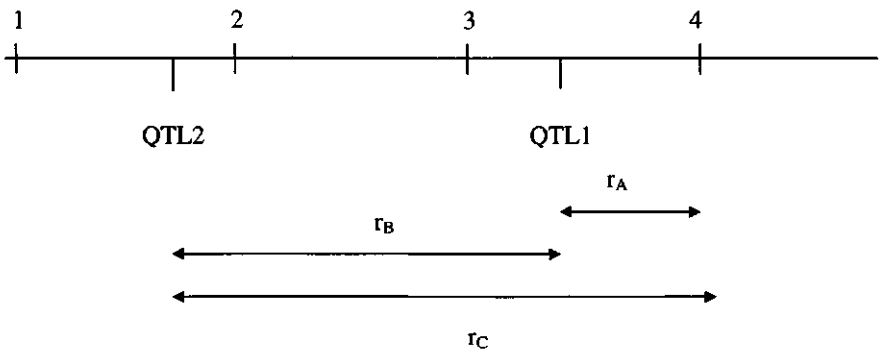
$$Y_{ijk1k2} = \mu + gs_i + b_{ik1}X_{ijk1} + b_{ik2}X_{ijk2} + e_{ijk} \qquad [4]$$

where k_1 and k_2 refer to the position of the first and second QTL. Other terms are as in [1].

The two QTL model was fitted by grid searching *i.e.* each combination of 1-cM positions was evaluated. However, HALEY and KNOTT (1992) observed that QTLs 20cM apart could not be differentiated. To ensure that the two postulated QTL had some distance between them, it was decided that only those positions where they were separated by an empty marker bracket would be evaluated. Having an empty marker bracket between postulated QTL was in agreement with ZENG (1993). He reported that two sample partial coefficients are generally uncorrelated unless the two markers are adjacent markers. An empty marker bracket between postulated QTL was not possible for all families as they were not informative at all marker loci (Table 1). For example, when a QTL was fitted in marker bracket one and the second QTL in marker

bracket three, some families did not have an empty marker bracket between both QTL (Figure 1) because they were not informative at one or both of markers 2 and 3.

FIGURE 1. Example of fitting a two-QTL model. Numbers 1-4 are the position of the markers and QTL1 and QTL2 are the positions of the postulated QTLs. Recombination rate between informative marker 4 and QTL1 and QTL2 are denoted r_A and r_C, respectively. Recombination rate between QTLs is r_B.



Further, if the family was not informative at markers 1, 2 and 3, the two QTL were placed to the left of the first informative marker, marker 4. The probabilities of transmission of the QTL were calculated from information derived from marker 4 (Figure 1) using the technique of KNOTT *et al.* (1994). Thus there are only two groups of progeny, depending on the allele that they inherit at marker 4. The probability for inheriting a given allele at each of the QTL is the same for all individuals within a group. This can be demonstrated mathematically. Utilising HALDANE'S (1919) equation we know that:

$$r_C = r_A + r_B - 2r_Ar_B \qquad [5]$$

and if the probability of receiving QTL1 (X_{ijk1}) is:

$$\text{probability QTL1} = (1 - r_A) \qquad [6]$$

and probability for QTL2 (X_{ijk2}) is:

$$\text{probability QTL2} = (1 - r_C) \quad [6a]$$

then utilising Equation 5:

$$\begin{aligned}\text{probability QTL2} &= 1 - (r_A + r_B - 2r_Ar_B) \\ &= (1 - r_A) - r_B + 2r_Ar_B \\ &= \text{probability QTL1} - r_B + 2r_Ar_B\end{aligned} \quad [6b]$$

which is equivalent to:

$$\text{probability QTL2} = \text{probability QTL1} + \text{constant.} \quad [6c]$$

This results in singularity. Three of the 20 families (E, K, S; Table 1) were uninformative for the first or last three markers and thus excluded from the two QTL across family analysis.

Two test statistics are calculated for the two-QTL model. One test statistic compares the fit of the two-QTL model to that corresponding with the position of the highest test statistic in the one-QTL model (Equation 7c). The second test statistic determined if neither, one or both positions explain a significant amount of the variance in the two-QTL model by the following method (Equation 7, a and b]. The following reduction in sums of squares were calculated.

$$R(X_{ijk1} | \mu, gs_i, X_{ijk2}) \quad [7a]$$

$$R(X_{ijk2} | \mu, gs_i, X_{ijk1}) \quad [7b]$$

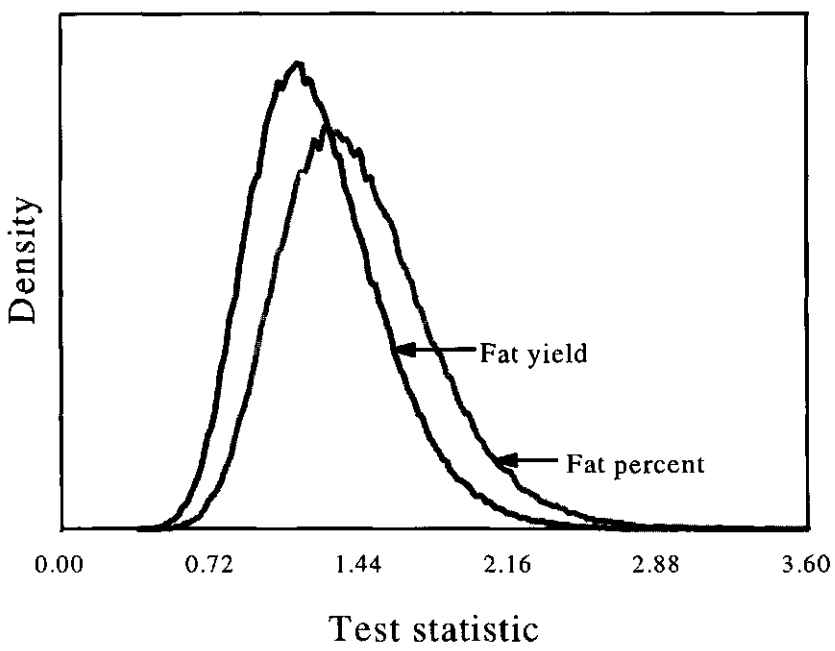
$$R(X_{ijk1}, X_{ijk2}, \mu, gs_i,) - R(X_{ijk}, \mu, gs_i) \quad [7c]$$

where k corresponds to the position with the highest explained sums of squares for the one-QTL model. If neither Equation 7a nor 7b is significant, then neither of the two QTL positions are significant, otherwise at least one of the two positions is significant.

RESULTS

Permutation Test: Distribution of test statistics between traits differed quite markedly. For example, fat percent had a larger proportion of higher permuted test statistics than fat yield (Figure 2). This is demonstrated numerically by the critical value at the 1% threshold level for fat yield being nearly equal to that at the 5% level for fat percent (Table 2). The mean of the test statistic distributions in Figure 2 are not one as would be expected with a F distribution. This is due to that the test statistic distributions in Figure 2 account for repeated testing across the chromosome and thus are not comparisonwise test statistics. CHURCHILL and DOERGE (1994) would refer to these distributions as chromosomewise but in this paper they are not as they do not account for repeated tests on the five correlated traits.

FIGURE 2: Approximate density function of test statistics for fat yield and fat percent derived from permutation test (150,000 shuffles).



To account for the five traits being analysed, the highest permuted value for the five traits from each shuffle (traits shuffled together) was used and combined critical values calculated (Table 2). However, with between trait differences, the highest value for each shuffle were dominated by traits that had higher absolute

critical values *i.e.*, fat percent in this study. Thus combined trait critical values were not applicable to the individual traits especially those traits with lower distribution of critical values.

TABLE 2: Chromosomewise threshold levels

Threshold level	Milk	Fat	Protein	Fat %	Protein %	Combined	F values
0.1%	2.95	2.67	2.79	3.11	3.00	3.04	2.33
1%	2.57	2.32	2.41	2.65	2.58	2.61	2.13
5%	2.26	2.03	2.11	2.32	2.26	2.28	1.82
10%	2.11	1.89	1.98	2.16	2.11	2.12	1.69

The chromosomewise critical values for the five milk production traits account for repeated testing over chromosome *six* and on the five correlated traits or equivalently three independent traits (150,000 shuffles). The critical values in the combined column are when the highest test value is taken from each shuffle of the five traits. The *F* values are tabulated values that have been adjusted with Bonferroni correction for testing on three independent traits.

The approach taken in this study was to estimate the equivalent number of independent traits tested. This was calculated by factor analysis (using SAS 1985) on a genetic correlation matrix for the five traits (VAN DER WERF and DE BOER 1989) and on the experimental phenotypic data. It was calculated that two factors account for approximately 90% of the variation and three factors account for some 99% on both the correlation matrix and DYDs. This was checked by analysing each shuffle of the permuted test statistics for the five traits. In each shuffle, it was determined whether the permuted test statistics for each trait was significant at a certain threshold level using individual trait critical values. It was assumed that if the data was equivalent to x independent traits, then at the 10% threshold level (for a single trait) from 10,000 shuffles there would be $10000 * 0.1^x$ occurrences where all five traits in the one shuffle were significant. Solving for x at the 10% threshold level with 10,000 permuted *F* values, gave 2.8 independent traits, which agrees with the results from the factor analysis. Based on these considerations three independent traits were taken.

Equation 8a calculates the probability (p) of false positives at a given type I error (α) with n independent tests:

$$p = 1 - (1 - \alpha)^n \qquad [8a]$$

which can be re-arranged to:

$$\alpha = 1 - \exp\frac{\log(1-\gamma)}{n} \qquad [8b]$$

where α is the threshold level to ensure γ significance level over the n independent tests.

Equation 8b is equivalent to the standard Bonferroni correction for multiple testing. The correction factor is applied to all five traits (Table 2).

To account for testing on 29 autosomal chromosomes, experimentwise critical values were calculated for each trait. It was assumed that the distribution of test statistics seen for chromosome *six* were very similar for all of the other chromosomes. This assumption was based on the knowledge that the length of chromosome *six* is representative of the average length of the 29 autosomal chromosomes and thus representative of the amount of repeated testing across a chromosome. Using Equation 8b with n = 87 (three independent traits analysed on 29 independent chromosomes) experimentwise critical levels were calculated (Table 3).

Comparisonwise critical values (not reported) were similar to tabulated F values. Comparisonwise critical values were relatively constant over the chromosome that is in agreement with the findings of CHURCHILL and DOERGE (1994) and VILKKI *et al.* (1996).

TABLE 3: Experimentwise threshold levels for the five traits (150,000 shuffles).

Threshold level	Milk	Fat	Protein	Fat (%)	Protein (%)
1%	3.05	2.86	2.92	3.29	3.17
5%	2.85	2.60	2.69	2.99	2.88
10%	2.73	2.48	2.59	2.87	2.77
15%	2.67	2.41	2.52	2.76	2.69

For the rest of the paper, experimentwise critical values are used for across-family analysis unless stated otherwise. At chromosomal areas of interest, based on significance levels, within family critical levels are tabulated F values unless stated otherwise. F values were chosen for ease of computation as the issue of repeated testing had been accounted in the across family analysis and comparisonwise values were similar to F values. All additive genetic effects are reported as half the difference in genetic value between homozygotes.

One-QTL Model: Across-family analysis for the five production traits revealed a possible QTL for protein percent positioned at 13cM, *i.e.*, the location of the second marker (Figure 3). The test statistic was significant at the 1% level for protein percent. The yield traits showed little indication of a QTL on chromosome *six* (Figure 4).

Two families were identified as having significant effects for protein percent at the mapped position of marker 2 (Table 4). The test statistics were significant at the 0.1% level for both families. Point estimates for the QTL effect for the two families were 1.12 and 0.68 genetic standard deviation, when using an estimate of protein percent genetic standard deviation of 0.136 (VAN DER WERF and DE BOER 1989) (Table 4).

Absolute marker readings for grandsires at marker position 2 (location of QTL) revealed that both families received a common marker allele (denoted as *X*). Grandsire B is one of grandsire A's six sons that are grandsires in the experiment. Grandsire B was the only son that received marker allele *X*. One other family in the

experiment also had marker allele *X*. This family, distantly related to families A and B, had no significant QTL effect for any of the traits. Marker allele *X* is associated with lower protein percent compared to the other marker allele for both families.

FIGURE 3: Test statistics for different positions on chromosome *six* from an across-family analysis for protein and fat percent (arrows indicate position of markers).

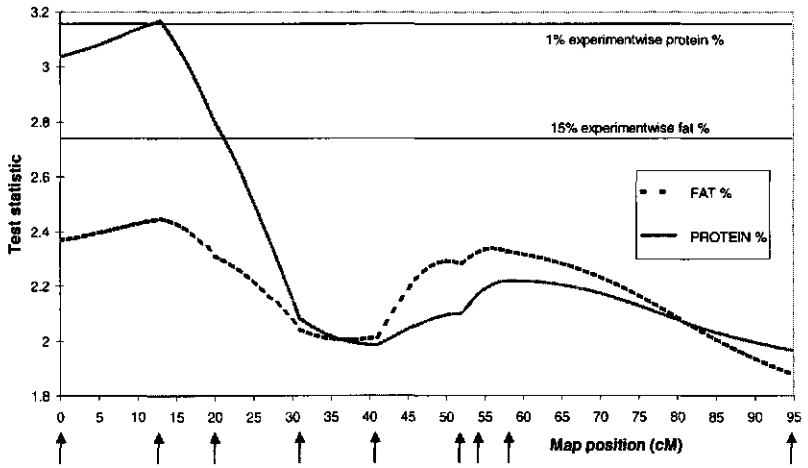
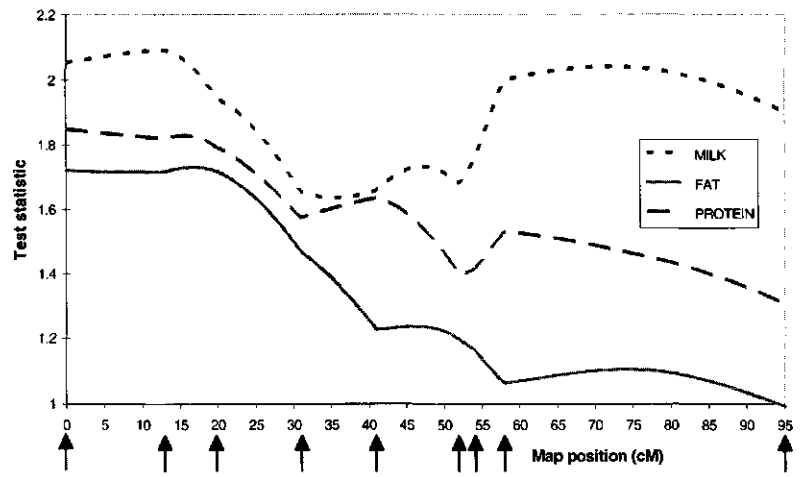


FIGURE 4: Test statistics for different positions on chromosome *six* from an across-family analysis for milk, fat and protein yield (arrows indicate position of markers).



Grandsire A had a significant effect for milk (1%) and protein yield (5%). The effect was an increase of 555 kg of milk and an increase of 8.46 kg of protein in DYDs for sons that received marker allele X. The corresponding increase in milk is approximately double that expected for an increase of 8.46 kg of protein based on average protein percent of 3.46% in the Netherlands (AGRA EUROPE 1995). Grandsire B had a significant effect for protein yield (5%) and no significant effect for milk yield. Protein yield DYDs were 5.36 kg less for sons that received marker allele X while there was no difference in milk yield.

TABLE 4: QTL effect for protein percent in families A and B at position 13 cM.

	Family A	Family B
Number of sons	13	40
F value	24.52[a]	15.85[a]
QTL effect (%)	0.15 ± 0.04	0.09± 0.02
QTL effect (σ_G)	1.12 ± 0.26	0.68 ± 0.14

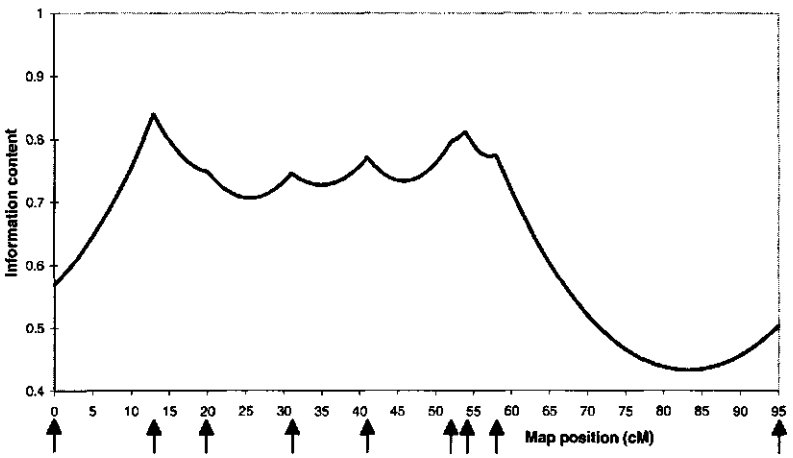
[a] Significance at 0.1% based on tabulated *F* values.

Information content: Seventy percent of peak test statistics derived in across and within-family one QTL analysis occurred at the position of a marker locus (105 observations: 20 families and one across family analysis for 5 traits). KNOTT and HALEY (1992) reported that when considering only flanking markers, the QTL position can be biased and placed in the marker brackets with higher information content. HALEY et al. (1994) reported that this problem can be overcome with the use of multimarker approach. However, the change in information content in HALEY'S *et al.* (1994) simulation of outbred line crosses was not as marked as seen in this study.

If true descent (maternal or paternal) of every centiMorgan of DNA was known, the distribution of the QTL conditional probabilities would be for both 0 and 1. This distribution has mean 0.5 and variance 0.25. True descent is generally unknown and has to be inferred from informative flanking markers. Following a similar application by KRUGLYAK and LANDER (1995) variance of QTL conditional probabilities was calculated for each centiMorgan and is reported in Figure 5 as a

fraction of maximum variance (0.25) which is used as a measure of information content.

FIGURE 5: Information content derived from chromosome *six*. Information content calculated from variance of QTL conditional probabilities at each centiMorgan as a proportion of the variance when true descent is known (arrows indicate position of markers).



Two-QTL model: The two-QTL models for all traits were not significant at the 15% threshold level when compared to the best one QTL model (test statistic 1, Table 5). However, using test statistic 2 there was a significant effect (2%) for one of the two positions for protein percent. The significant position for protein percent was at 1cM whereas in the one QTL model it was at 13cM (Figure 3).

TABLE 5: Results for the two-QTL analysis applied to the five milk production traits.

	Milk		Fat		Protein		Fat %		Protein %	
Positions (cM)	36	63	25	65	36	65	13	58	1	61
Test statistic 1	1.37		1.07		1.08		1.63		1.57	
Test statistic 2	1.72	2.07	1.48	1.28	1.48	1.47	1.98	1.63	3.16[a]	1.72

QTL positions are where the lowest RSS occurred. Test statistic one is comparing the two-QTL model to the best one QTL model. Test statistic two is when the other QTL in the two-QTL model position has been accounted for in a one-QTL model. Significance levels have been calculated from critical values presented in Table 3.

[a] Significance at the 2% level.

DISCUSSION and CONCLUSIONS

Permutation test: The permutation test is a quick method of calculating critical levels which takes into account repeated measures over the genome for individual traits. Individual trait distributions for test statistics differed and thus distributions had to be calculated for each trait. Different trait test statistic distributions are caused by differences in phenotypic distributions as the marker data is the same for all traits. The degree of non-normality of the individual traits did not seem to have a direct link with the observed test statistic distribution differences. However, normality was calculated for each trait over all families, whereas the RSS are calculated within each family and then summed across all families. Therefore the normality of the phenotypic distributions within family may be the cause. Degree of normality was not determined, as it would be calculated on <30 observations for over half of the families.

To account for repeated testing of the five traits analysed the number of independent traits were calculated using factor analysis and analysing the permuted test statistics. The chromosomewise critical levels were considerably higher than tabulated *F* values (Table 2); this reflects the repeated testing over the chromosome is accounted for in the permuted values.

The other method to account for testing on correlated traits; shuffling each trait and then taking the highest permuted test statistic from the five traits, had the effect that the combined test statistics were dominated by the traits with higher absolute values. Test statistics calculated in this manner were not applicable to the traits especially those with lower distributions of critical values. This is demonstrated by the combined critical values at the 5% level being equivalent to the critical value at the 1% level for fat yield (Table 2).

Experimentwise critical values were calculated on the assumptions that chromosome *six* was representative in length of the bovine chromosomes and thus the degree of repeated testing, and that the marker data has very little effect on critical values. The latter assumption was justified based upon the result that when altering the marker density for chromosome *six* to the extremes likely to be seen for the other

chromosomes in this experiment (five and 12 markers per chromosome), only minor differences in critical values occurred. CHURCHILL and DOERGE (1994) using simulated data found differences in the distributions of test statistics for a 100 cM chromosome of 'high marker density' (50 markers) compared to 'low marker density' (10 markers). The difference in the finding of this study and that of CHURCHILL and DOERGE (1994) may be due to the influence and peculiarities of real data and the smaller contrast in marker density in this study.

Experimentwise critical values were chosen as all autosomal chromosomes will be analysed for marker-QTL associations in this experimental design. However, the determination of which threshold level should be used is uncertain. If the objective of the experiment is to identify QTL that will be subsequently confirmed in a second study, an appropriate threshold level may be 15-20% on an experimentwise basis to ensure QTL are not missed. The experiment objective and the effect of utilising a false positive will determine appropriate threshold levels (for further discussion see LANDER and SCHORK 1994; LANDER and KRUGLYAK 1995).

Information content: Information content as measured by the variance of QTL conditional probabilities was not constant over the chromosome. Low heterozygosity at marker one affected information content for the first marker bracket. The large distance for the last marker bracket combined with below average heterozygosity of the last marker also resulted in lower information content in the last marker bracket. As a result of the below average heterozygosity at the chromosomal ends, nine of the 20 families could not have QTL position and effect separated in the first and last marker brackets. The information content peak at marker two was because 18 of the 20 families were informative at that position (Table 1). The information content peak at 50-60 cM was due to the high density of markers in that region. Improvement in information content will be achieved when the dam allele frequencies are used to calculate probabilities for animals in which transmission of alleles is uncertain.

The approach of having evenly spaced markers (*e.g.*, DARVASI and SOLLER 1994) is not the best approach to have information content equitable over the

chromosome. Once the postulated QTL is positioned beyond the last informative marker, information is coming from only a single marker and thus information content decreases. Highly informative markers at the end of the mapped chromosomes would overcome the decrease in information content at the boundaries. However, it is not possible to know marker heterozygosity before the experiment and thus the use of two markers closely positioned at either end of the chromosome may increase the heterozygosity and information content. Increased heterozygosity with closely placed markers is seen with markers six and seven (Table 1). The four families homozygous at marker seven are all heterozygous for marker six. Increased heterozygosity at chromosomal ends will ensure estimates of position and QTL effect can be separated for most families in the first and last marker bracket. In addition, increased heterozygosity at chromosomal ends will overcome the singularity problem for the two QTL model.

The observation that some 70% of peak test statistics occurred at the marker positions is mostly derived from within family analysis. The information content for each family will differ. The across-family information content has an averaging effect on information content in the individual families. The information content for an individual family will fluctuate more than that shown for across family. Local information content peaks at marker positions may be the cause of location of peak test statistics occurring at the marker.

Two-QTL model: Extending KNOTT'S *et al.* (1994) multimarker regression approach to a two-QTL analysis was limited as families were not informative at all markers. Homozygosity at the start or end of the mapped chromosome resulted in fitting two QTL using information from only one of the flanking markers. This resulted in singularity and therefore three families being excluded from across family analysis. The approach of fitting two QTL is similar to that of using of markers as cofactors in the analysis of inbred crosses as described by JANSEN (1993) and ZENG (1994). These authors in addition to marker genotypes use trait phenotypic values in assigning conditional probabilities and also weight the probability of QTL phase in contrast to the KNOTT *et al.* (1994) approach where phase is assumed to be known with certainty. However, not assigning probabilities to phase was not critical in this

study, as in most cases (80%) the probability of chosen phase was greater than 70%. The exceptions to this were for the last marker bracket where the distance is large and for the smaller families. Using all markers together instead of an individual marker haplotype may improve the determination of phase for the smaller families. The approach of JANSEN (1993) and ZENG (1994) may overcome the singularity problem due to using information in addition to that from the single marker and thus breaking the complete collinearity between postulated QTL.

The use of markers as cofactors (JANSEN 1993; ZENG 1994) in outbred populations may not be possible as markers are not uniformly informative in all families as found in crosses of inbred lines. The approach of fitting postulated QTLs as cofactors, within families, on the same and other chromosomes may overcome this.

Two test statistics for comparison of a two-QTL model to a one QTL were used. Comparing the two QTL model to the best one QTL model had the bias that the comparison between models was for different QTL positions. The one-QTL model may detect a ghost QTL in between the two QTLs (MARTINEZ and CURNOW 1992; HALEY and KNOTT 1992). If the two QTLs are in phase and of the same effect, the variance explained by a ghost QTL will be inflated and therefore not a good comparison for the two QTL model. Fitting the two-QTL model and then determining if neither, one or both positions explained a significant amount of the variance in the two-QTL model was the preferred option for this study, as the comparison is then between a two- and one-QTL model for the same QTL positions. However, it is acknowledged that for the second test statistic that when two QTL are in phase and of the same effect this will inflate the variance explained at both positions in the one QTL model. This will also reduce the significance of the two QTL model when compared to the one-QTL models. Further research is needed in this area.

Casein: The findings of earlier studies for effects at and linked to the casein locus (summarised by BOVENHUIS *et. al.* 1992) were not confirmed in this study. Non-significant peaks for the test statistic near the casein loci (marker six) were found for fat percent in the one-QTL model and for all traits for one of the two locations identified in the two-QTL model.

Protein percent QTL: A QTL for protein percent was identified in the across family study with a one-QTL model and was significant at the 1% level. Location of the protein percent QTL at marker two (13cM) is practically the end of the mapped chromosome as marker one was informative in only 4 of the 20 grandsire families (Table 1). Families A and B were not informative at the first marker. Therefore QTL location and effect can not be separated for a QTL located in the first marker bracket for these two families. Informative markers to the left of marker two may change the mapped position of the QTL.

Allele *X* at marker position two was associated with the change in protein percent. Relative to the other marker allele the effect was a decrease in protein percent. The protein percent effect was caused by an increase in milk yield in family A and a decrease in protein yield in family B

As described, family A and B are related. This is one of many relationships that exist within the data set but not utilised in this study. Accounting for the relationships within an animal model setting would most probably increase the power of the design. Methods to utilise these relationships are being investigated.

GEORGES *et al.* (1995) identified a QTL in one family on chromosome *six* that appeared to increase milk yield but not fat or protein yield and as a result fat and protein percent decreased. This family had two informative markers. The location of the QTL in GEORGES *et al.* (1995) is some 5-10 cM to the left of marker three used in this study. This is nearly the same QTL location found in this study. Family A and the family identified by GEORGES *et al.* (1995) have a common ancestor two and three generations back, respectively. The QTL found in this study and GEORGES *et al.* (1995) is very likely to be the same. A similar finding has been made in the Finnish Ayrshire population (R. Velmala, personal communication).

Although the QTL has the same effect on protein percent in both studies and all three families, the effect on the yield traits differ between families. This may reflect the power of the respective studies. Further investigation through additional

markers and more genotyping in the identified region may increase our understanding of the identified QTL on chromosome *six*.

ACKNOWLEDGEMENTS

We thank SARA KNOTT for the multimarker programme. We are grateful to Livestock Improvement and Holland Genetics for financial support of the experiment and data access. The authors acknowledge the excellent genotyping work of LAURANCE MOREAU, FABIENNE MARCQ and ANDREINA SHOEBERLEIN, and CHRIS SCHROOTEN for providing data.

LITERATURE CITED

AGRA EUROPE, 1995 Dairy Review. Special Report No. 82.

BISHOP, M. D., S. M. KAPPES, J. W. KEELE, R. T. STONE, S. L. SUNDEN, *et al.* 1994 A genetic linkage map for cattle. Genetics **136:** 619-639.

BOVENHUIS, H., J.A.M. VAN ARENDONK and S. KORVER, 1992 Associations between milk protein polymorphisms and milk production traits. J.Dairy.Sci. **75:** 2549-2559.

BOVENHUIS, H., and J. I. WELLER, 1994 Mapping and analysis of dairy cattle quantitative trait loci by maximum likelihood methodology using milk protein genes as genetic markers. Genetics **137:** 267-280.

CHURCHILL, G. A., and R. W. DOERGE, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963-971.

DARVASI, A., and M. SOLLER, 1994 Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. Theor. Appl. Genet. **89:** 351-357.

GELDERMANN H., 1975 Investigations on inheritance of quantitative characters in animals by gene markers I. methods. Theor. Appl. Genet **46:** 319-330.

GEORGES, M., D. NIELSEN, M. MACKINNON, A. MISHRA, R. OKIMOTO, *et al.* 1995 Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics **139:** 907-920.

HALDANE, J.B.S., 1919 The combination of linkage values and the calculation of distances between the loci of linked factors. J. Genet. **2:** 3-19.

HALEY, C. S., and S. A. KNOTT, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity **69:** 315-324.

HALEY, C. S., S. A. KNOTT and J. ELSEN, 1994 Mapping quantitative trait loci in crosses between outbred lines using least squares. Genetics **136:** 1195-1207.

JANSEN, R. C., 1993 Interval mapping of multiple quantitative trait loci. Genetics **135:** 205-211.

KNOTT, S. A., J. M. ELSEN and C. S. HALEY, 1994 Multiple marker mapping of Quantitative trait loci in half sib populations. Proceedings of the 5th World Congress on Genetics applied to Livestock Production, Guelph, Canada, 1994 **21:** 33-36.

KNOTT, S. A., and C. S. HALEY, 1992 Maximum likelihood mapping of quantitative trait loci using full-sib families. Genetics **132**: 1211-1222.

KRUGLYAK, L., and E. S. LANDER, 1995 Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am. J. Hum. Genet. **57:** 439-454.

LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185-199.

LANDER, E., and L. KRUGLYAK, 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genetics **11:** 241-247.

LANDER, E. S., and N. J. SCHORK, 1994 Genetic dissection of complex traits. Science **265:** 2037-2048.

MACKINNON, M. J., and M.A.J. GEORGES, 1992 The effects of selection on linkage analysis for quantitative traits. Genetics **132:** 1177-1185.

MARTINEZ, O., and R. N. CURNOW, 1992 Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. **85:** 480-488.

SAS INSTITUTE INC, 1985 SAS® User's Guide: Statistics, Version 5 edition, Cary, N.C. 956p.

VAN DER BEEK, S., J.A.M. VAN ARENDONK, and A. F. GROEN, 1995 Power of two- and three-generation QTL mapping experiments in an outbred population containing full-sib or half-sib families. Theor. Appl. Genet. **91:** 1115-1124.

VAN DER WERF, J.H.J., and W. DE BOER, 1989 Estimation of genetic parameters in a crossbred population of black and white dairy cattle. J. Dairy Sci. **72:** 2615-2623.

VAN RADEN, P. M., and G. R. WIGGANS, 1991 Derivation, calculation, and use of National Animal Model Information. J. Dairy Sci. **74:** 2737-2746.

VILKKI, H. J., D-J., DE KONING, K. ELO, R. VELMALA, and A. MAKI-TANILA, 1996 Multiple marker mapping of quantitative trait loci of Finnish dairy cattle by regression. J. Dairy Sci. **80:** 198-204.

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., Y. KASHI and M. SOLLER, 1990 Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J. Dairy Sci. **73:** 2525-2537.

ZENG, Z. B., 1993 Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. **90:**10972-10976.

ZENG, Z. B., 1994 Precision mapping of quantitative trait loci. Genetics **136:** 1457-1468.

CHAPTER THREE

EFFECT OF INACCURATE PARAMETER ESTIMATES ON GENETIC RESPONSE TO MARKER ASSISTED SELECTION IN AN OUTBRED POPULATION

RICHARD J. SPELMAN and JOHAN A.M. VAN ARENDONK

PUBLISHED IN JOURNAL OF DAIRY SCIENCE (1997) **80:** 3399-3410

ABSTRACT

The effect of inaccurate estimates of variance and location of quantitative trait locus on the genetic response to marker assisted selection was studied by simulation of an adult multiple ovulation and embryo transfer nucleus breeding scheme. Two genetic models were simulated for the quantitative trait locus: a total of 10 alleles or 2 distinct alleles per base parent. For both models, the locus explained either 5 or 10% of phenotypic variance. A polygenic component was simulated, and the two genetic components were summed to 35% heritability for a trait measured on females. Overestimation of variance of the quantitative trait locus had minimal effect on genetic gain for marker assisted selection over the short term, but decreased long-term response. The long-term loss was reduced when variance of the quantitative trait locus was re-estimated after four generations of marker assisted selection. Selection for favourable alleles at a non-existent quantitative trait locus resulted in first generation losses of 3 and 7% for postulated quantitative trait loci explaining 5 and 10% of variance, respectively. The larger the degree of error in location, the larger was the genetic loss compared with the correct location scenario. For the largest simulated location error of 15 cM, genetic superiority of marker assisted selection was reduced by 80% in the first generation. We concluded that studies should be undertaken to verify estimates of quantitative trait locus and location to make optimal use of marker assisted selection.

KEY WORDS: Marker assisted selection, Quantitative trait locus, Genetic parameter estimates, Breeding scheme.

INTRODUCTION

Recent scientific literature has been increasingly reporting results from experiments using dairy cattle to study quantitative trait loci (QTL) (GEORGES *et al.* 1995; SPELMAN *et al.* 1996) and the theoretical responses to marker assisted selection (MAS) in breeding schemes (KASHI *et al.* 1990; MEUWISSEN and VAN ARENDONK 1992; SPELMAN and GARRICK 1997). Many different genetic models for the QTL in MAS have been used; for example, the number of QTL alleles in the population range from two alleles (RUANE and COLLEAU 1995, 1996) to two unique alleles per base parent (MEUWISSEN and GODDARD 1996). The QTL parameters in the MAS studies have assumed to have been known without error when genetic and economic

responses to MAS were estimated. The parameters used to characterise the QTL are the size of allelic effects or variance and the location of the QTL relative to a single marker or marker bracket.

WANG (1995) demonstrated by simulation that, on average, when the test statistic exceeded a certain significance threshold, the QTL effect was overestimated, especially when analyses had low power. GEORGES *et al.* (1995) also showed through simulation that a very high significance threshold reduced the power of the design and resulted in overestimation of the effect of the QTL when the threshold was exceeded.

The accuracy of QTL location is dependent on many factors, including marker density and heterozygosity, number of meiosis observed in the experimental design, and size of QTL effect. Genomic linkage maps that are being published for cattle report average marker interval size of less than 3 cM (centiMorgans) (KAPPES *et al.* 1997). The average heterozygosity for bovine microsatellites is approximately 60% (GEORGES *et al.* 1995). The QTL experiments have been limited by genotyping costs and pedigree structure and thus are unable to genotype large number of animals and observe large numbers of meiosis. The accuracy of QTL location estimates may be poor because of relatively sparse genetic linkage maps, low heteroygosity of marker loci, and small experiments, and large confidence intervals for a granddaughter and F_2 backcross design (VAN OOIJEN 1992) have been reported.

SALES and HILL (1976) investigated the effect of parameter errors on selection response for one trait when an additional trait, which could be regarded as an indicator, was added to the selection index. Those researchers concluded that, when the second trait contributes no useful information, the loss in efficiency by including the trait is equal to the predicted benefit from its inclusion based on the assumed parameters.

The objective of this study was to ascertain through stochastic simulation the sensitivity of genetic response resulting from MAS to incorrect estimates of variance explained by the QTL and QTL location. Sensitivity to these parameter estimates is investigated for two genetic models that differ in the number of QTL alleles: 10 alleles with equal frequency (A10) or alleles equal to twice the number of base

parents (BP2). In addition, the effect of incorrect QTL location was investigated for different marker spacings in the QTL-marker haplotype.

MATERIALS AND METHOD

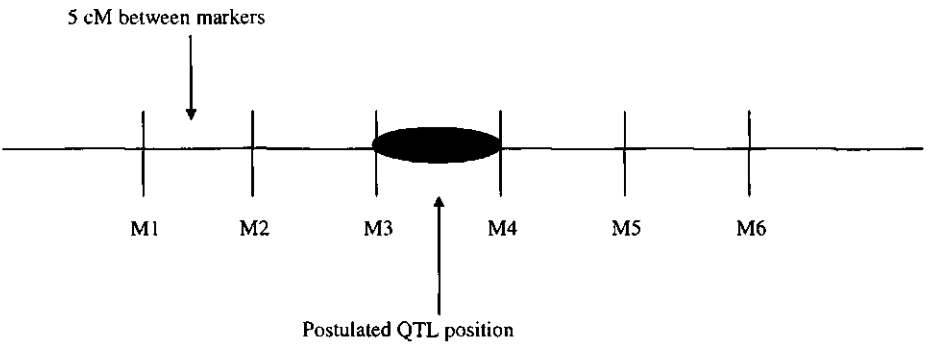
Simulation model: A stochastic simulation was developed that modelled an adult multiple ovulation and embryo transfer closed nucleus starting from an unselected, unrelated, and non-inbred population with discrete generations. Each generation had 1024 progeny and equal numbers of males and females. A single trait was simulated with base population heritability (polygenic and QTL) of 35%, which represents a milk production trait. The additive genetic variance was partitioned between unmarked additive polygenic variation (referred to as polygenic variance) and variation because of the marked chromosomal region (referred to as QTL). Phenotypic records were recorded for females only. The highest ranking 12.5% of males and 50% of females for estimated breeding values were selected as parents of the next generation. Because phenotypes were only available on the females, the male breeding values were estimated from pedigree information. Selection was undertaken after the single phenotypic record for females was available. Each sire was mated to 4 females (avoiding half-sib and closer matings), and each mating resulted in 4 offspring (2 male and 2 female). Each female was mated to one sire only.

