STRATEGIES FOR APPLYING MARKER ASSISTED SELECTION IN NUCLEUS BREEDING SCHEMES IN DAIRY CATTLE

A Thesis

Presented to

The Faculty of Graduate Studies

of

The University of Guelph

by

ALESSANDRA STELLA

In partial fulfilment of requirements

for the degree of

Doctor of Philosophy

December, 2000

ABSTRACT

STRATEGIES FOR APPLYING MARKER ASSISTED SELECTION IN NUCLEUS BREEDING SCHEMES IN DAIRY CATTLE

Alessandra Stella
University of Guelph, 2000

Advisor:
Gerald B. Jansen

The general goal of this thesis was to use simulation to examine practical issues for marker assisted selection (MAS) of dairy cattle. Markers were used to select within-family the bulls to enter progeny testing from a nucleus herd. The first study evaluated the effects of considering a confidence interval of the position for a quantitative trait loci (QTL) versus only the probable genotype at the predicted site of the QTL. The location of the QTL was estimated by interval mapping with a granddaughter design. Accounting for the confidence interval increased the response in all scenarios. The average true breeding value (TBV) of the selected bulls was increased 2.60% when the confidence interval was used, versus 2.00% when only the predicted location was considered. No differences were observed with respect to how the confidence interval was estimated.

The second study compared strategies for repeated application of QTL detection and MAS. Twenty QTL and 300 markers were randomly distributed across 30 chromosomes. A daughter design was used, every generation, to determine the associations between marker and QTL alleles. Maximum response was achieved by strategies that selected upon several markers flanking multiple QTL. The mean TBV of

selected bulls was increased by up to 12% when multiple loci were considered, versus $\leq$7% when only the best marker was used.

The third study examined MAS when the selection goal included two traits. Trait 1 had an economic weight and heritability three times greater than trait 2. Multiple trait MAS was compared to applying MAS for trait 1 only and conventional selection alone. Multiple trait MAS decreased response for trait 1 relative to both single trait MAS or conventional selection. However, response for trait 2 increased to a greater degree and, therefore, response for the final index was greater. This result was consistent whether the traits were positively or negatively correlated.

The final study examined how different assumptions about the underlying genetic model affected the long-term response to MAS. Models differed in terms of mutation rate and distributions of allelic effects and frequencies. The use of MAS was beneficial regardless of the genetic model.

Acknowledgments

During the past 4 years at Guelph I learned a lot of things about animal breeding and genetics, but the two things I'll remember most are that to obtain a Ph. D. can at times be difficult, and that these hard times become much more manageable when one can rely on the help of others. For this support, I owe a great deal of thanks.

First, I need to thank my advisor Dr. Gerald Jansen for his guidance and friendship. I also need to thank the other members of my advisory committee, Doctors Mike Lohuis, John Gibson, and Giulio Pagnacco. The value of their guidance and assistance can neither be easily measured nor sufficiently expressed in words.

I'd like to formally recognize the tangible and intangible contributions of the CGIL and APS communities. I thank the other graduate students with whom I shared our knowledge, ignorance, success, failure, disappointment, appreciation and lamentation. I'm also grateful for the contribution of the staff at APS, who helped keep the computers running and the asbestos relatively well isolated from human contact.

I am indebted for the financial support I received from TOP-ROYAL of Italy. Without their help this whole experience would have never started.

My fantastic friends in Italy also need to be mentioned. Their continual flow of calls and emails helped me keep in touch with them and reality.

Finally, I think I owe the most gratitude to my family. They supported me in every possible way when I made my decision to move away from them and spend four years in a place that was only half way around the world but probably seemed to be on another planet. A mention of family would not be complete without including Paul, for always being there for me.

Table of Contents

List of Tables

Chapter 4.

Chapter 5.

List of Figures

List of Abbreviations

AFLP	Amplified fragment length polymorphism
AI	Artificial insemination
BLUP	Best linear unbiased prediction
CI	Confidence interval
cM	Centimorgan
DA	Di-allelic
DNA	Deoxyribonucleic acid
EBV	Estimated breeding Value
IBD	Identical by descent
IVP	In-vitro embryo production
MAS	Marker assisted selection
MOET	Multiple ovulation and embryo transfer
PA	Poly-allelic
QTL	Quantitative trait loci
RAPD	Randomly amplified polymorphic DNA
REML	Restricted maximum likelihood
RFLP	Restriction fragment length polymorphisms
SNP	Single nucleotide polyorphism
SSCP	Single stranded conformational polymorphism
TBV	True breeding Value

1. General Introduction.

Many studies have examined the possible benefits to be gained in selection accuracy and response by applying technologies such as the nucleus herd (e.g. Nicolas and Smith 1988) and marker assisted selection (MAS) (e.g. Meuwissen and Goddard, 1996). The dairy industry is now beginning to adopt these technologies. The general focus of this thesis was on the use of MAS in the dairy industry. Specifically, several practical issues about the application of MAS when selecting young bulls to progeny test within nucleus herd of dairy cattle were addressed. Kashi et al. (1990), Mackinnon and Georges (1997) and Spelman and Garrick (1998) have already laid the groundwork and examined many of the general issues related to the application of MAS within the dairy industry. This intention of this work was to build upon those studies. All of the investigation was performed by using computer simulation.

One of the major issues examined was how to use the information from quantitative trait loci (QTL) detection and location experiments to select animals most efficiently. This question was investigated for both granddaughter and daughter designs and single and multiple trait selection goals. Another issue examined was the sustainability of response in MAS programs over several generations, relative to conventional selection. Finally, the sensitivity of simulation results to different genetic models was examined.

This thesis is divided into eight chapters:

Chapter One is a general introduction that outlines the general focus of the thesis and briefly explains the contents of each of the remaining chapters.

Chapter Two specifically states the basic objectives of the four different experiments undertaken.

Chapter Three is a review of the previous literature that is most pertinent for the practical aspects of MAS in dairy cattle. In addition, Chapters Four through Seven each include an introduction that provides further detail about previously published work that pertains specifically to each study. Comments and comparisons about these studies are also included throughout the four chapters, primarily in the Discussion sections.

Chapter Four examines issues related to the application of results from a granddaughter design to MAS of young bulls. Specifically, the efficacy of using a confidence interval to account for uncertainty in the location of a single QTL was investigated. Different approaches to estimate this confidence interval are proposed, implemented, and compared.

Chapter Five compares strategies to use full genome scans for MAS. A daughter design is used for estimation of marker effects. The different strategies vary in the number of markers used and their respective genomic locations. The effectiveness of MAS over several generations is also examined in the context of genetic response and accuracy of marker-QTL associations.

Chapter Six applies the most effective approaches from Chapter Five to MAS with a two-trait selection objective. Two strategies of selection are proposed and response to each is compared to conventional and single-trait selection. The experiment is applied to selection goals with negatively and positively correlated traits and traits that affect fitness.

Chapter Seven compares maintenance of genetic variance and response to MAS for a variety of genetic models. Genetic models differ in terms of the distribution of allelic effects and frequencies, number of segregating loci and mutation rates. A model for which genetic variance is maintained for 100 generations is proposed and benefits of MAS in the long term are evaluated for this model.

Chapter Eight is a general discussion that briefly summarizes the key results and conclusions from each of the preceding four chapters and proposes new studies that could be done to build upon the work presented in this thesis.

2. Objectives

The aim of this study was to develop strategies for the integration of marker assisted selection (MAS) in a nucleus breeding scheme in dairy cattle. Previous studies in the application of MAS have, for convenience, been restricted to relatively simple designs for implementation. The objective of this study is to investigate how the benefits of MAS are affected when more complex applications are considered.

The major objectives were as follow:

1. Determination of the optimal information used to track the transmission of the potential quantitative trait locus (QTL). In particular, the study evaluated whether the efficacy of MAS, in situations where the position of the QTL is uncertain, could be improved by considering a confidence interval of QTL position.

2. Development and comparison of several strategies for long-term application of MAS to dairy cattle nucleus breeding schemes.

3. Analysis of the efficiency of MAS for a two trait breeding objective with different methods to account for marker information.

4. Evaluation of the effects of different genetic models on results of MAS studies.

3. Literature review.

3.1 Introduction.

Dairy livestock, like many other domesticated species, have undergone selection for a number of generations and for a variety of different traits of economic importance. Most of these traits are quantitative in nature, i.e. their continuous variation is assumed to be the result of the effects of an unknown but large number of genes and influenced by environmental factors. The great achievements produced by selection in recent decades have relied on phenotypic measurement and pedigree recording under the general framework of the *infinitesimal model* (Falconer and Mackay, 1996). Based on this model, statistical tools to predict individual breeding values accounting for the non-genetic nuisance factors have been fully developed and applied (i.e. Best Linear Unbiased Prediction, BLUP, Henderson, 1984). The infinitesimal model describes the individual genetic merit as the sum of small effects contributed by a very large (effectively infinite) number of genes equally spread throughout the genome. The model implies that each chromosome or genomic region has equal importance in determining the individual genetic superiority or inferiority and no special value can be associated to a specific segment of DNA. As a consequence, covariances between animals are fully described by simple pedigree relationships.

The polygenic model described above can be extended to a mixed inheritance model by including the effect of one or more major genes. Developments in molecular genetic techniques have made it possible to identify differences between individuals at the DNA level. In the last decade, several major collaborative projects have started to

map an increasing number of genetic markers in the genomes of different livestock species (e.g. Barendse et al., 1994; http://www.ri.bbsrc.ac.uk/pigmap/ecpigmap.html accessed May 2000; http://bos.cvm.tamu.edu/bovarkdb.html, accessed May 2000; and http://www.genome.iastate.edu/chickmap/, accessed May 2000). Such genetic markers have provided the ability to track the inheritance of linked segments of the genome in suitable pedigrees. By following a large number of markers spread approximately evenly throughout the genome, in appropriate pedigrees, it is possible to identify the major quantitative trait loci (QTL) influencing variation of a quantitative trait of economic relevance in a specific population.

Once a genomic region is detected where a QTL is segregating, marker-assisted selection (MAS) can be implemented to accelerate the genetic gain otherwise achievable through conventional selection programs. Information on individual genes with effects on quantitative traits can affect the genetic gain by affecting selection intensities, accuracy of selection and generation interval. In conventional breeding schemes the generation interval is increased by the need for phenotypic records for each individual. If the genes and their effects were known, typing of animals at DNA the level at an early age would make it possible to reduce generation interval. Moreover, BLUP methods do not account for the actual contribution of a parent to an offspring and, thus, full-sibs with pedigree information only will have the same estimated breeding values (EBV). In dairy cattle breeding, to reduce inbreeding and risk, a limited number of full sibs should be selected and this selection can only be at random until phenotypic data is obtained. If information on genes was available, it could be possible to differentiate among EBV of full sibs and increase selection accuracy (Stam, 1987; Kashi et al., 1990; Meuwissen and

Van Arendonk, 1992; Bovenhuis and De Boer, 1994). Because selection response is a function of the accuracy of selection, information on QTL will be particularly advantageous when accuracy is low or when used in novel stages or new pathways of selection (Gomez-Raya and Gibson, 1993). Selection for traits with low heritabilities or presenting other difficulties (e.g. expense for recording or sex-limited expression) can take major advantage of information from molecular genetics (Smith and Simpson, 1986).

Based upon observation of current populations, a reasonable assumption is that most of the traits of interest in animal breeding are controlled by a relatively large number of loci, each with a small effect, and only a small number of loci with a large effect (Shrimpton and Robertson, 1988). Therefore, traditional selection methods will probably not be replaced by molecular genetics, but an integration of the two selection methods could be beneficial to the selection response.

The benefits of molecular genetic information will not be limited to the use of genetic markers to aid selection within a population. Genetic markers can also be used to help introgress an interesting gene from one population to another (Groen and Smith, 1995). Results from QTL studies can also provide a better understanding of the mode of inheritance of traits of interest (e.g. the *callipyge* gene as described by Cockett et al., 1996). The final aim of identification and positioning of genes affecting quantitative traits is the cloning of the gene of interest (Oliver, 1996), to perhaps be transferred into the genome of another species.

3.2 Genetic markers

The ability to undertake large-scale QTL studies depends strongly upon the development of genetic markers, whose technology is moving at a very fast pace. Required characteristics of marker loci are: 1. being evenly distributed in the genome, in order to be able to produce complete maps, 2. being highly polymorphic and being co-dominant to provide the maximum information with respect to the diversity among individuals, 3. being neutral with respect to traits of interest and fitness (for QTL mapping), 4. being reliable and repeatable, and 5. being easy to automate and low cost (Falconer and Mackay, 1996). The information content of a marker is relevant within the context of its use and relates to the level of polymorphism (number and frequency of alleles) and the ease with which the allelic information can be retrieved, both in terms of identification and of assigning alleles to the loci.

Restriction fragment length polymorphisms (RFLP) were initially the predominant DNA markers. More recently, many other types of markers such as minisatellite, microsatellite, randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), single stranded conformational polymorphism SSCP and single nucleotide polymorphism (SNP) have been implemented and widely used. In the current maps for livestock, microsatellites are the prevalent markers. The marker density achieved to date in the cattle maps is approximately 1 marker per cM (http://bos.cvm.tamu.edu/bovarkdb.html, accessed May 2000). However, molecular studies on the human genome have suggested an average density of 1 polymorphism per 1000 base pairs (Chakravarti, 1999). If so, the potential exists to increase the marker density by about 1000 times in the livestock maps. Most polymorphisms are likely the result of

single base-pair changes in DNA sequence (SNP). In the near future, DNA arrays and DNA chips are expected to allow rapid genotyping of a large number of individuals for thousands of polymorphic loci (Visscher and Haley, 1995; Hacia, 1999).

3.3 Mapping QTL

Information from phenotypes and marker genotypes can be used to pinpoint chromosome areas responsible for part of the genetic variation for a given trait. Two main approaches have been applied to map QTL: analysis of candidate genes and genome scans.

The candidate gene strategy relies on identifying polymorphisms in genes with functions that relate directly to the trait of interest. Association between the candidate polymorphism and the trait is determined on the basis of population-level linkage disequilibrium, provided that the sample of genotyped animals is truly representative of the general population. The advantage of this strategy is that it directly exploits the association in the breeding population and that the segregating gene is directly identified (Rothschild and Soller, 1997). Moreover, the candidate gene approach can be combined with positional information on a genetic or physical map, which then provides a solid base to identify the putative gene. The candidate gene analysis has been successfully applied to livestock. Examples are studies on litter size in swine (Rothschild et al., 1996), double muscling in cattle (Grobet et al., 1997) and porcine stress syndrome in swine (Fujii et al., 1991).

Genome scans using anonymous markers have been widely used to map QTL in livestock (e.g. Andersson et al., 1994; Georges et al, 1995) and they are considered the most reliable and effective technique by many (Haley, 1999). The detection of a QTL through its linkage with a marker requires linkage disequilibrium between the QTL and the marker. Many experimental designs can be used with the genome scan in QTL mapping studies and the most suitable approach is dependent upon the species characteristics (physiology, family structure, breeding strategy such as crossbreeding or outbreeding) and to the industry structure (Soller and Medjugorac,1999).

In swine, most of the current studies on QTL mapping are based on crosses (F2 or backcross) of highly divergent parental populations, with respect to the trait of interest (Andersson et al., 1994; Andersson et al., 1998; Knott et al., 1998; Roher and Keele, 1998, De Koning et al., 1999). The advantages of experiments for QTL detection based on crosses include the likelihood of QTL segregation, the power of detection of the QTL and the simplicity of the statistical analysis. However, a limitation of the designs based on crosses of lines is that the power of detection of QTL segregating within either of the parental populations is low. In contrast, experiments based on information from an outbred population can reveal the presence of QTL of direct relevance to the population under study.

Within an outbred population, recombination decreases linkage disequilibrium between genetic markers and QTL over generations and, therefore, the information is usually not often useful at a population basis. However, within families even loose linkage between marker and QTL will cause linkage disequilibrium. The basic principle underlying identification of QTL in an outbred population can be simplified as follows.

If an individual is heterozygous at a marker locus and at a linked QTL, then, recombination excepted, offspring receiving a particular marker allele from the individual will also tend to receive the corresponding allele of the linked QTL. Individuals are scored for their genotype at the marker locus and their phenotype for quantitative traits. If a difference exists in the mean phenotypes among marker genotype classes, based on the allele inherited from a common ancestor, then the presence of a QTL linked to the marker can be inferred. Marker loci can be considered singly or jointly, with single or multiple analyses.

In dairy cattle, crosses are not commonly produced commercially and, when made, they are time consuming to generate due to the long generation interval. However, the breeding structure of the most common populations of dairy cattle (e.g. Holstein) provides a good source of accurately recorded information for QTL detection. A QTL mapping experiment in an outbred population is usually based on information for genotypes from two generations of animals (parents and their offspring) and phenotypes recorded on the offspring or on their respective progeny.

In 1990, Lande and Thompson proposed a method to associate phenotype to marker information. A multiple regression of phenotype on the number of copies of a given marker allele was performed. In order to avoid bias due to the overestimation of marker effect, they suggested re-estimation for the most evident effects.

The *daughter design* has been proposed in several studies on QTL detection in dairy cattle (Kashi et al., 1990; Weller et al., 1990; Meuwissen and Van Arendonk, 1992). The flow of information in this design can be described as follows. The marker effect of an elite sire is estimated by using information on daughters. The sire and the

daughters are scored for their genotypes and the performance of daughters is recorded. Data from this design can be analyzed with a multiple regression model (Meuwissen and Van Arendonk 1992). Markers are then traced (if possible) to the grand offspring of the elite sires, which are assumed to have segregating markers linked to the QTL. Weller et al. (1990) introduced the use of a *granddaughter design* for the detection of marker-QTL associations in dairy cattle. With this design, sons of the heterozygous sire are scored for their genotype at the genetic marker and the granddaughters are evaluated for the quantitative traits. Records of granddaughters are then regressed on the genotypes of the sons. Weller et al. (1990) calculated the statistical power of the design and compared it with the daughter design as a function of heritability, size of the QTL effect and family structure.

Results for power calculations (based on a chi-square analysis of the linkage between marker and QTL) showed that, for a given number of animals scored in the daughter design, power decreased with more sires and fewer daughters per sire. The power of the granddaughter design increased with the number of grandsires, sons per grandsire, granddaughters and size of gene effect. In addition, the granddaughter design was shown to require fewer marker assays than did the daughter design for equivalent power. This advantage of the granddaughter design was later shown to decrease with the increase of the QTL effect (Van der Beek et al., 1995) because conventional selection is very effective on such alleles. One disadvantage of the granddaughter design is the extra generation of selection.

In conclusion, the power of an experiment to detect linkage between the marker and the QTL depends upon the recombination rate between these two loci, the number of

informative individuals for the marker, the heritability of the trait, the size of QTL effect and the frequency of alleles at the QTL. Using flanking markers instead of single markers reduces the number of uninformative families and offspring (Jansen, 1989; Van Arendonk et al., 1994a).

Coppieters et al. (1999) suggested that by accounting for all genetic relationships among sires and sons, rather than treating the sires as unrelated and ignoring material relationships among the sows, the power of a granddaughter design can be increased.

For dairy traits, another interesting approach to QTL detection and mapping is selective genotyping (Darvasi and Soller, 1992). Some traits are routinely recorded at a relatively low cost. All individuals may then be scored for a single trait and only a subset of them may be genotyped. The basic idea is that individuals with extreme phenotypes (in the extreme ends of a normal distribution) are more informative than the ones with a score around the mean. This strategy can provide a large increase in power for a fixed number of genotypings (Darvasi and Soller, 1992). The statistical model used for analysis of selective genotype data must correct for the sampling bias.

The use of new markers, such as SNP, may provide in the future the possibility for linkage disequilibrium mapping (Haley, 1999). With a very dense marker map, marker loci and genes of interest are non-randomly associated across the population and not only within family (Lynch and Walsh, 1998). Linkage disequilibrium can be used to fine map QTL in dairy cattle. Using the popular linkage methods and current populations of livestock, estimates of QTL location lack precision. Often the confidence intervals for QTL location are 20 to 30 cM (Riquet et al., 1999). The major obstacle to decreasing the support area of the QTL location is the typical family size in livestock species.

Recombination is obviously relatively rare within regions very close to a QTL and recombination is necessary to perform precise positional mapping to precisely localise a QTL. Single families are typically not large enough to have experienced enough random recombinations to localise the QTL in a small region. Although increasing family size is biologically possible, particularly in dairy cattle with artificial insemination, it is not economically justified. Therefore, a more feasible approach has been to increase the family size by redefining the family. A larger family can be formed by grouping the tested individuals descending from a past common ancestor and sharing genes through identity by descent. Rather than creating more offspring from a single individual, related individuals that seem to carry a similar QTL allele can be grouped and their pedigrees traced to find a common ancestor several generations previous. Then these related individuals can be genotyped for several markers spanning an interval in the imprecise QTL support region and the small region shared by all individuals may be identified as the most likely location of the QTL. Such a practice has been applied by Riquet et al. (1999) to dairy cattle. They were able to apply this technique to a family of 7 sires and fine map a QTL to within a region of 5 cM. Recently, furthermore, unexpected high levels of disequilibrium between pairs of markers, have been shown in a genome-wide scan in the Holstein population (Farnir et al., 2000).