Effects of the QTL alleles for the unselected base population were drawn from the distribution $N(0,½V_{QTL})$, where V_{QTL} is the variance explained by the QTL. Two QTL variances were used in this study: 5 and 10% of phenotypic variance. The number of QTL alleles in the base population was either BP2 or A10. Because of sampling variation, the variance explained by the QTL–especially for the A10 situation–often deviated substantially from the desired level. The actual variance of the effects of the sampled QTL alleles was calculated (with n degrees of freedom, where n is the number of alleles). The QTL effects were transformed by multiplying them by the inverse of the square root of the fraction of actual variance over desired variance. This procedure ensured that the variance of the QTL alleles equalled the desired level. Deviation from the desired variance for the A10 situation could still occur, because sampling may result in variable QTL allelic frequencies in the base population.

A polygenic effect for each base animal was sampled from the distribution $N(0, V_a)$, where V_a is the polygenic variance. In subsequent generations, the polygenic component was sampled from the distribution $N(\frac{1}{2}g_s + \frac{1}{2}g_d, \frac{1}{2}(1 - \frac{1}{2}(F_s+F_d))V_a)$, where s and d = sire and dam, g = true additive genetic value, and F = inbreeding coefficient that was calculated using the algorithm presented by TIER (1990). Residual components from the distribution $N(0, V_e)$, where V_e = residual variance, were sampled for females and added to the previously sampled polygenic and QTL effects to complete the phenotypic observations. Phenotypic variance in the base population, comprising of $V_a + V_{QTL} + V_e$, had an expected value of 100.

Marker alleles were simulated for all animals in the base population. It was assumed that the linkage map had six equally spaced markers (distance 5 cM) that bracketed the postulated QTL position (Figure 1). It was also assumed that each marker locus had five alleles with equal frequency and that the linkage phase and haplotype in the base population was known in order to simulate transmission of haplotype. The HALDANE (1919) mapping function was assumed for the construction of the marker-QTL haplotypes that were transmitted to the offspring. Therefore, the probability of recombination between adjacent loci is independent from other recombination events.

FIGURE 1: The marker haplotype that surrounds the postulated location of the quantitative trait locus (QTL).



The required number of sires and dams were simulated for the base population and the mating to produce the first generation. To move the population to equilibrium

selection response, three generations of selection were undertaken without using marker genotypes (conventional BLUP). MAS was undertaken for seven generations in total. The generation number for offspring born from the first application of MAS is termed generation 1 in this paper, therefore the base population generation is generation –4.

Breeding value estimation: Breeding value estimation of polygenic and marker-linked effects for MAS was undertaken using the model described by MEUWISSEN and GODDARD (1995):

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{ZQq} + \mathbf{e}$$

where

$\mathbf{y}$ = vector of phenotypic records,

$\mathbf{X}$ = incidence matrix linking fixed effects to records,

$\mathbf{b}$ = vector of fixed effects (only the mean in this study),

$\mathbf{Z}$ = incidence matrix linking animals to records,

$\mathbf{u}$ = vector of polygenic effects,

$\mathbf{Q}$ = incidence matrix linking QTL allelic effects to animals (every row has two elements equal to one and the other elements are zero),

$\mathbf{q}$ = vector of allelic effects for QTL, and

$\mathbf{e}$ = vector of residual effects.

Mixed model equations (HENDERSON 1984) are used to estimate $\mathbf{b}$, $\mathbf{u}$ and $\mathbf{q}$:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'ZQ} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\mathbf{A}^{-1}\lambda & \mathbf{Z'ZQ} \\ \mathbf{Q'Z'X} & \mathbf{Q'Z'Z} & \mathbf{Q'Z'ZQ}+\mathbf{G}^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{q}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{Q'Z'y} \end{bmatrix}$$

where

$\mathbf{A}^{-1}$ = inverse of numerator relationship matrix,

$\lambda = V_e/V_a$,

$\mathbf{G}^{-1}$ = inverse of gametic relationship matrix, and

$\alpha = V_e/½V_Q$.

This model is an extension of the method of FERNANDO and GROSSMAN (1989) that was developed for single markers and the method of GODDARD (1992) that adapted the previous model for marker haplotypes.

In brief, the computational method for marked QTL considers that in the base population the number of QTL alleles is equal to twice the number of base animals. Each base population QTL allele is arbitrarily assigned paternal or maternal descent. In the next generation, the transmission of QTL alleles is followed by inference on marker haplotype. When transmission of marker haplotype can be followed, the **Q** matrix links progeny's phenotype to the transmitted effect of parental QTL allele. When transmission cannot be followed by the flanking markers, an effect of a new QTL allele is formed. The progeny phenotype is linked via the **Q** matrix to the effect of the new QTL allele, and the effect of the new QTL allele is linked to its parents through the **G** matrix; that is, the expectation of the effect of the new QTL allele in the progeny is equal to mean of the effects of the parental QTL alleles. This model does not assume that the exact location of the QTL within a marker bracket is known but postulates that the QTL is within the marker bracket. Probability statements are that either QTL transmission can or cannot be followed by inference from marker haplotype. Thus, probability statements other than 0 or 1 are not made about transmission based on recombination events between flanking markers and postulated QTL position relative to markers. Double recombinants in the calculation of the probability of transmission are assumed to be absent in this technique. [for further description of model see (MEUWISSEN and GODDARD 1996)]. The described MAS breeding value estimation method is referred to as marker assisted (MA) BLUP for the rest of the paper.

For MA BLUP, when the origin of marker allele could not be established at the closest flanking markers around the postulated QTL bracket, then the next marker out was used. If allele origin could not be determined for at least one side of the marker haplotype, QTL transmission could not be determined according to the rules of MEUWISSEN and GODDARD (1996). Also, when recombination was observed between the informative markers bracketing the QTL, the parental QTL allele transmitted could not be determined. The linkage phase of the marker-QTL haplotype was assumed known in the parents over all generations.

From generation −3 onward, conventional mixed model equations were used to estimate **b** and **u**. The additive genetic variance in the mixed model was the sum of polygenic variation and QTL variation in the base population (35%). After three generations of conventional BLUP, MAS was undertaken from generation 0 using the MA BLUP model. In the MA BLUP, model the additive genetic variance was partitioned into the two components: polygenic variance and QTL variance. As a control, conventional BLUP was also continued for another seven generations, and the additive genetic variance in the base population remained at 35%.

Estimates of polygenic and QTL effects were obtained using iteration on the data (SCHAEFFER and KENNEDY 1986). Solutions were considered to be stable when convergence criterion, which equals the sum of squares of differences in solutions between iterations divided by the sum of squares of the most recent solutions, was less than 10^{-10}.

Different scenarios to study sensitivity of genetic response to errors in parameter estimates were evaluated. The scenarios broadly fell into two categories, QTL variance and position. Analysis was undertaken to study the genetic consequences of overestimation of QTL variance. The total additive genetic variance was always 35%. When the QTL was assumed to explain 15% of the phenotypic variance, the polygenic component in the MA BLUP was set to 20%, and the QTL component was set to 15%. Assumed variance components were used to calculate λ and α in the MA BLUP model. True QTL variance was always 5% of phenotypic variance for this scenario.

Analysis was also undertaken to study the genetic consequence of the true QTL location differing to that postulated. The sensitivity to inaccuracies of parameter estimates was studied for the two differing simulated genetic models. Eighty replicates were simulated for each differing scenario for both MAS and the control.

RESULTS

Genetic gain with the base model: The rate of genetic gain for the breeding scheme, which was modelled for a 35% heritability consisting solely of polygenic variance, was close to 0.3 of a phenotypic standard deviation per generation. The equilibrium response with this model was reached after three to four generations of conventional BLUP selection.

Genetic gain with correct parameter estimation: Genetic superiority of MAS over the control for the BP2 model was approximately 7 and 15% in the first generation of MAS for QTL with 5 and 10% of phenotypic variance respectively (Table 1). Less polygenic gain occurred with MAS than with the control, and this difference was larger for the QTL with 10% of phenotypic variance (10% QTL). The cumulative superiority of MAS for the 10% QTL decreased over the later generations as the QTL variance decreased due to selection. After the three initial generations of normal BLUP–to get the breeding program to equilibrium response–the QTL variance was 90 to 95% of that in the base population. Over the next seven generations for the QTL with 5% phenotypic variance (5% QTL), the QTL variance decreased to 35% of the original variance for MAS and to 70% for the control. For the 10% QTL, the QTL variance decreased to 10% of the original variance for MAS and to 50% for the control. As a result of the decrease in QTL variance, the rate of genetic gain for MAS was less than that achieved by the control over the last three generations for the 10% QTL. This result can be observed in Table 1 for the 10% QTL; the absolute overall genetic superiority of MAS over the control in generation 5 was less than that in generation 4.

For the A10 model, the percentage gains and absolute genetic gain for the first three generations of MAS were equal to the BP2 model for the 5% QTL but less for the 10% QTL (Tables 1 and 2). The QTL variance decreased quicker for the A10 model than for the BP2 model. For example, the QTL variance after four to five generations of MAS with the A10 model was equal to that of generation 7 with the BP2 model (results not shown). After only three generations of MAS, the rate of genetic gain for the control with the A10 model, for the 10% QTL, was greater than that achieved with MAS. Thus, the superiority of MAS at the end of seven

generations for the two QTL variances was 1% for the A10 model and 5% for the BP2 model. Inbreeding, calculated from the pedigree, was monitored over the simulation period and was slightly less for the MAS schemes for both genetic models, probably because the QTL information differentiates the genetic merit of close relations (e.g., full-sibs). Therefore, the correlation of estimated breeding values between related animals is less with QTL information and results in selection of animals from other families, instead of all the full sibs from one family. BRISBANE and GIBSON (1995) reported a similar result.

TABLE 1: Cumulative differences [phenotypic standard deviation (σ_P)] in genetic response between breeding programs using marker assisted selection (MAS) or not using MAS and a quantitative trait locus (QTL) that explains 5 and 10% of phenotypic variance.[1]

	5%				10%			
Generation	QTL	Polygenic	Overall[2]		QTL	Polygenic	Overall	
				(%)				(%)
1	0.048	-0.027	0.021	7	0.106	-0.058	0.048	15
2	0.091	-0.054	0.037	6	0.186	-0.098	0.088	14
3	0.133	-0.084	0.049	5	0.255	-0.129	0.126	14
4	0.162	-0.098	0.064	5	0.291	-0.150	0.141	11
5	0.184	-0.109	0.075	5	0.298	-0.159	0.139	9
6	0.193	-0.100	0.093	5	0.287	-0.177	0.110	6
7	0.197	-0.100	0.097	5	0.265	-0.172	0.093	5

[1]The number of QTL alleles in the base population is twice the number of base parents.
[2]Overall percentage superiority (inferiority if negative) of MAS over the control is {[(genetic gain from generation 0 to generation X for MAS)/(genetic gain for the same time period for the control)] $-1\} \times 100\%$. Standard errors (σ_P) for overall genetic difference are 0.004 (generation 1), 0.013 (generation 4), and 0.020 (generation 7).

TABLE 2: Cumulative differences [phenotypic standard deviation (σ_P)] in genetic response between breeding programs using marker assisted selection (MAS) or not using MAS and a quantitative trait locus (QTL) that explains 5 and 10% of phenotypic variance.[1]

	5%				10%			
Generation	QTL	Polygenic	Overall[2]		QTL	Polygenic	Overall	
				(%)				(%)
1	0.044	-0.023	0.021	7	0.087	-0.047	0.040	13
2	0.073	-0.039	0.034	5	0.125	-0.069	0.056	9
3	0.096	-0.047	0.049	5	0.133	-0.070	0.063	7
4	0.104	-0.059	0.045	4	0.127	-0.072	0.055	5
5	0.105	-0.068	0.037	2	0.115	-0.064	0.051	4
6	0.108	-0.074	0.033	2	0.098	-0.061	0.037	2
7	0.100	-0.079	0.021	1	0.087	-0.064	0.023	1

[1]The number of QTL alleles in the base population is 10 at equal frequency.
[2]Overall percentage superiority (inferiority if negative) of MAS over the control is {[(genetic gain from generation 0 to generation X for MAS)/(genetic gain for the same time period for the control)] −1} × 100%. Standard errors (σ_P) for overall genetic difference are 0.005 (generation 1), 0.011 (generation 4), and 0.017 (generation 7).

Selection for a non-existent QTL: Genetic response with MAS for a QTL postulated to explain either 5 or 10% of the phenotypic variance, when in reality there was no QTL, was less than that obtained with the control (Figure 2). For the 10% QTL, the loss in the first generation was 7%. The slower rate of genetic gain for MAS than for the control continued over the next six generations. At the end of the seven generations of MAS, the cumulative loss was approximately 3%. For the 5% postulated QTL, the loss in the first generation was approximately 3% and the genetic inferiority after seven generations of MAS was 1%.

The estimated rate of genetic gain by the MA BLUP method at the non-existent QTL was less than that with a QTL (Figure 3). The rate of change in estimated QTL effects slowed over time when no QTL was present. The estimated rate of genetic gain for the first generation of MAS was 0.08 phenotypic standard deviation, and, in the last four generations of MAS, the estimated rate of gain was 0.04 phenotypic standard deviation, which resulted in most of the loss associated with selection for a non-existent QTL in the first two to three generations of MAS (Figure 2).

FIGURE 2: Cumulative differences in genetic response (σ_P) between breeding programs using marker assisted selection (MAS) or not using MAS with a quantitative trait locus (QTL) that is estimated to explain 5% ([]) or 10% (■) of phenotypic variance (σ_P^2). In reality, there is no QTL.

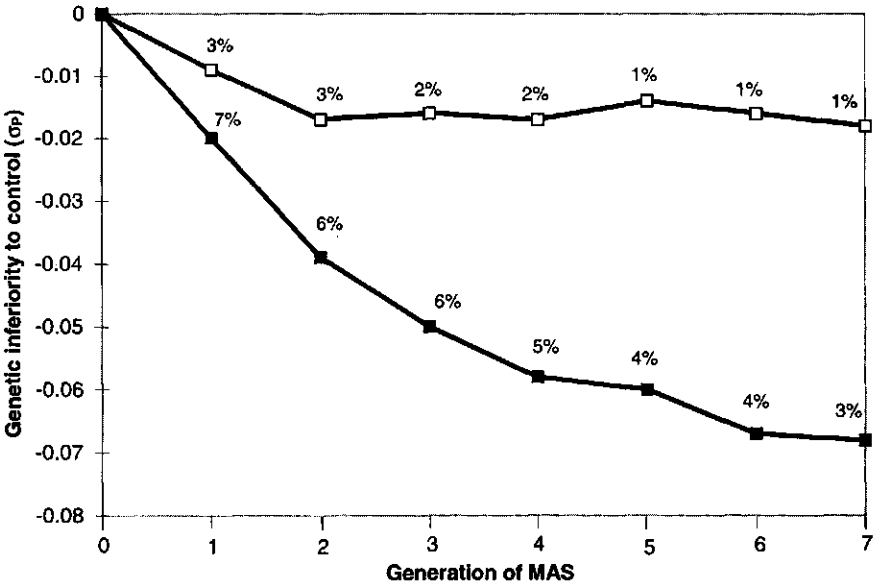
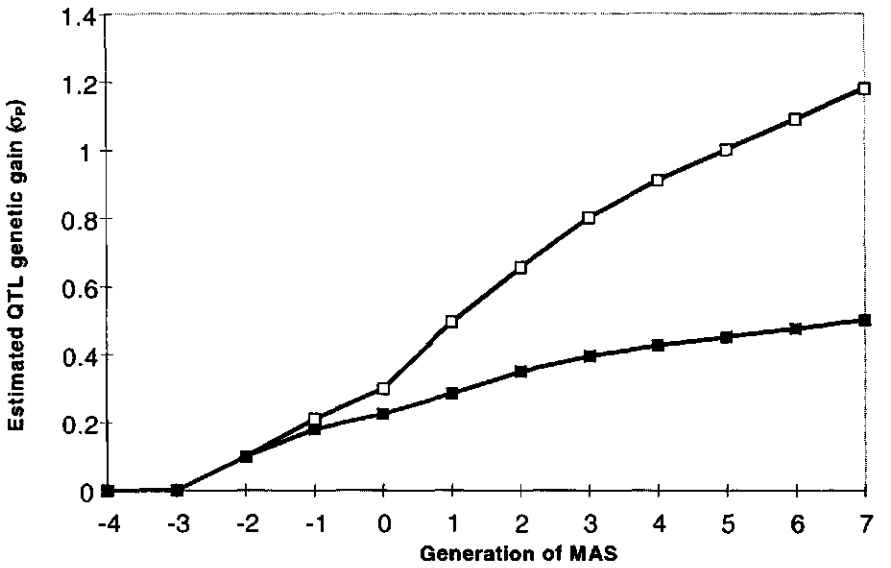


FIGURE 3: Estimated genetic gain (σ_P) at the quantitative trait locus (QTL) from the establishment of base population when it is estimated that the QTL explains 10% of phenotypic variance and the real QTL effect is either 10% ([]) or does not exist (■). Estimated values calculated in generation 7.



Overestimation of the variance explained by the QTL: The effect of overestimating the variance explained by a 5% QTL was evaluated with assumed QTL variances of 10 and 15% for both simulated genetic models. Greater genetic gain at the QTL was observed with the overestimated variance, but at the expense of lower polygenic gain. The overall rate of genetic gain compared with that using MAS with correct parameter estimates (5%) was inferior for all generations for the 10 and 15% assumptions (Table 3).

TABLE 3: Effect on cumulative genetic gain [phenotypic standard deviation (σ_P)] of overestimating variance of quantitative trait locus (QTL) compared with correct estimation (5%) of QTL.

	BP2[1]						A10[2]					
	10%			15%			10%			15%		
Generation	QTL	Overall	%[3]	QTL	Overall	%	QTL	Overall	%	QTL	Overall	%
1	0.014	-0.005	76	0.022	-0.011	77	0.013	-0.003	86	0.021	-0.010	75
2	0.021	-0.008	78	0.035	-0.017	81	0.017	-0.014	59	0.028	-0.024	57
3	0.024	-0.004	92	0.041	-0.012	90	0.015	-0.014	71	0.025	-0.026	59
4	0.025	-0.006	91	0.043	-0.019	87	0.015	-0.020	56	0.019	-0.027	51
5	0.024	-0.010	87	0.043	-0.027	81	0.013	-0.025	32	0.017	-0.034	33
6	0.023	-0.018	81	0.043	-0.032	71	0.010	-0.026	31	0.005	-0.045	–
7	0.016	-0.029	70	0.035	-0.042	55	0.006	-0.023	–[4]	0.005	-0.047	–

[1]The number of QTL alleles in the base population is twice the number of base parents.
[2]The number of QTL alleles in the base population is 10 at equal frequency.
[3]Percentage superiority of MAS over the control with overestimation of QTL variance compared with MAS with correct QTL estimation.
[4]Not able to calculate percentage as genetic response less than that in the control. Standard errors (σ_P) for overall genetic difference are 0.004 (generation 1), 0.011 (generation 4), and 0.019 (generation 7).

The difference in rate of genetic gain for the first three generations of MAS was minimal for the comparison between the correct 5% and incorrect 10% assumption for the BP2 and A10 models (Table 3). However, in the later generations, the genetic difference increased, and the incorrect variance scenario became increasingly inferior. Genetic gain was affected more for the 15% assumption than that for the 10% assumption. The genetic advantage over the control (without MAS)

for the BP2 model can be determined from Tables 1 and 3. For the 10% assumption, the overall inferiority compared to the correct variance is –0.005 in generation 1 (Table 3). The superiority of the correct variance (5%) over the control (without MAS) is 0.021 in generation 1 (Table 1). Therefore, the genetic advantage over the control, in generation 1, for the 10% assumption is 0.016 (i.e., 76% of that achieved with QTL variance correctly estimated at 5%). For the A10 model, genetic response with MAS with both incorrect variance estimates was inferior to the control in generation 7.

Incorrect positioning of postulated QTL: The MAS was undertaken when the true position of the QTL was 5, 10, or 15 cM away from the postulated QTL position. Thus, the QTL was one bracket away from the postulated position for an error of 5 cM, two brackets away for an error of 10 cM, and outside the marker haplotype for an error of 15 cM (Figure 1). The effect of poor location estimation was undertaken for a 10% QTL for both simulated genetic models.

TABLE 4: Effect on cumulative genetic gain [phenotypic standard deviation (σ_P)] of the incorrect position of 10% quantitative trait locus (QTL) compared with that of the correct position of QTL.[1]

	5 cM			10 cM			15 cM		
Generation	QTL	Polygenic	Overall	QTL	Polygenic	Overall	QTL	Polygenic	Overall
1	-0.013	0.006	-0.007	-0.035	0.013	-0.022	-0.054	0.017	-0.037
2	-0.017	0.006	-0.011	-0.062	0.023	-0.039	-0.095	0.040	-0.055
3	-0.031	0.008	-0.023	-0.083	0.023	-0.060	-0.135	0.060	-0.075
4	-0.036	0.013	-0.023	-0.090	0.029	-0.061	-0.153	0.069	-0.084
5	-0.035	0.016	-0.019	-0.083	0.026	-0.057	-0.142	0.073	-0.069
6	-0.037	0.019	-0.018	-0.076	0.027	-0.049	-0.133	0.079	-0.052
7	-0.033	0.014	-0.018	-0.069	0.030	-0.039	-0.116	0.075	-0.041

[1]The number of QTL alleles in the base population is twice the number of base parents. Standard errors (σ_P) for overall genetic difference are 0.004 (generation 1), 0.012 (generation 4), and 0.018 (generation 7).

Genetic gain with MAS with the three incorrect QTL positions was less than that achieved when the QTL position was estimated correctly (Table 4). The level of

the genetic inferiority compared with scenario using the correct location (Table 4) was less than the superiority of the 10% QTL over the control (without MAS) for the BP2 model (Table 1). Therefore, the rate of genetic gain when the location errors were included was greater than that achieved without MAS for the BP2 model.

Polygenic gain was greater with the inaccurate location estimates, but QTL genetic gain was inferior (Table 4). As the estimate of QTL location became less accurate, the loss in overall genetic gain increased. The largest differences between genetic response for correct and incorrect position were found for generation 4 for all location errors for the BP2 model (Table 4).

The results for incorrect QTL location estimates for the A10 model (not shown) were similar to that for the BP2 genetic model. However, in the last generation, for all incorrect estimates of location, the genetic level for MAS was less than that of the control (without MAS). For an error of 15 cM, the MAS superiority over the control was approximately 25% of that achieved with correct location after just one generation of MAS. This reduction in superiority is equivalent to that found with the BP2 model.

Effect of marker spacing outside flanking QTL-marker bracket: When choosing the markers around the postulated QTL position, one would assume that there would be a large number of microsatellites in the region of interest from which to choose. Also, only one gel lane would probably be run for each animal to keep costs to a minimum. In that case, 8 to 10 markers would likely be genotyped for each animal. A choice would need to be made on which markers to choose in addition to the markers that are predicted to bracket the QTL. Until now, in this study, it has been assumed that there are three markers on either side of the QTL and that all markers are spaced at 5 cM. The two markers closest to the QTL were identified as flanking markers and bracket the QTL. One could choose to have non-flanking markers closer than 5 cM to the flanking markers. The genetic effect of having markers spaced 1 cM outside the flanking markers and the sensitivity of this type of haplotype to parameter errors were evaluated. The flanking markers that bracketed the QTL remained at 5 cM.

The rate of genetic gain for MAS, with correct QTL location, differed when the second and third markers outside the flanking markers were either 1 or 5 cM (Table 5). The smaller marker spacing outside the flanking markers resulted in superior genetic gain from greater polygenic response but lower QTL gain for both genetic models. SPELMAN and BOVENHUIS (1998) also report a similar source of genetic improvement for some situations when the distance between the markers flanking the QTL was reduced. Those authors concluded that the greater accuracy in estimating QTL allelic effects results in the more accurate adjustment of phenotype for QTL effects, resulting in more accurate estimate of polygenic value.

TABLE 5: Effect on cumulative genetic gain (σ_P) for markers positioned 1 cM compared to 5 cM outside flanking markers; for correct location and for a 15 cM quantitative trait locus (QTL) location error, for a 10% QTL.[1]

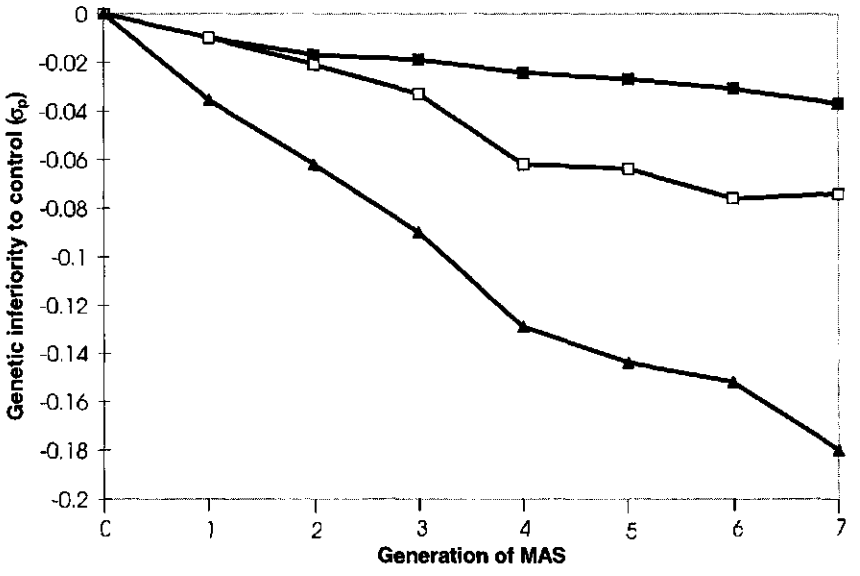
	No error			15 cM		
Generation	QTL	Polygenic	Overall	QTL	Polygenic	Overall
1	0.004	0.001	0.005	-0.004	0.003	-0.001
2	0.006	0.002	0.008	-0.005	-0.011	-0.016
3	-0.004	0.015	0.011	-0.002	-0.019	-0.021
4	-0.012	0.025	0.013	0.005	-0.023	-0.018
5	-0.013	0.029	0.016	-0.006	-0.023	-0.029
6	-0.017	0.039	0.022	-0.008	-0.032	-0.040
7	-0.017	0.044	0.027	-0.009	-0.039	-0.048

[1]The number of QTL alleles in the base population is twice the number of base parents. Standard errors for overall genetic difference are 0.004 (generation 1), 0.011 (generation 4) and 0.018 (generation 7).

When the location was in error by 5 cM, the rates of genetic gain with the closer marker spacing (1 cM) were still superior to gains with the 5-cM marker spacing. However, when the QTL location was incorrectly estimated by 10 or 15 cM, rate of genetic gain in the later generations was slower with the marker haplotype bracket with 1-cM spacing than with the 5-cM spacing (Table 5). The results for the A10 were very similar to those of the BP2 model (results not shown); the closer marker spacing haplotype was more sensitive to location errors of 10 or 15 cM.

Incorrect position and size estimation: For a 5% QTL, the effects were evaluated of a 15% variance estimate combined with a QTL location that was incorrect by 15 cM. This scenario combines the largest inaccuracies for location and effect studied. Figure 4 shows the individual effect of estimating the QTL variance to be 15% (identical to Table 3), location estimate to be incorrect by 15 cM, and the combined effect of these two parameter errors compared to MAS with correct parameter estimates. The level of inferiority for the combined effect of the two parameter estimates errors (Figure 4), is greater than the superiority of MAS for a 5% QTL, with correct parameter estimates, over the control (without MAS) for the BP2 model (Table 1). Therefore, the combination of the two parameter errors resulted in rates of genetic gain with MAS that were inferior to rates without MAS for all generations.

FIGURE 4: Effect of incorrect location ([]) of a 5% quantitative trait locus (QTL) and overestimation of the QTL effect (by a factor of three) (■) independently and together (▲) compared with marker assisted selection (MAS) using the correct location and phenotypic variance (σ_P^2) estimate for the genetic model where the number of QTL alleles in the base population is twice the number of base parents.



The combination of incorrect variance and location estimates resulted in the overall genetic inferiority of this scenario being marginally more than the sum of the cumulative losses of the location and variance errors occurring separately for the BP2 model as well as for the A10 model (not shown).

DISCUSSION

Rates of genetic gain: Genetic responses to MAS were superior to those achieved under the control scheme. The larger the QTL effect, the greater was the short-term genetic response to MAS. In the long-term (seven generations), the percentage of superiority of MAS over the control was similar for both proportions of QTL variance, but was greater for the model with more QTL alleles. The genetic responses to MAS for the 10% QTL was 15% in the first generation, decreasing to a cumulative superiority of 5% over the control. This level of response falls into the range of responses to MAS that have been previously reported; of previous studies, the model of MEUWISSEN and GODDARD (1996) is the closest to the model used in this present study. The genetic responses for the 10% QTL in this study are slightly higher than those observed by MEUWISSEN and GODDARD (1996) for a single QTL with BP2 alleles of similar size.

BLUP model for estimating QTL allelic effects: The MA BLUP method requires the knowledge of polygenic and QTL variance and methods have been developed to estimate these variances (CLARKE *et al.* 1997; UIMARI *et al.* 1996). The MA BLUP method gave unbiased estimates of the QTL allelic effects and polygenic effects for the 5% QTL over all MAS generations for the BP2 model (results not shown). For the 10% QTL, the QTL allelic effects were overestimated, and polygenic gain was underestimated, in the last two to three MAS generations. This result may be caused by the decrease in QTL variance through changes in QTL allelic frequencies. These changes in QTL allelic frequency changes violate the assumption for the QTL component of the mixed model that QTL variation is not affected by changes in allele frequency. For the A10 model, the QTL effects were underestimated in the early generations (results not shown). DE BOER and VAN ARENDONK (1992) also found an effect of changes in allelic frequency on estimates of polygenic variance.

Genetic model: The conclusion of MEUWISSEN and GODDARD (1996) that the simulated genetic model does not affect the rate of genetic gain in the first three generations of MAS, was observed in this study for the 5% QTL but not for the 10% QTL (Tables 1 and 2). Their (MEUWISSEN and GODDARD 1996) conclusion was based on similar rates of genetic gain for a BP2 genetic model and two allele QTL for the QTL of approximately 9% phenotypic variance. The QTL variance of 9% is closest to the 10% QTL in this study, which found differences in genetic response from the two genetic models. The finding of MEUWISSEN and GODDARD (1996) is likely to be sensitive to the frequency of the best allele, which was 25% in their study. There were no large differences between the simulated genetic models in this study in sensitivity to inaccurate estimates of QTL location or effect.

The lower genetic response for the A10 genetic model than for the BP2 model was primarily due to order statistics. The expectation of the allele with the biggest effect was smaller when 10 alleles were drawn from a normal distribution than when 640 (BP2) were drawn. Therefore, the BP2 model has many alleles that have larger effects than the largest allele in the A10 model. All of the BP2 alleles would occur at low frequencies, but, if the favourable alleles were retained through selection, this retention would result in greater rates of genetic gain than with the A10 model. In agreement with the results of MEUWISSEN and GODDARD (1996), the number of alleles that were simulated at the QTL affected longer term genetic response, because QTL variation decreased more quickly with the A10 model than with the BP2 model.

The large number of alleles under the BP2 model has been defined to represent the situation in which the assumed QTL effects are actually due to a cluster of closely linked QTL (FERNANDO and GROSSMAN 1989; MEUWISSEN and GODDARD 1996). The superiority of MAS for a cluster of 5 biallelic QTL of equal effect was investigated for a 5 and 10% QTL (results not shown). The superiority of MAS over the control was closer to that achieved by the BP2 model than the A10 genetic model.

Selection for a non-existent QTL: Genetic gain was less that that of the control when a type I error in a QTL experiment was made that identified and located a QTL that did not exist and then utilised it through MAS. The 7% loss for the 10% QTL in the first generation was less than the 14% reported by MEUWISSEN and

GODDARD (1996) when selecting for a QTL that explained approximately 9% of phenotypic variance. The lower loss in this present study is likely because of MEUWISSEN and GODDARD (1996) selected before records became available; thus, marker information was more important in that setting.

The degree of loss in this study is less than the predicted gain, which differs from the conclusion of SALES and HILL (1976) that the predicted benefit equalled the real loss in efficiency when a trait is included that in reality does not contribute. When marker data are only present on the parents and grandparents of the generation that MAS is to be undertaken, the genetic loss from selection for a non-existent QTL was approximately equal to that of the expected gain, which agreed with the finding of SALES and HILL (1976). The value of marker information on earlier generations is important because that information reduces the relative genetic loss when type I errors are made, and increases the rate of genetic gain when there are QTL, as shown by MEUWISSEN and GODDARD (1996) and as observed in this study (results not presented). Larger groups of full and half-sib in each generation also increases the accuracy of estimation of QTL effects and therefore affects genetic response to MAS and, most likely, sensitivity to errors.

The BLUP model partially self-corrected when QTL allelic effects were estimated over the seven generations of MAS when no QTL existed (Figure 3). This correction would be due to the expected genetic differences being non-existent on average between the two offspring groups that were presumed to have received different QTL alleles from parents with different predicted allelic effects. Consequently, the re-estimation of QTL allelic effects over time reduces the vulnerability of MAS to type I errors in QTL detection compared with schemes that do not re-estimate QTL allelic effects.

Overestimation of the phenotypic variance explained by the QTL: The effect of overestimation of the QTL variance was minimal during the early generations of MAS, but long-term response was affected. Overestimation of the QTL variance is equivalent to applying a larger weight to the QTL than is optimal for a single generation. The greater polygenic loss with overestimation of QTL size in

the long run resulted in poorer overall genetic response than found with correct parameter estimation (Table 3).

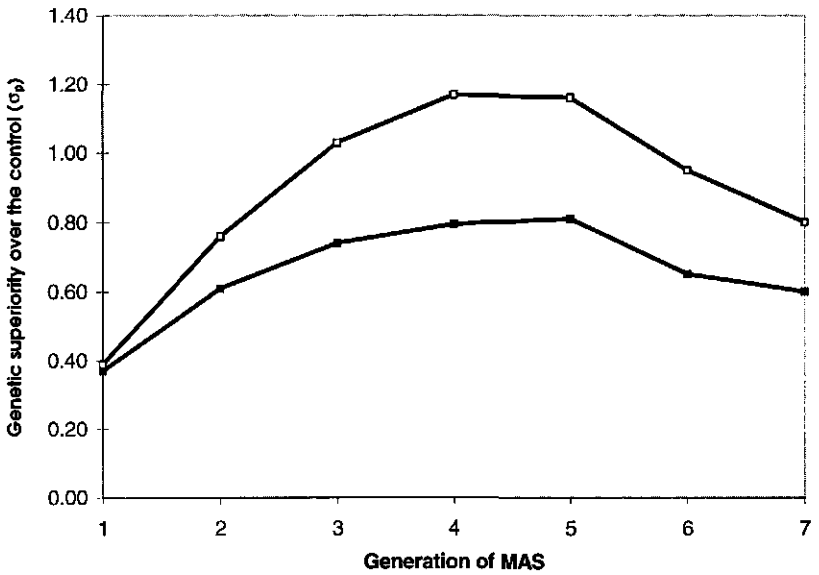
RUANE and COLLEAU (1996) investigated overestimation of QTL effect and observed genetic responses that were less than those achieved for the correct QTL variance. The loss in genetic response was similar to that found in this study.

After a number of generations, additional information will have been collected during the course of the MAS breeding scheme that allows re-estimation of QTL parameters. The scenario of correctly re-estimating QTL variance after three generations of MAS and then utilising this information in the procedure for breeding value estimation procedure was simulated for the BP2 model. For the first three generations of MAS, the 5% QTL variance was assumed to be 15%; in the last four generations, QTL variance was correctly assumed to be 5%, resulting in half the genetic loss at generation 7 that would exist after an incorrect estimate for all generations. This scenario is likely to be typical of MAS in practice (i.e., the ability to re-estimate QTL variance over time). Therefore, because overestimation of QTL variance did not have a very large effect in the initial generations of MAS, and because QTL variance could be re-estimated using the breeding population, MAS was relatively insensitive to QTL variance errors.

Incorrect positioning of postulated QTL: The degree of inaccuracy in QTL location estimates used in this study are probably indicative of the confidence intervals that can be expected from an initial genome scan. DARVASI *et al.* (1993) commented that a QTL of moderate effect can only be assigned to a map location in a rather broad chromosomal region with the usual experimental designs using F_2, backcross, half-sibs and full-sibs, even with an infinite number of markers, because of the limited meiosis observed.