Quantitative trait loci with an effect on production traits have been mapped to different chromosomes, in some cases with overlapping results in different studies (e.g. Georges et al., 1995; Spelman et al., 1996; Ashwell et al., 1997; Coppieters et al., 1998; Lipkin et al., 1998; Kuhn et al., 1999; Velmala et al., 1999). Studies have demonstrated

the existence of QTL for health and conformation traits (Vilkki et al., 1997; Zhang et al., 1998; Spelman et al., 1999; Schrooten et al., 2000).

3.4 Statistical methods.

A variety of statistical methods is available for mapping QTL in cattle. The choice of one method over another depends upon the data structure in the experiment, computational constraints and assumptions on the distributions.

The simpler methods are the ones based on (multiple) linear regression (Haley and Knott, 1992; Zeng, 1993; Spelman, 1996; Uimari et al., 1996). The interval mapping based upon least squares is a relatively straightforward method that can be applied either across the population (Haley and Knott, 1994) or within family (Knott et al., 1996; 1997). These methods include fixed regression and iteratively re-weighted regression (Dentine and Cowan, 1990). This regression method is computationally efficient, robust and, because of its simplicity, allows performing data permutation to determine genome-wide significance thresholds (Hoeschele et al., 1997). An added advantage is that standard software packages can be used. The method also allows for multiple QTL analysis (Jansen, 1993), multiple trait analysis (Knott and Haley, 2000) and composite interval mapping (Zeng, 1994). The main drawback of the regression methods is that, due to the approximations involved (e.g. unrelated parents), their use is limited to certain designs and population structures, such as large half-sib families.

Maximum Likelihood analysis has also been applied to half-sib designs in outbred populations, using assumptions similar to the ones for least squares methods (Lander and

Botstein, 1989; Weller, 1986; Bovenhuis and Weller, 1994; Mackinnon and Weller, 1995).

Restricted Maximum Likelihood (REML) based on the mixed model, in contrast to least squares, allows incorporation of full pedigree information to perform variance component estimation (Fernando and Grossman, 1989; Goddard, 1992; Van Arendonk et al., 1994a and 1994b; Grignola, 1996). The method uses marker information to estimate identity by descent (IBD) relationships between individuals at a given point of the genome. Using REML, the relationship matrix is used to calculate the variance due to the specific point in the genome as well as the remaining polygenic effect. For instances when the data includes many pedigrees, markers are not fully informative, or some individuals in the pedigrees are not genotyped, the computational demand is very high. Monte Carlo Markov Chain methods may be used to decrease memory requirements for calculating the IBD coefficients at a given position of the genome when dealing with complex pedigrees (Bink and Van Arendonk, 1999).

The Bayesian analysis is the most flexible approach and can account for multilocus marker-QTL genotypes and the variable numbers of QTL on each chromosome, as well as different models for QTL variation (biallelic or multiallelic) (Hoeschele and van Raden, 1993; Janss et al., 1995; Hoeschele et al., 1997). This general approach also provides exact posterior variances and covariances among parameters and exact confidence intervals. However, computational limitations make the application of this approach difficult.

A somewhat controversial issue surrounding the detection of potential QTL is the definition of the significance level at which to declare confidence in the presence of a

QTL. A common approach in nearly all fields of science is to apply nominal rates of significance as if only a single experiment was performed. However, most QTL detection experiments involve many individual significance tests. Often analyses are applied to a number of different marker loci, on several chromosomes, and within several families. The experiment is often repeated for multiple traits. If one performed a study on three traits, for five loci, on each of three chromosomes, in four half-sib families, then a total of 180 different tests would be applied. If a nominal significance level of 0.05 was used and the tests were assumed to be independent, one would expect to observe nine significant results, even in the absense of QTL effects.

If one can assume that the tests are independent, then the standard Bonferroni correction can be applied. With the Bonferroni correction, if an overall significance of α is desired across n tests, then a significance rate of approximately α/n can be used to account for multiple testing (Lynch and Walsh, 1998).

The multiple tests are rarely independent, however. Markers are often linked together on the same chromosome and traits are often genetically related. The application of a Bonferroni correction would be excessively conservative in such instances and would decrease the power of QTL detection. Lander and Kruglyak (1995) developed the theory for an adjustment to significance levels based on Chi-square statistics that accounted for multiple tests with markers on the same chromosome.

Another approach that requires fewer assumptions is the permutation test (Churchill and Doerge, 1994). This test involves repeatedly (thousands of times) shuffling the phenotypic values for a fixed set of genotypes and recording the resulting significance tests. Experiment-wise significance levels can then be chosen empirically

based on the distribution of the significance tests. This approach automatically accounts for factors such as missing markers and non-random segregation of marker alleles (Lynch and Walsh, 1998).

A variation of the permutation test can also be used to establish significance tests for multiple correlated traits (Georges and Coppieters, 2000). This approach is based on establishing an "effective" number of traits and using that value as n in a Bonferroni correction. This approach involves shuffling all trait values for individuals into one of two abritrary "treatment" groups and using a t-test to determine the significance of the treatment effect. The effective number of traits (e) is then determined by solving the following equation for e:

$$n_\alpha / N = \alpha^e$$

where n_α is the number of permuations that exceed (for all traits) the threshold associated with the desired significance level α and N is the total number of permutations.

The choice of significance levels to use may depend upon how the information from the QTL detection experiment is to be used. If the goal is to identify chromosome segments for fine mapping, then a stringent threshold may be desired, to help avoid wasting resources searching for a non-existent QTL. For MAS, applying intermediate thresholds may yield the greatest genetic response (Spelman 1996) by avoiding type II errors.

3.5 Application of marker-assisted selection to dairy cattle.

Two main strategies have been outlined in the literature for the implementation of MAS in dairy cattle.

The first strategy proposes that selection decisions be made on breeding values that combine QTL and polygenic components. Several studies have shown how information on a single marker can be used in a mixed model analysis by fitting additive effects for alleles at a QTL linked to the marker and additive polygenic effects for alleles at remaining quantitative trait loci (Fernando and Grossman, 1989; Goddard, 1991; Goddard, 1992; Van Arendonk et al., 1994a, Hoeschele et al., 1995; Kinghorn and Clarke, 1997). In 1989, Fernando and Grossman first presented a technique to include the information provided by a single marker closely linked to a QTL into the mixed model equations. Their model includes the usual covariances between the additive effects of the animals for polygenic background (additive genetic relationships), and a variance-covariance matrix between additive effects of marked QTL alleles. For this gametic relationship matrix the recombination rate between marker and QTL, i. e. the exact position of the QTL, is required. As pointed out by the authors and later by Van Arendonk et al., (1994a), in this model the number of equations can be very large.

With the large number of anonymous markers currently available, in practice several traceable markers could well flank a putative QTL. Goddard (1992) extended the model to flanking markers and Ruane and Colleau (1995) introduced the possibility of double recombination events. In these models, covariances between bracketed QTL alleles can be described just as the relationships between animals determine the usual relationship matrix (**A**). This approach has been subsequently developed by Meuwissen and Goddard (1996) and used for different purposes by Spelman and Bovenhuis (1998).

Meuwissen and Goddard (1996) assume that the inheritance of a bracketed QTL follows that of the marker haplotype, neglecting double recombination, and do not attempt to define the position of the QTL. They also noted that their gametic relationship matrix was easily inverted using Henderson's rules. The number of equations of their model depends on the recombination rate of the marker haplotype, but is still beyond the number of animals (approximately 3 times the number of animals).

An interesting alternative approach that reduces the number of effects in the model to an animal level was presented by Nejati-Javaremi et al. (1997). They suggest that, if all the QTL affecting a trait were known, a total allelic relationship matrix could replace **A** in the mixed model equations. The definition of a total allelic relationship coefficient between animal i and j is twice their coancestry (Malécot, 1984) and defines the probability of identity in state rather than identity by descent of the genes. As a consequence, the total allelic relationship matrix is trait-specific and traces relationships also among unrelated animals. The idea is related to the concept of identical by descent coancestry conditional on marker genotypes originally introduced by Chevalet et al. (1984) and later described by Ollivier (1998).

In the near future, relationship information based on marker genotypes or haplotypes will replace the standard additive relationship matrix and genomic information will impact the infinitesimal model. Ultimately a unified model can be envisaged where different genomic regions have appropriate weights to the variance they explain (Haley and Visscher, 1998; Pagnacco and Jansen, 2000).

In most of the studies where genomic information is included in genetic evaluation, all animals have generally been assumed to have genotypic information.

However, in most likely scenarios for MAS, marker genotypes will be available only on a limited number of individuals. Therefore, some procedures have been proposed (Wang et al., 1994; Van Arendonk et al., 1994a; Bink et al., 1998; Bink and Van Arendonk, 1998) that allow the inclusion of information from animals with unknown genotypes.

The second strategy includes selection decisions first made on conventional BLUP breeding values followed by within family decisions on QTL information (Kashi et al., 1990; Meuwissen and Van Arendonk, 1992).

The first MAS strategy is technically more demanding, but also provides, theoretically, a superior genetic response. The inclusion of QTL information in the breeding value estimation, in fact, results in more accurate estimation and, therefore, higher selection differentials. However, the dairy industry has hesitated to rely on this first strategy. This indecision is probably due to a number of reasons. First, and most obviously, genotyping is expensive and, therefore, scoring of only a small fraction of the population can be justified financially. Also, the risk of such a strategy probably seems high, because different studies have reported very different results. Smith and Smith (1993) and Spelman and Garrick (1997) demonstrated the risk of reduced genetic response (compared to conventional selection) that could result from MAS when the marker information is inaccurate. On the contrary, Van Arendonk et al. (1994b) found that this risk is relatively small.

Gibson (1994) showed that maximum accuracy selection, when the genotype is known, yielded more short term response, but lower long term response than did a control selection where the genotype was unknown. This reduction occurred because both types of selection eventually fixed the positive allele and achieved the maximum response at

the QTL. However, maximum accuracy selection reduced the pressure on polygenic selection, resulting in a lower long-term gain.

Over the past ten years many studies have used simulations to examine the efficiency of MAS in dairy cattle. Although these studies have consistently shown increased responses where MAS has been applied, the estimated benefits of MAS have varied greatly from study to study. Ruane and Colleau (1996) assessed a superiority of MAS of up to 15% more genetic progress. Mackinnon and Georges (1998), and Spelman and Garrick (1998) found that the genetic level of selected bulls was respectively 10 and up to 9 percent higher when MAS was applied.

The main reason for these differences is the difficulty in creating a simulation that is general enough to mimic and apply to all dairy cattle breeding schemes. Also, the true genetic model is unknown and the parameters used to simulate the underlying model have varied greatly. Spelman (1998) analysed the major determinants of the variability in estimates of genetic response from MAS.

The first major difference among studies has been in describing the distribution of the alleles of the QTL. Many studies have assumed a di-allelic genetic model (e.g. Ruane and Colleau, 1996; Villanueva et al., 1999). Others have simulated many alleles (e.g. Mackinnon and Georges, 1998; Spelman and Garrick, 1998). Responses with the di-allelic model have generally been less, particularly when more than one generation was simulated, because the population quickly approached fixation for the favorable allele, decreasing variance at the QTL. Even among instances where multiple alleles were simulated, the distributions of the alleles differed. Spelman and Garrick (1998) simulated alleles that were normally distributed and Mackinnon and Georges (1998) simulated

alleles with a double-exponential distribution. Also, the variance of the QTL effects has differed both across and within studies. Response to MAS has been greater for simulated genotypes with greater variability among alleles at the QTL.

Furthermore, with more animals genotyped in each generation and multiple generations considered, the increased accuracy of estimation of the QTL effect increased the advantage of MAS (Meuwissen and Goddard, 1996; Spelman and Van Arendonk, 1997). Lastly, the number of years or generations of simulated selection influences the MAS superiority. Superiority of MAS decreases over generations as the variance at the QTL decreases. This explains differences in the results of studies considering only one generation (Kashi et al., 1990; Mackinnon and Georges, 1998) from studies that have used MAS over several generations, accounting for reduction in QTL variance (Ruane and Colleau, 1995 and 1996).

Kashi et al. (1990) provided a theoretical analysis of the advantages of MAS on young bulls prior to entering the progeny testing. They assumed that di-allelic (DA) and poly-allelic (PA) markers linked to a QTL were mapped and evaluation of elite sires was used to estimate the marker-QTL association. Candidate bulls were selected, prior to progeny testing, on information on marker alleles derived from the elite sire. The analysis considered the possibility of recombination between marker and QTL and the number of marker-QTL couplings identified in an elite sire that can actually be traced to the grandson (i.e. the candidate bull). The MAS on candidate bulls was performed using an index ($I = P - Z$), which represented the difference between the number of marker alleles (or haplotype) associated with favorable QTL alleles and those associated with the unfavorable ones. Results obtained by Kashi in the study showed that the statistical

power of detection of a marker-QTL association increased with the number of dams and with the number of QTL assumed. The MAS on candidate bulls based on a single DA had a negligible effect on the rate of genetic gain. Significant increments were obtained when QTL were detected by use of a single PA marker, a haplotype of DA markers or, most effectively (20-30% higher than other schemes), a haplotype of PA markers. Analysis at a cost level showed an advantage for the use of PA markers or marker haplotypes over single markers.

Spelman and Garrick (1998) compared two within-family MAS schemes in terms of genetic progress and economic return, using a stochastic simulation. Schemes were identified as top-down (a granddaughter design) and bottom-up (daughter design). Different reproductive rates on the female were used: 1, 3 and 40 progeny per female were considered in order to simulate aritificial insemination (AI), multiplue ovulation and embryo transfer (MOET) and in-vitro embryo pick-up (IVEP). The top-down scheme provided no improvement in genetic gain (compared to the scheme where no account was taken of marker information) when only one offspring was produced. However, with 3 and 40 offspring per dam, the genetic gain using a top-down scheme increased by 1 and 2%, respectively. The bottom-up scheme provided increased (ranging from 1 to 5%) response in each of the reproductive situations. Spelman and Garrick (1998) also studied the possible benefits of tracking the QTL in the maternal path. The increased amount of information provided by this procedure increased response up to a total of 9%.

3.5 MAS in dairy cattle nucleus schemes

As stated before, markers used within families allow for determination of which of the two parent chromosome segments was inherited by an individual. In this way, marker information explains part of the single parent's Mendelian sampling variance (Dekkers and Dentine, 1991). This knowledge can be particularly useful in breeding schemes characterized by short generation intervals, such as juvenile MOET nucleus schemes, in which selection among full sibs is inaccurate or random.

In the study by Meuwissen and Van Arendonk (1992), a deterministic model was implemented to assess the value of MAS in progeny testing, and open and closed juvenile nucleus schemes in dairy cattle. Marker-QTL associations were found by multiple regression of performance data on the number of copies of marker alleles present for all marker loci (where marker effect was considered as random). They assumed that equidistant markers were available on each chromosome and simulated a cluster of closely linked QTL. The QTL effects could potentially explain all the genetic variance. Within family linkage disequilibria were used so that, by tracing markers from grandsire to grand-offspring, deviations of grand-offspring records from their full-sib family means were predicted. The fraction of the within family variance explained by the markers ranged from 4.1 to 13.3%, with the maximum corresponding to higher number of dams and minimum marker distance.

Prediction of within family genetic deviations gave a negligible contribution to genetic gain in progeny testing schemes, where most of the information on the candidate comes from individual and progeny performance. In nucleus schemes, where selection is based upon pedigree information, genetic gain increased from 9.5% to 25.8% and from 7.7% to 22.4% in open and closed nuclei, respectively. Although genetic gain increased,

the variance of the gain was also stable with use of MAS, meaning that MAS could help decrease risk. The use of MAS, in fact, reduces the probability of selecting relatives and reduces inbreeding. Ruane and Colleau (1996) calculated the effect of MAS on inbreeding coefficients for loci neutral with respect of the trait of interest (from the relationship matrix) and for the QTL (from number of individuals with QTL alleles identical by descent). Inbreeding coefficients were always much higher at the QTL than at the neutral loci. Restricted selection of full brothers, although decreasing inbreeding at the neutral loci relative to BLUP selection, had only very little effect on inbreeding coefficients at the QTL for MAS, since selection pressure for QTL remained high, even when only conventional selection was practiced.

In addition to potential advantages for genetic response (relative to conventional selection), a MOET nucleus scheme could have a more centralized structure, which could make it economically feasible to genotype all the selection candidates of both sexes for the markers of interest. In 1996, Ruane and Colleau analyzed the benefits of MAS within a closed MOET nucleus scheme. They used a Monte Carlo simulation of a nucleus breeding scheme in which the animals were typed for two markers flanking a QTL and evaluated candidates using the BLUP method of Fernando and Grossman (1989). Unlike Meuwissen and Van Arendonk (1992), they considered in the model the reduction in variance due to inbreeding. Selection was for a single trait measured only on females. The study showed that, when the favorable allele was initially at a frequency of 0.5, MAS increased the response at the QTL locus but decreased the polygenic response. Cumulative response was found to be superior to conventional BLUP by 3, 9, 12, and 6% at one, two, three, and six generations of selection, respectively. The reduction of

response in the polygenic selection was due to the competition between correlated sources of information. An increased correlation between the QTL and the index corresponded to a reduction in the correlation between the polygenic information and the index. Moreover, after the first round of selection, due to the negative correlation created between QTL and polygenic effects, animals with higher QTL values had, on average, lower polygenic values. Similar results were found by Spelman and Van Arendonk (1997): polygenic response was lower under MAS than under a control selection (BLUP) and this difference was greater when the QTL explained a large proportion of phenotypic variance (10%).

When the starting frequency of the favorable allele was low (0.1) the benefits of MAS were larger because of the increased probability of loss of that allele without MAS (Ruane and Colleau, 1996). When heritability was low, the effect of drift variance was relatively larger and similar for BLUP and MAS selection, so that the frequency of loss of the allele was increased for both schemes.

The model derived by Fernando and Grossman (1989) and modified by Goddard (1992) was used by Meuwissen and Goddard (1996) to study the effect of selection on marker haplotypes. The study assumed that a region where QTL were present had been mapped and many alleles were assumed at the QTL. This design reflected a situation where the observed QTL effect was due to a cluster of closely linked QTL. Ruane and Colleau (1995, 1996) based their calculation of breeding values on the expected inheritance derived from the recombination rate. However, in the model of Meuwissen and Goddard (1996), probability statements were that QTL transmission either could be

or could not be followed by inference from marker haplotype (the recombination event was known). Therefore, probability statements other than 0 and 1 were not made.

Because MAS can be performed on different traits, characteristics of the traits, such as heritability, availability of the records, time of recording, and information available on the QTL are very important factors in implementing a program for MAS. Meuwissen and Goddard (1996) simulated different situations with respect to the availability of records (before or after selection) and calculated the respective genetic gain expected in the short term. When records were available within the nucleus after selection and only on females the situation reflected a juvenile MOET scheme. In this case, MAS increased genetic gains by 38% and 21% in the first and fifth generations, respectively. In general, when records for traits were available only after selection decisions were made (e.g. fertility or longevity) genetic gains from MAS were substantially higher than in situations when records were collected before selection (+38% versus +8.8%).

Traits characterized by low heritability are also good candidates for MAS. Ruane and Colleau (1996) showed that the beneficial effect of MAS on QTL response and total response was higher for traits with progressively lower heritabilities. For the first three generations of MAS superiority to BLUP selection was 3, 5, 12% for traits with of heritabilities of 0.5, 0.25 and 0.1, respectively. Similarly, Meuwissen and Goddard (1996) found greater increases in response rate due to MAS for traits with lower heritabilities, but their model assumed that the QTL position and variance were known. In general, with lower heritability, the accuracy of selection decreases but the QTL effect can still be accurately estimated, but only with a large number of genotyped daughters

(such as available in a well-designed progeny test program). This result is explained by the fact that tracing copies of QTL alleles by markers leads to availability of multiple records to estimate the effects of QTL alleles. Therefore, accuracy is decreased less due to decreased heritability than is the accuracy of an EBV based on a single record. In 1998, Muir and Stick compared, at different levels of heritability, response to phenotypic selection (P), candidate gene selection (C), and selection on an index (B) combining both sources but favoring the candidate gene. Based on response in the long term, at a very low heritability (1%) B was superior to P and P was superior to C. When heritability was 10% the situation was the opposite for B and P (both were still superior to C).

When a closed nucleus scheme is analyzed, information on genotype and performance are typically only recorded within the nucleus. Meuwissen and Goddard (1996) investigated the advantage of also including information from outside the nucleus (records from 1000 commercial progeny). Relatively little advantage was found in genetic gain with that extra information (44% versus 38%). The accumulation of information from previous generations used in the model greatly decreased the need for further recording outside the nucleus. However, results showed that having limited marker information from previous generations could greatly decrease the extra genetic gain from MAS.