The effect of an incorrect estimate of location was reduced genetic gain. The lower rate of genetic gain at the QTL became more predominant as the distance increased between the true QTL location and the postulated position. Genetic loss from incorrect location for generation 1 of MAS was more pronounced than that experienced through error in QTL variance.

Figure 5. Cumulative genetic superiority of marker assisted selection (MAS), for a quantitative trait locus (QTL) that explained 10% of the phenotypic variance (σ_P^2), over the control (without MAS) for the flanking markers with bracket of 20 cM (■) and weighted mean of 5-cM flanking markers with location errors ([]). For the 5-cM bracket, the following were assumed: 45% probability that location was correct, 35% probability of an error of 5 cM, 15% probability of 10 cM error, and 5% probability of 15 cM error. The weighted mean of the respective responses was calculated.



A larger bracket of flanking markers than the 5 cM used in this study may be used when the location estimate of the QTL is poor. However, using a larger marker bracket reduces the superiority of MAS (SPELMAN and BOVENHUIS 1998). The weighted superiority of MAS for a 10% QTL over the control was calculated for 5-cM flanking markers using the following assumptions: 45% probability that location was correct, 35% probability of 5-cM location error, 15% probability of 10-cM location error, and 5% probability of 15-cM error. The weighted superiority was compared with flanking markers with bracket size of 20 cM, for which 95% of the above QTL positions are still between the flanking markers (Figure 5). The weighted response for generation 1 was calculated as 0.45 × 0.048 (Table 1) + 0.35 × (0.048 – 0.007), where 0.007 is the loss from having a 5-cM location error (Table 4) + 0.15 ×

$(0.048 - 0.022) + 0.05 \times (0.048 - 0.037) = 0.041$. The weighted mean of the 5-cM flanking markers is superior to that of the 20-cM flanking markers for all generations (Figure 5). This result indicates that the use of a smaller bracket, despite location errors, is superior to trying to ensure that the QTL is always within the bracket by using a large distance between flanking markers.

General: Application of stringent type I thresholds in QTL experiments or, equivalently, experiments with low power have overestimated the effect of the QTL (GEORGES *et al.* 1995; WANG 1995). The effect of overestimation of QTL effect or variance has been shown in this study to have a minor effect on genetic gain in the short term. Long-term loss associated with overestimation can be minimised by re-estimating QTL variance after some generations of MAS. In contrast, when a non-existent QTL is selected for (type I error), genetic gain is affected adversely in the first two to three generations of MAS. Therefore, it is better to have stringent threshold levels to reduce type I errors and to avoid using MAS on non-existent QTL.

Genetic response to underestimation of QTL variance was not investigated in this study. RUANE and COLLEAU (1996) reported that the impact on the genetic response was less substantial than for overestimation, which introduces the option of shrinking the QTL variance estimate in the MA BLUP procedure to reduce the risk of suboptimal genetic gain. A reduction in the QTL variance is equivalent to lower QTL weighting. It has been shown that the longer the time frame that genetic gain is to be optimised, the lower is the QTL weighting (DEKKERS and VAN ARENDONK 1998). Breeding organisations implementing this technology may be more interested in short term response than long term response. If the method to estimate the QTL variance is unbiased, there is no reason to shrink the estimate. The parameter estimate should be used and re-estimated when new data are available. However, for the situations where the method of estimation of QTL variance is known to result in overestimates, then shrinkage is a viable option.

The sensitivity of genetic gain with MAS to incorrect QTL location implies that it is important to improve location estimates from those that are achieved by the initial genome scan. There are fine mapping methods, such as identity-by-descent, that have been successfully used to locate single genes in linkage studies with humans

(PUFFENBERGER *et al.* 1994) and cattle (CHARLIER *et al.* 1996). The applicability to quantitative traits is still uncertain, but those fine mapping methods can probably be used to localise QTL better than is currently achieved by the first genome scan (GEORGES and ANDERSSON 1996).

The results from this study illustrate that verification studies should be undertaken to ensure that putative QTL are in fact segregating and to provide better QTL location estimates in order to make optimum use of MAS in breeding schemes.

ACKNOWLEDGEMENTS

The authors acknowledge the helpful comments of THEO MEUWISSEN, HENK BOVENHUIS, and SIJNE VAN DER BEEK in the preparation of this manuscript. Livestock Improvement Corporation is thanked for its financial support.

LITERATURE CITED

BRISBANE, J. R., and J. P. GIBSON, 1995 Accuracy of evaluation and correlation of estimated breeding values among relatives, with evaluation based on information from relatives or from identified loci. J. Anim. Breed. Genet. **112:** 17-32.

CHARLIER, C., F. FARNIR, P. BERZI, P. VANMANSHOVEN, B. BROUWERS, *et al.* 1996 Identity-by-descent mapping of recessive traits in livestock: application to map the bovine *syndactyly* locus to chromosome 15. Genome Res. **6:** 580-589.

CLARKE, B. E., J.A.M. VAN ARENDONK, and B. P. KINGHORN, 1997 Analysis of linkage between genetic markers and QTL using REML: The effect of selection on parameter estimates. Proc. Assoc. Advmt. Anim. Breed. Genet. **12:** 208-212.

DARVASI, A., A. WEINER, V. MINKE, J. I. WELLER, and M. SOLLER, 1993 Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics **138:** 1365-1373.

DE BOER, I.J.M., and J.A.M. VAN ARENDONK, 1992 Prediction of additive and dominance effects in selected or unselected populations with inbreeding. Theor. Appl. Genet. **84:** 451-459.

DEKKERS, J.C.M., and J.A.M VAN ARENDONK, 1998 Optimum selection for quantitative traits with information on an identified locus. Genet. Res. (in press)

FERNANDO, R. L., and M. GROSSMAN, 1989 Marker-assisted selection using best linear unbiased prediction. Genet. Sel. Evol. **21:** 467-477.

GEORGES, M., and L. ANDERSSON, 1996 Livestock genomics comes of age. Genome Res. **6:** 907-921.

GEORGES, M., D. NIELSEN, M. MACKINNON, A. MISHRA, R. OKIMOTO, *et al.* 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.

HALDANE, J.B.S., 1919 The combination of linkage values and the calculation of distances between the loci of linked factors. J. Genet. **2:** 3-19.

HENDERSON, C. R., 1984 Applications of Linear Models in Animal Breeding. Can Catal Publ Data, Univ, Guelph, Guelph, ON, Canada.

KAPPES, S. M., J. W. KEELE, R. T. STONE, T. S. SONSTEGARD, T.P.L. SMITH, *et al.* 1997 A second generation linkage map of the bovine genome. Genome Res. **7(3):** 235-249.

KASHI, Y., E. HALLERMAN, and M. SOLLER, 1990 Marker assisted selection of candidate bulls for progeny testing programmes. Anim. Prod. **51:** 63-74.

MEUWISSEN, T.H.E., and M. E. GODDARD, 1996 The use of marker haplotypes in animal breeding schemes. Genet. Sel. Evol. **28:** 161-176.

MEUWISSEN, T.H.E., and J.A.M. VAN ARENDONK, 1992 Potential improvements in rate of genetic gain from marker-assisted selection in dairy cattle breeding schemes. J. Dairy Sci. **75:** 1651-1659.

PUFFENBERGER, E. G., E. R. KAUFFMAN, S. BOLK, T. C. MATISE, M. WASHINGTON, *et al.* 1994 Identity-by-descent and association of a recessive gene for Hirschprung disease on human chromosome 13q22. Hum. Mol. Genet. **3:** 1217-1225.

RUANE, J., and J. J. COLLEAU, 1995 Marker assisted selection for genetic improvement of animal populations when a single QTL is marked. Genet. Res. **66:** 71-83.

RUANE, J., and J. J. COLLEAU, 1996 Marker-assisted selection for a sex-limited character in a nucleus breeding population. J. Dairy Sci. **79:** 1666-1678.

SALES, J., and W. G. HILL, 1976 Effect of sampling errors of efficiency of selection indices. 2. Use of information on associated trait for improvement of a single important trait. Anim. Prod. **23:** 1-14.

SCHAEFFER, L. R., and B. W. KENNEDY, 1986 Computing solutions to mixed model equations. Proceedings of the 3rd World Congress on Genetics applied to Livestock Production, Lincoln, Nebraska. **12:** 382-393.

SPELMAN, R. J., and H. BOVENHUIS, 1998 Genetic response from marker assisted selection in an outbred population for differing marker bracket sizes and with two identified quantitative trait loci. Genetics **148:** 1389-1396.

SPELMAN, R. J., W. COPPIETERS, L. KARIM, J.A.M. VAN ARENDONK, and H. BOVENHUIS, 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., and D. J. GARRICK, 1997 Utilisation of marker assisted selection in a commercial dairy cow population. Livest. Prod. Sci. **47:** 139-147.

TIER, B., 1990 Computing inbreeding coefficients quickly. Genet. Sel. Evol. **22:** 419-430.

UIMARI, P., G. THALLER, and I. HOESCHELE, 1996 The use of multiple markers in a Bayesian method for mapping QTL. Genetics **143:** 1831-1842.

VAN OOIJEN, J. W., 1992 Accuracy of mapping quantitative trait loci in autogamous species. Theor. Appl. Genet. **84:** 803-811.

WANG, Y., 1995. Detection and estimation of associations between genetic markers and quantitative trait loci in segregating populations. Ph.D. Diss., Univ, Guelph, Guelph, ON, Canada.

CHAPTER FOUR

GENETIC RESPONSE FROM MARKER ASSISTED SELECTION IN AN OUTBRED POPULATION FOR DIFFERING MARKER BRACKET SIZES AND WITH TWO IDENTIFIED QUANTITATIVE TRAIT LOCI

RICHARD J. SPELMAN and HENK BOVENHUIS

PUBLISHED IN GENETICS (1998) **148:** 1389-1396.

ABSTRACT

Effect of flanking quantitative trait loci (QTL)-marker bracket size on genetic response to marker assisted selection (MAS) in an outbred population was studied by simulation of a nucleus breeding scheme. In addition, genetic response with MAS from two QTL on the same and different chromosome(s) was investigated. QTL loci that explained either 5% or 10% of phenotypic variance were simulated. A polygenic component was simulated in addition to the QTL. In total, 35% of phenotypic variance was due to genetic effects. The trait was measured on females only. Having smaller flanking QTL-marker brackets increased the genetic response from MAS. This was due to the greater ability to trace the QTL transmission from one generation to the next with the smaller flanking QTL-marker bracket, which increased the accuracy of estimation of the QTL allelic effects. Greater negative covariance between effects at both QTL was observed when two QTL were located on the same chromosome compared to different chromosomes. Genetic response with MAS was greater when the QTL were on the same chromosome in the early generations and greater when they were on different chromosomes in the later generations of MAS.

KEY WORDS: Marker assisted selection, Quantitative trait loci, Genetic response, Marker bracket, Breeding scheme.

INTRODUCTION

Quantitative trait loci (QTL) are being detected in many species using many different experimental designs. In outbred livestock populations, half-sib experimental designs (WELLER *et al.* 1990) have been successfully used to identify QTL (e.g. GEORGES *et al.* 1995; SPELMAN *et al.* 1996). Similar experimental designs have also been used successfully for QTL detection in forest trees (GRATTAPAGLIA *et al.* 1996). In other livestock species such as poultry and pigs crosses between divergent lines have been used in two and three generation experimental designs (ANDERSSON *et al.* 1994; VAN DER BEEK *et al.* 1995). Utilisation of QTL detected in these livestock and forest populations through marker assisted selection (MAS) is still at the theoretical level but will most probably be applied in the near future.

Theoretical evaluation of MAS in breeding schemes has been undertaken starting with the work of NEIMANN-SORENSON and ROBERTSON (1961). Sporadically over the next 25 years further papers addressed MAS (e.g. SMITH 1967; SOLLER 1978; SOLLER and BECKMANN 1982; SMITH and SIMPSON 1986; STAM 1986). In the 1990's, there have been many papers evaluating MAS. These studies have investigated MAS for dairy cattle (e.g. KASHI *et al.* 1990; MEUWISSEN and VAN ARENDONK 1992; BRASCAMP *et al.* 1993; SPELMAN and GARRICK 1997), forestry (e.g. WILLIAMS and NEALE 1992; STRAUSS *et al.* 1992), poultry (e.g. VAN DER BEEK and VAN ARENDONK 1995) and for other situations (e.g. LANDE and THOMPSON 1990; GIMFELARB and LANDE 1994a,b; RUANE and COLLEAU 1995, 1996). The theoretical genetic responses from MAS in these studies have varied among studies as many different QTL sizes, genetic models and breeding schemes have been modelled. However, the near unanimous conclusion from these studies is that extra genetic responses through utilising MAS can be made. Larger increases in genetic response with MAS are seen for low heritability traits (SMITH 1967), and for traits where selection is undertaken before the phenotype is observed on selection candidates, or the trait is sex-limited or carcass limited (MEUWISSEN and GODDARD 1996).

Molecular geneticists are continually developing and applying different methods in trying to get closer to the QTL of interest (GEORGES and ANDERSSON 1996). From a scientific point of view, this is important and interesting. However, for the application of MAS in a breeding programme the benefits from this extra work has not been quantified. SMITH and SMITH (1993) advocated the need to have close marker QTL linkages in outbred populations (1-2 cM) so that selection could exploit linkage disequilibrium between marker and QTL. However, the benefits of this were not quantified and has been questioned by others (e.g. VAN ARENDONK *et al.* 1994a).

Genome scans have identified multiple QTL that affect the same trait (e.g. GEORGES *et al.* 1995). Plant breeding programmes have not limited themselves to using MAS for only one QTL but have selected for many QTL at the same time (STUBER and EDWARDS 1986). However, STUBER and EDWARDS' (1986) MAS selection was solely on marker information and did not account for the genetic variation not explained by the markers. Livestock and forestry breeding programmes

are also likely to implement MAS for multiple QTL that affect the same trait. To date the utilisation of more than one QTL in MAS has not been extensively investigated.

The objective of this study is to quantify the effect of differing sizes of flanking QTL-marker brackets on genetic response from MAS. In addition, genetic response from two QTL on the same and different chromosome(s) is investigated. Furthermore for the two QTL situation, genetic responses are investigated for two QTL of the same size, and also one large QTL and one small QTL.

METHOD

Simulation model: A stochastic simulation modelling a closed nucleus breeding scheme with discrete generations (each animal present as parent for only one generation) was developed. The initial generation of animals (termed base population) were unselected, unrelated and non-inbred. Each generation had 1024 animals with equal numbers of males and females. A single trait was simulated with base population heritability of 0.35, where heritability is the additive genetic variance divided by the phenotypic variance. The additive genetic variance was divided between unmarked additive polygenic variation (which will be referred to as polygenic variance) and variation due to the marked chromosomal region(s) (which will be referred to as QTL). Phenotypic records were recorded on females only. The highest ranking 12.5% of males and 50% of females for estimated genetic merit were selected as parents of the next generation. As phenotypes were only available on females, estimates of male genetic merit were calculated from pedigree information (e.g. sire, dam and full- and half-sib information). Selection of males and females was undertaken after the single phenotypic record for females was available. Each sire was mated to four females (avoiding half-sib and closer matings) and each mating resulted in four offspring (two male and two female). Each female was mated to one sire only.

QTL alleles for the unselected base population were drawn from the distribution $N(0,\frac{1}{2}V_{QTL})$, where V_{QTL} is the variance explained by the QTL. Two QTL variances were used in this study: 5% and 10% of phenotypic variance. The additive genetic variance (polygenic variance plus QTL variance) was constant at 35% for both

QTL variances. The number of QTL alleles in the base population was twice the number of parents selected from this generation. The large number of alleles represents the situation where the assumed QTL affect is actually due to a cluster of closely linked QTL.

A polygenic effect for each animal in the base population was sampled from the distribution $N(0, V_a)$, where V_a is the polygenic variance. In subsequent generations, the polygenic component was sampled from the distribution $N(\frac{1}{2}a_s + \frac{1}{2}a_d, \frac{1}{2}(1-\frac{1}{2}(F_s+F_d))V_a)$, where s and d denote sire and dam, a is the true polygenic value, and F is the inbreeding coefficient that was calculated using the algorithm presented by TIER (1990). The inbreeding coefficient is the probability that the two genes at any locus in an individual are identical by descent (FALCONER and MACKAY 1996 pp.52). Residual components from the distribution $N(0, V_e)$, where V_e is the residual variance, were sampled for females and added to the previously sampled polygenic and QTL effects to complete the phenotypic observations. Phenotypic variance in the base population, that comprised of $V_a + V_{QTL} + V_e$, had an expected value of 100, and $V_a + V_{QTL}$ had an expected value of 35.

Marker alleles were simulated for all animals in the base population. It was assumed that the linkage map had six markers that bracketed the postulated QTL position (Figure 1). For the individuals in the base population, marker genotypes were simulated for each of the marker loci assuming five alleles with equal frequency. HALDANE (1919) mapping function was assumed for the construction of the marker-QTL haplotypes transmitted to the offspring.

The required number of sires (64) and dams (256) were simulated for the base population and mated to produce the first generation. Three generations of selection were undertaken without using marker genotypes in the estimation of an animal's genetic merit. Polygenic variance decreases while selection is undertaken because of induced negative covariance between polygenes (BULMER 1971). The level of polygenic variance stabilises over time and the three generations of conventional breeding (without markers) were undertaken to enable this to occur before using MAS. MAS was introduced after the three generations of conventional breeding and,

therefore, the MAS genetic responses represent MAS in an ongoing breeding programme. MAS was undertaken for seven generations in total. The generation number for offspring born from the first application of MAS will be termed generation one in this paper. Therefore the base population is generation −4.

Breeding value estimation: Breeding value estimation (estimation of genetic merit) of polygenic and marker linked effects for MAS was undertaken using the model described by MEUWISSEN and GODDARD (1996):

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \Sigma_i \mathbf{Z}\mathbf{Q}_i\mathbf{q}_i + \mathbf{e}$$

where

$\mathbf{y}$ = vector of phenotypic records,

$\mathbf{X}$ = incidence matrix linking fixed effects to records,

$\mathbf{b}$ = vector of fixed effects,

$\mathbf{Z}$ = incidence matrix linking animals to records,

$\mathbf{u}$ = vector of polygenic effects,

$\mathbf{Q}_i$ = incidence matrix linking allelic effects for the *i*th QTL to animals (every row has two elements equal to one and the other elements are zero),

$\mathbf{q}_i$ = vector of allelic effects for *i*th QTL, and

$\mathbf{e}$ = vector of residual effects.

Mixed model equations (HENDERSON 1984) are used for best linear unbiased predictions (BLUP) of $\mathbf{b}$, $\mathbf{u}$ and $\mathbf{q}$ (for one QTL):

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'ZQ} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\mathbf{A}^{-1}\lambda & \mathbf{Z'ZQ} \\ \mathbf{Q'Z'X} & \mathbf{Q'Z'Z} & \mathbf{Q'Z'ZQ}+\mathbf{G}^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{q}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{Q'Z'y} \end{bmatrix}$$

where

$\mathbf{A}^{-1}$ = inverse of numerator relationship matrix,

λ = residual variance / polygenic variance,

$\mathbf{G}^{-1}$ = inverse of the matrix that describes the relationship between the QTL alleles, and

α = residual variance / half the QTL variance.

This model is an extension of the methods of FERNANDO and GROSSMAN (1989) that was developed for single markers and GODDARD's (1992) method that adapted the previous model for marker haplotypes.

In brief, the computational method for marked-QTL considers that in the base population the number of QTL alleles is equal to twice the number of base animals. In the next generation, the transmission of the parental QTL alleles is followed by inference on marker haplotype. When transmission of marker haplotype can be followed, the **Q** matrix links the progeny's phenotype to the transmitted parental QTL allelic effect. When it is uncertain which QTL allele was transmitted a new QTL allelic effect is formed in the evaluation procedure. The progeny's phenotype is linked via the **Q** matrix to the new QTL allelic effect and the new QTL allelic effect is linked to its parents through the **G** matrix i.e. the expectation of the new QTL allelic effect is equal to mean of the parental QTL allelic effects.

The evaluation model does not assume that the exact location of the QTL within a marker bracket is known, but postulates that it is within the marker bracket. Probability statements are either that QTL transmission can be followed by inference from marker haplotype, or it can not. Thus probability statements, other than 0 or 1, are not made about transmission based on recombination events between flanking markers (double recombination) and postulated position relative to single markers (for further description of model see MEUWISSEN and GODDARD, 1996). The described MAS breeding value estimation method will be referred to as MA-BLUP for the rest of the paper.

If origin of marker allele could not be established at the closest flanking markers around the postulated QTL, based on parental and offspring marker genotypes, then the next marker in the haplotype was used. If allele origin could not be determined for at least one side of the marker haplotype, QTL transmission could

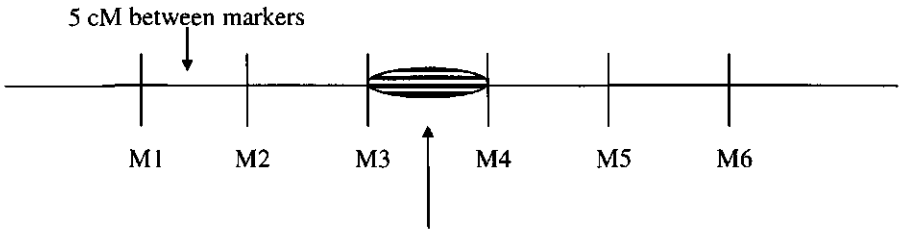
not be determined according to the rules of MEUWISSEN and GODDARD (1996). Also if a recombination was observed between markers, QTL transmission could not be determined.

From generation −3 conventional mixed model equations (marker information not used) (HENDERSON 1984) were used to estimate **b** and **u**. After three generations of conventional selection, MAS was undertaken, using marker information and phenotypic observations, from generation zero with the aforementioned MA-BLUP model. Markers were available on all animals. As a control, conventional selection was also continued for seven generations from generation zero. The additive genetic variance used in solving the mixed model equations for situations without MAS was the sum of polygenic variation and QTL variation in the base population.

Estimates of polygenic and QTL effects were obtained using iteration on the data (SCHAEFFER and KENNEDY 1986). Iterations were continued until solutions were stable, i.e. when convergence criterion, which equals the sum of squares of differences in solutions between iterations divided by the sum of squares of the most recent solutions, was less than 10^{-10}.

Differing flanking marker-QTL size: The size of the interval between the two flanking QTL-markers was varied to determine the genetic benefit for MAS of localising a QTL to a small chromosomal area. The four distances studied were; 15 cM, 10 cM, 5 cM, and 2 cM. Distance to markers outside the flanking QTL-markers was kept constant in all simulations at 5 cM (Figure 1). One hundred and sixty replicates were simulated for both MAS and the control for each scenario investigated.

FIGURE 1: Marker haplotype that surrounds the postulated location of the QTL.



Two QTL: Two QTL were simulated either on the same or on different chromosomes. The number of alleles simulated for both QTL was twice the number of base parents. The variance due to QTL were either the same size or one accounted for 75% of the QTL variance and the other 25%. The combined variance of the two QTL was either 5% or 10% of the phenotypic variance i.e. the same levels as used for the one QTL models. When the two QTL were placed on the same chromosome the distance between the two QTL was 30 cM. Thirty centiMorgans was chosen as this distance is the approximate level of resolution that one can identify two separate QTL in current livestock QTL experiments (HALEY and KNOTT 1992) The flanking QTL-marker distance was 5 cM in all cases. QTL allelic effects were estimated separately for both QTL by extending the MA-BLUP model. Negative covariance generated by selection, between the two QTL, and also between the polygenic and QTL components was evaluated. The negative covariance between the two QTL was calculated as half of the difference between total QTL variance, less the sum of the two individual QTL variances. The negative covariance between the QTL component and polygenic component was calculated each generation as half of the difference between total additive genetic variance less the sum of the QTL variance and polygenic variance.

The control for the two QTL scenarios was conventional selection on the genetic model of polygenic variance and variance at two QTL. One hundred and sixty replicates were simulated for both MAS and the control for each scenario investigated.

RESULTS

Genetic gain with base model: The rate of genetic gain for the breeding scheme modelled for a trait of 35% heritability that consisted solely of polygenic variance was close to 0.3 σ_P per generation. Equilibrium response with this model was reached after three to four generations of conventional BLUP selection, confirming that three generations of conventional breeding was sufficient to mimic the introduction of MAS in to an ongoing breeding scheme.

Flanking QTL-marker size: The smaller the flanking QTL-marker bracket the greater the cumulative superiority of MAS over the control (Tables 1 and 2). The

5-cM bracket had, on average, 90% and 85% of the genetic superiority of MAS (over the control) which was achieved with the 2-cM bracket for the 5% and 10% QTL, respectively. The 10-cM bracket achieved an average genetic response of some 80% relative to that of the 2-cM bracket for both sized QTL (results not shown). For the 5% QTL and 15-cM bracket, the MAS superiority was quite variable, relative to the 2-cM bracket (Table 1) and lower than that of the 10% QTL (Tables 1 and 2). The relative superiority of the 5% QTL for the 15-cM bracket is similar to that of the 20-cM bracket for the 10% QTL (results not shown).

TABLE 1: Effect of differing flanking QTL-marker bracket size on cumulative superiority of MAS over the control for a QTL that explains 5% of phenotypic variance (σ_P).

	Flanking marker bracket size							
	2 cM		5 cM			15 cM		
Generation	QTL	Polygenic	QTL	Polygenic	%	QTL	Polygenic	%
1	0.049	-0.026	0.046	-0.024	96	0.036	-0.021	70
2	0.090	-0.055	0.088	-0.053	95	0.068	-0.051	45
3	0.135	-0.081	0.131	-0.085	85	0.099	-0.079	36
4	0.166	-0.095	0.158	-0.097	83	0.122	-0.091	43
5	0.194	-0.110	0.182	-0.111	85	0.143	-0.104	46
6	0.208	-0.117	0.192	-0.106	95	0.154	-0.105	54
7	0.212	-0.122	0.195	-0.112	92	0.160	-0.115	49

Cumulative overall genetic superiority for the 5 cM and 15 cM QTL-marker brackets is presented as percentage of that achieved with the 2 cM bracket.
Standard errors for the QTL component are 0.003 (generation 1), 0.007 (generation 4) and 0.010 (generation 7). Standard errors for the polygenic component are 0.002 (generation 1), 0.010 (generation 4) and 0.014 (generation 7).

The difference in relative response of the 15-cM bracket to the 2-cM bracket between the 5% and 10% QTL, after generation one, may reflect that the value of phenotypes is a curvilinear function, i.e. the first phenotypes per QTL allelic effect have a larger effect on accuracy than the additional ones. The number of phenotypes needed per QTL allelic effect to get a certain accuracy will be larger for the 5% QTL

than the 10% QTL, since the 5% QTL explains less of the phenotypic variance. Thus, for the 5% QTL the 15-cM flanking QTL-marker bracket may move the accuracy of QTL estimation off the plateau-like level of the curvilinear slope. However, for the 10% QTL, the reduction in number of phenotypes per allelic effect when going from a 10-cM bracket to a 15-cM bracket may only reduce accuracy a little. This was observed with the reduction in QTL accuracy decreasing more for the 5% QTL than the 10% QTL when going from a 10-cM bracket to 15-cM (not shown).

Table 2: Effect of differing flanking QTL-marker bracket size on cumulative superiority of MAS over the control for a QTL that explains 10% of phenotypic variance (σ_P).

	Flanking marker bracket size							
	2 cM		5 cM			15 cM		
Generation	QTL	Polygenic	QTL	Polygenic	%	QTL	Polygenic	%
1	0.113	-0.054	0.106	-0.054	90	0.086	-0.044	71
2	0.205	-0.096	0.189	-0.093	88	0.157	-0.077	73
3	0.274	-0.122	0.258	-0.131	83	0.213	-0.107	70
4	0.308	-0.144	0.295	-0.151	88	0.244	-0.124	73
5	0.310	-0.142	0.301	-0.158	85	0.252	-0.121	77
6	0.292	-0.140	0.291	-0.167	82	0.244	-0.126	78
7	0.265	-0.134	0.267	-0.163	79	0.224	-0.128	73

Cumulative overall genetic superiority for the 5 cM and 15 cM QTL-marker brackets is presented as percentage of that achieved with the 2 cM bracket.
Standard errors for the QTL component are 0.002 (generation 1), 0.009 (generation 4) and 0.011 (generation 7). Standard errors for the polygenic component are 0.003 (generation 1), 0.008 (generation 4) and 0.013 (generation 7).

The source of the extra genetic gain with the smaller marker brackets was from extra gain made at the QTL when moving from a 15-cM bracket to a 5-cM bracket for the 5% and 10% QTL (Tables 1 and 2). Moving from a 5-cM to a 2-cM bracket, for the 10% QTL, the increase in overall genetic gain was from extra QTL response in the first two generations. In the next three generations the extra gain was from both QTL and polygenic and in the last two generations it came from a reduction in polygenic

loss (Table 2). For the 5% QTL the extra genetic gain from going from a 5-cM bracket to a 2-cM bracket came from primarily an increase in QTL response with polygenic loss staying stable (Table 1).

Ability to follow transmission of QTL: The ability to unambiguously follow QTL transmission from parent to offspring based on marker haplotype decreased over generations (Table 3). The size of the flanking QTL-marker bracket affected the ability to follow QTL transmission in the first four to five generations of MAS but after seven generations there were only minor differences (Table 3). Reduction in ability to follow QTL transmission was greater for the 10% QTL compared to the 5% QTL due to greater QTL selection pressure and therefore faster fixation (results not shown).

TABLE 3: Effect of flanking QTL-marker bracket size on the ability to determine parental origin of QTL allele based on marker genotypes for 5% QTL (%).

Generation	15 cM	10 cM	5 cM	2 cM
0	82	86	90	92
1	81	85	89	91
2	79	82	86	88
3	77	79	82	85
4	75	76	78	80
5	71	73	73	75
6	68	68	69	70
7	65	66	64	66

Correlation of estimated and true QTL effects: The smaller the flanking QTL-marker bracket the higher the correlation between estimated and true QTL effects for the 5% QTL (Table 4). This was also observed for the 10% QTL where the correlation between estimated effects and true effects was higher than that for the 5% QTL (results not presented). The correlation increased in the first three to four generations of MAS as more information (phenotypes) accumulated for the estimation of QTL allelic effects. In the last three to four generations of MAS, the correlation decreased as the ability to follow QTL transmission decreased and, therefore, new QTL allelic effects were formed in the evaluation method. The new allelic effects were allocated the average of the parental effects that resulted in lower accuracy.

For the 10% QTL, the BLUP evaluation method was slightly biased in the later generations and genetic gain at the QTL was over-estimated. This is probably due to the decrease in QTL variation through changes in allele frequencies, which violates the assumptions of the model. MAKI-TANILA and KENNEDY (1986) commented that this type of bias can occur when fixation or equivalently a selection limit is reached. Accuracy of polygenic estimates increased slightly as the QTL-marker bracket size decreased. This may be due to the greater accuracy of estimated QTL allelic effects. When estimating the polygenic value the phenotype is adjusted for the fixed effect and the QTL allelic effects. With greater accuracy for QTL effects the phenotype will be adjusted more correctly, resulting in more accurate estimate of polygenic value.

TABLE 4: Effect of flanking QTL-marker bracket size on the average correlation between estimated allelic effects and true effects for a QTL that explains 5% of the phenotypic variance.

Generation	15 cM	10 cM	5 cM	2 cM
0	0.53	0.56	0.58	0.66
1	0.57	0.61	0.63	0.71
2	0.59	0.63	0.68	0.71
3	0.59	0.64	0.70	0.74
4	0.57	0.66	0.70	0.75
5	0.56	0.61	0.68	0.73
6	0.53	0.58	0.64	0.70
7	0.49	0.55	0.60	0.65

Two QTL: For the two QTL that together explained 10% of the phenotypic variance, the genetic response was similar regardless of the relative size of the two QTL (Table 5). In the early generations of MAS, the genetic response with MAS was greater when the two QTL were located on the same chromosome than when they were on different chromosomes. In the later generations, the rate of genetic gain when the two QTL were on the same chromosome was less than when they were on different chromosomes. Comparing the two QTL which had a cumulative variance of

10% to one 10% QTL, the genetic superiority over no MAS was nearly the same for the first five generations. In the last two generations, the two QTL model had greater superiority over the control compared to the one QTL model. This was due to there being more QTL variance for the two QTL genetic model in the later generations compared to the single 10% QTL.

For the 5% QTL, the relative size of the two QTL had an effect on the percentage superiority of MAS over the control (Table 5). Having two QTL that were unequal in size resulted in lower percentage superiority in the later generations than that achieved with QTL of equal size. The lower response for the unequal QTL size for the 5% QTL was due to the size of the smaller QTL explaining only 1.25% of the phenotypic variance. MAS with a single QTL of this size (1.25%) was not superior to that without MAS (results not shown) as the accuracy of the QTL allelic effects was low for the breeding scheme structure simulated.

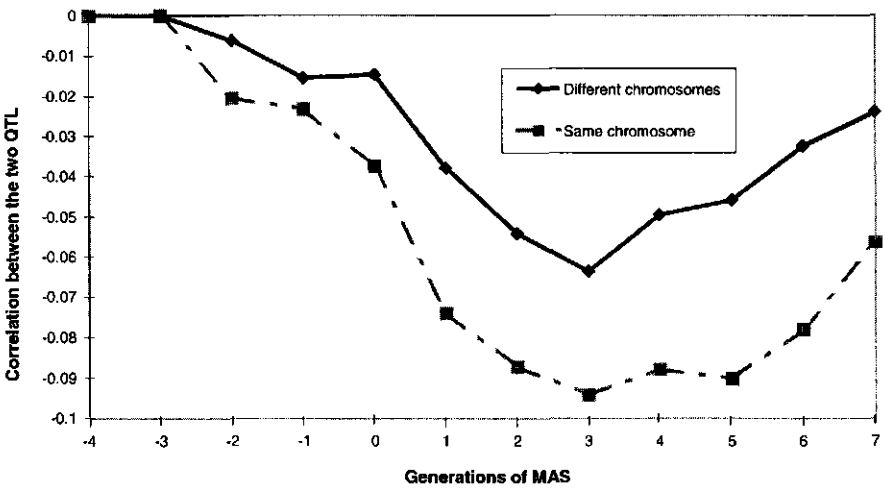
TABLE 5: Cumulative percentage difference in genetic response between MAS and non-MAS breeding programmes with two QTL that explain 5% and 10% of phenotypic variance.

	5%					10%				
Generation	One QTL	Same equal	Different equal	Same unequal	Different unequal	One QTL	Same equal	Different equal	Same unequal	Different unequal
1	7.1	6.0	4.9	5.5	5.3	15.3	16.8	14.4	14.9	16.2
2	5.4	5.9	5.7	5.0	5.6	14.4	13.8	12.8	15.0	14.7
3	4.8	5.6	6.1	4.9	5.4	13.1	12.5	11.6	13.7	13.3
4	4.8	5.7	6.3	4.8	5.2	10.7	11.5	12.0	11.3	12.3
5	4.5	5.5	5.8	4.3	4.8	8.6	10.3	11.3	10.0	11.3
6	4.7	5.3	5.6	4.1	4.8	6.6	9.8	10.9	8.3	9.5
7	3.9	5.0	5.3	4.0	4.4	5.0	8.2	9.4	7.2	8.4

Same = two QTL positioned on the same chromosome; Different = two QTL positioned on the different chromosome; Equal = two QTL explain the same amount of phenotypic variation; Unequal = one QTL explains more of the phenotypic variance than the other QTL. Standard errors are 1.1% for generation 1, 0.7% for generation 4 and 0.5% for generation 7 for the 5% QTL and 1.3% for generation 1, 0.9% for generation 4 and 0.7% for generation 7 for the 10% QTL.

When the two QTL were positioned on the same chromosome, the level of negative covariance between the two QTL was greater than when the QTL were on different chromosomes (Figure 2). The negative covariance increased in the generations previous to the introduction of MAS. With the introduction of MAS, the level of negative covariance between the QTL increased and the negative covariance remained at a higher level when the two QTL were on the same chromosome. When one QTL comprised 75% of the QTL variance and the other 25%, the level of negative covariance was less than that observed for two QTL of equal size (not shown). The level of negative covariance between the polygenic component and the QTL component, was not affected by the relative location of the two QTL nor relative size (not shown). The same trends were observed for two QTL that had a cumulative variance of 5%.

FIGURE 2: Negative covariance between the two QTL and between the polygenic component and the QTL component. Two equally sized QTL that explain 10% of phenotypic variance cumulatively and are located on the same or different chromosomes.



When the two QTL were of unequal size (75% and 25%), greater selection response was made at the larger QTL, as was expected. The level of contribution to the QTL variance from the two QTL changed over the generations. For the 10% QTL, the QTL variance in generation four comprised of 66% from the larger QTL and 34%

from the smaller, and by generation seven it was 50:50. For the 5% QTL the QTL variance in generation seven comprised of 60% from the larger QTL and 40% from the smaller QTL. In comparison, the level of variance contributed in the control was some 70:30 after seven generations for both sized QTL.