Although it was monitored in the simulation by Ruane and Colleau (1996), none of the studies reviewed adequately accounted for the reduction in QTL variances due to changes in allelic frequencies. Ruane and Colleau showed that when a scheme with 8 sires and 64 dams is considered, fixation for the beneficial allele was reached in only

three generations. The model, however, continued to estimate QTL differences that eventually did not exist, reducing the advantage of MAS over BLUP selection.

Furthermore, the models assumed in the previously described studies generally require the use of known values for the proportion of genetic variance explained by the QTL. When the test statistic exceeds a certain significance level, the QTL effect can be easily overestimated (Wang, 1995). In 1997, Spelman and Van Arendonk used stochastic simulation to investigate the effect of inaccurate estimates of genetic variance associated with QTL on genetic response in an adult, closed MOET nucleus scheme. In their model, a QTL actually accounting for 5% of the genetic variance, was evaluated with assumed QTL variances of 10% and 15%. Greater genetic gain at the QTL was observed when the variance was overestimated and MAS was used, but polygenic response was severely reduced. As a result, the overall genetic gain was decreased, particularly in the long term (7 generations). The long term loss of response was reduced when the variance of the QTL was re-estimated after four generations of MAS. In the same study, the relative advantage of MAS over BLUP selection on polygenic effect only was estimated for two levels of variance explained by the QTL (5% or 10%) and two genetic models: a total of 10 alleles at the QTL (A10) or 2 distinct alleles per each base parent (BP2). As expected, the greater the QTL effect the larger was the response. The two models gave similar rate of response for the 5% QTL but when the QTL explained 10% of the variance, the model with more alleles (BP2) was superior.

In summary, the results from the studies published on the use of MAS within a MOET nucleus scheme show sizeable effects of the assumptions made in the model on the resulting estimates of genetic response achievable.

3.6 Moet nucleus breeding schemes.

The introduction of MOET techniques in animal breeding made it possible to overcome the restriction on female reproduction, i.e. low reproductive rate and long generation interval. Until MOET, the genetic gain achieved through selection was mostly due to selection on male pathways.

To better capitalize on the genetic opportunity offered by MOET, Nicholas and Smith (1983) proposed the implementation of nucleus schemes by artificial insemination (AI) organization, operating in a large population of dairy cattle. Because MOET costs are high, AI firms are most likely to be the organization prepared to implement and exploit the genetic improvement generated by MOET and other reproductive techniques such as ovum pick up, IVEP, embryo sexing, splitting and cloning.

Several studies have investigated the various structures possible in nucleus programs. In an early study, Nicholas and Smith simulated various situations of nucleus selection, following the rules of selection index theory (deterministic simulation) and comparing the different breeding schemes in terms of annual genetic gain. In these schemes, a separate herd is established, with the ultimate goal of production of young bulls. The scheme is designed to reduce the generation interval while tolerating lower accuracy and requiring a smaller number of recorded cattle. Two types of mating design were described: a juvenile scheme, with selection at one year of age based on pedigree information and an adult scheme, in which females were selected after completion of first lactation. The results obtained by the simulation showed superiority, in terms of annual

genetic change, of the juvenile scheme applied to both sexes. All these schemes predicted a higher (up to 80%) genetic response compared to conventional progeny testing at a national level.

In 1987, Juga and Maki-Tanila used a stochastic model to simulate an adult nucleus scheme. The genetic gain was substantially lower than deterministic predictions (33% less), and adult nucleus herds were not expected to be competitive with an efficient progeny testing scheme. Similar results were found by Woolliams and Smith (1988). In their deterministic study, adult schemes used high MOET rates (16 progeny per donor) to compete with an efficient progeny testing scheme, while juvenile schemes gave a better response at any rate of MOET. The use of MOET within a nucleus herd proposed by Nicholas and Smith was modified in several studies.

In the 80's, Christensen and Liboriussen (1986) and Colleau (1985, 1986) proposed "MOET x conventional" hybrid nucleus scheme, in which all the sires used for breeding were progeny tested. Furthermore, open nucleus schemes were considered in which genetically superior animals from the commercial population could be selected to breed nucleus replacements.

Meuwissen (1990) compared expectation and variance of the steady state genetic gain in open and closed MOET nucleus and progeny testing schemes. Genetic gains of conventional (fixed generation interval) and modern progeny testing and open nucleus schemes were about 19, 13 and 3% lower than closed nucleus schemes, which gave the best response.

Colleau (1986) simulated the effect of opening the nucleus to foreign genetic material. The nucleus studied was open to genetically superior bulls progeny tested in the

population. Opening the nucleus in the female paths allows the best females in the nucleus and across the commercial population to be used as donors. This would lead, according to Colleau (1986) to 6% higher genetic gain in the short term, but to lower response in the long term with most of the selected females (70%) coming from the commercial population.

In 1990, Meuwissen suggested that opening the nucleus could provide a higher response by increasing the number of selection candidates. However, intense selection on base females can decrease the genetic variance in the nucleus and this can cause lower gain in an open versus a closed nucleus. Closed nuclei benefit more from an increased female reproductive rate, because selection differential of female nucleus replacements selected from the nucleus are low when the reproductive rate is low. In an open nucleus scheme, the selection differential of replacements selected from the base is high even when reproductive rate is low (i.e. progeny testing scheme).

Dekkers and Shook (1990) used a semi-stochastic model to simulate five competing AI firms. Changes in the breeding scheme of one of the firms were compared in terms of genetic and economic response to an efficient conventional program using MOET to produce young bulls. The effect of opening the nucleus to donors from the commercial population (open versus closed nucleus) was investigated. When a closed adult nucleus was analyzed, the genetic gains obtained were not competitive to conventional MOET used as a comparison base, with the exception of very a large nucleus, (which are also associated with very high costs). Opening the nucleus to superior animals from an elite population had a positive effect on the genetic response, more so for adult than for juvenile schemes. The model implemented by Dekkers and Shook

(1990) was partially followed by Schrooten and Van Arendonk (1992). A hybrid nucleus scheme was simulated stochastically (progeny test of all the bulls). The annual genetic response at equilibrium was compared for open and closed nuclei with 120 or 240 heifers entering the nucleus every year. Compared to the progeny testing scheme, response in open nucleus schemes with respectively 120 and 240 replacements per year were 4.3% (0.244 sa) and 5.6% (0.247 sa) higher. In the first years after the establishment of the nucleus, a relatively small share of dams was selected from the nucleus. The proportion increased for 30 years of simulation and then stabilized with 60% of dams from the nucleus, of which 36% were heifers. Closed nucleus schemes gave a lower response to selection than open schemes and this difference was larger in a small nucleus.

4. Accounting for uncertainty in QTL location in marker-assisted preselection of young bulls prior to progeny test.

4.1 Summary

The objective of this study was to evaluate whether the efficacy of MAS could be improved by considering a confidence interval (CI) of the QTL position. Specifically, MAS was applied for within-family selection in a stochastic simulation of a closed nucleus herd. The location and effect of the QTL were estimated by least squares interval mapping with a granddaughter design and marker information was used in a top down scheme. Three approaches were used to select the best bull within fullsibships of 3 or 40 bulls. All three were based on the probability of inheriting the favorable allele from the grandsire (PROB). The first method selected the sib with the highest PROB at the location with the highest F-ratio (MAXF). The other two approaches were based on sums of estimated regression coefficients weighted by PROB at each cM within a 95 % CI based on either bootstrapping (BOOT) or approximate LOD scores (LOD).

Accounting for CI increased the relative genetic gain in all scenarios. The average TBV of the selected bulls was increased by 2.00, 2.60 and 2.59 % when MAS was applied using MAXF, BOOT and LOD, respectively, compared to random selection (h^2 = .30). Selected bulls carried the correct allele in 63.0, 68.5, 67.6 and 50.1% of the cases for MAXF, BOOT, LOD and random selection respectively.

4.2 Introduction

The potential genetic improvement from marker assisted selection (MAS) in dairy cattle has been evaluated in many studies. These studies have consistently shown that genetic progress can be improved by the implementation of MAS. Breeding schemes in dairy cattle that employ multiple ovulation and embryo transfer for genetically superior females provide the opportunity for the use of marker information in selection among full sibs prior to progeny testing (Kashi et al., 1990; Meuwissen and Van Arendonk, 1992; Ruane and Colleau, 1996; Spelman and Garrick, 1997). Many of the studies examined the maximum responses possible with MAS. These studies often assumed that location of the QTL relative to markers was known with certainty and, in most cases, recombination between flanking markers and the QTL was not accounted for (e.g. Mackinnon and Georges, 1998; Spelman and Garrick 1998). Future advances in molecular genetics will almost surely increase the resolution of genetic maps and identify some of the actual sites of polymorphism responsible for phenotypic differences in economic traits, but currently much uncertainty about QTL position remains. Thus, MAS decisions can be based only on the genotypes of candidate animals at the most probable location of the QTL, with respect to anonymous genetic markers.

Various methods have been proposed to calculate the confidence interval for the QTL location in mapping experiments (Visscher et al., 1996; Mangin and Goffinet, 1997). The objective of this study is to evaluate if the efficacy of MAS, in situations where the position of the QTL is uncertain, can be improved by considering a confidence interval of the QTL position.

4.3 Material and Methods

Marker-assisted selection within a adult closed nucleus population was simulated. Selection for a sex-limited trait was modeled. We assumed that results from previous studies had suggested the presence of a QTL on a given chromosome but its location was undefined. Therefore, a single chromosome was targeted and simulated both QTL detection and MAS based on the result of the mapping study. A top down scheme was used (Mackinnon and Georges, 1998).

Three discrete generations of truncation selection were performed. Grandsires and granddams were selected from an unrelated base population and were used to establish the nucleus. They were mated to produce a second generation of nucleus sires and dams. Estimated breeding values of the sires were generated and used in a granddaughter design to estimate the location of the QTL and substitution effect for each grandsire. Then a third generation of bulls was generated and results from QTL detection were used for within family selection of the bulls to be admitted into a progeny test program.

Several different scenarios were simulated. The standard scenario was based on the work of Spelman and Garrick (1998) and simulated a trait with true heritability of 0.30 and a QTL that accounted for 17% of the genetic variance. Young bulls were selected from full-sib families of 40 males. The parameters that were varied to define the alternate scenarios are shown in Table 4.1.

Population

The scheme of the population structure is shown in Figure 4.1. The base population consisted of 750 males and 25,000 females. To form the nucleus, five

grandsires and 250 grandams were selected from the base population on their EBV. For all individuals, EBV were simulated by adding random, normally distributed, prediction errors to the respective genetic components, using the same approach as Spelman and Garrick (1998):

$$EBV_i = r_{TI}^2(BV_i - MBV)\, z_i\sqrt{(V_p(h^2 r_{TI}^2 - h^2 r_{TI}^4) + MBV}, \qquad [1]$$

where, r_{TI}^2 was the squared accuracy of selection, BV_i was the true breeding value, MBV was the population mean breeding value, z_i was a standard normal deviate, and V_p was the phenotypic variance. The standard deviation of the prediction error was varied by altering r_{TI}^2 according to sex and generation to reflect the amount of information typically available for prediction of TBV.

Each selected grandsire was mated to 50 granddams and from each of these families one sire and three dams were replacements in the nucleus. Estimated breeding values were generated for these replacements by assuming that the sires were progeny tested (85 daughters each) and the dams produced a single lactation record.

The best 5 sires and 250 dams from the second generation were chosen and mated to produce the third generation of young bulls. Multiple ovulation and embryo transfer or *in vitro* embryo production were used to produce three or forty young bulls, respectively, for each dam (as in Spelman and Garrick, 1998), generating either 750 or 10,000 young bulls in total.

Genetic model

Genetic variation was contributed by a single QTL and an unlinked polygenic effect such that total heritability of the trait was either 0.10 or 0.30. The polygenic effect was sampled from a normal distribution. Only additive effects were modeled.

Ten alleles were generated for the QTL. Allelic effects in each replicate were drawn from a normal distribution N (0, σ^2_{QTL}). This approach for modeling allelic effects simulates a situation in which several sites of polymorphism are present within the same gene (or closely linked genes) and different animals carry alleles defined by different combinations of the polymorphisms. Two QTL variances were used in the study: 17% and 35% of the total genetic variance. In the base population all alleles had the same frequency.

The QTL was surrounded on the chromosome by six markers. Six alleles, all at the same frequency, were assigned to each marker in order to mimic microsatellite markers. In the base scenario the markers were evenly spaced across a chromosome of 130cM (Coarse). This design represented the base situation, for which we assumed that previous knowledge (such as previous genome scans or comparative mapping with another species) provided reason to believe that a QTL may exist on the chromosome of interest, but little indication of the QTL location. The QTL was located either midway between the two central markers or midway between the first two markers.

To compare to the base situation with results from previous studies, we also simulated finer mapping for which a narrow chromosomal region was targeted and the QTL was positioned relative to 6 markers spaced at 1 cM (Fine).

QTL Mapping and MAS

Genotypic information from the grandsire and the progeny tested sires and the EBV of the sires were used in the granddaughter design. Interval mapping was performed using multimarker regression as developed by Knott et al. (1996). Marker assisted selection was applied to young bulls in the third generation. Only one bull was selected per fullsibship. Two basic strategies for MAS were practiced.

The first strategy was the simplest and involved selecting bulls based only on their genotype at the chromosomal location with the maximum probability of being the QTL (MAX). This approach consisted of first using across-sire least squares analysis (Knott et al., 1996) to estimate the most probable location of the QTL. The regression approach of Knott et al. (1996) uses the genotypic information to calculate the probability that sires and sons share a given haplotype at every centiMorgan along the chromosome of interest. This information was then extended to a second generation to estimate the probability of each young bull sharing the same haplotype as the grandsire at the chromosomal location most likely to be the QTL.

When this haplotype was estimated to include the superior QTL allele, the full sib with the greatest probability of inheriting the haplotype at the predicted best location was selected. The converse was applied when the reference haplotype contained the inferior allele. In other words, the young bull with the lowest probability of inheriting the grandsire haplotype at the most probable QTL location was selected.

The other basic strategy involved calculating a confidence interval (CI) for the QTL location and choosing the bull with the best haplotype spanning this CI, rather than simply concentrating on the best location. Within this strategy, several different

approaches were used to define the selection criterion and the performance of the different strategies was compared.

These strategies differed with respect to two different variables, 1) the length of the CI, and 2) the relative weighting placed upon centiMorgans within the CI. Two methods were used to establish the length of the CI.

One approach (BOOT) used a bootstrapping technique based on work by Visscher et al. (1996) to derive a 95% confidence interval of QTL location. For this approach, 50 random samples of EBV and genotype information of sons were drawn (with replacement) from the data. Then the method of Knott et al. (1996) was used to identify the most probable location of the QTL for each of these 50 samples. The lowest and highest estimates were discarded, and the extreme values from the remaining 48 samples were used as the endpoints of the CI. Only 50 bootstrap samples were generated for the sake of computing efficiency and because the results of Visscher et al. (1996) showed that increasing the number of samples had very little effect on the estimated CI.

The second approach estimated the length of the confidence interval based on approximate LOD scores (LOD) (Lander and Botstein, 1989). The LOD score at each location along the chromosome was estimated by the following equation (Lynch and Walsh, 1998):

$$LOD_{approx} = [n \times \ln(RSS_{reduced} / RSS_{full})] / 4.61, \quad [2]$$

where, LOD_{approx} is approximate LOD score at a given cM, n is the number of sires, $RSS_{reduced}$ is the residual sum of squares for the reduced model (which included only

grandsire effects), RSS_{full} was the residual sum of squares from the full model (which included effects of grandsire and the genotypes of the sires). The CI included all adjacent cM for which the approximate LOD score was within 1 of the LOD of the site that was the most probable location of the QTL.

For each centiMorgan (cM) within the CI the probability of the young bull sharing the reference haplotype with the grandsire was multiplied by one of three different weights.

For the first approach (Within grandsire), the probability of transmission to the young bull at each cM was multiplied by the standardized estimate, for the respective grandsire, of the QTL allele effect at that location. The sum of these values was then used as the selection criterion:

$$I_{WITHIN} = \sum_{i=1}^{m} p_i \beta_{ij} , \qquad [3]$$

where, m is the length of the CI in cM, p_i is the probability that the son inherited the reference grandsire haplotype at cM i, and β_{ij} is the regression coefficient for grandsire j at cM i (standardized by dividing by its standard error).

The second approach (Across grandsire) multiplied the approximate likelihood ratio (approximate LOD * 4.61) at each cM, calculated across grandsires, by the maximum regression coefficient for each sire and then summed up values across the CI. Then, similar to the first approach, the bull with the greatest sum was selected. Thus,

$$I_{ACROSS} = \sum_{i=1}^{m} p_i \beta_{kj} T_i \,, \qquad [4]$$

where, m and p_i were as defined in Equation [3], β_{kj} was the regression coefficient for grandsire j at the most probable location of the QTL, and T_i was the approximate likelihood ratio at cM i.

The final approach (Uniform) simply summed the probabilities of transmission at each site, without weighting locations with respect to the predicted location of the QTL. Following the notation of [3] and [4],

$$I_{UNIFORM} = \sum_{i=1}^{m} p_i \beta_{kj} \,. \qquad [5]$$

With these methods no discrimination was necessary with respect to whether the reference haplotype contained the favorable or the unfavorable allele because this factor implicitly accounted for by the sign of the estimated QTL effect.

In applying mapping results for MAS, no consideration was given to the significance level of the regression either between sire or pooled across sires.

The average TBV of young bulls selected prior to progeny test were calculated for the different MAS strategies and compared to random within family selection of young bulls. The following formula was used for the comparison:

$$(TBV_{MAS}\text{-}TBV_{Random})/(TBV_{Random}\text{-}TBV_{Comm})*100 \qquad [6]$$

where TBV_{MAS} is the average TBV of young bulls selected by each of the MAS approaches, TBV_{Random} is the average TBV of full sibs selected at random and TBV_{Comm} is the average TBV of the commercial population.

4.4 Results

The simulation program was validated by comparing results for random and index selection with selection index predictions. In the base situation with heritability = 0.30, observed mean genetic values of males were 0.09, 22.10, and 35.85 for generations 1, 2, and 3, versus expected values of 0.00, 22.30, and 35.95. For selected females, observed values were 17.28 and 28.40 in generations 1 and 2 and expected values were 17.48 and 28.47.

Table 4.2 shows the respective probabilities of selecting a young bull that carried the desired QTL allele when using the three different approaches for weighting chromosomal locations within a confidence interval. The desired allele was defined as the paternal grandsire allele if the sire inherited the superior allele from the grandsire or the paternal granddam allele if the sire inherited the inferior allele from his grandsire. If either the grandsire or the sire were homozygous at the QTL then all the young bulls were not considered to have inherited the favorable allele.

Regardless of which method was used to establish the CI, the approach that weighted each location based on regression coefficients within grandsire was superior over the approaches that used weights estimated across grandsires or a uniform weight for all locations within the CI. The table gives results for $h^2 = 0.30$, σ^2_{QTL}= 17% and

family size of 40, but the superiority of using the estimates within grandsire was consistent across all combinations of parameters simulated. Because the approach of weighting location based on regression coefficient within grandsires was consistently superior, all subsequent results reported will be based on this approach.

As expected, all strategies of MAS increased the probability of selecting a young bull that carried the desired allele and, therefore, increased the average breeding value of the group of selected young bulls. Table 4.3 compares the efficacy of MAS relative to random selection when young sires are selected based on their genotype at the single most probable location of the QTL (MAX) or for their haplotypes for the region bounded by a CI calculated by bootstrapping (BOOT) or approximate LOD score (LOD). Results for random selection are given for reference. On average either approach using the confidence interval was superior to considering only the most probable location.

The overall advantage of using MAS was decreased when the size of the full-sib families was smaller. The average gain for MAS with family size of 3 was 1.08%, 1.56%, 1.50% for MAX, BOOT and LOD, respectively, relative to random selection. This decrease was due to the lower selection intensity and reduced selection opportunities in such small sibships. In one third of the families, the randomly selected and the marker assisted selected sons were the same. The two methods considering the confidence interval were still significantly different from MAX.

No statistically significant differences were observed in the efficacy of MAS when the CI was determined by BOOT versus LOD. As a rule, the CI estimated by BOOT tended to be longer than with LOD. For the base scenario the average lengths of the CI were 105 cM and 70 cM, respectively, for bootstrapping and LOD approximation.

The reason little difference was observed in the genetic gain between methods was probably that, despite their differences in length, the CI from both methods nearly always spanned most of the chromosome.

Table 4.4 shows the relative increases in the average TBV (over random selection) of young bulls from each method of MAS for four different combinations of true h^2 (0.10 or 0.30) and σ^2_{QTL} (17% or 35%). The benefits of using MAS were increased when σ^2_{QTL} was greater. Improvement relative to random selection was greatest when true $h^2 = 0.30$ and $\sigma^2_{QTL} = 35\%$. For all combinations, the methods using the CI were significantly ($P \leq 0.05$) superior over MAX. The difference between the two CI methods and MAX was greatest at true $h^2 = 0.30$ and $\sigma^2_{QTL} = 17\%$.