DISCUSSION and CONCLUSIONS

Negative covariance between two QTL was maintained at a higher level when the two QTL were on the same chromosome in contrast to being on different chromosomes. This is to be expected as the decay of negative covariance is slowed by linkage (BULMER 1971). That is, the unfavourable linkages between QTL alleles can only be broken by recombination when the QTL are on the same chromosome. It is interesting to note that the genetic response was higher in the early generations of MAS for the situation where the two QTL were on the same chromosome despite the higher negative covariance. In the later generations, the genetic response was greater when the QTL were situated on differing chromosomes, which would be expected. The level of negative covariance is affected by population size, selection intensity and mating structure (WEIR and HILL 1980). Therefore, the results presented on the effect of negative covariance may alter for different breeding scheme structures.

The accuracy of allelic effect estimates was reasonably high at the start of MAS (Table 4). This was due to marker genotypes being present on all five generations prior to the start of MAS. When MAS started with fewer previous generations of marker genotypes and phenotypes the genetic response to MAS was reduced (MEUWISSEN and GODDARD 1996; SPELMAN and VAN ARENDONK 1997) as the accuracy of estimated allelic effects was lower. Increasing the accuracy of QTL allelic effects can also be achieved by genotyping and phenotyping more full and half-sibs. This may be important for QTL that only explain a small percentage of the variance, as the breeding structure simulated in this study did not have enough observations to accurately estimate QTL effects and use them successfully via MAS for a QTL that explained 1.25% of phenotypic variance. Therefore, breeding schemes may have different optimal sizes for QTL of differing variances. This will also depend on how many previous generations of phenotypes and genotypes are available.

Therefore for a given breeding scheme you may decide not to select for a QTL below a certain size.

The greater accuracy in estimation of QTL effects with the smaller flanking brackets resulted in greater gain at the QTL when reducing bracket size from 15 cM to 10 cM and subsequently to 5 cM as would be expected. However, the greater polygenic response, or equivalently the reduction in polygenic loss, when reducing the bracket from 5 cM to 2 cM for the 10% QTL, was not expected. In the last two generations the greater response from the smaller bracket was solely from the polygenic component. The polygenic response may be due to the QTL allele being more accurately estimated in the 2-cM bracket situation and, therefore, the adjustment of phenotype in estimation of polygenic value is more correct. In the last two generations, when one QTL allele may be predominant, the same QTL allele may be selected for both bracket sizes but it is selected in animals with better polygenic value for the 2-cM bracket situation.

The genetic evaluation system used in this study (MEUWISSEN and GODDARD 1996), may be slightly more sensitive to flanking QTL-marker bracket sizes than other MAS evaluation methods proposed. This is due to the model in this study requiring that the marker haplotype is informative on both sides of the QTL location. Other methods (FERNANDO and GROSSMAN, 1989, VAN ARENDONK *et al.* 1994b; WANG *et al.* 1995) make probability statements about QTL transmission from single markers. Therefore, when markers on one side of the haplotype could not be followed, probability statements about QTL transmission would be made from a single marker rather than forming a new QTL effect. Making the probability statements from one side of the haplotype requires an estimate of the QTL location within the QTL-flanking marker bracket. However, by simulating relatively informative markers and three marker loci on each side of the QTL the effect of non-informity has been reduced in this study.

In the MA-BLUP method that was used in this study, a shortcoming was when the two QTL effects for a parent were the same and QTL transmission from the marker haplotype could not be followed. In this situation, a new QTL effect was

formed in MA-BLUP for the offspring. An improvement would be to identify via the evaluation method if two QTL effects were presumed to be the same in a parent and offspring of this parent get allocated this QTL effect in the **Q** matrix regardless of the marker haplotype information. This may have improved the accuracy of estimation of QTL effects in later generations.

EDWARDS and PAGE (1994) showed through simulation that the benefits for MAS when using flanking markers instead of single markers was 11% for markers close to the QTL and 38% for markers loosely linked to the QTL. This study has demonstrated and quantified that getting closer to the QTL and having smaller flanking QTL-marker brackets further increases the genetic response from MAS. The close flanking markers used in this study for MAS is different from the MAS scheme outlined by MEUWISSEN and VAN ARENDONK (1992). Those authors had only two markers on each chromosome forming the marker haplotype for estimation of QTL effects. As shown by this study, MAS schemes will genetically benefit from getting closer to the QTL or chromosomal segment. The improved genetic responses should be balanced against the costs of achieving it, particularly as the amount of work and cost required to get another centiMorgan closer is invariably more than it was for the previous centiMorgan.

ACKNOWLEDGEMENTS

The authors acknowledge the helpful comments of JOHAN VAN ARENDONK in the preparation of this manuscript. RJS thanks Livestock Improvement Corporation for financial support.

LITERATURE CITED

ANDERSSON, L., C. S. HALEY, H. ELLEGREN, S. A. KNOTT, M. JOHANSSON, *et al.* 1994 Genetic mapping of quantitative trait loci for growth and fatness in pigs. Science **263:** 1771-1774.

BRASCAMP, E. W., J.A.M. VAN ARENDONK, and A. F. GROEN, 1993 Economic appraisal of the utilization of genetic markers in dairy cattle breeding. J. Dairy Sci. **76:** 1204-1213.

BULMER, M. G., 1971 The effect of selection on genetic variability. Am. Naturalist **165(943):** 201-211.

EDWARDS, M. D., and N. J. PAGE, 1994 Evaluation of marker-assisted selection through computer simulation. Theor. Appl. Genet. **88:** 376-382.

FALCONER, D. S., and T.F.C. MACKAY, 1996 Introduction to Quantitative Genetics. 4th edition. Longman Group Ltd., Essex United Kingdom.

FERNANDO, R. L., and M. GROSSMAN, 1989 Marker-assisted selection using best linear unbiased prediction. Genet. Sel. Evol. **21:** 467-477.

GEORGES, M., and L. ANDERSSON, 1996 Livestock genomics comes of age. Genome Res. **6:** 907-921.

GEORGES, M., D. NIELSEN, M. MACKINNON, A. MISHRA, R. OKIMOTO, *et al.* 1995 Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics **139:** 907-920.

GIMFELARB, A., and R. LANDE, 1994a Simulation of marker assisted selection in hybrid populations. Genet. Res. **63:** 39-47.

GIMFELARB, A., and R. LANDE, 1994b Simulation of marker assisted selection for non-additive traits. Genet. Res. **64:** 127-136.

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

GRATTAPAGLIA, D., F.L.G. BERTOLUCCI, R. PENCHEL, and R. R. SEDEROFF, 1996 Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family an RAPD markers. Genetics **144:** 1205-1214.

HALDANE, J.B.S., 1919 The combination of linkage values and the calculation of distances between the loci of linked factors. J. Genet. **2:** 3-19.

HALEY, C. S., and S. A. KNOTT, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity **69:** 315-324.

HENDERSON, C. R., 1984 Applications of linear models in animal breeding. Can Catal Publ Data, Univ Guelph, Guelph.

KASHI, Y., E. HALLERMAN, and M. SOLLER, 1990 Marker assisted selection of candidate bulls for progeny testing programmes. Anim. Prod. **51:** 63-74.

LANDE, R., and R. THOMPSON, 1990 Efficiency of marker assisted selection in the improvement of quantitative traits. Genetics **124:** 743-756.

MAKI-TANILA, A., and B. W. KENNEDY, 1986 Mixed model methodology under genetic models with a small number of additive and non-additive loci. Proceedings of the 3rd World Congress on Genetics applied to Livestock Production, Lincoln, Nebraska **12:** 443-448.

MEUWISSEN, T.H.E., and M. E. GODDARD, 1996 The use of marker haplotypes in animal breeding schemes. Genet. Sel. Evol. **28:** 161-176.

MEUWISSEN, T.H.E., and J.A.M. VAN ARENDONK, 1992 Potential improvements in rate of genetic gain from marker assisted selection in dairy cattle breeding schemes. J. Dairy Sci. **75:**1651-1659.

NEIMANN-SORENSON, A., and A. ROBERTSON, 1961 The association between blood groups and several production characteristics in three Danish cattle breeds. Acta. Agric. Scan. **11:** 163-196.

RUANE, J., and J. J. COLLEAU, 1995 Marker assisted selection for genetic improvement of animal populations when a single QTL is marked. Genet. Res. **66:** 71-83.

RUANE, J., and J. J. COLLEAU, 1996 Marker -assisted selection for a sex-limited character in a nucleus breeding population. J. Dairy Sci. **79:** 1666-1678.

SCHAEFFER, L. R., and B. W. KENNEDY, 1986 Computing solutions to mixed model equations. Proceedings of the 3rd World Congress on Genetics applied to Livestock Production, Lincoln, Nebraska **12:** 382-393.

SMITH, C., 1967 Improvement of metric traits through specific genetic loci. Anim. Prod. **9:** 349-358.

SMITH, C., and S. P. SIMPSON, 1986 The use of genetic polymorphism in livestock improvement. J. Anim. Brdg and Genet. **103:** 205-217.

SMITH, C., and D. B. SMITH, 1993 The need for close linkages in marker assisted selection for economic merit in livestock. Anim. Brdg Abstr. **61(4):** 197-204.

SOLLER, M., 1978 The use of loci associated with quantitative effects in dairy cattle improvement. Anim. Prod. **27:** 133-139.

SOLLER, M., and J. S. BECKMANN, 1982 Restriction fragment polymorphisms and genetic improvement. Proceedings of the 2nd World Congress on Genetics applied to Livestock Production, Madrid, Spain **6:** 396-404.

SPELMAN, R. J., W. COPPIETERS, L. KARIM, J.A.M. VAN ARENDONK, and H. BOVENHUIS, 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., and D. J GARRICK, 1997 Utilisation of marker assisted selection in a commercial dairy cow population. Livest. Prod. Sci. **47:** 139-147.

SPELMAN, R. J., and J.A.M. VAN ARENDONK, 1997 Effect of inaccurate parameter estimates on genetic response to marker assisted selection in an outbred population. J. Dairy Sci. **80:** 3399-3410.

STAM, P., 1986 The use of marker loci in selection for quantitative characters. In: C. Smith, J.W.B. King and J.C. McKay (Ed.) Exploiting new technologies in animal breeding - genetic developments. p 170. Oxford University Press, New York.

STRAUSS, S. H., R. LANDE and G. NAMKOONG, 1992 Limitations of molecular-marker aided selection in forest tree breeding. Can. J. For. Res. **22:** 1050-1061.

STUBER, C. W., and M. D. EDWARDS, 1986 Genotypic selection for improvement of quantitative traits in corn using molecular marker loci. Proc. 41st Annual Corn and Sorghum Research conference., Am. Seed Trade Association. **41:** 70-83.

TIER, B., 1990 Computing inbreeding coefficients quickly. Genet. Sel. Evol. **22:** 419-430.

VAN ARENDONK, J.A.M., H. BOVENHUIS, S. VAN DER BEEK, and A. F. GROEN, 1994 Detection and exploitation of markers linked to quantitative traits in farm animals. Proceedings of the 5th World Congress on Genetics applied to livestock production. Guelph, Canada **21:** 193-200.

VAN ARENDONK, J.A.M., B. TIER, and B. P. KINGHORN, 1994 Use of multiple genetic markers in prediction of breeding values. Genetics **137:** 319-329.

VAN DER BEEK, S., and J.A.M. VAN ARENDONK, 1995 Marker assisted selection in an outbred poultry breeding nucleus. Anim. Sci. **62:** 171-180.

VAN DER BEEK, S., J.A.M. VAN ARENDONK, and A. F. GROEN, 1995 Power of two- and three-generation QTL mapping experiments in an outbred population containing full-sib or half-sib families. Theor. Appl. Genet. **91:** 1115-1124.

WANG, T., R. L. FERNANDO, S. VAN DER BEEK, M. GROSSMAN, and J.A.M. VAN ARENDONK, 1995 Covariance between relatives for a marked quantitative trait locus. Genet. Sel. Evol. **27:** 251-274.

WEIR, B. S., and W. G. HILL, 1980 Effect of mating structure on variation in linkage disequilibrium. Genetics **95:** 477-488.

WELLER, J. I., Y. KASHI, and M. SOLLER, 1990 Power of daughter and granddaughter designs for determining linkage between marker loci an quantitative trait loci in dairy cattle. J. Dairy Sci. **73:** 2525-2537.

WILLIAMS, C. G., and D. B. NEALE, 1992 Conifer wood quality and marker-aided selection: a case study. Can. J. For. Res. **22:** 1009-1017.

CHAPTER FIVE

MOVING FROM QTL EXPERIMENTAL RESULTS TO THE UTILISATION OF QTL IN BREEDING PROGRAMMES

RICHARD J. SPELMAN and HENK BOVENHUIS

PUBLISHED IN ANIMAL GENETICS (1998) **29:** 77-84.

ABSTRACT

Results from quantitative trait loci studies can not be readily implemented into breeding schemes through marker assisted selection, due to uncertainty about whether the quantitative trait loci identified are real, and whether the identified quantitative trait loci are segregating in the breeding population. The present paper outlines and discusses strategies to reduce uncertainty in the results from quantitative trait loci studies. One strategy to confirm results from quantitative trait loci studies is to combine p-values from many quantitative trait loci experiments, while another is to establish a confirmation study. The power of a confirmation study must be high to ensure that the postulated quantitative trait loci can be verified. In the calculation of the experimental power there are many issues that have to be addressed: size of the quantitative trait loci to be detected, significance level required, experimental design and expected heterozygosity for the design. To ensure marker assisted selection can be quickly implemented once quantitative trait loci are confirmed, DNA samples should be retained from daughters, and the sires and dams of elite sires.

KEY WORDS: Quantitative trait loci, Marker assisted selection, Confirmation study, Replication.

INTRODUCTION

Quantitative trait loci (QTL) experiments are being undertaken in many livestock species, for example, cattle (GEORGES *et al.* 1995), pigs (ANDERSSON *et al.* 1994), poultry (GROENEN *et al.* 1997) and sheep (CRAWFORD *et al.* 1997). Results from these experiments and others are being reported on a regular basis.

A major objective of the QTL studies is to find QTL that can be implemented in to breeding schemes via marker assisted selection (MAS). The theory and application of MAS have been investigated for many species; dairy cattle (MEUWISSEN and VAN ARENDONK 1992; SPELMAN and GARRICK 1997), poultry (VAN DER BEEK and VAN ARENDONK 1996) and more generally (LANDE and THOMPSON 1990). The near unanimous finding from these studies and others is that the application of MAS has the potential to increase the rate of genetic gain.

In most cases, results from QTL experiments are not directly applicable to the current breeding schemes. Results from studies of dairy cattle using granddaughter experimental designs (WELLER *et al.* 1990) are at least two generations away from the current breeding stock. In the experimental designs commonly used in poultry and pigs, such as divergent crosses or inbred crosses, the reported QTL results are mainly from lines or breeds other than those in the present breeding population. To implement the identified QTL in MAS strategies, further genotyping and analysis in the appropriate breeding population will most probably have to be undertaken.

SPELMAN and VAN ARENDONK (1997) have shown that the implementation of MAS with a postulated QTL that was falsely identified causes genetic loss compared to a breeding scheme that has no knowledge of the QTL. The authors concluded that the QTL should be verified in a further study before MAS is implemented. Furthermore, in order to make optimal use of detected QTL through MAS, accurate estimates of QTL location and effect are required (SPELMAN and VAN ARENDONK, 1997).

The present paper discusses and outlines strategies in moving from the initial QTL experiment results, to being able to utilise QTL in breeding programmes. The strategies discussed include possible experimental structures for verification or confirmation studies.

PRESENT SITUATION

Currently there are many QTL experimental results being published from genome scans. The method of reporting results from QTL studies varies from experiments that only report "significant" findings based on experimentwise threshold values (SPELMAN *et al.* 1996) to others that report significant findings based on comparisonwise threshold levels (ASHWELL *et al.* 1996). The difference in critical values between comparisonwise and experimentwise threshold levels can be quite marked because of the extent of multiple testing. There are many methods that have been proposed to address the issue of multiple testing (CHURCHILL and DOERGE 1994; BENJAMINI and HOCHBERG 1995). LANDER and KRUGLYAK (1995) recommended the use of experimentwise threshold critical values in order to restrict the number of type I errors. These authors proposed a classification based on the significance of the

QTL, ranging from suggestive linkage, when the QTL is not significant at the 5% experimentwise level, to confirmed linkage, where the QTL has been confirmed in another study. LANDER and KRUGLYAK (1995) further commented that linkage results must be replicated to be credible.

The first step in extending the results from the initial QTL genome scan may be to confirm the QTL results in an independent study. This type of confirmation study has already been reported for a QTL identified in dairy cattle (WELLER *et al.* 1996) and is also being undertaken by other groups (M. GEORGES, personal communication; M. GROENEN, personal communication).

The first step for the confirmation study is to formulate the objective of the study. Two possibilities for the objective are:

1) to confirm that the QTL is a real effect in the family or line it was initially identified in; and/or
2) to confirm that the QTL is present in the current breeding population.

The first and second objectives may be achieved in the same experiment depending on whether the families that have been identified as segregating for the QTL in the initial study have descendants in the current breeding population. The risk of only undertaking objective two, is that if the QTL is not detected in the current breeding population, doubt arises about the existence of the original QTL. Therefore confirming that the QTL is a real effect in the family it was first identified in, should be undertaken to gain confidence in the QTL results and the analytical methods used in the initial genome scan. After confirming some QTL in this manner a degree of confidence will be attained in the QTL results and analytical methods employed in the initial genome scan. When this confidence has been attained, the QTL should be confirmed directly in the breeding population (objective two) when the objective is to use the QTL in MAS.

Further objectives of the confirmation study may be:

3) to estimate the QTL effect on correlated traits;
4) to estimate the QTL frequency in the breeding population;

5) to estimate the number of alleles segregating at the QTL; and

6) to use the genetic material in the confirmation study for fine mapping purposes.

EXPERIMENTAL STRUCTURE FOR CONFIRMATION

It may be possible to confirm QTL from results from other QTL studies reported in the scientific literature, and therefore, save the expense of setting up a confirmation study. Confirmation of QTL from literature reports is based on calculating a test statistic from the different experiments. Calculating an overall significance test from the experiments could be simply achieved by combining the p-values from the individual studies, as outlined by FISHER (1946). There are also other methods that can be used to combine results from independent studies (outlined by ROSENTHAL 1978). FISHER'S (1946) method is based on the fact that the sum of a number of values of χ^2 is itself a χ^2 distribution with the appropriate degrees of freedom. To transform the p-values to the equivalent χ^2 for 2 degrees of freedom the natural logarithm of the p-value should be taken and this value should have its sign changed and then be doubled [1].

$$\chi^2 = -2 \times \text{LN(p-value)} \qquad [1]$$

Any number of χ^2 values can be added together to give a composite test. For example, if experiment one has a p-value of 0.08 (χ^2=5.06), and experiment two has a p-value of 0.20 (χ^2=3.22), and experiment three has a p-value of 0.11 (χ^2=4.42), the overall χ^2 values is 12.70 and the overall significance is approximately 0.05 for 6 degrees of freedom.

In the above examples of FISHER'S (1946) method the number of degrees of freedom for each experiment is two. This may not be appropriate as JANSEN (1994) reported that when no QTL are segregating, the asymptotic distribution is expected to be between the χ_1^2 and χ_2^2 distribution. The χ_2^2 distribution is justified by the difference in the number of parameters; QTL size and QTL location and the χ_1^2 distribution is justified by the fact that the null hypothesis is defined by the single constraint that the QTL effect is equal to zero.

There are at least two areas of concern with combining p-values from literature reports. First, there is the high probability of publication bias; only QTL that are significant are published. Second, there is the problem of what type of significance criteria (comparisonwise or experimentwise) has been used in the individual experiments. MORTON (1955) suggested that LOD scores be combined from different studies, as an alternative to χ^2. This may be an option for studies that report results from a likelihood type of analysis.

If there are no appropriate literature reports to confirm the QTL of interest, then a confirmation experiment may be undertaken. The main criterion of the confirmation experiment is that it must have high power to detect the postulated QTL. To calculate the power of the confirmation experiment the following are required:

i) the size of the QTL effect to be detected,
ii) the type I error that is acceptable in the confirmation study;
iii) the experimental design and number of animals available for the confirmation study; and
iv) the expected heterozygosity of the experimental design.

QTL effect: The estimated QTL effect from the original QTL study should not be used in calculation of the power of a confirmation study, as the effect is likely to be overestimated; when the test statistic exceeds a certain significance threshold, the QTL effect is over-estimated, especially in analyses with low power (GEORGES *et al.* 1995; WANG 1995). GEORGES *et al.* (1995) showed that the degree of over-estimation could be by a factor of three for situations of low power. The degree of overestimation is increased as the significance threshold is increased to account for multiple hypothesis testing. Based on type I error and power in the original QTL experiment, it may be possible to derive an adjusted effect that is closer to the true effect. However, a rule of thumb may be to calculate the power for a QTL that is half of the estimated effect from the original QTL study.

Type I error: An approach to calculate an appropriate type I error in the confirmation study may be to have an overall type I error, *i.e.* the overall type I error

would combine the error rates from the original QTL study and the confirmation study. If the significance level from the original QTL experiment is 10% and an overall significance of 5% is desired, then the type I error that is acceptable in the confirmation study can be calculated using FISHER'S (1946) method. The χ^2 value for the original QTL study is 4.60 for the p-value of 10%. The combined χ^2 value required for significance at the 5% threshold level from the original QTL study and the confirmation study is 9.50 for 4 degrees of freedom (two degrees of freedom from both QTL experiments). Therefore, the required χ^2 value for the confirmation study is 4.90 (9.50 – 4.60) which is equivalent to a p-value of 0.087.

A similar approach for setting the type I error in the confirmation experiment is to use a posterior type I error (SOUTHEY and FERNANDO 1998). The posterior type I error can be defined as the number of false positives occurring in the results that are deemed significant. To calculate the posterior type I error three pieces of information are required:

i) the prior probability that there is a QTL in the chromosomal area;
ii) the type I error accepted in the confirmation study; and
iii) the type II error, and thus, the power of the confirmation experiment.

The posterior or likelihood distribution from the original QTL experiment can be used as an indication of the probability that there is a QTL segregating, *i.e.* the prior probability for the confirmation study. For given type I (α) and type II (β) error rates in the confirmation study and prior probability of a QTL segregating, the frequency of no errors, type I errors and type II errors can be calculated. A type I error will occur when it is concluded that a QTL exists when in reality there is no QTL. A type II error will occur when there is a QTL but it is not detected. No error occurs in two situations; when there is no QTL and the experiment's conclusion is that there is no QTL, and when there is a QTL and the QTL is detected in the experiment. For an 80% prior probability of a QTL in the chromosomal area of interest and the type I error in the confirmation experiment set to 0.05 and power ($1-\beta$) of 75%, the probability that a QTL that exists is detected is 0.75 ($1-\beta$) $\times$ 0.8 (prior probability of QTL) = 0.6. Probability of correctly not detecting the QTL is 0.95 (1-

$\alpha) \times 0.2$ (prior probability of no QTL) = 0.19. Type II error will be 0.25 $(\beta) \times 0.8$ (prior probability of QTL) = 0.2 and type I error will be 0.05 $(\alpha) \times 0.2$ (prior probability of no QTL) = 0.01. The posterior type I error can be calculated as:

$$\frac{\text{P(QTL detected when no QTL)}}{\text{P(QTL detected when no QTL) + P(QTL detected when a QTL exists)}}$$

where P = probability. Using the values from the above example, the posterior type I error is:

$= 0.01 / (0.01 + 0.60) = 0.0164$

In this example, there is a 1.64% chance that the QTL detected in the confirmation study is not real. To obtain a 5% posterior type I error the confirmation type I error rate should be relaxed. Relaxing the confirmation study type I error rate will increase the power in the confirmation study. Therefore, to solve the above equations in ascertaining the appropriate type I error rate in the confirmation study, an iterative approach will be applied.

Significance levels from the original QTL experiment cannot be used as prior probabilities for the confirmation study because the p-values represent the probability that the null hypothesis (usually that there is no QTL) is correct, and not the probability that the QTL is the size that it was found in the original QTL study. However, the posterior distributions from likelihood and Bayesian QTL analysis of the original QTL experiment give prior probabilities for the confirmation study.

Experimental design: The confirmation study should be independent of the original QTL experiment. Therefore when estimating the QTL effect, one should use different animals than those used in the original QTL study to estimate a family's QTL effect. In theory, there are many groups of animals that can be used in the confirmation study. The following is a list of options for the confirmation of a QTL that has been identified in a grandsire in a dairy cattle population.

i) *Daughters of the grandsire:* For this situation it is preferable that the daughters should be in their first lactation so culling has not occurred, as MACKINNON and GEORGES (1992) showed that selection reduces the power to detect QTL. However, the selection intensity from first to second lactation is unlikely to be high and therefore the power to detect a QTL should not be reduced dramatically if daughters from later lactations are used. Analysis tools that sample the missing genotypes, given the estimated QTL size and the phenotype of the animals that have no recorded genotype (JOHNSON *et al.* 1998), may minimise the effects on power of detection from production based culling.

ii) *Daughters of postulated heterozygous sons of grandsire segregating for the QTL:* The identification of sons that are heterozygous for the QTL could prove to be difficult. The sons from the extreme of the trait distribution could be chosen but this raises the question about whether these individuals are more likely to be homozygous for the QTL than the sons that are in the middle of the trait distribution. With no QTL genotype probabilities for the dams, there is a chance that sons are not the predicted QTL genotype.

iii) *Progeny tested grandsons of the grandsire:* This design has been termed the grand^2-daughter design (COPPIETERS *et al.* 1998). The grandsons are more likely to be linked to the grandsire through his daughters, as it is unlikely that a grandsire will have many sons which themselves are sires of sons.

iv) *Other related animals through sire or dam lines:* The more generations that separate the related animals from the grandsire of interest the greater the chance that the identified QTL is not segregating.

v) *Unrelated animals:* This group of animals can be used when the objective is to identify whether the QTL detected in the original QTL experiment is segregating in the breeding population. In some situations, the breeding population will be related to the grandsire that was originally identified as segregating for the QTL.

Expected heterozygosity: Detecting a QTL in a granddaughter design means that there is a difference between the two grandsire alleles, e.g. A and B. When confirming the QTL effect in the grandsire's daughters or by the grand^2-daughter design, the same contrast as that seen in the granddaughter design is tested. Attempting to confirm the QTL in progeny tested sons of the grandsire may not be testing the same allelic contrast if there are more alleles segregating in the population,

e.g. C. There is no guarantee that the difference between A and C, or B and C, will result in a detectable contrast. Therefore, "heterozygous" in the power calculations means heterozygous with respect to the allelic effects. A question arises about which of the grandsire's two alleles (A and B) should be traced in subsequent generations. Is it more probable that a son that receives allele A from the grandsire will be heterozygous (i.e., a detectable difference between allelic effects) than a son that receives allele B? The grandsire allele with the largest deviation from the population mean may be the best allele to follow as this allele is more likely to give a detectable contrast with alleles the sons receive from the dam population. However, this strategy may result in tracing an allele that is not attractive from a commercial point of view, that is, an allele with a detrimental effect.

Once the size of the QTL effect to be detected, confirmation study type I error, experimental design and number of animals available for the confirmation study, and the expected heterozygosity have been ascertained, the power of the experiment can be calculated using the methodology presented by WELLER *et al.* (1990). In the confirmation experiment, it is likely that markers will be used that cover the majority of the chromosome, as the confidence intervals from the genome scans are usually very large (VAN OOIJEN 1992). Thus, repeated testing being undertaken across the chromosome should be taken into account when deriving the critical values for the confirmation experiment. It is not necessary to account for multiple testing over a genome, as the confirmation study involves testing an established prior hypothesis (LANDER and KRUGLYAK 1995). The estimate of the QTL effect in the confirmation study is likely to be an unbiased estimator of the real QTL effect as there is no bias caused by threshold levels. Therefore, the confirmation study QTL effect estimate should be used in calculating response from MAS. This is contradictory to traditional meta-analysis where the treatment effect is calculated from pooled experiments (WOLF 1986).

IDENTIFYING QTL IN THE BREEDING POPULATION

Once there is sufficient confidence that the QTL is real, the application of the QTL by MAS relies on identifying families in the breeding population that are segregating for the QTL. Larger increases in genetic gain with MAS occur when selection is undertaken before the phenotype is observed on selection candidates

(animals that are eligible for selection) (NEIMANN-SORENSON and ROBERTSON 1961; MEUWISSEN and GODDARD 1996). For dairy cattle breeding, examples of selection before phenotypic records become available are the selection of bulls entering the progeny test scheme and selection of non-lactating cows as bull dams.

Identification of selection candidates that carry the favourable QTL allele requires the knowledge of whether their parents are segregating at the QTL locus. The selection candidate's sire and dam will have information from their ancestors and the sire will have additional information from his progeny test daughters. The progeny test daughters provide the best source of pedigree information in establishing heterozygosity of the sire at the QTL loci of interest. The identification of heterozygosity in this manner is the basis of the "bottom-up" MAS approach of MACKINNON and GEORGES (1998). The number of daughters in a dairy cattle progeny test is approximately 100. Based on this number of daughters, the power of QTL detection will not be high. To detect whether a sire is heterozygous for the chromosomal segment of interest, the daughters can be divided into two groups depending on the marker haplotype they received from the sire. The mean phenotype for the trait of interest can be calculated for the two daughter groups and if it is larger than a pre-defined threshold criterion, the sire is deemed heterozygous. However, the criterion to decide whether the sire is heterozygous for the QTL of interest does not have to be strict. MACKINNON and GEORGES (1998) have shown that assuming a sire is heterozygous when there are only small differences between the two marker allele daughter groups results in greater increases in genetic gain with the "bottom-up" MAS scheme than when using a stricter criteria.

Dams are usually limited by their reproductive capacity and will not have a large number of offspring to estimate if they are heterozygous at the QTL. Therefore, heterozygosity of the dam of the selection candidate is primarily estimated based on the QTL status of her ancestors. To estimate the QTL status of the dam this requires that the QTL status of her sire or grandsire is known which is the basis of the design described by KASHI *et al.* (1990) for MAS.

To improve the power for both sires and dams, information from progeny and full and half-sibs could be combined with ancestor information. The most formal

setting for combining this information would be with best linear unbiased prediction (BLUP) that incorporates marker information (FERNANDO and GROSSMAN 1989, MEUWISSEN and GODDARD 1996).

DISCUSSION

LANDER and KRUGLYAK (1995) state that linkage results need to be replicated to be credible. The replication is proposed to ensure that expensive fine-mapping or positional cloning studies are not for phantom loci. The need for replication is also similar for the two broad objectives of the genomic studies undertaken in livestock species. If the objective is to clone the QTL, and the research centres on a chromosomal segment that does not contain a QTL, the consequences are that are a lot of time and money will be wasted in chasing the phantom locus. If the objective is to use the QTL in MAS, breeding companies will not want to select for a falsely identified QTL as this will lower the rate of genetic gain (SPELMAN and VAN ARENDONK 1997). The significance level required before utilisation of the QTL will depend on the risk adversity of the breeding company and this significance level could possibly be calculated using methods such as those outlined by MEUWISSEN (1991).

It is proposed that QTL results should be confirmed to gain confidence in the analytical methods. This assurance is needed because of the different assumptions made on the underlying genetic model in the statistical methods used in the QTL analysis. Some of the more common assumptions which are made are: no segregation distortion, and usually a model that fits a single QTL. At present, little is known about the actual behaviour of genes affecting quantitative traits. It has been shown that some single genes have rather complicated patterns of inheritance, such as polar overdominance (COCKETT *et al.* 1996).

The ability to confirm QTL from literature reports requires that p-values and QTL effects are published in literature. When results are published, all of the chromosomes that have been evaluated should be presented to ensure an unbiased sample of experimental results are in the scientific literature (e.g., CHARLIER *et al.* 1996). In addition to QTL results being published in scientific journals, it may be beneficial to have the most recent results on the internet and thus reduce the time lag from analysis to publication. With this quantity of information, there is potential for

meta-analysis to be undertaken to confirm QTL, without the need for further experiments.

The value of literature studies might be limited due to low probability that the same QTL is segregating in another study, as a result of isolated or genetically different populations. However, in dairy cattle the concern is lessened as the Holstein population has many links between countries. Using results from another breed, for confirmation, depends on the probability that the same QTL is segregating. The genetic links between different breeds may be poor, therefore reducing the probability of QTL segregating across breeds. However, GEORGES and ANDERSSON (1996) reported that the same QTL for milk production might have been identified in the Holstein and the Finnish Ayrshire breeds.

The effects of multiple testing should be taken into account when obtaining critical values in QTL experiments. For genome scans, this will involve accounting for testing on all chromosomes and for each independent trait. In the confirmation experiments, testing across the whole chromosome or however much the marker coverage is across the chromosome should be accounted for. If p-values are presented on a comparisonwise basis, then they should be converted to experimentwise basis if they are to combined with p-values from other studies.

Selective genotyping is a viable option for confirmation studies as it reduces the number of genotypes needed for a given power. As a rule of thumb the percentage of animals to be selected is approximately 40%, with 20% from each end of the trait distribution without losing much power (DARVASI and SOLLER 1992). DARVASI and SOLLER (1992) presented formulae to calculate the power of QTL experiments for the trait that selective genotyping has been undertaken on. BOVENHUIS and SPELMAN (1998) have developed formulae to calculate the power for traits correlated to the trait that selective genotyping was applied to. In addition, BOVENHUIS and SPELMAN (1998) have described an algorithm that ensures unbiased estimates of QTL effects for all traits analysed in a selectively genotyped experiment. RONIN *et al.* (1998) and JOHNSON *et al* (1998) have also developed methods that ensure unbiased estimates for all traits analysed in a selectively genotyped QTL experiment.

With the ever-increasing number of QTL results being published, breeding organisations want to apply QTL results in their breeding schemes as quickly as possible. The major step before application will be to identify sires that are heterozygous for the QTL. To ensure that one can quickly identify sires that are heterozygous for the QTL of interest, DNA samples from progeny test daughters should be retained for all sires that are or have been used as proven bulls. In addition, DNA samples should be retained from all bull parents, or all animals in the case of a nucleus scheme. MEUWISSEN and GODDARD (1996) have shown that having generations of genotypes for ancestors increased the rate of genetic gain with MAS as QTL allele effects can be more accurately evaluated. Therefore the parental DNA samples will be beneficial in the application of MAS.

The implementation of identified QTL requires that DNA samples be stored or the animals needed to verify QTL and/or identify QTL status of animals of interest be identified, to ensure that QTL can be quickly utilised in breeding schemes. Using the systems mentioned above, QTL reported in the literature can be studied in the current breeding population to identify heterozygous sires. Following this analysis, the QTL can be immediately utilised within a MAS breeding programme.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the helpful comments of MICHEL GEORGES in the preparation of this manuscript. RJS thanks Livestock Improvement Corporation for financial support.

LITERATURE CITED

ANDERSSON, L., C. S. HALEY, H. ELLEGREN, S. A. KNOTT, M. JOHANSSON, *et al.* 1994 Genetic mapping of quantitative trait loci for growth and fatness in pigs. Science **263:** 1771-1774.

ASHWELL, M. S., C. E. REXROAD JR, R. H. MILLER, and P. M. VAN RADEN, 1996 Mapping economic trait loci for somatic cell score in Holstein cattle using microsatellite markers and selective genotyping. An. Genet. **27:** 235-242.

BENJAMINI, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. **57(1):** 289-300.

BOVENHUIS, H., and R. J. SPELMAN, 1998 Selective genotyping to detect QTL for multiple traits in outbred populations. Proceedings of the 6th World Congress on Genetics applied to Livestock Production, Armidale, Australia, 1998. **26:** 241-244.

CHARLIER, C., F. FARNIR, P. BERZI, P. VANMANSHOVEN, B. BROUWERS, *et al.* 1996 Identity-by-descent mapping of recessive traits in livestock: application to map the bovine *syndactyly* locus to chromosome 15. Genome Res. **6:** 580-589.

CHURCHILL, G. A., and R. W. DOERGE, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963-971.

COCKETT, N. E., S. P. JACKSON, T. D. SHAY, F. FARNIR, S. BERGHMANS, 1996 Polar overdominance at the ovine callipyge locus. Science **273:** 236-238.

COPPIETERS, W., A. KVASI, J.-J. ARRANZ, B. GRISART, J. RIQUET, *et al.* 1998 The grand^2-daughter design: a simple strategy to increase the power of the grand-daughter design. (submitted).

CRAWFORD, A. M., J. C. MCEWAN, K. G. DODDS, S. A. BISSETT, P. A. MACDONALD, *et al.* 1997 Parasite resistance: A genome scan approach to finding markers and genes. Proceedings New Zealand Society of Animal Production. **57:** 297-300.

DARVASI, A., and M. SOLLER, 1992 Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theor. Appl. Genet. **85:** 353-359.

FERNANDO, R. L., and M. GROSSMAN, 1989 Marker-assisted selection using best linear unbiased prediction. Genet. Sel. Evol. **21:** 467-477.

FISHER, R. A., 1946 Statistical methods for research workers (10th edition). Oliver and Boyd, London.

GEORGES, M., and L. ANDERSSON, 1996 Livestock genomics comes of age. Genome Res. **6:** 907-921.

GEORGES, M., D. NIELSEN, M. MACKINNON, A. MISHRA, R. OKIMOTO, *et al.* 1995 Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics **139:** 907-920.