The relative increases in the average TBV (over random selection) of young bulls from each method of MAS for three different distributions of markers and QTL across the chromosome is shown in Table 4.5. Logically, the greatest benefits of MAS were observed when the markers were located closer to the QTL. For example, using the CI from bootstrapping, the improvement over random selection was 2.59% with coarse mapping (10 cM from QTL), and 3.86% for Fine (1 cM). The benefits of MAS were greater when the QTL was located near the center of the chromosome, versus near one of the telomeres. Accounting for the CI provided the greatest relative gain in all instances. The difference between the MAX and CI approaches was not significant, however, when low recombination rate was generated between the markers and the QTL. In this case an implicit assumption was that the QTL location was well known (within the marked 5 cM region), so the success of selection was not highly dependent upon the precision of the estimated location of the QTL within the interval.

Marker assisted selection is costly to apply to a breeding program. Many factors other than which potential allele at a single QTL was inherited by a bull will have significant influences on the TBV or EBV of that bull. Among the major factors are the size of polygenic effect or other QTL on unmarked chromosomes, random effects on the daughters of bulls and even the alleles at the QTL inherited from dams. Intense selection limits the variability of allelic effects at a QTL. These considerations limit the value of applying full scale MAS.

The sources of variances other than the QTL introduce error into the interval mapping procedure. For this reason, sometimes a Type I error occurs and a given QTL allele of a grandsire is estimated to be superior when the opposite is the case. In some case the grandsire can be homozygous. Type I errors are more likely to occur when the difference between alleles is small, so that application of MAS has relatively little value. Therefore, a breeding company may wish to apply a critical value (based on F-ratio for example) to help guard against Type I errors. Then MAS can be applied to families for which the F value exceeded the threshold. One potential problem with this approach is that some families with truly large allelic differences will be overlooked (Type II error) and the overall gain from selective application of MAS will more likely be less than with full scale MAS.

The proportion of correct contrasts increased when restrictions were applied to the F-ratio, estimated within family. Table 4.6 shows the proportion of correct contrasts and of selected young bulls inheriting the desired allele and the average TBV of the selected group for different levels of F-ratio of the significance test. When the F-ratio within a family was lower than the threshold, MAS was not applied and young bulls were selected

at random in those families. Average breeding values are presented for the families where MAS was used and for all families. The stricter the constraint on the F-ratio, the fewer grandson families are genotyped (Table 4.6) and the greater the average TBV of the young bulls selected by MAS. When the F-ratio was required to be greater than 10, only about 5% of the families were selected using MAS. This yielded to relatively high average TBV of the group selected by MAS (15.32) while the average TBV of the total group of selected young bulls was only slightly higher than when using random selection in all families (14.00 versus 13.94).

4.5 Discussion

In nearly all situations simulated, MAS schemes that considered a confidence interval were superior to selecting sons based only on the most probable location of the QTL. This difference tended to increase when family sizes were larger and when the QTL was less precisely localized.

The methods employing the CI seemed to be more robust than was MAX, probably because they considered a wider range of cM across the chromosome and were therefore less subject to inaccuracies in the estimate of the QTL location. The MAX method selected sons with the greatest probability (or lowest if the grandsire allele was unfavorable) of inheriting the grandsire allele at the most probable location. This practice simply favored the selection of bulls that had no recombination between the markers that flanked the most probable location, without regard to other regions of the chromosome. Therefore, some precision was lost if the true position of the QTL was outside the

markers flanking the most probable location. The CI methods, on the other hand, favored the selection of sons with little or no recombination in the entire region of the CI, which almost always included the true QTL position. The CI methods were particularly superior when selecting among full-sib sons of bulls who had a recombination event between the two markers that directly flanked the QTL. This difference was amplified when the estimated location of the QTL was relatively inaccurate.

To illustrate this problem with an example, Figure 4.2 has the genotypes of a grandsire and one of his sons that was chosen as a sire of sons in generation 2. As the figure shows, a recombinant haplotype was received by the son. The recombination occurred between the QTL and the estimated location of the QTL. In this case, the MAX method performed poorly as none of the selected sons carried the desired allele. On the other hand, by considering an interval of loci, the CI methods performed far better, selecting 48 (BOOT) and 49 (LOD) sons with the desired QTL allele out of 50 fullsibships.

The CI approach was superior to MAX regardless of the method used to form the CI and to the weight values within the CI. Although differences among CI methods were small and in some cases not significant, the methods that weighted each chromosome position based on results within grandsire family tended to yield the best results. A plausible explanation is that MAS, in general, was most effective within families of grandsires that had large differences between the QTL alleles. Including information from other families possibly only tended to add noise to the within family prediction of QTL location for these grandsires.

The estimates of CI by using LOD score approximation were shorter than CI calculated by bootstrapping for all scenarios. Van Ooijen et al. (1992) and Mangin et al. (1994, 1997) pointed out that the LOD score drop-off method can be biased downward for populations of small or medium size. This is because in those situations the distribution of the test statistic does not follow a chi-square distribution.

The genetic gains from MAS estimated in this study are relative to a single-generation application of MAS. If MAS was to be applied to the same locus in following generations, a further decrease in QTL variance would reduce the genetic gain. The decrease in QTL variance after three generations of selection was 47% for the base scenario. Selection based on EBV chose sires with the better alleles at the QTL, which may, therefore, be homozygous or have lower difference between the two QTL alleles. For example, the average contrast between alleles was shown to be decreased by nearly 75% after two rounds of selection in a similar simulation by Spelman and Garrick (1998). In contrast, in another simulation model, within-family MAS was applied to an unselected population, the relative genetic gain was approximately 10% (Mackinnon and Georges, 1998).

Spelman and Garrick (1998) achieved an increase in genetic progress of 0.3% to 1.6% by applying MAS by using the top down scheme in a population for a trait with h^2 = .30 for which 16.5% of the genetic progress was controlled by the QTL. The responses we observed with similar genetic parameters were greater than 2%. The probable reason for difference was the increased number of sires (50 versus 28) used to estimate the difference in effect of the QTL alleles within each grandsire. This resulted in greater accuracy and power in our study. Evidence of increased accuracy was observed when

comparing the percentage of correct contrasts, when no significance test was applied to the contrast estimates to restrict the use of MAS. In these situations (F-ratio > 0) in our study, the proportion of incorrect contrasts was approximately 32% versus 41% for Spelman and Garrick (1998).

Results from this study were based on a genetic model with 10 alleles at a single QTL. If a di-allelic model was used the increase in genetic progress would be lower (Spelman 1998).

Table 4.1 **Input parameters for the simulated population.**

Parameter	Values
σ_P	20
h^2	0.10, 0.30
σ^2_{QTL}	17%, 35%[1]
Family Size	3, 40
QTL location	Center, Telomere
Marker distance	20cM, 1cM

[1] percentage of the total additive genetic variance

Table 4.2 **Percentage of bulls chosen based on MAS that inherited the desired allele for three methods of weighting the probability of transmission of the grandsire haplotype at each centiMorgan within the confidence interval[1].**

	Confidence interval	
Weights	**BOOT**	**LOD**
Uniform	64.9[a]	64.8[a]
Across Grandsires	65.7[a]	65.8[a]
Within Grandsires	68.5[b]	67.6[b]

[1] $h^2 = 0.30$, $\sigma^2_{QTL} = 17\%$, family size = 40

[a,b] Values in the same column with different superscripts are significantly different ($p < 0.05$).

Table 4.3 **Percentage of selected young bulls that inherited the desired grandsire allele and the average and relative[1] increase in breeding value when young bulls were selected randomly or with three schemes of MAS[2].**

Selection	Successes		ΔBV	
Method	Overall %	Relative %	Mean[3]	%
MAX	63.0[a]	25.7	14.23[a]	2.00
BOOT	68.5[b]	36.7	14.27[b]	2.60
LOD	67.6[b]	34.9	14.30[b]	2.59
RANDOM	50.1[c]	...	13.94[c]	...

[1] Relative to random selection

[2] $h^2 = 0.30$, $\sigma^2_{QTL} = 17\%$, family size = 40

[3] Expressed as deviation of sire TBV from the parental mean

[a,b,c] Values in the same column with different superscripts are significantly different ($p < 0.05$).

Table 4.4 **Increase in average TBV[1] of bulls selected by three methods of MAS with low and moderate h^2 and low and high QTL variance[2].**

	Heritability			
	10		30	
	$\sigma^2_{QTL} = 17\%$	$\sigma^2_{QTL} = 35\%$	$\sigma^2_{QTL} = 17\%$	$\sigma^2_{QTL} = 35\%$
	----------	----------	% ----------	----------
MAX	2.09[a]	2.48[a]	2.00[a]	2.80[a]
BOOT	2.28[b]	3.01[b]	2.60[b]	3.20[b]
LOD	2.39[b]	3.01[b]	2.59[b]	3.11[b]

[1] Relative to random selection

[2] Family size = 40

[a,b] Values in the same column with different superscripts are significantly different ($p < 0.05$).

Table 4.5 **Increase in average TBV[1] of bulls selected by three schemes of MAS with different distribution of markers and QTL along the chromosome[2].**

	Distribution		
	Corse	Fine	Telomere
	----------%----------		
MAX	2.00[a]	3.64[a]	1.70[a]
BOOT	2.60[b]	3.79[a]	2.22[b]
LOD	2.59[b]	3.86[a]	2.01[b]

[1] Relative to random selection

[2] $h^2 = 0.30$, $\sigma^2_{QTL} = 17\%$, Family Size = 40

[a,b] Values in the same column with different superscripts are significantly different ($p < 0.05$).