GROENEN, M.A.M, R.P.M.A. CROOIJMANS, T. VEENENDAAL, J.B.C.H.M. VAN KAAM, A.L.J. VEREIJKEN, 1997 QTL mapping in chicken using a three generation full sib family structure of an extreme broiler x broiler cross. An. Biotechnology **8(1):** 41-46.

JANSEN, R. C., 1994 Controlling the type I and type II errors in mapping quantitative trait loci. Genetics **138:** 871-881.

JOHNSON, D. L., R. C. JANSEN, and J.A.M. VAN ARENDONK, 1998 Mapping quantitative trait loci in a selectively genotyped outbred population using a mixture model approach. (submitted to Genetical Research)

KASHI, Y., E. HALLERMAN, and M. SOLLER, 1990 Marker assisted selection of candidate bulls for progeny testing programmes. Anim. Prod. **51:** 63-74.

LANDE, R., and R. THOMPSON, 1990 Efficiency of marker assisted selection in the improvement of quantitative traits. Genetics **124:** 743-756.

LANDER, E., and L. KRUGLYAK, 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genetics **11:** 241-247.

MACKINNON, M. J., and M.A.J. GEORGES, 1992 The effects of selection on linkage analysis for quantitative traits. Genetics **132:** 1177-1185.

MACKINNON, M. J., and M. GEORGES, 1998 Marker-assisted preselection of young dairy bulls prior to progeny testing. Livest. Prod. Sci. **54:** 229-250.

MEUWISSEN, T.H.E., 1991 Expectation and variance of genetic gain in open and closed nucleus and progeny testing schemes. Anim. Prod. **53:** 133-141.

MEUWISSEN, T.H.E., and M. E. GODDARD, 1996 The use of marker haplotypes in animal breeding schemes. Genet. Sel. Evol. **28:** 161-176.

MEUWISSEN, T.H.E., and J.A.M. VAN ARENDONK, 1992 Potential improvements in rate of genetic gain from marker-assisted selection in dairy cattle breeding schemes. J. Dairy Sci. **75:** 1651-1659.

MORTON, N. E., 1955 Sequential test for the detection of linkage. Am. J. Hum. Genet. **7:** 277-318.

NEIMANN-SORENSON, A., and A. ROBERTSON, 1961 The association between blood groups and several production characteristics in three Danish cattle breeds. Acta. Agric. Scan. **11:** 163-196.

RONIN, Y. I., A. B. KOROL, and J. I. WELLER, 1998 Selective genotyping to detect quantitative trait loci affecting multiple traits: interval mapping analysis. Theor. Appl. Genet. (in press)

ROSENTHAL, R., 1978 Combining results of independent studies. Psychological Bulletin **85(1):** 185-193.

SOUTHEY, B .R. and R. L. FERNANDO, 1998 Controlling the proportion of false positives among significant results in QTL detection. Proceedings of the 6th World Congress on Genetics applied to Livestock Production, Armidale, Australia **26:** 221-224

SPELMAN, R. J., W. COPPIETERS, L. KARIM, J.A.M. VAN ARENDONK, and H. BOVENHUIS, 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., and D. J. GARRICK, 1997 Utilisation of marker assisted selection in a commercial dairy cow population. Livest. Prod. Sci. **47:** 139-147.

SPELMAN, R. J., and J.A.M. VAN ARENDONK, 1997 Effect of inaccurate parameter estimates on genetic response to marker assisted selection in an outbred population. J. Dairy Sci. **80:** 3399-3410

VAN DER BEEK, S., and J.A.M. VAN ARENDONK, 1995 Marker assisted selection in an outbred poultry breeding nucleus. Anim. Sci. **62:** 171-180.

VAN OOIJEN, J. W. 1992 Accuracy of mapping quantitative trait loci in autogamous species. Theor. Appl. Genet. **84:**803-811.

WANG, Y., 1995 Detection and estimation of associations between genetic markers and quantitative trait loci in segregating populations. Ph.D. thesis, University of Guelph.

WELLER, J. I., Y. KASHI, and M. SOLLER, 1990 Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J. Dairy Sci. **73:** 2525-2537.

WELLER, J. L., G. R. WIGGANS, P. M VAN RADEN, and M. RON, 1996 Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multi-trait experiment. Theor. Appl. Genet. **92:** 998-1002.

WOLF, F. M., 1986 "Meta-Analysis Quantitative methods for research synthesis." Sage University Paper series on Quantitative Applications in the Social Sciences, 07-001, Beverly Hills: Sage Pbns.

CHAPTER SIX

GENETIC AND ECONOMIC RESPONSES FOR WITHIN-FAMILY MARKER ASSISTED SELECTION IN DAIRY CATTLE BREEDING SCHEMES

RICHARD J. SPELMAN and DORIAN J. GARRICK

JOURNAL OF DAIRY SCIENCE (ACCEPTED)

ERRATUM

The diagram in Figure 1 (p. 100) belongs to Figure 2 (p. 112), and vice versa.

ABSTRACT

Marker assisted selection schemes that utilise information about quantitative trait loci information to pre-select progeny test bulls within a family are the most practical application of quantitative trait loci results in the short-term. Technical difficulties exist for across-family marker assisted selection using BLUP procedures. Two within-family marker assisted selection schemes were evaluated genetically and economically using stochastic simulation for a locus that explained 5% of phenotypic variance. The genetic and economic impacts of variation in the number of offspring per bull-dam were evaluated. The 'top down' marker assisted selection scheme identifies sires that are heterozygous for the locus based on the granddaughter design and uses the quantitative trait locus information in the pre-selection of grandsons entering progeny test. The 'bottom up' marker assisted selection scheme identifies quantitative trait locus heterozygous sires based on the daughter design and uses the information in the pre-selection of sons entering progeny test. The top down scheme with one progeny per bull-dam reduced the rate of genetic gain compared with that from a breeding scheme that ignored knowledge of the quantitative trait locus. The top down scheme with reproductive performance of 3 or 40 progeny per bull-dam, increased genetic gain by 1 to 2%. The bottom up scheme increased the rate of genetic gain by 1.5, 3.5 and 5% for 1, 3, and 40 progeny per bull-dam respectively. When the top down scheme was used on the maternal path and the bottom up scheme on the paternal path, increases were 9% with 40 progeny per bull-dam. The use of reproductive technologies on bull-dams is imperative to prevent gains from marker assisted selection being eroded by the loss in polygenic selection differential that results when more bull-dams are required to enable pre-selection of sons using markers.

KEY WORDS: Marker assisted selection, Dairy cattle, Genetics.

INTRODUCTION

Quantitative trait loci (QTL) for milk production traits have been identified in dairy cattle (GEORGES *et al.* 1995; SPELMAN *et al.* 1996; VILKKI *et al.* 1997). The major objective of most, if not all, studies of dairy cattle QTL studies is to identify the QTL that can be utilised in marker assisted selection (MAS) breeding schemes, and MAS for dairy cattle has been evaluated in many studies (BRASCAMP *et al.* 1993;

KASHI *et al.* 1990; MACKINNON and GEORGES 1998; RUANE and COLLEAU 1996; SPELMAN and GARRICK 1997a). Those studies and others have shown that the rate of genetic gain can be increased with the implementation of MAS.

Two broad categories of MAS schemes have been evaluated; those based on within-family selection and those incorporating marker information in BLUP evaluations. Within-family MAS involves selection decisions first made on conventional EBV, and QTL information used for within-family selection. The BLUP-based MAS involves the use of mixed models that incorporate effects for individual QTL alleles and selection decisions are made on EBV that combine QTL and polygenic components (FERNANDO and GROSSMAN 1989). The BLUP-based algorithms method initially presented by FERNANDO and GROSSMAN (1989) had the requirement that all animals must have marker information, which is not practical for most commercial dairy cattle populations comprising >1 million milking cows. HOESCHELE (1993) presented an algorithm that eliminates equations for animals without marker data and not providing relationship ties among genotyped descendants. However, an approximation to that of polygenic inheritance is made for the ungenotyped animals that do provide relationship ties among genotyped descendants, and where marker genotype can not be determined unequivocally from progeny or parents (HOESCHELE 1993). Another method has been developed that samples missing marker genotypes in a Markov chain Monte Carlo setting (BINK *et al.* 1998). However, this method is computer intensive and time consuming and is not currently practical for routine application to an entire population. Within-family selection incorporating marker information is one practical option for implementation of MAS for dairy cattle breeding schemes in the immediate future.

Two different types of MAS schemes have been described for within-family selection. The first scheme, the 'top down' scheme described by KASHI *et al.* (1990) is based on the granddaughter design and involves identifying whether a grandsire is heterozygous for a QTL based on his progeny-tested sons that are genotyped for the area of interest (Figure 1). The QTL information from the grandsire is used in the pre-selection of his grandsons entering progeny test. The second scheme, the 'bottom up' scheme, is that of MACKINNON and GEORGES (1998). This scheme is similar to the top down scheme but is based on the daughter design in which the sires' of the

candidate progeny test bulls are assessed for the presence of a segregating QTL allele, based on marker contrasts in their progeny test daughters, which are genotyped for the chromosomal areas of interest (Figure 2). The QTL information for the sires is used in the pre-selection of their sons entering progeny test.

KASHI *et al.* (1990) reported increases in the rate of genetic gain of approximately 20% with the top down MAS scheme that they proposed, with several identified QTL, and MACKINNON and GEORGES (1998) reported increases about 10% with the bottom up scheme. MACKINNON and GEORGES (1998) also evaluated the top down scheme (KASHI et al. 1990) and reported increases in rate of genetic gain that were equivalent to or slightly less than those of their own bottom up scheme.

MACKINNON and GEORGES (1998) evaluated the use of marker information from the sire or paternal grandsire to help in selection decisions for the candidate progeny test bulls but did not use QTL information from the dam or maternal granddam. Neither study (KASHI *et al.* 1990, MACKINNON and GEORGES 1998) accounted for the selection differential reduction on the bull-dam pathway, which may occur as more bulls are generated to enable pre-selection.

The objectives of this study are to evaluate the use of QTL information on both the sire and dam sides in the selection of candidate bulls for progeny testing, and to evaluate the impact of reproductive technologies such as embryo transfer and in vitro fertilisation to overcome the loss of selection differential on the dam pathway with the utilisation of MAS.

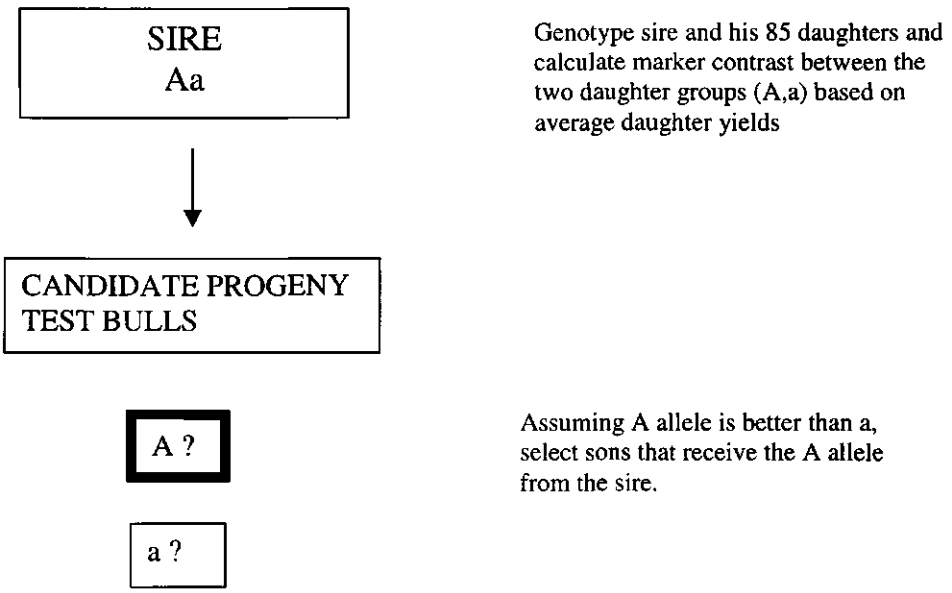
MATERIAL and METHODS

Outline of Breeding Schemes

Top down: The basis of the top down scheme (KASHI *et al.* 1990) is that loci of interest, which were identified in earlier studies, are investigated in the current population of elite sires. To identify whether the current elite sires are segregating for the QTL of interest, their progeny test sons are grouped on the basis of which marker haplotype they received from their sire. Segregation for the QTL is identified when a significant difference exists in the average EBV (or daughter yield deviations) between the two groups of progeny-tested sons. If there is a significant difference

between the two haplotype groups, the QTL information for that grandsire is used in selection. The QTL information cannot be used in selection decisions for the next generation (sons of the grandsires), because the sons have already been progeny tested. However, the QTL information can be used in the following generation, in the selection of the grandsons of the grandsires in which the QTL contrast has been identified (Figure 1).

FIGURE 1: A top down scheme for marker assisted selection. The two grandsire quantitative trait loci (QTL) alleles are A and a, and the unknown QTL allele from the dam population is ?.



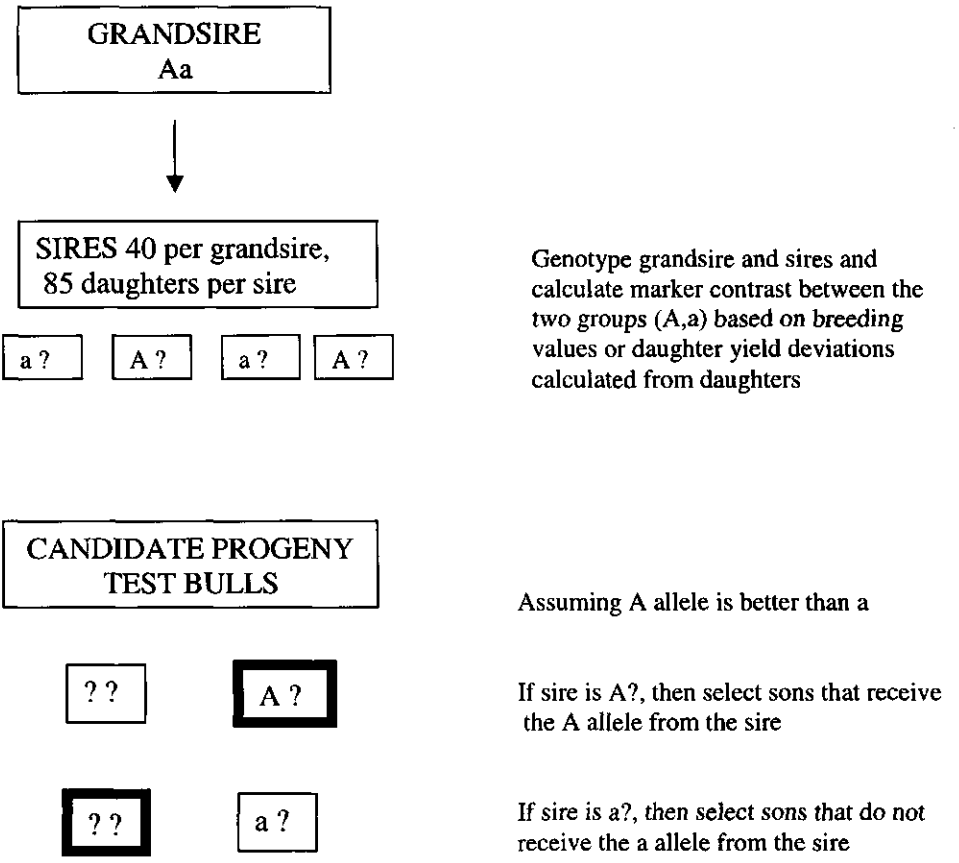
A grandsire will transmit each haplotype to half of his offspring, on average. For males (sons of the grandsire with a haplotype difference) that are used as sires of sons for the next generation, QTL information can potentially be used in selection of which of their sons (grandsons) are progeny tested. For the sires that have received the better haplotype from the grandsire, selection within the grandsons is for those that also received the better haplotype. For the sires that have received the poorer haplotype from the grandsire, selection on which of grandsons to progeny test is for

those grandsons that have not received the poorer grandsire haplotype. This selection based on QTL information assumes that the better haplotype from the grandsire is better than the haplotype that the sire received from the dam population and the poorer haplotype from the grandsire is worse than the haplotype from the dam population. The top down scheme can also be undertaken on the dam path of the progeny test bull when QTL heterozygosity is identified in the maternal grandsire of the bull entering progeny testing.

Bottom up: The basis of the bottom up design (MACKINNON and GEORGES 1998) is that sires are evaluated for pre-identified QTL by genotyping their progeny test daughters in the regions of interest. Daughters of the sires are grouped, depending on which sire haplotype they received, and the magnitude of the contrast in the average yield deviation for the two daughter groups is used to determine whether the sire is heterozygous for the QTL locus. When a sire is determined to be heterozygous for the QTL, only the sons that receive the better haplotype are progeny tested (Figure 2).

Determining heterozygosity: The sires are evaluated for heterozygosity at the QTL based on the difference between the means of the two haplotype progeny groups. For the top down scheme, the contrast is calculated in the same manner as in the granddaughter design and, for the bottom up scheme, the contrast is calculated in the same manner as the daughter design. To determine whether a (grand)sire, is heterozygous the difference between the two progeny groups has to be bigger than the pre-defined threshold level. The threshold level can range from any difference to a large required difference (e.g. 1 to 2 genetic standard deviations (σ_G)) between the two haplotype groups for a (grand)sire to be identified as heterozygous at the QTL.

FIGURE 2: A bottom up scheme for marker assisted selection scheme. The two sire quantitative trait loci (QTL) alleles are A and a, and the unknown QTL allele from the dam population is ?.

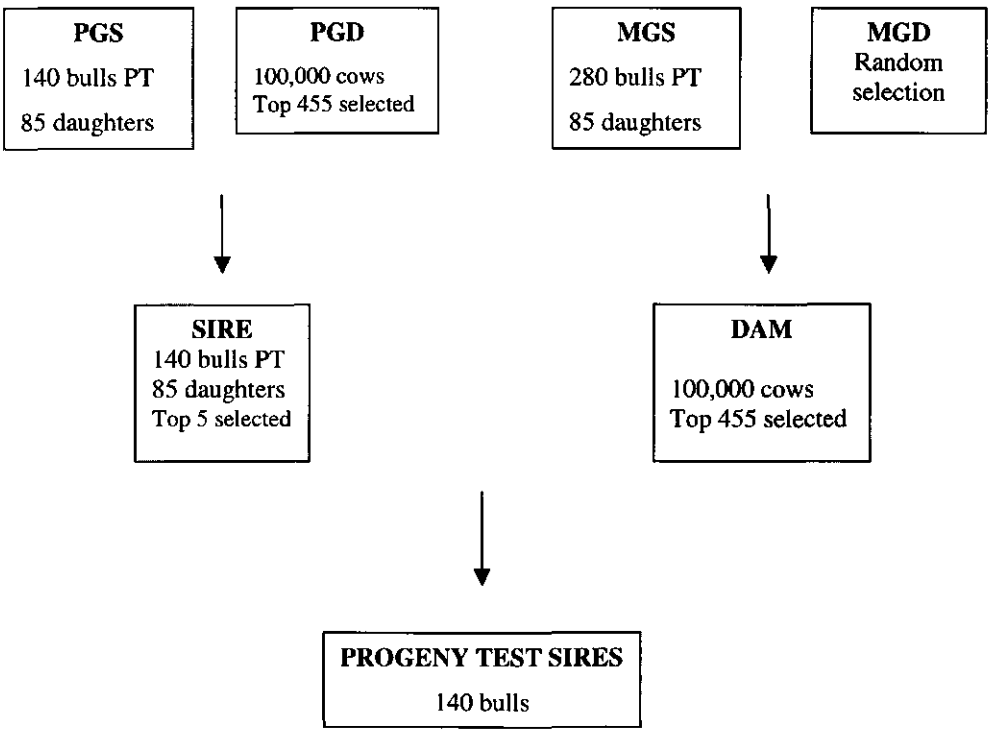


Simulation Model

Population structure: The New Zealand dairy cattle breeding scheme for the Holstein-Friesian breed was modelled using stochastic simulation. A base population for sires and dams was simulated (Figure 3). The paternal grandsire population comprised 140 sires that were progeny tested on 85 daughters. The 5 highest ranked sires on EBV from the 140 sires were selected as sires of sons. The paternal granddam population was simulated to include 100,000 cows that were eligible to be bull-dams. In New Zealand, a cow can be a bull-dam if she is included in a milk recording system and has three generations of artificially bred parents to the same breed. The highest ranked 455 cows based on EBV from the 100,000 cows were

selected as bull-dams. One hundred and forty bulls were produced from the selected animals from the paternal granddam and paternal grandsire populations.

FIGURE 3: Simulated population structure (PGD = paternal granddam, PGS = paternal grandsire, MGS = maternal grandsire, MGD = maternal granddam, and PT = progeny tested).



The maternal grandsire population comprised 280 males (two years of progeny tested bulls) from which 14 were selected on EBV as sires to breed cows. The 14 bulls were mated to an unselected base population of cows to produce 100,000 potential bull-dams. The appropriate numbers of bull-dams were selected and mated to the sires of sons to produce the progeny test bulls. Three levels of reproductive performance were assumed for the bull-dam; 1 calf (representing normal reproductive performance), 3 calves (representing embryo transfer), and 40 calves (representing a reproductive tool such as in vitro fertilisation). For all scenarios, only one bull per full-sib group was progeny tested, which ensured that comparable rates of genome-wide inbreeding were achieved.

Genetic model: A single trait with heritability of 30% was simulated. The additive genetic variance was divided between unmarked additive polygenic variation (which is referred to as polygenic variance) and variation from the marked chromosomal region (which will be referred to as QTL). The QTL component explained 5% of the phenotypic variance or, equivalently, 16.7% of genetic variance. The QTL had 10 alleles at equal frequency, and allelic effects were drawn from a normal distribution.

Estimated breeding values: EBV were simulated as follows.

$$EBV_i = r_{TI}^2(BV_i - \overline{BV.}) + z_i\sqrt{Vp}\sqrt{h^2 r_{TI}^2 - h^2 r_{TI}^4} + \overline{BV.} \quad [1]$$

where BV_i = true breeding value of the animal i, which includes polygenic and QTL effects, $\overline{BV.}$ is the population mean for true breeding value, r_{TI}^2 = squared correlation between estimated and true breeding values (reliability), h^2 = heritability, V_P = phenotypic variance, and z_i is a standard normal deviate. The accuracy of evaluation was based on pedigree information and the first lactation records of 85 effective daughters for sires and pedigree information and one lactation record for cows.

The effect of selection in reducing the genetic and therefore also the phenotypic variance, known as the BULMER effect (1971), was accounted for as outlined by FALCONER and MACKAY (1996). Heritability was updated with the new genetic and phenotypic variances as was the accuracy of evaluation (r_{TI}^2). Inbreeding was ignored.

Evaluation of MAS schemes: The two MAS schemes were compared with a breeding scheme that ignored any knowledge of segregating QTL (termed the control). Therefore, all selection decisions in the control were made solely on EBV. The control had the same reproductive performance for the bull-dams as the MAS schemes. The MAS breeding schemes and the control were evaluated in genetic terms based on comparison of the average genetic merit of bulls entering the progeny test and the average genetic merit of the top 10 bulls graduating from the progeny test. The required significance thresholds that are used to identify whether the sires are

segregating for the QTL (i.e. heterozygous at the QTL) were increased by 0.1 σ_G increments between the levels 0 σ_G to 1.2 σ_G for both MAS schemes. Five hundred simulations were undertaken for each scenario.

Economic evaluation was undertaken by financially quantifying the benefit of a one off response from MAS. Returns from MAS were extra milk returns over 20 years plus the terminal value of the genetically improved cows. The returns were evaluated on the basis that the trait with the QTL information was protein, the most economically important trait in New Zealand. The economic value (US$) of protein is $2.70/kg of protein increased. The value of $2.70 accounts for the selection response being re-expressed in later generations over a 20-yr period and the terminal value of the increase (HARRIS 1998). A cow population of 1.5 million cows was assumed to benefit from the increase in protein selection response. Because each generation is not replaced every year the increase in protein yield is expressed in 1.5 million cows multiplied by the average replacement rate [0.21; (SPELMAN and GARRICK 1997b)].

The costs associated with MAS were the costs of sampling the blood from the daughters ($3/daughter) and 6 markers being genotyped ($2 each) in the sires and daughters. Six flanking markers were chosen to ensure that marker haplotype transmission could be followed. Other costs were $500 for each bull produced and $200 for each cow under going embryo transfer. As costs occur in differing years, a discount rate of 5% was used to calculate the present value of the costs.

RESULTS

Top down scheme: When 1 calf was produced per bull-dam, the genetic level of bulls entering progeny testing with top down MAS was less than the level of the bulls when no selection was undertaken at the QTL loci (Table 1). For the scenarios of 3 calves and 40 calves per bull-dam, the genetic level with top down MAS was greater than the control. The level of improvement was 1 to 1.5% when the threshold level was 0 to 0.5 σ_G and 0.5 to 1% for stricter thresholds.

TABLE 1: The effect of bull-dam reproductive performance and threshold level on the increase in genetic level[1] of bulls being progeny tested for top down marker assisted selection.

Threshold	Calves		
level (σ_G)	1	3	40
0.0	-1.5	1.4	1.3
0.1	-1.2	0.9	1.6
0.2	-0.9	0.9	1.6
0.3	-0.6	0.8	1.5
0.4	-0.5	1.0	1.3
0.5	-0.2	0.7	1.2
0.6	-0.1	1.0	0.9
0.7	-0.1	0.9	0.9
0.8	0.0	0.4	0.6
0.9	0.1	0.4	0.4
1.0	0.1	0.2	0.4
1.1	0.0	0.2	0.3
1.2	0.0	-0.1	0.3

[1]Increase in genetic level is the average percentage superiority compared with values for the control without QTL information, for all progeny-tested bulls.

Bottom up scheme: The bottom up MAS scheme increased the genetic level of bulls being progeny tested by 1 to 1.5% when the reproductive performance for the bull-dams was one calf (Table 2). All of the sires of sons were deemed to be heterozygous at the QTL when the required threshold was 0 σ_G. The percentage of errors at this threshold was 41%. An error occurred when the poorer haplotype (QTL allele) was incorrectly determined to be the better haplotype or when the sire was actually homozygous at the QTL loci. For all threshold levels, approximately 60% of the errors were for the sire being homozygous at the QTL. As the threshold level increased, the percentage of sires of sons determined to be heterozygous decreased, as did the error rate. The effect of fewer sires of sons identified as QTL heterozygous at the higher threshold levels ($\geq$0.8 σ_G) resulted in smaller superiority in genetic level over the control compared to lower threshold levels.

TABLE 2: The effect of different threshold levels in identifying heterozygosity of quantitative trait loci (QTL) in sires of sons for bottom up marker assisted selection on the genetic level of bulls being progeny tested, the percentage of bulls identified as heterozygous, and the percentage of the bulls for which a QTL allele is incorrectly identified as the better QTL.

Threshold level (σ_G)	Increase in genetic level[1]	Bulls with significant contrast	Errors in detecting better allele
		%	
0.0	1.5	100	41
0.1	1.8	85	37
0.2	1.3	71	33
0.3	1.6	58	29
0.4	1.5	46	26
0.5	1.4	35	21
0.6	1.2	27	17
0.7	1.2	20	13
0.8	1.4	15	11
0.9	0.6	11	7
1.0	0.4	7	8
1.1	0.1	5	4
1.2	0.0	3	0

[1]Increase in genetic level is the average percentage of superiority compared with values for the control without QTL information for all progeny-tested bulls.

TABLE 3: Effect of bull-dam reproductive performance and threshold level on the increase in genetic level[1] of bulls being progeny tested for bottom up marker assisted selection.

Threshold level (σ_G)	Calves		
	1	3	40
0.0	1.5	4.2	5.2
0.1	1.8	3.6	5.1
0.2	1.3	3.8	5.0
0.3	1.6	3.1	4.6
0.4	1.5	2.5	3.8
0.5	1.4	3.3	3.4
0.6	1.2	2.7	2.8
0.7	1.2	2.0	2.4
0.8	1.4	1.5	1.9
0.9	0.6	1.1	1.6
1.0	0.4	0.7	1.5
1.1	0.1	0.2	0.8
1.2	0.0	0.3	0.9

[1]Increase in genetic level is the average percentage superiority compared with values for the control without QTL information, for all progeny-tested bulls.

The use of reproductive technologies on the bull-dams increased the genetic superiority of the bottom up MAS scheme compared with that of the control (Table 3). For a threshold level of 0 σ_G, the increase in the genetic level of bulls entering progeny testing was 1.5% for one progeny per bull-dam, 4.2% for 3 progeny per bull-dam, and 5.2% for 40 progeny per bull-dam.

Top down and bottom up scheme: The bottom up MAS scheme had greater increases in genetic level than did the top down scheme and, therefore, would be the preferred MAS scheme for the sire path. On the dam path, bottom up MAS is not possible because the dams did not have enough progeny permit to estimation of QTL heterogosity. It is possible to apply the top down scheme to the dam path and combine it with the bottom up scheme to the sire path. The maternal grandsire QTL heterozygosity can be determined from progeny test sons, if any, or from his daughters. In this simulation, the maternal grandsires were assumed to have 200 daughters each. If a sire and maternal grandsire were both deemed to be heterozygous, sons had to receive the better alleles from both sides of the pedigree to be progeny tested.

TABLE 4: The effect of bull-dam reproductive performance and threshold level on the increase in genetic level[1] of bulls being progeny tested in bottom up marker assisted selection (MAS), for bottom up MAS on the paternal path and top down MAS on the maternal path.

Threshold	Calves		
level (σ_G)	1	3	40
0.0	0.7	5.0	9.5
0.1	1.2	5.8	8.8
0.2	1.9	6.1	8.1
0.3	2.6	5.5	7.4
0.4	1.8	4.7	6.7
0.5	2.2	3.8	5.7
0.6	1.4	3.1	4.9
0.7	0.9	2.4	4.5
0.8	0.7	2.1	3.5
0.9	0.3	1.2	2.5
1.0	0.5	1.1	2.1
1.1	0.3	0.3	1.5
1.2	0.4	0.2	0.8

[1]Increase in genetic level is the average percentage superiority compared with values for the control without QTL information, for all progeny-tested bulls.

The increase in genetic level of the bulls entering the progeny test was 9.5% when the threshold of 0 σ_G was used and when the bull-dams produced 40 calves (Table 4). When 1 calf was produced per bull-dam, the increase in genetic level was 2% and 5 to 6% when the bull-dams produced 3 progeny each.

Genetic response: The increase in genetic level for both the bottom up and the top down MAS schemes (Tables 1 to 4) is the average increase of all of the bulls being progeny tested. It is the bulls graduating from the progeny test that contribute to the rate of genetic gain and influence the resulting gains in milk production through the pathways of bull to cow and bull to bull. The genetic level of the top 10 bulls graduating from the progeny test based on EBV was evaluated for the MAS schemes and for the control. The superiority in the genetic level for the MAS schemes over that of the control was less for the top 10 bulls than for all of the progeny-tested bulls (Table 5). The reduction in genetic level superiority between all progeny-tested bulls and the top 10 bulls was greater when the threshold level was low because of lower variance in breeding value in the progeny-tested bulls when low threshold levels were used. The variance of the bull breeding value is the variance of the QTL allele received from the sire, the variance of the QTL allele received from the dam, and the polygenic variance. At low threshold levels, the variance at the sire QTL was less than that at the higher threshold levels because all of the sires of sons are deemed to be heterozygous at the QTL and, therefore, their progeny-tested sons are selected to have the same QTL allele. The dam QTL variance and the polygenic variance were unaffected by the threshold level. As a result of the lower sire QTL variance, the breeding value variance was lower for the lenient threshold levels. The lower superiority in the top 10 bulls compared with that of all of the progeny-tested bulls was observed for both top down and bottom up schemes.

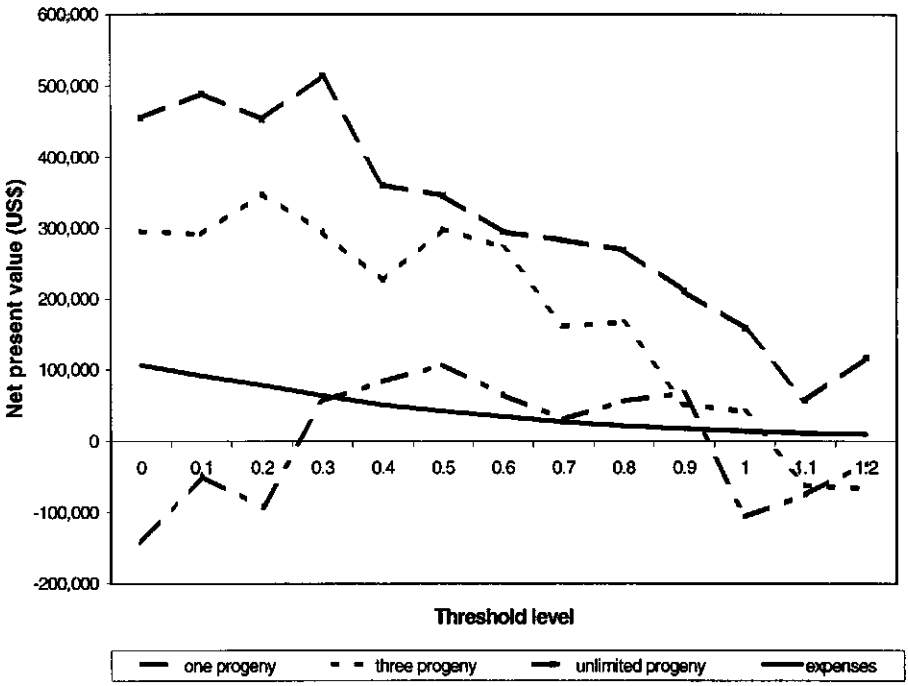
TABLE 5: The effect of different threshold levels in identifying heterozygosity of quantitative trait loci (QTL) on the QTL and breeding value variances and the genetic level[1] of bulls entering and leaving the progeny test.

Threshold level (σ_G)	Sire QTL variance	Dam QTL variance	Breeding value variance	Genetic increase in all bulls	Genetic increase in top 10 bulls
0	0.65	1.93	18.50	0.52	0.43
0.4	0.85	1.92	18.64	0.38	0.31
0.8	1.15	1.93	18.75	0.19	0.21
1.2	1.35	1.92	18.80	0.09	0.09

[1]Increase in genetic level is the average percentage superiority compared with values for the control without QTL information, for all progeny-tested bulls.

Economic response: The increase in genetic level for the top 10 bulls graduating from the progeny test was used to calculate the economic response from MAS. Figure 4 outlines the net present values for the different threshold levels for the three levels of reproductive performance for the bull-dams. Expenses were approximately $105,000 for a threshold of 0 σ_G and progressively decreased as the threshold level increased and as fewer sires of sons were deemed to be heterozygous; therefore, fewer bulls had to be produced for pre-selection (Figure 4). For one calf per bull-dam, the MAS scheme was unprofitable at low and high thresholds and was marginally profitable at the intervening threshold levels. For threshold levels up to 0.5 σ_G the net present value was $300,000 for 3 progeny born per bull-dam and $0.5 million for 40 progeny per bull-dam. For threshold values above 0.5σ_G, the net present value decreased as the threshold levels increased.

FIGURE 4: Net present value of a bottom up scheme for marker assisted selection (MAS) for different levels of bull-dam reproductive performance and thresholds in identifying heterozygosity of quantitative trait loci. Expenses are shown for the bottom up scheme for MAS with one progeny per bull-dam.



The highest net present value for the use of bottom up selection on the sire path and top down selection on the dam path was approximately $1 million when there was no restriction on female reproduction in the bull-dams (40 progeny) and $0.5 million for 3 progeny per bull-dam (results not shown). The costs for this scheme were higher than for bottom up alone, as on average, 4 bulls had to be produced to get one bull with the desirable QTL alleles from the sire and dam paths.

DISCUSSION

The average genetic superiority of the bulls being progeny tested with MAS over that of the control was greater than the genetic superiority of the elite bulls graduating from the progeny test because of reduced genetic variance from selection at the QTL. The genetic level of bulls graduating from the progeny test is more important than the average level of the progeny test because the elite bulls determine the rate of genetic gain. However, in this study, the genetic superiority of the average level of the progeny test bulls is used, because it is easy to interpret as the number of bulls differs for the pathways of sire to sire and sire to cow and also between breeds and countries.

The rates of genetic gain with top down and bottom up MAS breeding schemes are dependent on the use of reproductive techniques being used on the bull-dams. Without reproductive techniques, the loss in selection differential caused by the extra sons required for pre-selection negates nearly all of the genetic benefits of bottom up and is greater than the benefits of top down. The utilisation of reproductive technology is even more important when bottom up MAS is used on the paternal path and top down MAS is used on the maternal path. When reproductive technology is not used, the genetic response of bottom up and top down schemes utilised together is less than that of solely using the bottom up scheme on the paternal side. This result is due to the increased selection differential loss as, on average, 4 bulls are required for 1 bull to be progeny tested when both the sire and maternal grandsire are identified as segregating for the QTL.

When the requirement is to progeny test only bulls that carry all of the favourable QTL alleles requirements will be greater for the reproductive performance of bull-dams if >1 QTL is identified and used in MAS. This requirement would probably be detrimental to the rate of genetic gain when many QTL are identified and selected for because of the selection differential loss on the bull-dam path. A better option would be to use an index that details the number of favourable QTL alleles minus the number of unfavourable QTL alleles and then to progeny test the bulls with the highest index (KASHI *et al.* 1990).

KASHI *et al.* (1990) reported gains of some 20% and MACKINNON and GEORGES (1998) reported gains of 10%, for top down MAS. The scenario in this study that is most comparable to the previous two is that of 40 progeny born (no selection loss) for each bull-dam that had an increase in genetic level of about 2%. The lower genetic response with the top down scheme in this study is primarily because of the violation of the assumption of top down that the poorer or better marker haplotype (allele) for the paternal grandsire is better or poorer than the QTL allele received from the paternal granddam population. For a threshold of 0.3 σ_G, the assumption that the better or poorer allele in the grandsire was also, respectively, better or poorer than the allele from the dam population did not hold in 53% of the sires selected as sire of sons. Eleven percent of the time the sire was homozygous for the QTL, and 42% of the time the QTL allele from the dam population was superior to the better paternal QTL allele or inferior to the poorer paternal QTL allele. Contributing to the 42% error rate is a 17% error rate in identifying which of the grandsire alleles was the best for the threshold level of 0.3 σ_G.

The top down scheme on the maternal side resulted in greater increases in genetic gain than the top down scheme on the paternal path (results are not shown but can be calculated from responses for bottom up and top down schemes together minus the response from the bottom up scheme alone). The increased response was because the assumption about the superiority or inferiority of the maternal grandsire QTL allele was violated in 40% of the cases, which is less than the 53% for the paternal side. Eighteen percent of the time the bull-dam was homozygous for the QTL, and 22% of the time the QTL allele from the dam population was superior to the better paternal QTL allele or inferior to the poorer paternal QTL allele. Homozygosity at the QTL does not affect genetic response when progeny per bull-dam is unlimited because the QTL allele transmitted is the same, and no loss of selection differential occurs.

In addition, the average contrast in the sires of sons and bull-dams was less than the average contrast for the maternal and paternal grandsires with significant effects. Therefore, even when the superior or inferior QTL allele was correctly

identified in the grandsires, the difference between that allele and the allele from the dam population was less, and, therefore, the QTL information had less value.

The larger error rate in top down schemes (40 to 50% for 0.3 σ_G threshold) than in bottom up schemes (29%) (Table 2) and the reduced inferiority or superiority of the paternal QTL allele highlight that the main advantage of the bottom up scheme is that the marker contrast is observed in the parents of the selection candidates and, therefore, is not affected by another generation of selection.

MACKINNON and GEORGES (1998) reported that the genetic response to the bottom up scheme was approximately 10%, but in this study, the response was 5%. MACKINNON and GEORGES (1998) applied MAS to an unselected population, but, in this study, selection had been undertaken for one to two generations. When MAS was applied to an unselected population in this study, the increase in genetic level was 10% for a QTL that explained 5% of phenotypic variance and 15% for a QTL that explained 10% of the phenotypic variation (results not shown). Selection in the previous generations resulted in an average contrast between QTL alleles in the sires of sons that was approximately 75% of that when no selection was undertaken. This result is because EBV selection chose the sires that had the better QTL alleles and, therefore, were homozygous at the QTL or had smaller differences between the two QTL alleles.

Lower thresholds to identify whether a sire is heterozygous for the QTL were the genetic and economic optima when reproductive technologies were used in conjunction with MAS. In agreement with the results of MACKINNON and GEORGES (1998), the number of daughters used in the identification of whether a sire of sons was heterozygous did not affect the increase in genetic level at the three levels investigated: 60, 85, and 150 daughters (results not shown).

The genetic responses in this paper are from a one-off, first generation use of MAS. The longer-term response for the MAS schemes presented will be lower than that in the first generation for continued MAS at the same locus (GIBSON 1994). In addition, 10 alleles were arbitrarily chosen for the QTL. Genetic response for the two

MAS schemes would be lower if the QTL had less than 10 alleles segregating (SPELMAN 1998)

The genetic and economic gains outlined in this paper demonstrate that within-family MAS has the potential to have a reasonable impact on the rate of genetic level, especially when QTL information is utilised on both the paternal and maternal paths. However, the use of reproductive technologies on the bull-dams is imperative; otherwise, the gains from MAS are eroded by the selection differential loss of selecting more bull-dams to enable pre-selection.

ACKNOWLEDGEMENTS

The authors thank MARGARET MACKINNON and MICHEL GEORGES for making their then unpublished manuscript available and thus motivating this study.

REFERENCES

BINK, M.C.A.M., J.A.M. VAN ARENDONK, and R. L. QUAAS, 1998 Breeding value estimation with incomplete marker data. Genet. Sel. Evol. **30:** 45-58.

BRASCAMP, E. W., J.A.M. VAN ARENDONK, and A. F. GROEN, 1993 Economic appraisal of the utilization of genetic markers in dairy cattle breeding. J. Dairy Sci. **76:** 1204-1214.

BULMER, M. G., 1971 The effect of selection on genetic variability. Am. Naturalist **165(943):** 201-211.

FALCONER, D. S., and T.F.C. MACKAY, 1996 Introduction to Quantitative Genetics. 4th edition. Longman Group Ltd., Essex, United Kingdom. pg 201-202.

FERNANDO, R. L., and M. GROSSMAN, 1989 Marker-assisted selection using best linear unbiased prediction. Genet. Sel. Evol. **21:** 467-477.

GEORGES, M., D. NIELSEN, M. MACKINNON, A. MISHRA, R. OKIMOTO, *et al.* 1995 Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics **139:** 907-920.

GIBSON, J. P., 1994 Short-term gain at the expense of long-term response with selection of identified loci. Proceedings of the 5th World Congress on Genetics applied to Livestock Production, Guelph, Canada **21:** 201-204.

HARRIS, B. L., 1998 Breeding cattle for economic efficiency: A New Zealand pasture-based system. Proceedings of the 6th World Congress on Genetics applied to Livestock Production, Armidale, Australia **25:** 383-386.

HOESCHELE, I., 1993 Elimination of quantitative trait loci equations in an animal model incorporating genetic marker data. J. Dairy Sci. **76:** 1693-1713.

KASHI, Y., E. HALLERMAN, and M. SOLLER, 1990 Marker assisted selection of candidate bulls for progeny testing programmes. Anim. Prod. **51:** 63-74.

MACKINNON, M. J., and M. GEORGES, 1998 Marker-assisted preselection of young dairy bulls prior to progeny testing. Livest. Prod. Sci. **54:** 229-250.

RUANE, J., and J. J. COLLEAU, 1996 Marker-assisted selection for a sex-limited character in a nucleus breeding population. J. Dairy Sci. **79:** 1666-1678.

SPELMAN, R. J., 1998 Major factors in marker-assisted selection genetic response in dairy cattle populations. Proceedings of the 6th World Congress on Genetics applied to Livestock Production, Armidale, Australia **26:** 365-368.

SPELMAN, R.J., W. COPPIETERS, L. KARIM, J.A.M. VAN ARENDONK, and H. BOVENHUIS. 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., and D. J. GARRICK, 1997 Utilisation of marker assisted selection in a commercial dairy cow population. Livest. Prod. Sci. **47:** 139-147.

SPELMAN, R. J., and D. J. GARRICK, 1997 Effect of live weight and differing economic values on milk fat, protein, volume and live weight selection response. J. Dairy Sci. **80:** 2557-2562.

VILKKI, H. J., D.-J. DE KONING, K. ELO, R. VELMALA, A. MÄKI-TANILA, 1997, Multiple marker mapping of quantitative trait loci of Finnish dairy cattle by regression. J. Dairy Sci. **80:** 198-204.

CHAPTER SEVEN

UTILISATION OF GENETIC VARIATION BY MARKER ASSISTED SELECTION IN COMMERCIAL DAIRY CATTLE POPULATIONS

RICHARD J. SPELMAN, DORIAN J. GARRICK and JOHAN A.M. VAN ARENDONK

SUBMITTED TO LIVESTOCK PRODUCTION SCIENCE

ABSTRACT

Potential genetic benefits of marker assisted selection (MAS) were evaluated by calculating selection response resulting from four pathways of selection. Genetic variation was partitioned into polygenic and loci that were in linkage disequilibrium with marker loci or haplotypes. The percentage of genetic variation that was marked, was varied from 0 to 100 percent. These assumptions describe the degree of genetic knowledge that may be available in ten years. Three breeding strategies with markers were evaluated: progeny test scheme (BMARK); progeny test scheme but unproven bulls allowed on the bull to bull selection path (YBULL); and a breeding programme where cows without lactation information and bulls without progeny information were eligible for selection (OPEN). Rates of genetic gain (per year) with no marked genetic variance were 0.26 σ_G for the BMARK and YBULL schemes and 0.28 σ_G for the OPEN scheme. On average, an increase of one percent marked genetic variance resulted in an increase in genetic gain of approximately 0.25% for the BMARK scheme, 0.5% for the YBULL scheme and 1% for the OPEN scheme. Maximum genetic response (100% marked genetic variance) for the BMARK scheme was 1.24 times that achieved with no marked genetic variance, 1.52 times for the YBULL scheme, and 2.05 times for the OPEN scheme. Changes in the structure of the breeding scheme are needed to fully gain the benefits of identified loci especially for medium to large proportions of marked genetic variance.

INTRODUCTION

Quantitative trait loci (QTL) experiments in dairy cattle using granddaughter and daughter designs are successfully detecting QTL (GEORGES *et al.* 1995). However, the proportion of genetic variation for individual traits that has been explained to date in these experiments is usually less than 15%. For milk production traits with heritabilities of some 30%, this is equivalent to approximately 5% of the phenotypic variance. The expected genetic improvement from marker assisted selection is some 2-10% given up to 15% genetic variance identified, (SPELMAN and GARRICK 1997, 1998; MACKINNON and GEORGES 1998).

Experimental techniques such as selective DNA pooling (DARVASI and SOLLER 1994) applied to large half-sib families, which exist in commercial dairy

populations, have the power to explain more genetic variation than granddaughter and daughter designs with less genotyping effort (SPELMAN *et al.* 1998). The potential of selective DNA pooling has been demonstrated in the Israeli dairy population with a large proportion (0.5-0.75) of genetic variation for one milk production trait being explained (LIPKIN *et al.* 1998, M. SOLLER personal communication).

A limitation for MAS is that linkage phase has to be estimated for each family and confirmed in subsequent generations. To overcome this problem, the QTL themselves would have to be identified or markers or marker haplotypes identified that are in linkage disequilibrium with the QTL (SMITH and SMITH 1993). Linkage disequilibrium mapping has been successfully applied to identify single genes in livestock (CHARLIER *et al.* 1996) and its application to complex traits in humans is viewed positively (RISCH and MERIKANGAS 1996). However, BARET and HILL (1997) report that the application of linkage disequilibrium mapping in livestock is limited. Linkage disequilibrium mapping is being currently applied to complex traits in livestock (M GEORGES personal communication) and in ten years time there is the possibility that in dairy cattle populations a large proportion of identified genetic variance will be in linkage disequilibrium with marker loci (haplotypes).

SMITH (1967) and LANDE and THOMPSON (1990) among others have studied genetic responses to a single generation of marker assisted selection with the assumption of genes in disequilibrium. However, to date the utilisation of disequilibrium by MAS in a dynamic cattle breeding scheme has not been investigated.

The objective of this study is to identify the possible genetic responses that could be achieved with MAS assuming a large proportion of the genetic variation is in disequilibrium with markers. The study investigates the gains from MAS that can be made in a dynamic breeding scheme where the age at selection is not fixed.

METHOD

Simulation model: A deterministic simulation model accounting for four pathways of selection and overlapping generations was developed. Population parameters were based on the New Zealand Holstein-Friesian breed and are outlined in Table 1.

TABLE 1: Population parameters used to calculate annual genetic gain.

Cow population
Age structure; 22% calves, 16% 1 year olds (yo's), 14% 2 yo's, 13% 3 yo's, 11% 4 yo's, 9% 5 yo's, 8% 6 yo's and 7% 7 yo's.
272,000 calves to 7 yo cows eligible for selection as bull dams
455 selected each year as bull dams.
No selection on the cow to cow pathway.
Bull population
140 bulls progeny tested per year.
Two percent death rate (three bulls) per year.
Receive progeny test proof as five year olds on 85 daughters.
Ten bulls selected for bull to cow pathway each year from live 5, 6 and 7 yo's.
Two bulls selected for bull to bull pathway each year from 5 and 6 year olds (dead bulls eligible for selection as frozen semen held).

The base breeding scheme comprises one hundred and forty bulls that are progeny tested on 85 daughters each and receive their progeny test proof at the age of 5 years. To be eligible for selection in the base scheme a female had to be at least one year of age and a male had to have 85 daughters with at least one complete lactation record. Ten bulls were selected from the live 5, 6 and 7 year old bulls for the bull to cow path and three bulls for the bull to bull path from live or dead (frozen semen stored) 5 and 6 year old bulls. The genetic contribution of young bull inseminations undertaken in the cow population for progeny testing was ignored. Four hundred and fifty lactating cows were selected from a population of some 272,000 females for the cow to bull path. No selection was undertaken on the cow to cow path; females were

produced from the 30% of one year olds that were artificially inseminated and from all older age groups (Table 1). For the MAS strategies the requirements to be eligible for selection were relaxed (outlined later).

Selection indices were developed to calculate the accuracy of evaluation (variance of EBV) for the selection candidates, for a trait or index with heritability of 0.25 and repeatability of 0.6. The information sources in the base selection index were dam (2 lactations), paternal halfsibs (85), paternal grand halfsibs (1000), maternal grand halfsibs (1000), paternal granddam performance (4 lactations) and maternal granddam performance (4 lactations). Selection indices for female selection candidates included lactation information (0-6 lactations) on the animal itself, and for the male selection candidates information from female progeny (0 or 85) was included.

The additive genetic variance was partitioned between unmarked additive polygenic variation (which will be referred to as polygenic variance) and variation because of the QTL in linkage disequilibrium (referred to as marked genetic variance). The marked genetic variance was not assumed to be one QTL but many QTL with effects ranging from large to small as described by SHRIMPTON and ROBERTSON (1988). The marked genetic variance was included in the above selection index as an information source for the selection candidate only in the selection index. The molecular marker data for the relatives provides no more information on the selection candidate's markers, as they are in disequilibrium. However, the data does allow more accurate estimation of the relatives and indirectly the breeding value of the selection candidate (LANDE and THOMPSON, 1990). In this study the marker data was not included as an information source for the relatives as the effect on accuracy of the selection candidate's estimated breeding value is minor.

Mean and variance of EBV was calculated for males and females for each age group (0-7 years). Truncation selection was undertaken across age groups eligible for selection (DUCROCQ and QUAAS, 1988). Based on normal distribution theory, the selected fraction and the standardised selection differential were calculated for each age group eligible for selection. Selection differentials were adjusted to account for finite population size using the approximation of BURROWS (1972);

$$i_f = i_\infty - \frac{(1-p)}{(2i_\infty p(N_{tot}+1)} \quad [1]$$

where i_f is the finite selection intensity, i_∞ is the infinite selection intensity, p is the fraction selected from the number of individuals available for selection (N_{tot}) in that age group.

The effect of selection on the (co)variances of the information sources of the selected animals was calculated using a generalisation of COCHRAN'S (1951) formula.

$$\sigma_{jk}^* = \sigma_{jk}(1-r_{TI}^2\, i_\infty\, (i_\infty - t)) \quad [2]$$

where σ_{jk}, and σ_{jk}^* are the covariance of j and k before and after selection respectively (when j=k it is the variance of j), r_{TI}^2 is the squared correlation between index and objective for the selected animals and t is the truncation point.

Selection was undertaken for fifty years to ensure the breeding program reached equilibrium. Equilibrium rates of genetic gain are reported. Reduction in genetic variance through selection [equation 2], (BULMER 1971) was modelled over the fifty years. The marked genetic variance as a proportion of the additive genetic variation was maintained at the same level over all years. Maintaining the marked genetic variation at the same proportion over the years was based on the assumption that there were no alterations in genetic variance through allele frequency changes. Inbreeding was ignored in the model.

Marker assisted selection strategies: Three different breeding schemes with varying proportions of genetic variance explained by the markers (0 to 100%) were investigated. The control for each of the MAS schemes was when marked genetic variance was 0% for that scheme. All the genetic responses are in terms of annual gains in genetic standard deviations in the base population and percentage gains are relative to the appropriate control.

a) Base breeding program with marker information (BMARK): The breeding scheme is the same as that described as the base breeding program but with the

additional information of markers. Marker information was exploited for all selection paths.

b) Young unproven bulls selected for the bull to bull path (YBULL): Bulls selected for the bull to bull pathway could be either selected as one year olds (pedigree and marker information) or as five year olds (pedigree, progeny and marker information). Bulls that were selected as yearlings were also available for selection as five year olds. For the bull to cow selection path, only bulls with progeny records were eligible for selection. Marker information was used on all selection paths.

c) Open scheme (OPEN): Bulls were eligible for selection from one year of age to seven years of age for both bull to cow and bull to bull selection paths. Bulls could be selected in more than one year. Selection of young bulls for the bull to cow pathway is limited by semen production constraints. In the simulated breeding scheme it is assumed that all breeding occurs in a 3 month period (New Zealand dairy production system). Mature bulls (≥3 years) are assumed to produce an average of 200,000 doses of semen over this period and yearling bulls 30,000 doses and 2 year old bulls 150,000 doses (D. HEMARA personal communication). The truncation procedure applied across bull ages accounted for the lower semen capabilities of the younger bulls for the bull to cow path. Age constraints on the cow to bull path were relaxed to allow for selection from calves (5 months of age) and older animals. Reproduction from calves is dependent on techniques such as *in vitro* fertilisation. Replacements on the cow to cow path are as in the base scheme. Marker information is used on all selection paths.

For MAS schemes b) and c) the age groups were divided into subgroups that reflected the amount of pedigree information. For instance in the open MAS scheme a young bull may be either sired by a bull that was selected at one year of age through to a bull that was selected on progeny information. Likewise the number of lactations of dams of the young bulls may vary from none through to six. With differing amounts of pedigree information the accuracy of selection (variance of EBV) will differ and was accounted for.

The contribution of each selection path to the increase in genetic gain, compared with the situation when there was no marked genetic variance, was calculated. This was determined by only including the marker information on the selection path of interest (e.g. bull to cow), and then comparing the genetic response to that when marker information was used on all selection paths.

RESULTS

The rate of genetic gain in the base breeding program with no marked genetic variance was 0.258 σ_G (Table 2). The rate of genetic gain increased as the proportion of marked genetic variance increased in the BMARK scheme (Table 2). The maximum percentage increase was 24% (0.26 to 0.32 σ_G) when all of the genetic variance was marked (Table 2). Eighty to ninety five percent of the increase in genetic gain was from the cow to bull path. The increase in genetic gain for each unit increase in marked genetic variance was approximately linear up to 50% marked genetic variance and then it increased in an exponential manner (Figure 1). The greater rate of improvement in genetic gain for the larger proportions of marked genetic variance is due to the accuracy of evaluation increasing more per unit marked genetic variance than at lower proportions of marked genetic variance, especially for sire evaluation as most of the genetic variance was explained through the progeny test of 85 daughters (Figure 1). The marked genetic variance had little effect on the age of animals selected: the largest decrease in generation interval being 0.05 years for the cow to bull path over the range of 0-100% marked genetic variance.

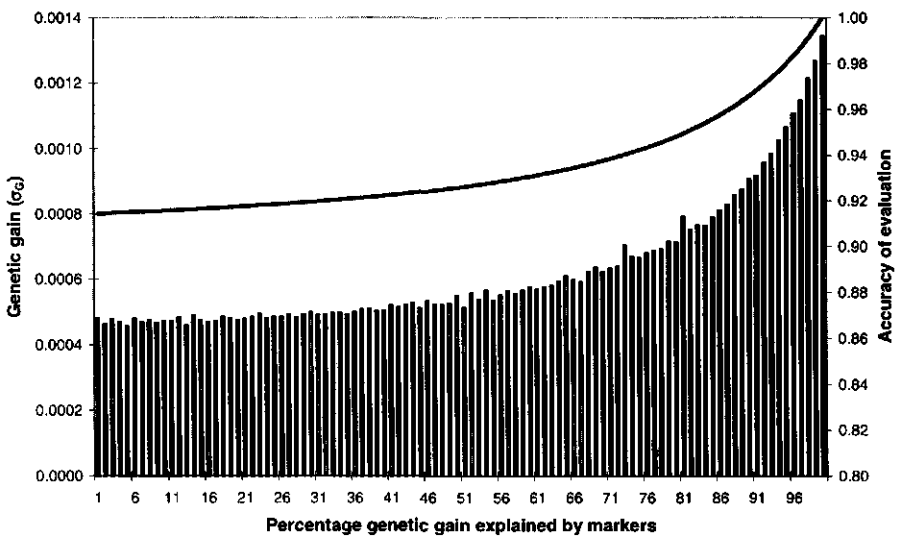
The rate of genetic gain for the breeding scheme with young bulls being used for the bull to bull path (YBULL) for 0% marked genetic variance was approximately the same as that when only proven bulls were used for that selection path (Tables 2 and 3). The proportion of young bulls selected for this path was 0.24 resulting in a generation interval of 5.04 years compared to 6.28 years for the proven bull scheme.

TABLE 2: Rates of genetic gain in the BMARK[1] breeding program with different proportions of genetic variance explained by markers.

Marked genetic variance	Genetic gain [2] (σ_G/year)	Percent increase [3]
0.0	0.258	-
0.1	0.263	1.82
0.2	0.268	3.67
0.3	0.273	5.56
0.4	0.278	7.49
0.5	0.283	9.52
0.6	0.288	11.66
0.7	0.295	13.99
0.8	0.301	16.62
0.9	0.309	19.78
1.0	0.320	24.05

[1]Sires must have a progeny test proof to be eligible for selection and cows must be one yr of age.
[2]Genetic standard deviation in the base population.
[3]Compared to no marked genetic variance.

FIGURE 1: Incremental increase in genetic gain (bars) and accuracy of sire evaluation (line) for each additional percent of marked genetic variance.



The rate of genetic improvement in the YBULL scheme increased by 52% when all of the genetic variance was marked. A linear increase in genetic gain for

extra unit of marked genetic variance was observed (~0.0013 σ_G / % marked genetic variance). As the percentage of marked genetic variance moved from zero to 10% the proportion of bulls selected as young bulls for the bull to bull path increased from 24% to 56% (Table 3) and surpassed 90% when marked genetic variance reached 36%. Therefore the rate of genetic gain in a breeding scheme where only young bulls are eligible for selection for the bull to bull path is nearly equal to that of young and old bulls when marked genetic variance is more than 30%.

TABLE 3: Rates of genetic gain in the YBULL[1] breeding program with different proportions of genetic variance explained by markers.

Marked genetic variance	Genetic gain [2] (σ_G/year)	Percent increase [3]	Percent young bulls [4]
0.0	0.260	-	24.1
0.1	0.271	4.49	56.0
0.2	0.284	9.42	74.6
0.3	0.300	15.46	86.0
0.4	0.313	20.45	92.3
0.5	0.326	25.45	95.8
0.6	0.339	30.48	97.7
0.7	0.352	35.55	98.8
0.8	0.365	40.72	99.2
0.9	0.379	46.12	99.5
1.0	0.395	52.09	99.6

[1]Sires for the bull to bull selection path are eligible for selection as either 2 year olds (unproven) or as 5 year olds (proven). Sires must have a progeny test proof to be eligible for selection for the bull to cow path and cows must be one year of age.
[2]Genetic standard deviation in the base population.
[3]Compared to no marked genetic variance.
[4]Percentage of sires selected for the bull to bull selection path that are 2 years of age.

The contribution to improvement in genetic gain from each of the three selection paths was, on average (over all proportions of marked variance), approximately 45% from the cow to bull path, 5% from the bull to cow path and 50% from the bull to bull path in the YBULL scheme.

The rate of genetic gain with the OPEN scheme with no marked genetic variation was 0.28 σ_G (Table 4). The proportion of unproven bulls selected for the bull to cow pathway was 0.67 and 0.37 for the bull to bull pathway, when no genetic variation was marked. Seventy percent of the cows selected as bull mothers did not

have lactation information. The average generation interval in the OPEN scheme for the bull to cow path was 3.67 years and 4.79 years for the bull to bull path and 2.30 years for the cow to bull path when there was no marked genetic variance.

TABLE 4: Rates of genetic gain in the OPEN[1] breeding program with different proportions of genetic variance explained by markers.

Marked genetic variance	Genetic gain [2] (σ_G/year)	Percent increase [3]
0	0.282	-
0.1	0.301	6.73
0.2	0.330	17.28
0.3	0.360	27.85
0.4	0.406	44.09
0.5	0.437	55.16
0.6	0.467	65.88
0.7	0.497	76.33
0.8	0.525	86.23
0.9	0.551	95.69
1.0	0.577	104.71

[1]Sires are eligible for selection for both sire selection paths from one year of age and cows are eligible for selection as calves and older age groups.
[2]Genetic standard deviation in the base population.
[3]Compared to no marked genetic variance.

The rate of genetic gain for the OPEN scheme when all of the genetic variance was marked was 0.57 σ_G, which is an increase of 105% over the scheme with no marked genetic variance. The rate of increase in genetic gain when marked genetic variance increased from 0 to 10% was 7%. The largest percentage increase in genetic gain was when marked genetic variance increased from 30 to 40%. On average, an increase of one percent marked genetic variance resulted in an increase of approximately one percent in genetic gain. Sixty percent of the increase in genetic gain originated from the cow to bull selection path when up to 40% marked genetic variance was marked, and reduced to 50% for higher levels of marked genetic variance. The marker information benefited the bull to bull pathway more than the bull to cow selection path with some 30% of the extra genetic response from the bull to bull path and 20% from the bull to cow path.

When all of the genetic variance was marked the proportion of unproven bulls for the bull to cow pathway and for the bull to bull pathway was more than 0.99, as was the proportion of cows selected without lactation information for the cow to bull selection path. The average generation interval for the bull to cow path was 2.34 years and 2.18 years for the bull to bull path and 1.6 years for the cow to bull path when all of the genetic variance was marked.

DISCUSSION

Simulation model: It has been shown in MAS simulation studies, with a genetic model comprising of polygenic and a small number of marked loci, that selection alters the allele frequency at the marked loci, which decreases the variation at the loci as they near fixation (MEUWISSEN and GODDARD 1996; RUANE and COLLEAU 1996). In this study it was assumed that the variance of the marked QTL was not affected by allele frequency changes, but only reduced by the negative covariance between loci generated by selection (BULMER 1971). Therefore the marked genetic variance as a proportion of the additive genetic variance was static over years. This was assumed as this study concentrated on the genetic response when large proportions of marked genetic variance were identified. In this situation it is likely that there will be many loci marked and changes in allele frequency at each locus will be small. In addition, it would be inconsistent to assume there are only allele frequency changes when the genetic variance is marked and ignore changes when the genetic variance is treated as polygenic. The fixed proportion of additive genetic variance being marked may be realistic as a result of new QTL being identified over time.

In this study the percentage of marked genetic variance was varied from 0 to 100%. Although the true underlying genetic model is unknown (e.g. number of loci, distribution of effects, and interaction of loci) it is unlikely that all of the genetic variance will be able to be identified as experimental power will be too low to detect loci with small effects, and with epistatic effects. The proportion of genetic variance that the loci with small effects constitute is unknown. However, with experimental techniques such as selective DNA pooling up to 50-75% of the genetic variance for one trait has been identified (M. SOLLER personal communication). Therefore it can

be expected that large proportions (over 50%) of genetic variance should be identified in dairy cattle populations in future.

The accomplishment of identifying loci that are in linkage disequilibrium with marker loci (haplotypes) will be challenging for molecular and quantitative geneticists. RISCH and MERIKANGAS (1996) report that for a complex disease trait in humans, the statistical power of linkage disequilibrium mapping is greater than that of linkage analysis. These authors see the primary limitation of genome-wide association tests as not a statistical one but a technological one. The technological limitations that they saw, were the identification of a large number of polymorphisms and the testing of these polymorphisms on a large number of individuals. These technological limitations may be overcome with the development of single nucleotide polymorphisms (SNPs) as biallelic markers. KRUGLYAK (1997) reported that the SNPs are highly abundant with classic estimates of 1 per 1000 base pairs or more than 3 million in the genome. KRUGLYAK (1997) also reported that the use of these highly abundant markers with non-gel based assays (DNA chips) is promising. This technology has the potential to enable genome scans for linkage disequilibrium and association studies. However, BARET and HILL (1997) state that the application of linkage disequilibrium mapping to livestock will be limited due to insufficient knowledge of the genetic history of the population and the operation of disruptive factors such as selection and drift. These authors see the application of linkage disequilibrium to livestock being limited to discrete traits in specific populations (isolated populations or populations stemming from an admixture event). Further research in this area is needed.

Many alternative breeding schemes could have been investigated, but of the three chosen, BMARK represents a traditional breeding, and the OPEN scheme represents the other end of the spectrum, and the YBULL scheme an intermediary scheme. Other MAS strategies such as pre-selection of bulls entering progeny test were not investigated.

Genetic response: The levels of genetic response that are presented in this study are not what could be achieved today but possibly in 10 years time. If the assumptions used in this study of large proportions of marked genetic variation in

linkage disequilibrium with markers are realised, marker assisted selection has the potential to have a considerable impact on dairy cattle breeding schemes.

MEUWISSEN and VAN ARENDONK (1992) reported similar increases in rates of genetic gain for a progeny test scheme that is comparable to the BMARK scheme. The rates of improvement in genetic gain of approximately 4% for the progeny test based MAS scheme (BMARK) when 20% of the genetic variance is explained is similar to gains outlined in other studies on progeny test based breeding programs (SPELMAN and GARRICK, 1997, 1998; MACKINNON and GEORGES 1998). For the OPEN and YBULL schemes the most comparable study is that of MEUWISSEN and GODDARD (1996) where they simulated a closed nucleus breeding scheme. Those authors reported an increase in genetic gain of 38% when 33% of the genetic variance was marked, which is a similar result to that of the OPEN scheme.

Rates of genetic gain presented in this study are at equilibrium, which are only reached after 20-30 years of selection. When changing the breeding program the genetic responses in the immediate years can fluctuate quite dramatically (DUCROCQ and QUAAS 1988). If a MAS breeding scheme is implemented that is quite different from the current scheme the genetic response in years immediately after implementation should be investigated, because it will be of importance to the breeding company in terms of retaining and increasing market share until equilibrium response is reached. Another aspect of the breeding scheme to be investigated would be the variance of genetic response. MEUWISSEN (1991) reported that breeding schemes with the shortest generation intervals had the highest variance of response. The increase in variance of response will be less than that reported by MEUWISSEN (1991), as marker information increases the accuracy of the genetic merit estimate for the younger animals compared to the situation of no marker information (MEUWISSEN 1991). Marker assisted selection also has the additional risk factor of errors in estimation of location and size of the marked genetic variance. The decision on implementation of a genetically superior breeding scheme with larger variance of genetic response will depend on the degree of risk aversion.

In all three of the MAS schemes it was assumed that 140 bulls were progeny tested each year. For the open scheme when 40% of the genetic variance was marked

95% of the bulls selected for the two sire paths were unproven bulls. In this situation progeny testing the bulls adds little to the genetic response and only adds expense to the breeding scheme. However, if progeny testing is required as a tool to market bulls, then one could pre-select the young bulls to progeny test when marked variance is ≥40% (also possible at lower levels), which would reduce the cost. For the OPEN scheme without progeny testing the genetic response is near equal to that with progeny testing when marked variance is 20% and greater. In addition, costs for the breeding scheme with the young bull breeding scheme will be reduced as no progeny testing is undertaken.

In the BMARK scheme, marked genetic information added little to the bull selection paths as the progeny test on 85 daughters explained most of the genetic variance. If an organisation was not willing to change its breeding program away from proven bulls another strategy may be to undertake the progeny test on fewer daughters. Assuming the total number of daughters in the progeny test program is fixed, more sires could be progeny tested. For example, for marked genetic variance of 20% and 198 bulls progeny tested on 60 daughters, genetic gain increased by 6.7% over the base situation of 140 bulls progeny tested on 85 daughters and no marked genetic variance. This increase is nearly double the percentage increase for 20% marked genetic variance and 140 bulls and 85 daughters (Table 2). However, the genetic advantage may not be an economic advantage once the costs of producing and feeding/housing the extra bulls is accounted for (MEUWISSEN 1997).

Relaxing the constraint on age of selection for the bull to bull path, resulted in a breeding scheme (YBULL) that was able to benefit more from the identification of QTL. On average, the percent increase in genetic gain was double for the YBULL scheme compared to the BMARK scheme (Tables 2 and 3). When the age of selection was relaxed further on the bull to cow and cow to bull pathways (OPEN) the marked genetic information was utilised more efficiently again. On average, the percentage increase in genetic gain for the OPEN scheme was double that of the YBULL scheme (Tables 3 and 4). Therefore, to optimally use the marked genetic information, breeding programs will have to move away from the progeny test system and select animals without phenotypic and progeny information.

Non-acceptance of young bulls for the bull to cow path by the semen users could hinder implementation of a breeding program such as the OPEN scheme. This would not be the case for the YBULL scheme as decisions for the bull to bull path are made by the breeding organisations and the sires selected by the semen users will still have approximately the same reliability as the BMARK scheme. There is increasing interest from in the use of young bulls with 0% marked genetic variance, and as shown in this study the benefits of using young bulls improves as the rate of genetic gain increases.

This study shows that with medium to large proportion of genetic variance identified and being in linkage disequilibrium with marker loci, MAS can substantially increase the rate of genetic gain. To utilise the marker information, breeding companies will have to alter their breeding schemes from the traditional progeny test system to schemes selecting animals without lactation and progeny information. However this may not be appropriate at low proportions of marked genetic variance, as the cost of altering the breeding scheme may be greater than the benefits.

LITERATURE CITED

BARET, P. V., and W. G. Hill,. 1997 Gametic disequilibrium mapping: potential applications in livestock. An. Brdg Abstr. **65(5):** 309-318.

BULMER, M. G., 1971 The effect of selection on genetic variability. Am. Naturalist **165(943):** 201-211.

BURROWS, P. M., 1972 Expected selection differentials for directional selection. Biometrics **28:** 1091-1100.

CHARLIER, C., F. FARNIR, P. BERZI, P. VANMANSHOVEN, B. BROUWERS, *et al.* 1996 Identity-by-descent mapping of recessive traits in livestock: application to map the bovine *syndactyly* locus to chromosome 15. Genome Res. **6:** 580-589.

COCHRAN, W. G., 1951 Improvement by means of selection. Proc. 2nd Berkley Symp. Math. Stat. Prob. (Ed J Neyman), pg 449-470.

DARVASI, A., and M. SOLLER, 1994 Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics **138:** 1365-1373.

DUCROCQ, V., and R. L. QUAAS, 1988 Prediction of genetic response to truncation selection across generations. J. Dairy Sci. **71:** 2543-2553.

GEORGES, M., D. NIELSEN, M. MACKINNON, A. MISHRA, R. OKIMOTO, *et al.* 1995 Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics **139:** 907-920.

KRUGLYAK, L., 1997 The use of a genetic map of biallelic markers in linkage studies. Nature Genetics **17:** 21-24.

LANDE, R., and R. THOMPSON, 1990 Efficiency of marker assisted selection in the improvement of quantitative traits. Genetics **124:** 743-756.

LIPKIN, E., M. O. MOSIG, A. DARVASI, E. EZRA, A. SHALOM, *et al* 1998 Mapping loci controlling milk protein percentage in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers. Genetics (in press).

MACKINNON, M. J., and M. GEORGES, 1998 A bottom up approach to marker assisted selection. Livest. Prod. Sci. (in press).

MEUWISSEN T.H.E., 1991, Expectation and variance of genetic gain in open and closed nucleus and progeny testing schemes. Anim. Prod. **53:** 133-141.

MEUWISSEN, T.H.E., 1997 Optimization of progeny tests with prior information on young bulls. Livest. Prod. Sci.:**52(1):** 57-68.

MEUWISSEN, T.H.E., and M. E. GODDARD, 1996 The use of marker haplotypes in animal breeding schemes. Genet. Sel. Evol. **28:** 161-176.

MEUWISSEN, T.H.E., and J.A.M. VAN ARENDONK, 1992 Potential improvements in rate of genetic gain from marker assisted selection in dairy cattle breeding schemes. J. Dairy Sci. **75:** 1651-1659.

RISCH N., and K. Merikangas, 1996 The future of genetic diseases of complex human diseases. Science **273:** 1516-1517.

RUANE, J., and J. J. COLLEAU, 1996 Marker-assisted selection for a sex-limited character in a nucleus breeding population. J. Dairy Sci. **79:** 1666-1678.

SHRIMPTON, A. E., and A. ROBERTSON, 1988 The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster.* II. Distribution of third chromosome bristle effects within chromosome sections. Genetics **118:** 445-459

SMITH C., 1967 Improvement of metric traits through specific genetic loci. Anim. Prod. **9:** 349-358.

SMITH, C., and D. B. SMITH, 1993 The need for close linkages in marker assisted selection for economic merit in livestock. Anim. Breed. Abstr. **61(4):** 197-204.

SPELMAN, R. J., and D. J. GARRICK, 1997 Utilisation of Marker Assisted Selection in a commercial dairy cow population. Livest. Prod. Sci. **47:** 139-147.

SPELMAN, R. J., and D. J. GARRICK, 1998 Genetic and economic responses for within-family marker-assisted selection in dairy cattle breeding schemes. J. Dairy Sci. (accepted)

SPELMAN, R. J., N. LOPEZ-VILLALOBOS, and D. J. GARRICK, 1998 Experimental designs for quantitative trait loci detection in the New Zealand dairy industry. Proceedings New Zealand Society of Animal Production. **58:** 6-9.

CHAPTER EIGHT

GENERAL DISCUSSION

The implementation of marker assisted selection (MAS) breeding schemes firstly relies on the ability to correctly identify quantitative trait loci (QTL). This chapter discusses different methods that have been put forward to calculate critical values for rejecting or failing to reject the alternative hypothesis that a QTL exist. In addition, the chapter outlines different experimental designs that could be applied in the New Zealand dairy industry for further detection of QTL. Furthermore three aspects of MAS are discussed: the reasons for variable responses with MAS from simulations for dairy cattle, the first implementation of MAS in the New Zealand dairy industry, and completing the chapter some thoughts on future scenarios for MAS in dairy cattle breeding.

CALCULATION OF CRITICAL VALUES

The issue of determining and setting a threshold level, which when exceeded the experimenter will accept the presence of a QTL, has been the centre of considerable debate. A variety of different threshold levels have been used in QTL experiments, thus causing a degree of confusion in determining which QTL results are significant and which are not. This section reviews three different approaches to determining significance thresholds.

The distribution of the null hypothesis has been approximated through theoretical methods (LANDER and BOTSTEIN 1989, 1994; FEINGOLD *et al.* 1993; DUPUIS 1994; REBAI *et al.* 1994), as well as empirical methods (CHURCHILL and DOERGE 1994; DOERGE and CHURCHILL 1996). DOERGE and REBAI (1996) reported that when trait distributions deviate from normality, and/or the sample sizes are small, approximate values based on the asymptotic (theoretical) distribution properties of the test statistics may not be appropriate, and empirical approaches should be considered. The empirical method of CHURCHILL and DOERGE (1994), applied in Chapter 2, also inherently accounts for the characteristics of the data set (e.g. missing phenotypic or genotypic data, segregation distortion, distribution of trait data), whereas the theoretical approximations are based upon perfect data.

Given the distribution of the null hypothesis (theoretical or empirical), the degree of repeated hypothesis testing; testing at many points over the genome (e.g.

every centiMorgan) and also testing many traits, must be acknowledged in the calculation of critical values. Three methods that address the effect of multiple testing have been put forward for calculating critical values in QTL detection experiments.

Experimental type I error rate: LANDER and BOTSTEIN (1989) and LANDER and KRUGLYAK (1995) stated that an experimental type I error rate should be used, and using a nominal significance level of 5% would not be appropriate when an entire genome scan was undertaken. LANDER and KRUGLYAK (1995) went one step further and stipulated that even if only one chromosome had been analysed, threshold levels should be adjusted to account for all of the genome, as all of the genome will be analysed in the duration of most experiments. In addition, they stated that threshold levels should be set using the assumption of a dense map even if the current map has sparse marker coverage. This is because additional markers will be placed in the region of interest when fine mapping is undertaken. This rationale could also be applied to adjusting for all traits measured on the animals, because in the future one will analyse all measured traits for QTL-marker associations. The level of adjustment for markers is finite, because once you have identified all recombination events; further markers do not add any additional information for linkage analysis. However, the number of traits that one could measure is potentially very large, although the independence of the additional traits may be minimal when applying factor analysis to determine the number of independent traits analysed (as in Chapter 2). It has been cynically suggested that using this rational it could be extended to a laboratory type I error, and thus adjust for all experiments in the laboratory, or lifetime type I error, and adjust for all experiments one will undertake in a lifetime.

Some authors (VAN KAAM *et al.* 1998; DE KONING *et al.* 1998) support the concept of accounting for the repeated hypothesis testing over the genome, but present significance on the basis of only one trait being analysed, regardless of how many are analysed. This approach has been adopted on the basis that it aids the comparison of results between experiments. The purpose of comparing results between experiments is to confirm QTL, although maybe not in the strict statistical sense as presented in Chapter 5. LANDER AND KRUGLYAK (1995) showed that by using nominal p-values in an initial experiment and in a confirmation experiment, false positives could be

confirmed (i.e. also found in the confirmation experiment). The genome-wide thresholds for one trait are certainly more strict than that of nominal threshold levels but the p-values presented from this method are inflated as they have not corrected for all of the multiple testing. Therefore they should not be combined as detailed in Chapter 5, as they do not reflect the complete experiment. The use of genome-wide threshold levels corrected for one trait when combining results from different experiments would only be appropriate if the experimenter was solely interested in that trait.

When undertaking confirmation experiments, as outlined in Chapter 6, the critical levels do not have to account for a genome scan but just for the chromosomal segment that is being investigated (LANDER and KRUGLYAK 1995). Also if an experimenter is only testing a part of the genome (e.g. one chromosome) for only one trait, and has no intention of further testing they should only account for the multiple testing that they have undertaken.

LANDER and KRUGLYAK (1995) proposed four levels of significance: *suggestive linkage* – statistical evidence that would be expected to occur one time at random in a genome scan; *significant linkage* – statistical evidence expected to occur 0.05 times in a genome scan; *highly significant linkage* – statistical evidence expected to occur 0.001 times in a genome scan; *confirmed linkage* – significant linkage confirmed in a subsequent independent study. These recommendations were met with some resistance (WITTE *et al.* 1996; CURTIS 1996) primarily on the basis that these authors thought that everybody should be able to interpret the genome-wide significance of nominal p-values.

Chapter 2 demonstrates the use of threshold levels that were calculated accounting for repeated testing over the genome, and for different traits, although not for a dense marker map. Adjustment to a dense marker map was not calculated, as an empirical distribution of the null hypothesis was being used that accounts for the characteristics of the marker data. Therefore deterministic adjustments, as used on theoretical null hypothesis distributions, could not be used. However, one could have simulated a very dense map with all the other characteristics of the experiment, but

this was not undertaken. On the basis of LANDER and KRUGLYAK'S (1995) guidelines one QTL was identified at the significant linkage level in Chapter 2. The 5% experimentwise significance level in Chapter 2 was equivalent to a nominal significance level of 0.00014.

The stringent threshold level will control the type I error rate, but it will also reduce the power (1–type II error rate) of the experiment. The loss of power from accounting for repeated testing has motivated two other methods to be proposed for calculation of significance levels in QTL detection: false discovery rate (BENJAMINI and HOCHBERG 1995; WELLER *et al.* 1998) and posterior type I error (SOUTHEY and FERNANDO 1998). Both techniques are based on controlling the number of false positives in the rejected null hypotheses.

False discovery rate: BENJAMINI and HOCHBERG (1995) define the false discovery rate (FDR) as "the expected proportion of true null hypotheses within the class of rejected null hypotheses", which is equivalent to the proportion of false positives in the tests deemed significant. WELLER *et al.* (1998) described the FDR as follows: "assume that m multiple comparisons are tested, and for each null hypothesis; H_1, H_2...Hm; a test statistic and the corresponding p-values; P_1, P_2...P_m are computed and ordered, $P_{(1)} \leq P_{(2)} \leq \ldots, \leq P_{(m)}$ for the respective null hypotheses $H_{(i)}$. If all null hypotheses are true, but k hypotheses are rejected, then the expectation of the number of hypotheses rejected should be approximately equal to the actual number of hypotheses rejected for any value of k. If in fact some of the null hypotheses are false, then the expectation of the number of hypotheses rejected should be less than k. The expectation of the number of hypotheses rejected assuming that all of the null hypotheses are true is $mP_{(k)}$. Defining $q = mP_{(i)}/i$, BENJAMINI and HOCHBERG (1995) prove that the FDR can be controlled at some level q*, by determining the largest i for which; $q^* = mP_{(i)}/i$. That is, out of the k hypotheses rejected, it is expected that the proportion erroneously rejected is no greater than q*".

Applying the FDR to chromosome 6 data used in Chapter 3, 480 hypothesis tests are undertaken (5 traits and map length of 95cM). The top ten p-values for chromosome 6 and their q-values are given in Table 1, and the q-values for the 480

hypothesis tests are given in Figure 1. For the most significant point the p-value is 0.0000044, and the expectation for the highest test statistic is mP (480 tests × 0.0000044) which is equal to 0.0021. This value is then divided by i (1) to calculate the q-value of 0.0021.

TABLE 1: The false discovery rate for the 5 traits analysed on chromosome *six*.

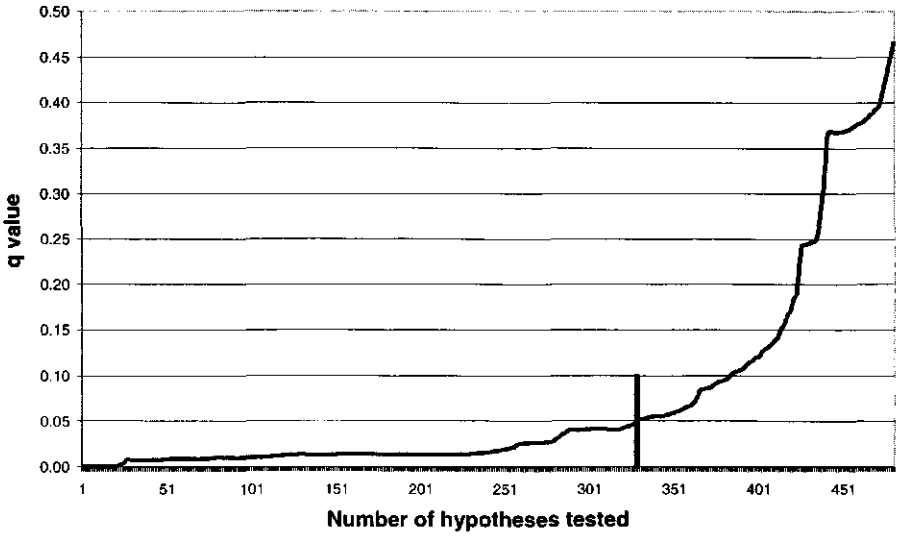
i	Trait	Position (cM)	Test statistic	p-value	Expectation[1]	q[2]
1	Protein %	13	3.17	0.0000044	0.0021	0.0021
2	Protein %	12	3.16	0.0000047	0.0022	0.0011
3	Protein %	11	3.15	0.0000050	0.0024	0.0008
4	Protein %	10	3.14	0.0000054	0.0026	0.0006
5	Protein %	14	3.13	0.0000057	0.0028	0.0006
6	Protein %	9	3.13	0.0000058	0.0028	0.0005
7	Protein %	8	3.11	0.0000063	0.0030	0.0004
8	Protein %	7	3.10	0.0000068	0.0033	0.0004
9	Protein %	6	3.09	0.0000073	0.0035	0.0004
10	Protein %	15	3.08	0.0000077	0.0037	0.0004

[1]Expectation for the number of hypotheses rejected under the null hypothesis.
[2]The expectation (mP) divided by the i^{th} ranked p-value.

Protein percent had the highest 26 test statistics (0-25 cM) for the five traits tested on chromosome *six*. Three hundred and thirty of the 480 hypothesis tests had a q value of less than 0.05. That is, if 330 of the null hypotheses are rejected, then it is expected that approximately 16 of them will be false. All of the five traits analysed were represented in the 330 null hypotheses that had q-values less than 0.05 (Figure 1). Therefore, using the FDR, one would expect that there was a QTL segregating for all of the five traits on chromosome *six*. However, from Chapter 2, there is little evidence for a QTL segregating for fat yield. The shortcoming of the FDR is that the correlated multiple hypothesis tests for protein percent (every 1 cM), had very high test statistics and therefore dominated the highest 100 test statistics. The expectations for the other traits were then divided by large i values, which resulted in small q-values. This shortcoming of the FDR is further accentuated when moving to a genome scan, and therefore combining results (p-values) from different chromosomes. Applying the FDR method to another chromosome, that has little evidence for QTL segregation, resulted in q values above 0.5 for most of the chromosome, and thus confirming little evidence of QTL segregation. However, when combining this

chromosome with chromosome *six*, the q-values were below 0.1 for segments of the chromosome. Chromosome *six* dominated the highest test statistics positions, and thus the expectations for results from the other chromosome were divided by large i values, resulting in low q-values for the chromosome that previously showed little evidence for QTL segregation.

FIGURE 1: The false discovery rate for the five milk production traits analysed on chromosome *six*.



WELLER *et al.* (1998) successfully demonstrated the application of the FDR on real data from a granddaughter design. However, in their case, single marker analysis (26 markers) was undertaken. It appears that the application of the FDR to genome analysis with interval mapping (highly correlated hypotheses tests) is limited. An approach to negate this problem may be to take the lowest p-value from each marker interval (J. WELLER personal communication). However, as more markers are added to the linkage map, the marker intervals will become smaller and the problem of highly correlated tests will re-occur. Dividing chromosome *six* into 4 equal intervals (approximately every 25 cM was identified as an independent test in this data set), and taking the highest test statistic from each interval, for each trait, resulted in very low q-values for the two percentage traits and milk yield. However, the q-values for the peak values for protein and fat yield were still below 0.2.

Posterior type I error: The method outlined by SOUTHEY and FERNANDO (1998) – termed here as the posterior type I error – is based on the null hypothesis of the absence of QTL in an interval flanked by markers. They state that this null hypothesis is more appropriate than the null hypothesis of no QTL in the genome (LANDER and BOTSTEIN 1989), which is inconsistent for traits that are heritable. The proportion of false positives (p_i) among significant results for interval i can be written as (based on MORTON (1955));

$$p_i = \frac{\alpha_i \Pr(H_{oi})}{\alpha_i \Pr(H_{oi}) + \Pr(H_{ai}) P_i}$$

where, for an interval i, α_i is the significance level, $\Pr(H_{oi})$ is the prior probability of the null hypothesis, $\Pr(H_{ai})$ is the prior probability of the alternative hypothesis, and P_i is the power of the test, averaging over all QTL locations within the interval (also outlined in Chapter five).

If m intervals are tested in a genome-wide scan for QTL, the genome wide proportion of false positives among significant results (p) is;

$$p = \frac{\sum_{i=1}^{m} \alpha_i \Pr(H_{oi})}{\sum_{i=1}^{m} (\alpha_i \Pr(H_{oi}) + \Pr(H_{ai}) P_i)}$$

If interval specific prior information is not available, the same values of the prior probabilities, significance level, and average power are used for each interval and p reduces to;

$$p = \frac{m\alpha \Pr(H_o)}{m(\alpha \Pr(H_o) + \Pr(H_a)\overline{P})}$$

and thus p does not depend on the number of tests undertaken.

SOUTHEY and FERNANDO (1998) demonstrated the use of the method on simulated backcross data. To calculate $Pr(H_a)$ they assumed that there were 10 QTL randomly placed in 150 intervals, and the prior probability of the null hypothesis ($Pr(H_o)$) being $1-Pr(H_a)$. In calculating statistical power it was assumed that the QTL were of equal size, there was no interference, and all QTL locations within an interval were equally likely. Application of this method is hindered by the assumptions that have to be made. The number of QTL segregating is unknown, as are QTL effects, and the power of detection is dependent on QTL size. Therefore, different significance levels (α) will be appropriate for different QTL sizes, if the posterior type I error is to be equal for all QTL sizes. As outlined in Chapter 5, the posterior type I error can be utilised in confirmation studies when there is more prior information on the probability of the null and alternative hypotheses, and QTL size for estimating the power of detection.

Conclusion: All three methods to set significance levels have their pitfalls, but the method of setting experimentwise levels (LANDER and BOTSTEIN 1989; LANDER and KRUGLYAK 1995) appears to be most applicable to genome scans with interval mapping.

It should be noted that there is no one correct method for setting threshold levels, and the two following quotes outline this: "...although the proposed genome-wide statistical significance criteria might appear to add objectivity to the evaluation of linkage, one must be cautioned that scientific inference is never objective" (WITTE *et al.* 1996) and, "...thresholds are not absurd – people who use them foolishly are" (LANDER and KRUGLYAK 1996).

In my opinion, the experimenter should set experimentwise significance levels (LANDER and BOTSTEIN 1989; LANDER and KRUGLYAK 1995) that account for the repeated testing over the genome, and for all traits investigated, as in Chapter 2. The experimenter must acknowledge that correcting for multiple testing in the experimentwise setting will reduce the power of detection. Therefore, one has to make a subjective judgement on what threshold to use. This will probably depend on the risk of making a type I error compared to the risk of making a type II error. I am

comfortable with relaxing the required significance level from the traditional 5% to some 20-30% experimentwise level, especially as QTL results should be replicated before MAS implementation or fine mapping projects start. The 20-30% experimentwise threshold is stricter than the suggestive linkage of LANDER and KRUGLYAK (1995) which rapidly asymptotes to about the 63% experimentwise level (exact asymptote value of $1-\frac{1}{e}$, where e is the exponential constant). The difference between significant linkage (5% experimentwise) and suggestive linkage (63% experimentwise), as suggested by LANDER and KRUGLYAK (1995), is quite large, and therefore I subjectively chose a slightly stricter type I error than that of suggestive linkage.

EXPERIMENTAL DESIGNS TO DETECT MORE QTL IN THE NEW ZEALAND DAIRY INDUSTRY

Significant genetic responses from computer simulated marker assisted selection in dairy cattle breeding schemes have been identified when a reasonably large percentage (e.g. 30%) of genetic variance is explained by markers (Chapter 7). The granddaughter and daughter designs both have medium to high power to detect QTL with large effects. It is unlikely that genetic variance is predominantly comprised of large sized QTL (SHRIMPTON and ROBERTSON 1988) and therefore these designs will not explain a large proportion of the genetic variance. Utilisation of the QTL that explain a small proportion of the genetic variance in marker assisted selection, will increase the rate of genetic gain and increase dairy industry returns, but the impact will not be large (Chapter 6). To identify more QTL, two prospective QTL experimental designs: selective DNA pooling to identify more within-breed QTL, and a Holstein-Friesian × Jersey cross to identify genetic differences (QTL) between breeds, are available to the New Zealand dairy industry (SPELMAN *et al.* 1998).

Selective DNA pooling: Selective DNA pooling is an extension of selective genotyping, with the DNA from the daughters from each tail of the distribution are pooled within sire (DARVASI and SOLLER 1994). The two pools are genotyped, and the marker allele frequencies are estimated for each pool. When there is a significant

difference in sire marker allele frequency, this indicates the presence of a linked QTL. Selective DNA pooling has been successfully applied in the Israeli dairy industry (LIPKIN *et al.* 1998).

The extensive use of elite bulls in the New Zealand dairy cattle population results in sires having up to 100,000 lactating daughters. The large family size has an impact on the power of detection, as the larger the half-sib family, the greater the power (Figure 2). The optimum percentage of daughters to select decreases as the size of the half-sib family increases. However, more daughters have to be sampled for the larger half-sib families. For example, for a family of 5,000 or 10,000 the optimum number of daughters to be selected from each tail is 400-500, whereas, for half-sib families of 50,000 or 100,000 the optimum number of daughters is 1,000 from each tail.

The power for a heterozygous half-sib family of 100,000 for an additive QTL effect of 0.15 σ_P is some 95% (Figure 2). To attain 95% power for a heterozygous sire with total genotyping, some 5,000 daughters would be required, and some 3,000 daughters with selective genotyping. For a genome scan with 200 markers this would require 400 (2 pools $\times$ 200 markers) genotypes for selective DNA pooling, 600,000 for selective genotyping and 1,000,000 for total genotyping. However, individual pools have to be formed for each trait analysed with selective DNA pooling. Thus for the three milk production traits (milk yield, milk fat and milk protein) and the two percentage traits (milk fat %, milk protein %), the required number of genotypes for pooling would increase 5-fold, but would still only be some 3% of that required with selective genotyping. It is likely that there will be statistical development in the analytical methods used for selective DNA pooling that will enable QTL effects to be estimated for traits correlated to the selectively pooled trait, as has occurred with selective genotyping (BOVENHUIS and SPELMAN 1998; JOHNSON *et al.* 1998).

Power of detection with selective DNA pooling is far superior to that of the designs that have been used to date; granddaughter and daughter design (Table 2). In addition, selective DNA pooling has reasonable power to detect small QTL (0.1 σ_P).

Figure 2: Power of selective DNA pooling for a heterozygous sire with a QTL effect of 0.15 σ_P (a in FALCONER terms) for different proportion of daughters selected and for different sized half-sib families.

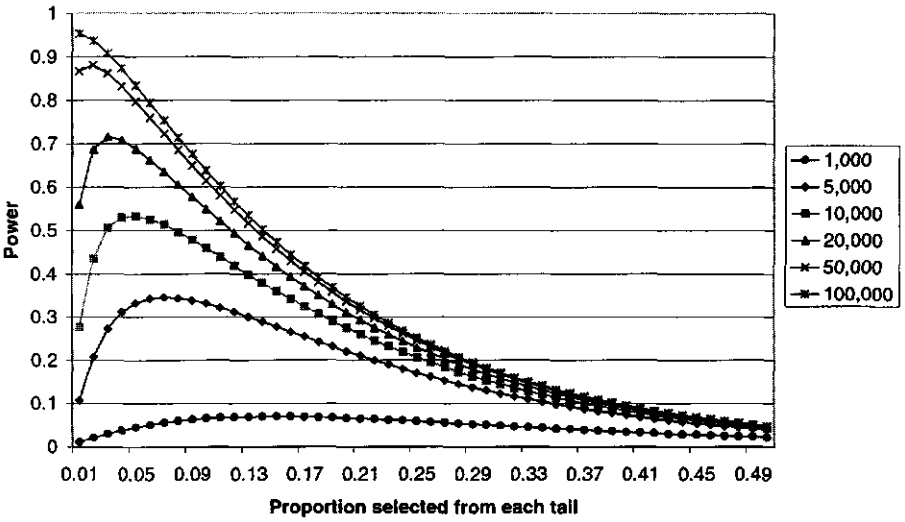


TABLE 2: Power of granddaughter, daughter, and selective DNA pooling designs for different QTL effects, nominal type I error of 0.001, trait heritability of 25%, 0.5 QTL allele frequency and 0.05 recombination rate between marker-QTL.

QTL effect[1]		Experimental design		
σ_P	σ_G	Gddtr	Dtr	Pooling
0.10	0.20	0.01	0.06	0.40
0.15	0.30	0.05	0.33	0.92
0.20	0.40	0.19	0.68	1.00
0.25	0.50	0.48	0.88	1.00
0.30	0.60	0.78	0.95	1.00
0.35	0.70	0.94	0.98	1.00
0.40	0.80	0.99	0.99	1.00

[1]QTL effect is equivalent to a in FALCONER terms.
Gddtr = granddaughter; thirty-two sires with an average of 35 progeny tested sons each.
Dtr = daughter; eight sires with an average of 800 daughters each with the top and bottom 25% of daughter selectively genotyped.
Pooling = Selective DNA pooling; 1 heterozygous sire with 75,000 daughters with the top and bottom 1% separately pooled.

The use of milk as the source of the DNA (LIPKIN *et al.* 1998) removes the cost of collecting blood from the identified daughters that could be spread through many hundreds of herds throughout New Zealand. Over 90% of the New Zealand dairy cattle are milk recorded and all of this milk is analysed at the one site. Equal amounts of DNA are required in the pool from each cow. This is achieved through the measurement of somatic cells in each milk sample (LIPKIN *et al.* 1998). The PCR reaction is undertaken directly on the pools of milk samples. There will be some error in estimation of allele frequencies from sources including unequal proportions of DNA from animals and inaccuracies in reading from densitometric intensities. LIPKIN *et al.* (1998) reported correlations of 0.88 to 0.94, between allele frequencies estimated through DNA pooling and from genotyping the individual animals comprising the pools. Another error introduced with selective DNA pooling is through parentage errors because individual samples are not analysed and the animals that fail parentage can not be removed from the analysis. This will have the effect of reducing the allele frequency differences and the estimated QTL effects. If the parentage error rate is similar in both pools, and not too large, the effect on the power of detection will be negligible, because it only reduces the effective number of animals that are in both pools.

Holstein-Friesian and Jersey QTL experiment: The primary objective of a QTL experiment with the two breeds would be to identify chromosomal regions that contribute to the genetic differences between the Holstein-Friesian and Jersey breeds. The design would first involve purebred Jersey and Holstein-Friesian animals being interbred to form F_1 bulls and cows. The F_1 animals will be heterozygous at all loci that are fixed with different allelic forms in the two breeds. Gametes produced by the F_1 animals will segregate the alternative forms of alleles. The options for the experiment are either a backcross (F_1 bull mated to HF and J cows) and/or F_2 experiment (F_1 bulls interbred with F_1 cows). In general more animals are required for the backcross design to have equivalent power to the F_2 design, and therefore the F_2 design is the preferred option (SOLLER *et al.* 1976).

Three possible scenarios that may occur when detecting differences between the two breeds need to be considered when calculating the power of a F_2 design;

i) The two breeds are homozygous for different allelic forms at the QTL loci, and marker haplotype origin can be identified unambiguously for the two breeds.

ii) The two breeds are not homozygous at the QTL and share QTL alleles in common, but at different frequencies, and marker haplotype origin can be identified unambiguously for the two breeds.

iii) The two breeds are not homozygous at the QTL and share QTL alleles in common but at different frequencies, and they have some marker allele sharing which reduces the ability to identify the breed origin for marker haplotypes.

Microsatellite markers are usually highly informative and therefore breed origin should be able to be ascertained for the markers. Furthermore, the development of new markers in the next five years should enable the selection of breed specific markers (M. GEORGES personal communication). Therefore the first two scenarios are used for power calculations.

The number of F_2 offspring required to attain 90% power (assumptions as in Table 2) increases as the degree of allele sharing increases between the two breeds (Table 3). For example, when the breeds are fixed for different alleles, 958 F_2 offspring are required to attain 90% power for a 0.2 σ_P QTL. A gene of 0.2 σ_P is some 4.5 kg for fat, 3.3 kg for protein, 105 litres for milk and 6.9 kg for live weight. The breed differences between Holstein-Friesian and Jersey are some 10 kg for fat, 17 kg for protein, 850 litres for milk and 89 kg for live weight. When the allele frequency in one breed is 90% and 10% in the other breed, 1497 F_2 offspring are required, and when the allele frequency in one breed is 80% and 20% in the other breed, 2662 F_2 offspring are required (Table 3).

TABLE 3: Required number of F_2 offspring to have 90% power for different QTL sizes and different degrees of QTL allele sharing between the two breeds.

	Degree of QTL allele sharing[1]		
QTL effect[2] (σ_P)	1:0	0.9:0.1	0.8:0.2
0.10	4732	7394	13145
0.15	1704	2662	4732
0.20	958	1497	2662
0.25	613	958	1703
0.30	426	665	1183
0.35	313	489	869
0.40	240	374	665

[1]The degree of allele sharing is the frequency of the allele in one breed and the frequency of the allele in the other breed. Note: the sum of the allele frequencies in the two breeds do not have to equal 1, but have arbitrarily been chosen in this way.
[2]The QTL effect is a in FALCONER terms.

Marker assisted selection has the potential to be utilised with crossbreeding in the New Zealand dairy industry. To identify the QTL that constitute the genetic differences between the two breeds, the F_2 design with approximately 1,000 female progeny would be very powerful.

Once the chromosomal regions have been identified in the two breeds, there is potential to generate crossbred individuals that have a combination of the best QTL alleles from the two breeds. If the two breeds are fixed for alternative QTL alleles at a locus, then marker information will not be helpful in the generation of F_1 animals, as they will be automatically be heterozygous at the QTL locus. However, crossing the F_1 animals to form F_2 progeny would benefit from marker information in endeavouring to produce homozygote animals for the favourable QTL allele. These crossbred individuals would probably be bulls to be progeny tested, and then the genes disseminated through semen to the population, or a proportion of it. Genes from crossbred females could be disseminated through cloning, if it is operational in the dairy industry in the future. There is also the potential that a new synthetic from the

two breeds with the help of markers could be produced or marker assisted introgression of QTL alleles from one breed to the other.

The greatest potential for MAS in the dairy industry is for traits that are not under efficient selection to date, i.e. not the milk production traits, but fertility, mastitis, and other health traits. These traits are not recorded well in the New Zealand dairy industry and in other countries, with the exception of the Nordic countries. Therefore it is very difficult to map QTL for these traits with the scarcity of reliable phenotypes. If the F_2 progeny were farmed at a small number of locations there is the potential to phenotype the animals for many traits and identify QTL for these traits. However, the breeds may not be as divergent for these traits as they are for the production traits. The QTL that are found from the F_2 cross may also give indications of segregating QTL loci within breed. TAYLOR *et al.* 1998 reported that from a cross between Brahman and Angus, a breed specific difference for one trait was identified at a chromosomal region. When analysing the trait within breed, the same location was identified as segregating for another trait, which was not detected in the breed specific analysis. It also should be noted that the breed, which is inferior for a trait, might have alleles segregating at loci that are better than those alleles in the superior breed (cryptic alleles). For example, the Jersey breed has lower live weight than Holstein-Friesian, but for a locus that affects live weight, the allele that increases it may come from Jersey. This phenomena has been reported by TANKSLEY (1996) for tomatoes, DE KONING (1998) for pigs and was also outlined by SAX (1923; described in Chapter 1)

Selective DNA pooling and a F_2 design involving Holstein-Friesian and Jersey are QTL experimental designs that could utilise unique aspects of the New Zealand dairy industry; large half-sib families and crossbreeding. Selective DNA pooling can be applied to traits that are routinely collected as part of the national animal evaluation, whereas the F_2 design can also be used for traits that are not routinely measured. These two QTL experimental designs have the potential to identify large proportions of the genetic variance, which will enable marker assisted selection to have a considerable impact on the New Zealand dairy breeding scheme.

MARKER ASSISTED SELECTION

Genetic response: Many authors have looked at the implications of MAS on dairy cattle breeding programmes through simulation. The results from these simulations are near unanimous in that MAS can increase the rate of genetic gain, but the degree of improvement in genetic gain is extremely variable. The following outlines some of the key assumptions in the simulations that are the major determinants of genetic response with MAS (SPELMAN 1998).

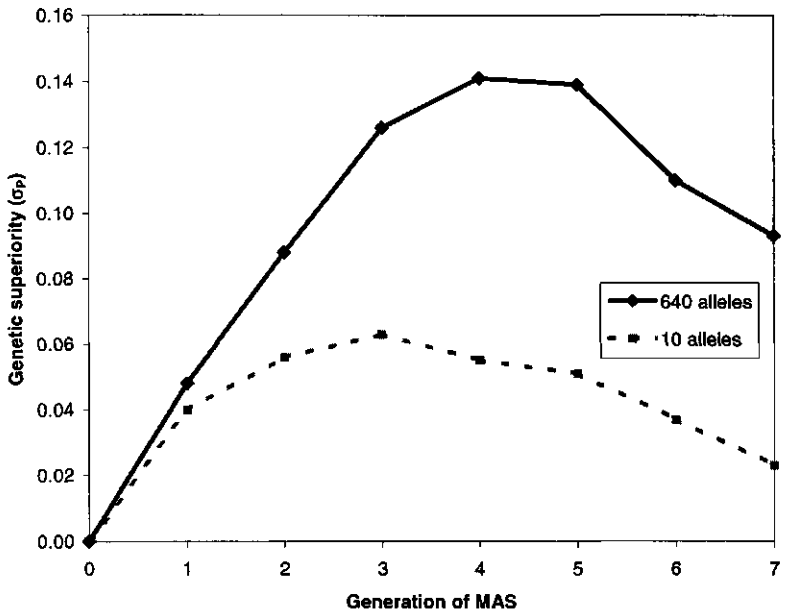
One element is the type of MAS scheme simulated. The majority of MAS simulations are either within-family MAS, where selection decisions are first made on conventional breeding values followed by within family decisions on QTL information, or across-family MAS, where selection decisions are made on breeding values that combine QTL and polygenic components (FERNANDO and GROSSMAN 1989). As previously outlined in this thesis across-family MAS is technically more demanding than within-family MAS, but genetically superior to within-family MAS. This is because the inclusion of QTL information in the breeding value estimation results in more accurate estimation and therefore higher selection differentials as differences between families can be exploited as well as within.

The simulated genetic model has a large effect on the resulting genetic improvement achieved with MAS. To-date the underlying genetic model is unknown i.e. the number of alleles, distribution of effects, and interactions between loci. This has led to differing genetic models being simulated in MAS studies: ranging from bi-allelic QTL to many alleles at the QTL. Authors have justified the simulation of many alleles at a QTL, by stating that it describes the possible situation where many QTL, each with a small number of alleles, are clustered together (e.g. MEUWISSEN and GODDARD 1996). The larger the number of alleles simulated at the QTL (assuming uniform frequency), the greater the superiority of MAS (Figure 3, adapted from Chapter 3, Tables 1 and 2).

A bi-allelic QTL under similar selection pressure will decrease in variance quicker again than that of the 10 allele QTL (Figure 3). This explains, in part, why authors such as MEUWISSEN and GODDARD (1996) who have simulated many alleles at

the QTL, have reported larger genetic gains than RUANE and COLLEAU (1995, 1996) and SPELMAN and GARRICK (1996), who both simulated bi-allelic QTL.

FIGURE 3: The effect of number of alleles simulated at a QTL, which explains 5% of phenotypic variance, on the cumulative genetic superiority of MAS over a breeding scheme with no knowledge of the QTL.



The population structure simulated also influences the MAS genetic response. The more animals genotyped in each generation and the more generations with genotypic information increases the accuracy of estimation of QTL effects and therefore MAS superiority (MEUWISSEN and GODDARD 1996; Chapter 3). Prior generations of animals that have genotypic information for the QTL of interest should be available in practice as semen from bulls is now retained and stored by most AI organisations and to a lesser degree, blood (or DNA) samples from bull-dams.

The number of years or generations that MAS is undertaken affects the genetic response. Superiority of MAS over breeding schemes without marker knowledge decreases over generations as the variance at the QTL decreases (Figure 3). The largest increase in rate of genetic gain in Figure 3 is in the first generation. Some

studies have reported responses for a one off selection step with MAS (KASHI *et al.* 1990); others have reported equilibrium response not accounting for the reduction in QTL variance (MEUWISSEN and VAN ARENDONK 1992; Chapter 7). The aforementioned genetic responses have been greater than the studies where authors have used MAS over many generations and have accounted for the reduction in QTL variance (RUANE and COLLEAU 1995, 1996; SPELMAN and GARRICK 1997).

One must be aware that MAS genetic responses are strongly influenced by simulation parameters such as genetic model, population structure and the number of generations that MAS is evaluated over. In truth, using different 'validated' assumptions one can objectively simulate a wide variety of genetic responses to MAS.

Implementation of MAS in the New Zealand dairy industry: Livestock Improvement Corporation is implementing MAS in its breeding scheme in 1998. Currently, four chromosomal areas that affect milk production have been identified and confirmed in the Livestock Improvement/Holland Genetics QTL experiment. Protein yield is the most important trait in the New Zealand dairy industry. The relative economic value of protein is about 6 times more than that of the second most important trait, fat yield. Therefore, to successfully implement MAS (economically), with the current state of knowledge, the MAS programme will focus on protein yield. Of the 4 chromosomal areas identified, one or two of these regions influence protein yield and thus are applicable to MAS. However, the power of the granddaughter design is not high, and there is some suggestion in the families genotyped in the daughter design, that protein yield could be affected by all of the identified chromosomal regions and, therefore all 4 regions will be further investigated in the MAS procedure.

Across-family MAS would be the most genetically beneficial MAS scheme to implement, but this could not be easily accomplished in the New Zealand dairy scheme. One reason for this is that the bull-dam population is spread throughout the commercial cow population and not in a central nucleus. In a central nucleus, genotype information can be routinely collected, and through the use of reproductive technologies each bull-dam will have a reasonable number of full and half –sibs, and

progeny that will be genotyped. The large number of relationships in the nucleus (dependent on how open the nucleus is) will be beneficial in estimating marker-QTL linkage phase and QTL allelic effects. It is also easy to accumulate marker information over generations on the elite breeding stock, which will improve the genetic response from MAS. In a nucleus breeding scheme where stock are owned by the breeding organisation, they also own the genotype information on the animals of interest. Breeding values that incorporate marker information can be estimated by the breeding organisation and they would not be under any obligation to publish the new estimate of genetic merit. In New Zealand the bull-dam population is owned by the farmers, which raises issues over who would own the genotype information on the animals and their new breeding values if estimated, and would they have to be publicly released? If released, then the breeding organisation that has undertaken the genotyping loses its advantage over competitors. In addition, there are also technical issues to resolve with regard to missing marker data before implementing QTL information in a national breeding value evaluation procedure that would allow across-family MAS.

The MAS scheme that is most applicable to Livestock Improvement, at the moment, is "bottom-up" (MACKINNON and GEORGES 1998; Chapter 6). In this MAS scheme, sire of sons and their progeny test daughters are genotyped for the identified areas of interest, and if the sire is heterozygous at a locus then only the sons carrying the favourable allele are progeny tested. Reproductive technology must be used for bottom-up MAS to be economically profitable (Chapter 6).

The New Zealand dairy production system is primarily based around seasonal calving to enable efficient utilisation of the seasonal pasture growth. The mating season is undertaken primarily in October and November. To ensure implementation at this date, firstly, markers that flank the chromosomal areas of interest need to be identified. Chapter 3 identified that QTL location error for MAS reduces the rate of genetic gain. Confidence intervals are constructed for each of the four chromosomal regions using bootstrapping methods (VISSCHER *et al.* 1996) to ensure (attempt at least) that the markers indeed do flank the QTL. At least two markers are placed at

each flanking boundary, to ensure that sires and daughters are informative for at least one of the flanking markers.

Sires that will be used to generate sons (sires of sons) in the 1998 mating season are identified based on the near complete daughter lactations that started in 1997. DNA from their daughters must be collected and sent to the genotyping service laboratory. Livestock Improvement already has the DNA at the laboratory because of the parentage testing (for sire) that they are conducting for all daughters in the progeny test scheme. The collection of DNA for parentage testing ensures that all of the daughters of the sires of sons are genotyped. If the DNA was collected at the end of the production season, when sires of sons are identified, there is the possibility that the lower producing daughters will have been culled. This selection bias would reduce the power of identifying whether the sire is heterozygous for the chromosomal regions (MACKINNON and GEORGES 1992).

The first analysis of the data will be simply contrasting the two groups of daughters that received different marker haplotypes from the sire. Only QTL for protein yield will be used in MAS, but analysis will be undertaken for all of the five milk production traits and also the non-production traits (e.g. farmer opinion, live weight, body and udder conformation) to identify correlated effects. To ensure that implementation is as easy as possible, as there are many new activities involved with MAS, probably only one of the sires of sons, for one heterozygous chromosomal area, will be used for MAS in the first year. This is assuming that there is a sire that is heterozygous for one of the four regions. In the future, MAS for more than one sire of son, and for more than one chromosomal region, is envisaged.

Chapter 6 identified that for bottom-up MAS to improve the rate of genetic gain, reproductive technology has to be applied on the bull-dams. About 90 sons need to be generated for the sire to have a 90% probability that 40 sons (required number for progeny testing) will have the desired QTL genotype. With *in vitro* fertilisation, two ovum pickups can be undertaken on a cow per week. One transferable embryo results from each pickup, on average (D. SELLARS personal communication). Assuming a 40% pregnancy rate, and 1:1 sex ratio, it will take, on average, 11 ovum

pickups per bull-dam to produce the 90 sons from 40 bull-dams. This requires on-farm ovum pickup for at least five weeks. The recipients for the embryos have to be synchronised to be available for transplanting fresh embryos. To reduce the number of embryos to be transplanted, and consequently the associated costs, the embryos will be genotyped before implantation, if the technology is available. If the technology is available, and the genotyping time requires that the *in vitro* embryos are frozen, the pregnancy rate will be reduced (TERVIT 1997).

The genetic and economic gains with the bottom-up MAS scheme are not large (Chapter 6), but the implementation of MAS will genetically and economically benefit the New Zealand dairy industry. The MAS implementation by Livestock Improvement is a very important starting point in which a number of logistic issues (reproductive technologies, genotyping) have to be resolved. The experience that Livestock Improvement will attain with this technology will enable it to be in the position to readily implement MAS in different types of breeding schemes when more QTL are identified.

Future MAS schemes: The future of MAS relies on the ability to identify further genetic variance. Two schemes: selective DNA pooling and a F_2 cross between the Holstein-Friesian and Jersey breeds have been described in this chapter, and if implemented will ensure more genetic variation is marked. Another possibility within the New Zealand population is the utilisation of the 12,000 progeny test daughters that are parentage tested every year. This resource will accumulate over years and will have pedigree links between years. If marker technology such as DNA chips (SOUTHERN 1996) enables large scale genotyping at low cost, this population could be a very good resource for further QTL detection. The pedigree structure would be appropriate to analytical methods that utilise all relationships (e.g. BINK 1998).

The assumed large number of loci with individually small effects (SHRIMPTON and ROBERTSON 1988) will probably prohibit the identification and marking of all of the genetic variation. This may lead to BLUP models that include QTL as random effects and then use allelic relationships based on marker information, instead of the

currently used additive genetic relationships, for the unmarked polygenic variation (JORGENSEN and JENSEN 1996; NEJATI-JAVAREMI *et al.* 1997).

Novel MAS schemes have been outlined such as velogenetics (GEORGES and MASSEY 1991) and whizzogenetics (HALEY and VISSCHER 1998). Velogenetics is based on harvesting oocytes from calves whilst still *in utero*. The harvested oocytes are matured and fertilised *in vitro* and transferred to recipients. This process can be repeated by harvesting oocytes from the resultant *in utero* calf for many generations and would reduce the generation interval to as little as 3 to 6 months (GEORGES and MASSEY 1991). The *in utero* calves may be selected on the basis of marker data for introgressing a gene into another genetic background. HALEY and VISSCHER (1998) suggest an enhancement to velogenetics, using nuclear transfer technologies that are currently being developed (CAMPBELL *et al.* 1996; WILMUT *et al.* 1997). Embryos are cultured *in vitro* and if selected on marker genotype, nuclear transfer the remaining embryonic cells into an enucleated oocyte to regenerate one or more of the desired embryos for transfer into recipient females. HALEY and VISSCHER (1998) state the need to harvest oocytes from calves *in utero* is a major difficulty, cost and potentially raises welfare issues. They proposed another scheme that has been termed whizzogenetics, which is based on nuclear transfer technologies. If one can imagine the technology will develop to a stage where cell differentiation can be controlled *in vitro*, then *in vitro* meiosis followed by fertilisation may become possible. In this case, the step requiring transfer to recipient female would become redundant. Cell cultures derived from fertilised oocytes could be selected using markers and then induced to undergo meiosis. This scheme would enable very rapid gene introgression aided by markers, and many generations of backcrossing could occur *in vitro*, and the final product could be grown into an animal. The viability of both of these MAS schemes will be dependent on technological advances and the strength of ethical objections.

As stated earlier in this chapter, there are currently technical issues to be resolved over the ownership of QTL genotype information and the estimated breeding values that incorporate marker information. Contracts with agreements regarding ownership of QTL information or purchasing the bull-dams will have to take place if

MAS is to be applied to both the sire and dam selection paths. This will lead to breeding populations becoming more protected and more nucleus type breeding programmes. Competitors will still be able to sample a company's QTL enhanced genetics via semen from proven bulls, but will be unable to distinguish allelic effects. If the breeding organisation regards the genetic superiority achieved to be very important, crossbreeding may be used to totally protect the genes from competitors that are breeding from purebred stock.

Genetic improvement from MAS is enhanced when young bulls are used on the sire selection paths. For MAS to have an appreciable impact this will have to occur. LANDE and THOMPSON (1990) stated that single genes of large effects may have deleterious pleiotrophic effects or be tightly linked to other genes with detrimental effects. These authors concluded that monitoring of the offspring's phenotype will have to continue. This would then stop the use of young bulls on the bull to cow path because progeny testing is required to observe phenotypes on a small number of daughters to ensure that there are no pleiotrophic or tightly linked deleterious effects. However, in my opinion, the progeny testing system struggles to detect deleterious autosomal recessive genes (e.g. BLAD), but is useful for the identification of deleterious dominant genes. The deleterious dominant genes will have a large negative economic impact if they are disseminated throughout the commercial cow population, and the risk of this occurrence will probably restrict the heavy use of young bulls on the bull to cow pathway. The use of marker assisted selected young bulls for the bull to bull path will be able to be implemented as only a small number of inseminations are undertaken and therefore the impact of deleterious genes is minor.

QTL detection and MAS in dairy cattle is primarily concentrating on milk production traits, as these are the traits that are conducive to QTL detection; routinely collected on large number of animals. It has been shown that the relative genetic response from MAS is greater for low heritability traits (e.g. fertility, mastitis) (SMITH 1967). As stated the identification of QTL for these traits is difficult due to the low level of recording and the low trait heritability. Emphasis in QTL mapping will move towards these traits in the future as the importance of secondary traits increases. If

QTL are found for these traits, there will be a greater need for these QTL to be in linkage disequilibrium with markers or marker haplotypes. This is based on the assumption of continuing low level of recording for these traits. If this is so, the QTL effects will not be able to re-estimated over time within a BLUP setting, due to lack of phenotypic measurements, unlike milk production traits. Traits such as specific milk characteristics (e.g. particular protein variants) may become more important in breeding programmes as manufacturers try to differentiate their products. The application of marker assisted introgression, possibly in a velogenetics setting, is a method that could be used to generate the desired proportion of the cow population with the particular variant.

These are some of the possible systems that I think MAS may be applied in the future. I am optimistic that as we identify and mark more genetic variation, and unravel the genetic model, MAS will become an integral component of dairy cattle breeding schemes.

ACKNOWLEDGEMENTS

This chapter has greatly benefited from the suggestions, ideas and corrections from DORIAN GARRICK, JOHAN VAN ARENDONK, HENK BOVENHUIS, MARCO BINK and PIM BRASCAMP.

LITERATURE CITED

BENJAMINI Y., and Y. HOCHBERG, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. **57(1):** 289-300.

BINK, M.C.A.M., and J.A.M. VAN ARENDONK, 1998 Detection of quantitative trait loci in outbred pedigrees with incomplete marker data. (submitted to Genetics).

BOVENHUIS, H., and R. J. SPELMAN, 1998 Selective genotyping to detect QTL for multiple traits in outbred populations. Proceedings of the 6th World Congress on Genetics applied to Livestock Production, Armidale, Australia **26:** 241-244.

CAMPBELL, K.H.S., J. MCWHIR, W. A. RITCHIE, I. WILMUT, 1996 Sheep cloned by nuclear transfer from a cultured cell line. Nature **380:** 64-66

CHURCHILL, G. A., and R. W. DOERGE, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963-971.

CURTIS, D., 1996 Genetic dissection of complex traits-correspondence. Nature Genetics **12:** 366-367.

DARVASI, A., and M. SOLLER, 1994 Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics **138:** 1365-1373.

DE KONING, D-J., L.L.G. JANSS, A. RATTINK, P.A.M. VAN OERS, B.J. DE VRIES, *et al.* 1998 Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (*Sus scrofa).* (submitted to Genetics).

DOERGE, R. W., and G. A. CHURCHILL, 1996 Permutation test for multiple loci affecting a quantitative character. Genetics **142:** 285-294.

DOERGE, R. W., and A. REBAI, 1996 Significance thresholds for QTL interval mapping tests. Heredity **76:** 459-464.

DUPUIS, J., 1994 Statistical problems associated with mapping complex and quantitative traits from genomic mismatch scanning data. Ph.D thesis, Stanford University.

FEINGOLD, E., P. O. BROWN, and D. SIEGMUND, 1993 Gaussian models for genetic linkage mapping using complete high-resolution maps of identity by descent. Am. J. Hum. Genet. **53:** 234-251.

FERNANDO, R. L., and M. GROSSMAN, 1989 Marker-assisted selection using best linear unbiased prediction. Genet. Sel. Evol. **21:** 467-477.

GEORGES, M., and J. M. MASSEY, 1991 Velogenetics, or the synergistic use of marker assisted selection and germ-line manipulation. Theriogenology **35(1):** 151-159.

HALEY C. S. and P. M. VISSCHER, 1998 Strategies to utilize marker-QTL associations. J. Dairy Sci. (in press).

JOHNSON, D. L., R. C. JANSEN, and J.A.M. VAN ARENDONK, 1998 Mapping quantitative trait loci in a selectively genotyped outbred population using a mixture model approach. (submitted to Genetical Research)

JORGENSEN, J. N. and P. JENSEN, 1996 Genome relationship in breeding value estimation. 47th meeting European Association for Animal Production.

KASHI, Y., E. HALLERMAN, and M. SOLLER, 1990 Marker assisted selection of candidate bulls for progeny testing programmes. Anim. Prod. **51:** 63-74.

LANDE, R., and R. THOMPSON, 1990 Efficiency of marker assisted selection in the improvement of quantitative traits. Genetics **124:** 743-756.

LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185-199.

LANDER, E. S., and D. BOTSTEIN, 1994 Corrigendum. Genetics **136:** 705.

LANDER, E., and L. KRUGLYAK, 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genetics **11:** 241-247.

LANDER, E., and L. KRUGLYAK, 1996 Genetic dissection of complex traits-correspondence. Nature Genetics **12:** 367-368.

LIPKIN, E., M. O. MOSIG, A. DARVASI, E. EZRA, A. SHALOM, *et al.* 1998 Mapping loci controlling milk protein percentage in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers. Genetics (in press).

MACKINNON, M. J., and M.A.J. GEORGES, 1992 The effects of selection on linkage analysis for quantitative traits. Genetics **132:** 1177-1185.

MACKINNON, M. J., and M. GEORGES, 1998 Marker-assisted preselection of young dairy bulls prior to progeny testing. Livest. Prod. Sci. **54:** 229-250.

MEUWISSEN, T.H.E., and M. E. GODDARD, 1996 The use of marker haplotypes in animal breeding schemes. Genet. Sel. Evol. **28:** 161-176.

MEUWISSEN, T.H.E., and J.A.M. VAN ARENDONK, 1992 Potential improvements in rate of genetic gain from marker-assisted selection in dairy cattle breeding schemes. J. Dairy Sci. **75:** 1651-1659.

MORTON N. E., 1955 Sequential test for the detection of linkage. American Journal of Human Genetics **7:** 277-318.

NEJATI-JAVAREMI, A., C. SMITH, and J. P. GIBSON, 1997 Effect of total allelic relationship on accuracy of evaluation and response to selection. J. Anim. Sci. **75:** 1738-1745.

REBAI, A., B. GOFFINET, and B. MANGIN, 1994 Approximate thresholds of interval mapping tests for QTL detection. Genetics **138:** 235-240.

RUANE, J., and J. J. COLLEAU, 1995 Marker assisted selection for genetic improvement of animal populations when a single QTL is marked. Genet. Res. **66:** 71-83.

RUANE, J., and J. J. COLLEAU, 1996 Marker-assisted selection for a sex-limited character in a nucleus breeding population. J. Dairy Sci. **79:** 1666-1678.

SAX, K., 1923 The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics **8:** 552-560.

SHRIMPTON, A. E., and A. ROBERTSON, 1988 The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. II. Distribution of third chromosome bristle effects within chromosome sections. Genetics **118:** 445-459

SMITH, C., 1967 Improvement of metric traits through specific genetic loci. Anim. Prod. **9:** 349-358.

SOLLER, M., T. BRODY, and A. GENIZI, 1976 On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor. Appl. Genet. **47:** 35-39.

SOUTHERN, E. M., 1996 DNA chips: analysing sequence by hybridisation to oligonucleotides on a large scale. Trends Genet. **12(3):** 110-115.

SOUTHEY, B .R. and R. L. FERNANDO, 1998 Controlling the proportion of false positives among significant results in QTL detection. Proceedings of the 6th World Congress on Genetics applied to Livestock Production, Armidale, Australia **26:** 221-224

SPELMAN, R. J., 1998 Major factors in marker-assisted selection genetic response in dairy cattle populations. Proceedings of the 6th World Congress on Genetics applied to Livestock Production, Armidale, Australia **26:** 365-368.

SPELMAN, R. J., and D. J. GARRICK, 1997 Utilisation of Marker Assisted Selection in a commercial dairy cow population. Livest. Prod. Sci. **47:** 139-147.

SPELMAN, R. J., N. LOPEZ-VILLALOBOS, and D. J. GARRICK, 1998 Experimental designs for quantitative trait loci detection in the New Zealand dairy industry. Proceedings New Zealand Society of Animal Production. **58:** 6-9.

TANKSLEY, S. D., S. GRANDILLO, T. M. FULTON, D. ZAMIR, Y. ESHED, *et al.* 1996 Advance backcross QTL analysis in a cross between elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor. Appl. Genet. **92(2):** 213-224.

TAYLOR, J. F., L. L. COUTINHO, K. L. HERRING, D. S. GALLAGHER JR., R. A. BRENNEMAN, 1998 Candidate gene analysis of GH1 for effects on growth and carcass composition of cattle. Anim. Genet. (in press).

TERVIT, H. R., 1997 *In vitro* production of cattle embryos. In: Milk composition, production and biotechnology. CAB International, Wallingford. pg 341-355.

VAN KAAM, J.B.C.H.M., J.A.M. VAN ARENDONK, M.A.M. GROENEN, H. BOVENHUIS, A.L.J. VEREIJKEN, *et al.* 1998 Whole genome scan for quantitative trait loci affecting body weight in chickens using a three generation design. Livest. Prod. Sci. **54:** 133-150.

VISSCHER, P. M., R. THOMPSON, and C. S. HALEY, 1996 Confidence intervals in QTL mapping by bootstrapping. Genetics **143:** 1013-1020.

WELLER, J. I., J. Z. SONG, D. W. HEYEN, H. A. LEWIN, and M. RON, 1998 A new approach to the problem of multiple comparisons for detection of quantitative trait loci. Proceedings of the 6th World Congress on Genetics applied to Livestock Production, Armidale, Australia **26:** 229-232

WILMUT, I., A. SCHNIEKE, J. MCWHIR, A KIND, and K.H.S. CAMPBELL, 1997 Viable offspring derived from fetal and adult mammalian cells. Nature **385:** 810-813.

WITTE, J. S., R. C. ELSTON and N. J. SCHORK, 1996 Genetic dissection of complex traits-correspondence. Nature Genetics **12:** 365-366.

SUMMARY

This thesis focuses on the detection of quantitative trait loci (QTL) and the potential genetic response when implemented in a marker assisted selection (MAS) dairy cattle breeding scheme.

In Chapter 2, analysis of chromosome *six*, the first chromosome to be genotyped in the Livestock Improvement/Holland Genetics QTL experiment, was undertaken. Twenty Dutch Holstein-Friesian families, with a total of 715 sires, were evaluated for marker-QTL associations for five traits; milk, fat and protein yield and fat and protein percent. Across-family analysis was undertaken using multimarker regression principles. The regression procedure was extended to fit two QTL on the same chromosome, which itself could be easily extended to fit co-factors on the same and/or different chromosomes. The permutation test was used to calculate critical values, and its application to multiple correlated traits was studied. Experimentwise critical values, which accounted for evaluating marker-QTL associations on all 29 autosomal bovine chromosomes and for five correlated traits, were calculated. A QTL for protein percent was identified in the one and two QTL models and was significant at the 1 and 2% level, respectively. This QTL was in a similar position to that previously reported by MICHEL GEORGES and co-workers.

Through the analysis of chromosome *six* it was observed that the degree of precision in estimating QTL location and size (or variance) was poor. In Chapter 3, the effect of inaccurate estimation of QTL variance and location on the genetic response to MAS was studied by stochastic simulation of an adult multiple ovulation and embryo transfer nucleus breeding scheme. Over-estimation of variance of the QTL had minimal effect on genetic gain for MAS in the short term, but decreased long-term response. The long-term loss was reduced when variance of the QTL was re-estimated after four generations of MAS. Selection for favourable alleles at a non-existent QTL resulted in first generation losses of 3 and 7% for postulated QTL explaining 5 and 10% of phenotypic variance, respectively. The larger the degree of error in QTL location, the larger was the genetic loss compared with the correct location scenario. For the largest simulated location error of 15 cM, genetic superiority of MAS was reduced by 80% in the first generation. It was concluded from this chapter that studies should be undertaken to verify estimates of QTL and location to enable optimal use of MAS.

The question arose 'how close should one get to the QTL before using the QTL in a MAS scheme?', or equivalently, 'what is the extra genetic benefit of getting close to the QTL?'. Using the same stochastic model as that in Chapter 3, the effect of flanking QTL-marker bracket size on genetic response to MAS in an outbred population was studied (Chapter 4). In addition, genetic response with MAS from two QTL, on the same and different chromosome(s), was investigated. Having smaller flanking QTL-marker brackets increased the genetic response from MAS. Moving from a 15 cM flanking QTL-marker bracket size to a 2 cM bracket approximately doubled the percentage increase in genetic gain from MAS over the control, for a QTL that explained 5% of the phenotypic variance. This was due to the greater ability to trace QTL transmission from one generation to the next with the smaller flanking QTL-marker bracket. Accurately tracing QTL transmission enabled more phenotypic records to be recorded for each QTL allele and increase the accuracy of estimation of the QTL allelic effects. Greater negative covariance between effects at both QTL was observed when two QTL were located on the same chromosome compared to different chromosomes. Genetic response with MAS was greater when the QTL were on the same chromosome in the early generations and greater when they were on different chromosomes in the later generations of MAS.

Chapter 3 concluded that QTL results should be confirmed before implementation in a MAS breeding scheme. Chapter 5 outlined and discussed two strategies to reduce uncertainty in the results from QTL studies. The first strategy was to combine p-values from multiple QTL experiments to confirm QTL results, and the second strategy was to establish a confirmation study. The size and structure of a confirmation study is dependent on the power of the design, which must be high to ensure that the postulated QTL can be verified. The chapter outlined the many issues that have to be addressed in the calculation of the experimental power; size of the quantitative trait loci to be detected, significance level required, experimental design and expected heterozygosity for the design.

Once the QTL are verified, MAS can be implemented. Chapters 6 and 7 investigated the genetic response from MAS in two different settings. Chapter 6 investigated MAS as it could be applied today, with the knowledge that we currently have available. Chapter 7 took a futuristic outlook and assumed that a large

proportion of the genetic variance will have been identified and the QTL will be in linkage disequilibrium with markers.

Chapter 6 evaluated through stochastic simulation, two MAS schemes that utilise QTL information to pre-select progeny test bulls within a family. The two within-family MAS schemes were: 'top down' MAS which identifies heterozygous sires for a locus of interest based on the granddaughter design, and uses the QTL information in the pre-selection of grandsons entering progeny test, and the 'bottom up' MAS scheme which identifies QTL heterozygous sires based on the daughter design, and uses the information in the pre-selection of sons entering progeny test. Bottom-up had greater genetic and economic responses than the top down scheme. The genetic response from the two MAS schemes was dependent on the reproductive performance of the bull-dams. The bottom up scheme increased the rate of genetic gain by 1.5, 3.5 and 5% for 1, 3, and 40 progeny per bull-dam, respectively. The maximum economic response was US$500,000 over a 20-year time horizon, from the bottom-up scheme with 40 progeny per bull-dam. When the top down scheme was used on the maternal path and the bottom up scheme on the paternal path, increases were 9% with 40 progeny per bull-dam. Chapter 6 concluded that the use of reproductive technologies on bull-dams was imperative to prevent gains from MAS being eroded by the loss in polygenic selection differential, which results when more bull-dams are required to enable pre-selection of sons using markers.

Three MAS schemes were evaluated through deterministic simulation for the futuristic situation presented in Chapter 7. The three MAS breeding strategies evaluated were: a progeny test scheme with markers (BMARK); a progeny test scheme with markers and unproven bulls allowed on the bull to bull selection path (YBULL); and a breeding program with markers where cows without lactation information and bulls without progeny information were eligible for selection (OPEN). On average, the percentage increase in genetic gain from MAS for the OPEN scheme was twice that for the YBULL scheme, and the YBULL increases were two and a half times greater than those for the BMARK scheme. The results showed that breeding companies must be willing to change the structure of the breeding scheme to fully gain the benefits of identified loci especially when a medium to large proportion of the genetic variance is marked.

In the general discussion, three different approaches to calculating critical values for QTL analysis were outlined and discussed. It was concluded that the approach used in Chapter 2 was more applicable and appropriate than the two alternatives. Secondly, two experimental designs were described and evaluated for the detection of QTL in the New Zealand dairy industry. It was shown that selective DNA pooling has very high statistical power to identify a large proportion of the within-breed genetic variation when applied to the large half-sib families that exist in the New Zealand dairy industry. Furthermore, a QTL experiment with 1000 F_2 progeny from a Holstein-Friesian and Jersey cross has high statistical power to identify QTL differences between the two breeds. QTL identified from this type of experiment have the potential to be used in a MAS setting that utilises the crossbreeding that is undertaken in the New Zealand dairy industry. Reasons for different simulated genetic responses from MAS for dairy cattle were outlined; genetic model, population structure simulated and the number of generations that MAS is evaluated over. The first implementation of MAS in the New Zealand dairy industry by Livestock Improvement is described and completing the thesis are some thoughts on how MAS will be applied in the future.

SAMENVATTING

Dit proefschrift richt zich op de detectie van genen die kwantitatieve kenmerken beïnvloeden (QTL) en op de potentiële genetische respons wanneer informatie over QTL wordt gebruikt bij de selectie in een melkvee-fokprogramma.

In hoofdstuk 2 wordt de analyse van chromosoom zes beschreven. Dit is het eerste chromosoom dat getypeerd is in het kader van het Livestock Improvement/Holland Genetics QTL experiment. Twintig Nederlandse Holstein-Friesian families, met in totaal 715 stieren, zijn voor vijf kenmerken op merker-QTL associaties onderzocht: kilogram melk, vet en eiwit en het vet en eiwit percentage. Bij de analyses is informatie van verschillende merkers gelijktijdig gebruikt (multi-merker analyse) en is het bewijs voor de aanwezigheid van een QTL geaccumuleerd over de verschillende families. De regressie procedure is uitgebreid om te kunnen onderzoeken of er zich mogelijk twee QTL op chromosoom zes bevinden. De gebruikte methode kan eenvoudig uitgebreid worden naar situaties waarin co-factors op hetzelfde of op verschillende chromosomen worden meegenomen. De permutatie test is gebruikt om kritische waarden uit te rekenen. De toepassing daarvan in een situatie met meerdere kenmerken is onderzocht. Kritische waarden voor het experiment zijn berekend door rekening te houden met 29 autosomale runder-chromosomen en vijf gecorreleerde kenmerken. Met behulp van één- en twee-QTL modellen is een QTL gevonden met een effect op het eiwit percentage. De significantie van het effect was 1% voor het één-QTL model en 2% voor het twee-QTL model. Dit QTL lag op een vergelijkbare positie als dat van een eerder door MICHEL GEORGES en medewerkers beschreven QTL.

De analyse van chromosoom zes wijst erop dat de positie van een QTL niet bijzonder nauwkeurig kan worden bepaald. Hetzelfde geldt voor de grootte van het effect dat het QTL heeft (of de variantie verklaard door het QTL). In hoofdstuk 3 is het effect van deze onnauwkeurige schatting van QTL locatie en variantie op de genetische respons van merker-ondersteunde-selectie onderzocht. Dit is gebeurt door middel van de stochastische simulatie van een nucleus fokprogramma waarin gebruik word gemaakt van multipele ovulatie en embryo transplantatie technieken. Overschatting van de QTL-variantie heeft een minimaal effect op de korte termijn genetische vooruitgang van merker-ondersteunde-selectie. Op de langere termijn

neemt de extra genetische vooruitgang tengevolge van merker-ondersteunde-selectie af. De lange termijn verliezen worden gereduceerd wanneer de QTL variantie na vier generaties van merker-ondersteunde-selectie opnieuw geschat word. Selectie op een niet bestaand QTL resulteert in de eerste generatie in een verlies van 3% wanneer in de fokwaardeschatting word aangenomen dat het QTL 5% van de fenotypische variantie verklaart. Het verlies is 7% wanneer in de fokwaardeschatting wordt aangenomen het QTL 10% van de fenotypische variantie verklaart. Wanneer een fout wordt gemaakt bij de schatting van de locatie van het QTL dan zullen de verliezen groter zijn wanneer een grotere fout wordt gemaakt bij de schatting van de locatie. In het geval van een gesimuleerde locatie fout van 15 cM wordt de genetische superioriteit van merker-ondersteunde-selectie in de eerste generatie met 80% gereduceerd. Op basis van dit hoofdstuk kan geconcludeerd worden dat voor een optimaal gebruik van merker-ondersteunde-selectie de schattingen van QTL variantie en locatie geverifieerd moeten worden aan de hand van vervolgonerzoek.

Hoe dicht moeten de merkers bij het QTL liggen voordat de QTL informatie in een merker-ondersteund-selectie programma gebruikt kan worden of anders geformuleerd, wat is het voordeel wanneer merkers beschikbaar zijn die dicht bij het QTL liggen. ? In hoofdstuk 4 wordt het effect van de lengte van het flankerende merker interval op de genetische respons van merker-ondersteunde-selectie in een melkveepopulatie bestudeerd. Hiertoe wordt gebruik gemaakt van hetzelfde stochastische model als in hoofdstuk 3. Ook is in hoofdstuk 4 de genetische respons met merker-ondersteunde-selectie voor twee QTL op hetzelfde en op verschillende chromosomen onderzocht. Kleinere flankerende merker intervallen verhoogen de genetische respons van merker-ondersteunde-selectie. Bij vergelijk van een flankerend merker interval met een lengte van 15 cM en 2 cM voor een QTL dat 5% van de fenotypische variantie verklaart verdubbelt het procentuele voordeel van merker-ondersteunde-selectie ten opzichte van de controle. Dit wordt veroorzaakt doordat voor een situatie met een kleiner flankerend QTL-merker interval, het in een groter aantal gevallen mogelijk is om het QTL te traceren van de ene generatie naar de volgende. Het nauwkeurig traceren van het QTL resulteert in meer fenotypische waarnemingen per QTL allel hetgeen resulteert in een nauwkeuriger schatting van de effecten van de QTL allelen. Wanneer twee QTLs op hetzelfde chromosoom liggen

dan is de negatieve covariantie tussen de beide QTLs groter dan wanneer ze op twee verschillende chromosomen gelokaliseerd zijn. De genetische respons met merker-ondersteunde-selectie is in de eerste generaties groter wanneer de QTLs op hetzelfde chromosoom liggen. In latere generaties is de respons echter hoger wanneer de QTLs op verschillende chromosomen liggen.

In hoofdstuk 3 is geconcludeerd dat de resultaten van QTL experimenten moeten worden bevestigd voordat de QTL informatie in een merker-ondersteunde-selectie-programma worden gebruikt. In hoofdstuk 5 worden twee strategieën om de onzekerheid omtrent de resultaten van QTL-studies te reduceren beschreven en bediscussieerd. De eerste strategie is om de p-waarden van meerdere QTL experimenten te combineren en de tweede strategie is het uitvoeren van een vervolgstudie. De omvang en de structuur van zo'n vervolg studie is afhankelijk van de statistische power. De statistische power moet groot genoeg zijn om er zeker van te zijn dat het veronderstelde QTL kan worden bevestigd. Hoofdstuk 5 beschrijft de vele zaken die in beschouwing genomen moeten worden bij de berekening van de statistische power van het experiment: grootte van het QTL effect, benodigde significantie niveau, ontwerp van het experiment en de verwachtte heterozygotie van het QTL.

Wanneer het QTL bevestigd is kan merker-ondersteunde-selectie worden geïmplementeerd. In de hoofdstukken 6 en 7 is de genetische respons van merker-ondersteunde-selectie in twee verschillende omstandigheden onderzocht. Hoofdstuk 6 handelt over merker-ondersteunde-selectie zoals deze heden ten dage toegepast zou kunnen worden. In hoofdstuk 7 is een futuristisch uitgangspunt gekozen en wordt aangenomen dat een groot deel van de genetische variantie geïdentificeerd is en dat de QTL in linkage disequilibrium zijn met de merkers.

In hoofdstuk 6 zijn twee merker-ondersteunde-selectie programma's geëvalueerd door middel van stochastische simulatie. In beide fokprogramma's wordt QTL informatie gebruikt om binnen families stieren te selecteren die na deze voor-selectie aan een nakomelingenonderzoek worden onderworpen. De twee binnen-familie merker-ondersteunde-selectie programma's kunnen worden aangeduid als 'top

down' en 'bottom up'. Top down merker-ondersteunde-selectie identificeert stieren die heterozygoot zijn voor een interessant locus op basis van het granddaughter design. De QTL informatie wordt gebruikt tijdens de voor-selectie van kleinzonen. Het bottom up merker-ondersteunde-selectie programma identificeert stieren die heterozygoot zijn voor het QTL gebaseerd op het daughter design. De informatie wordt gebruikt tijdens de voor-selectie van de zonen. Bottom up heeft een grotere genetische en economische respons dan het top down programma. De genetische respons van de twee merker-ondersteunde-selectie programma's zijn afhankelijk van de vruchtbaarheid van de stiermoeders. Het bottom up programma vergroot de genetische vooruitgang met 1,5 3,5 en 5% voor respectievelijk 1, 3 en 40 nakomelingen per stiermoeder. De maximale economische respons over een periode van 20 jaar voor het bottom up programma met 40 nakomelingen per stiermoeder is US$500 000. Bij gebruik van het top down programma voor het maternale pad en het bottom up programma voor het paternale pad en met 40 nakomelingen per stiermoeder is een toename in de genetische vooruitgang van 9% mogelijk. In hoofdstuk 6 wordt geconcludeerd dat merker-ondersteunde-selectie gecombineerd moet worden met het gebruik van reproductie technologieën bij stiermoeders. Wanneer dit namelijk niet gebeurt dan neemt de superioriteit van de geselecteerde stiermoeders af aangezien er bij een voor-selectie van stieren op basis van genetische merkers meer stiermoeder nodig zijn.

In hoofdstuk 7 zijn met behulp van deterministische simulatie een drietal merker-ondersteunde-selectie programma's geëvalueerd voor een mogelijke toekomstige situatie. De drie merker-ondersteunde-selectie strategieën die geëvalueerd zijn, zijn: een nakomelingen onderzoek programma met gebruik van merkers (BMARK); een nakomelingen onderzoek programma met gebruik van merkers en proefstieren die worden ingezet als stiervader (YBULL); en een fokprogramma met gebruik van merkers, waarin koeien zonder een eigen lactatie en stieren zonder een nakomelingen onderzoek selectiekandidaten zijn (OPEN). De procentuele toename van het merker-ondersteunde OPEN selectie programma is gemiddeld twee keer zo groot als dat voor het YBULL-programma. Voor het YBULL-programma is de genetische vooruitgang gemiddeld twee en een half keer zo groot als voor het BMARK programma. De resultaten geven aan dat fokkerij instellingen bereid moeten

zijn om de structuur van hun fokprogramma te veranderen om volledig profijt te kunnen trekken van de geïdentificeerde QTLs. Dit geldt vooral wanneer een middelmatig tot groot percentage van de genetische variantie gemarkeerd is.

In de algemene discussie worden drie verschillende benaderingen voor de berekening van kritische waarden voor een QTL analyse beschreven en bediscussieerd. De conclusie is dat de aanpak beschreven in hoofdstuk 2 beter geschikt is dan de dan de twee alternatieven. Vervolgens worden in de algemene discussie twee ontwerpen voor QTL detectie experimenten beschreven en geëvalueerd. Uitgangspunt is daarbij de Nieuw-Zeelandse melkveefokkerij. Er wordt aangetoond dat het selectief typeren van dieren in combinatie met het samenvoegen van DNA monsters (selective DNA pooling) een grote statistische power heeft om een groot gedeelte van de aanwezige genetische variatie te identificeren. Dit is voor een belangrijk deel toe te schrijven aan de aanwezigheid van grote half-sib families in de Nieuw-Zeelandse melkveepopulatie. Verder heeft een QTL experiment met 1000 F2 nakomelingen van een Holstein-Friesian en Jersey kruising een grote statistische power om QTLs te identificeren die verschillen tussen deze twee rassen. QTL die in dit experiment geïdentificeerd worden kunnen in potentie worden gebruikt voor merker-ondersteunde-selectie in de Nieuw Zeelandse kruisingspopulatie. In de algemene discussie worden verder redenen aangedragen voor de verschillen in genetische respons van merker-ondersteunde-selectie voor melkvee die wordt gevonden in simulatie studies. Mogelijke oorzaken voor de verschillen zijn het genetische model, de gesimuleerde populatie structuur en het aantal generaties waarover merker-ondersteunde-selectie is geëvalueerd. Ter afsluiting van dit proefschrift wordt eerste toepassing van merker-ondersteunde-selectie in de Nieuw Zeelandse melkvee fokkerij door Livestock Improvement beschreven en worden enkele gedachten gewijd aan de toekomstige toepassing van merker-ondersteunde-selectie.

ACKNOWLEDGEMENTS

Thanks to LIESBETH VAN DER WAAIJ and HENK BOVENHUIS for the Dutch summary.

CURRICULUM VITAE

Richard John Spelman was born on November 2nd 1968 in Palmerston North, New Zealand. I attended Awahou Primary School and then undertook my secondary school education at Saint Peters College, Palmerston North. In 1987, I started my tertiary education at Massey University, Palmerston North. I completed my Bachelors of Agricultural Science with honours in 1990, and commenced working for Livestock Improvement Corporation at the start of 1991 as a Livestock Analyst. After three years in this position, Livestock Improvement Corporation offered to sponsor me in furthering my education. I returned to Massey University in 1994 to undertake my Masters in Agricultural Science, which I completed with honours. On June 1st 1995, I started my PhD at Wageningen Agricultural University, The Netherlands in combination with Massey University, New Zealand. Upon completion of the PhD, I have undertaken a five-month post-doctoral period at the Wageningen Agricultural University, and I return to Livestock Improvement Corporation and start work as a scientist in October 1998.