Table 4.6 **Effect of selectively applying MAS to families exceeding F-ratio threshold[1,2].**

F-ratio threshold	Correct contrasts (%)	Selection success[3] (%)	Average TBV[4]		Families genotyped (%)
			MAS[5]	All	
0	68.5	78.5	14.30	14.30	100.0
1	70.8	80.1	14.33	14.26	83.4
2	73.5	81.6	14.40	14.22	61.8
5	79.9	85.2	14.58	14.09	25.0
10	88.3	94.7	15.32	14.00	4.9

[1] MAS was performed using LOD

[2] $h^2 = 0.30$, $\sigma^2_{QTL} = 17\%$, Family Size = 40

[3] Percentage of selected young bulls inheriting the correct allele

[4] Expressed as deviation of sire TBV from the parental mean

[5] MAS is applied only to families for which F values was greater than threshold

Figure 4.1 **Structure of the nucleus population.**

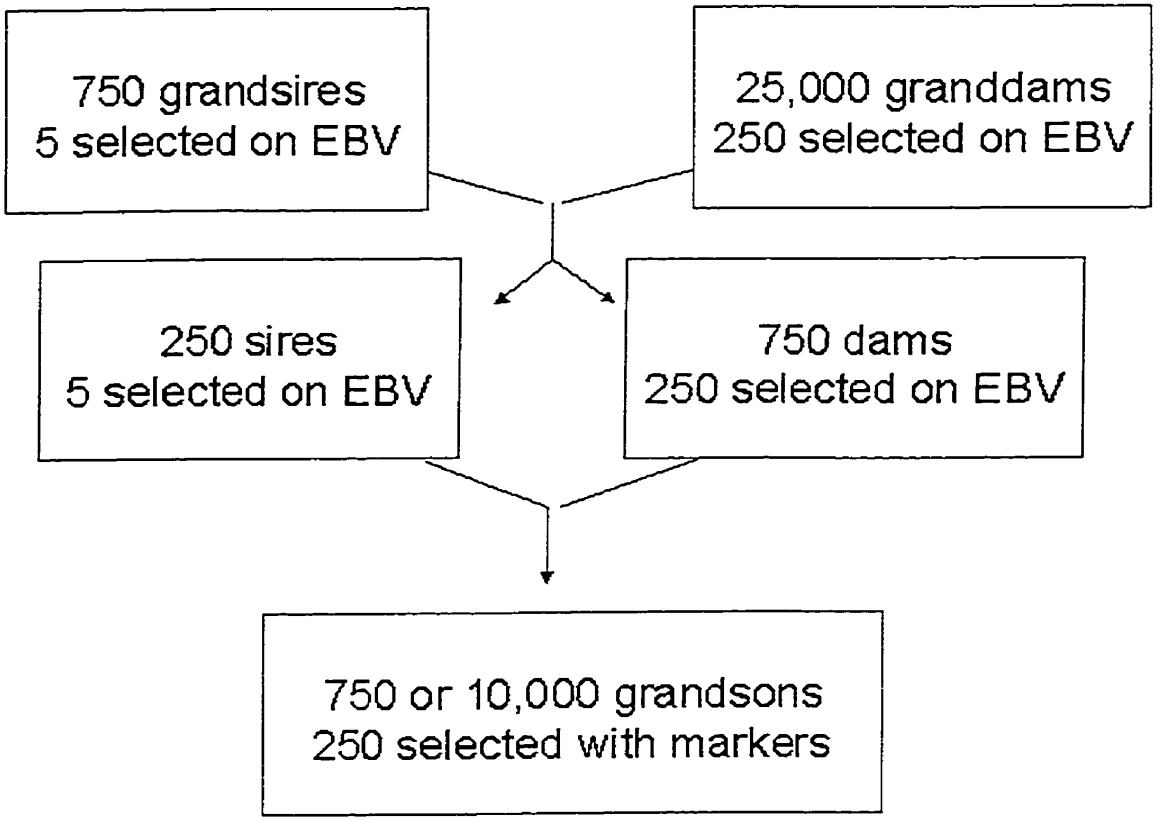
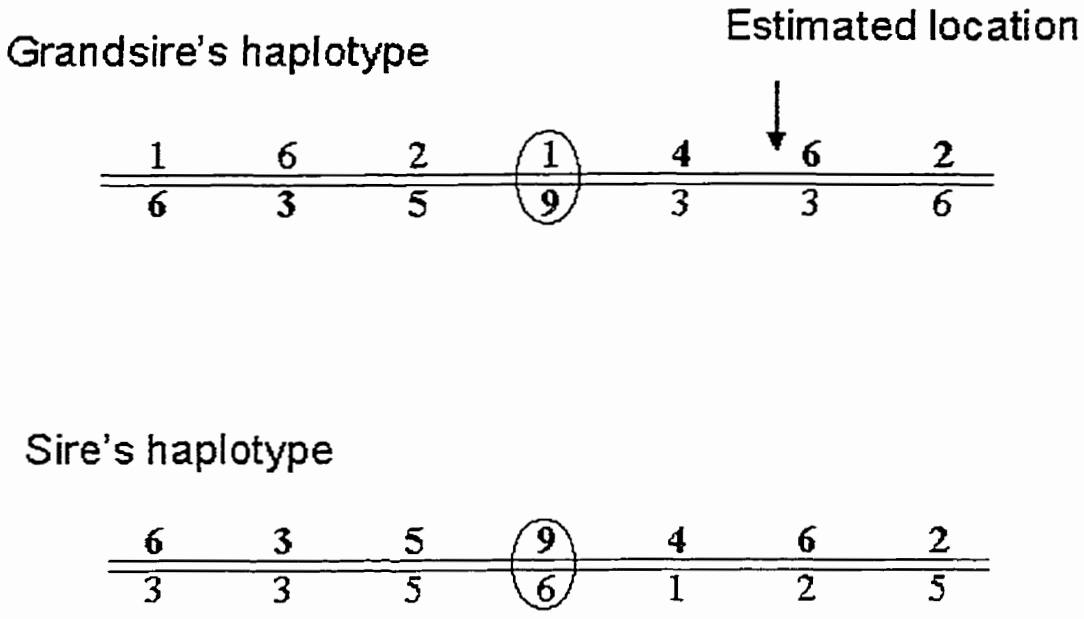


Figure 4.2 **Example of a situation where the confidence interval methods excel.**



5. Strategies for continual application of MAS in an open nucleus population.

5.1 Summary

The objectives of this study were to develop and simulate the implementation of several strategies for repeated application of QTL detection and MAS and to compare the short-term and continual genetic responses. A finite locus model was simulated with 20 QTL randomly distributed across 30 chromosome. Three hundred markers were evenly spaced across the genome. Allelic effects were sampled from a double exponential distribution. A daughter design was used, every generation, to determine the marker alleles favorably associated to QTL alleles. The MAS was applied within family to young bulls, prior to progeny testing, as part of an open nucleus. Young bulls were selected using strategies based on a) the single marker with greatest contrast (BEST*1*), b) the sum of *n* greatest contrasts (BEST*n*), c) the best *n* contrasts, limited to one per chromosome (LIM*n*), d) the sum of all contrasts exceeding a given threshold (THRES), and e) the sum of contrasts exceeding a threshold, but limited to one per chromosome (LIMT). The maximum progress was achieved by strategies that selected upon several markers flanking multiple QTL in each generation. When THRES was applied, the mean TBV of selected bulls was increased by 11.98% (over conventional selection) versus 6.73% for BEST*1* in the first generation. Applying a full genome scan in each generation allowed selection for different QTL across time. By selecting for multiple QTL over time, MAS maintained superiority over conventional selection for many generations.

5.2 Introduction

Numerous studies have demonstrated that marker assisted selection (MAS) is effective in increasing genetic response in the short term (Whittaker et al., 1995; Ruane and Colleau, 1996; Mackinnon and Georges, 1998; Spelman and Garrick, 1998). These studies have generally examined selection at a single locus across the population. However, when selection occurs, the variability of alleles at a QTL decreases from generation to generation. This decrease may be especially large at loci where direct MAS is applied (Chapter 4). Advantages of marker assisted selection programs can, therefore, be sustained in the long term only if new QTL are continually discovered and selected upon. Re-testing for QTL may also be necessary to account for recombination between the QTL and the markers (Zhang and Smith, 1993; Whittaker et al., 1995). These "new" QTL may be smaller than the "original" QTL, because the larger QTL are the most likely to be found early or moved to fixation via the forces of conventional selection. Thus, the gains in later generations may be reduced because the maximum potential (allelic substitution effect) of successive QTL is likely to be smaller and the type I error of the QTL detection test is likely to be larger (Whittaker et al., 1995; Meuwissen and Goddard, 1996; Spelman and Van Arendonk, 1998; Villanueva et al., 1999).

The objective of this study was to develop and simulate several strategies for repeated application of QTL detection and MAS and compare the strategies for short-term and continual genetic response. The sustainability of genetic response over several generations of QTL detection and MAS in a dairy nucleus herd was compared to genetic progress achieved with conventional selection. Sustainability was evaluated by the

number of QTL exploited in later generations compared to early generations. Stochastic simulations were developed based on a finite locus model.

5.3 Methods

The simulated population included an open nucleus and a commercial herd. Selection for a sex-limited trait was modeled. The MAS was applied based on a complete genome scan during each generation, using the bottom-up design by Mackinnon and Georges (1998). Associations between QTL and markers were identified based on genotypes and phenotypic records of progeny test daughters. This information was used for within family selection of young bulls entering the progeny test program.

Population

The structure is shown in Figure 5.1. The base population consisted of 1500 and 50,000 unrelated and unselected males and females, respectively. To form the nucleus, the top ten sires (0.66%) and top 500 dams (1%) were selected from the base population according to their EBV. For all individuals, EBV were simulated by adding random, normally distributed, prediction errors to the true breeding values (TBV), using the approach of Spelman and Garrick (1998):

$$EBV_i = r_{TI}^2(BV_i - MBV)\, z_i\sqrt{(V_p(h^2 r_{TI}^2 - h^2 r_{TI}^4) + MBV}, \qquad [1]$$

where, r_{TI}^2 was the squared accuracy of selection, BV_i was the true breeding value, MBV was the population mean breeding value in the current generation, z_i was a

standard normal deviate, and V_p was the phenotypic variance. The standard deviation of the prediction error was varied by altering r_{TI}^2 according to sex and generation to reflect the amount of information typically available for prediction of BV. The EBV of the sires were assumed to have accuracies corresponding to a progeny test with 100 daughters. Dams for the nucleus were selected based on EBV with accuracy corresponding to one-lactation record. Each of the 10 sires was mated randomly to 50 nucleus dams to generate full-sib families of three males and three females. Within each family, the best male was selected based on his genotype, considering results of a daughter design conducted on each of the ten sires. All 1500 females produced were eligible for selection on their EBV, based on one completed lactation. No additional fixed effects (such as age, parity, etc) were generated. Among these females, the top 400 were selected as part the next generation of dams. In addition, 100 dams were randomly selected from the top 2% of a commercial population. The commercial population was comprised of 50,000 cows that were sired by the top 50 bulls in the previous generation. These cows had one gamete from one of the sires and the other gamete was generated based on allelic frequencies in the commercial population. The allelic frequencies in the commercial population were computed in each generation as the average of previous generation and the average of the 50 selected bulls. For the base scenario 5 generations were simulated.

Genetic model

The TBV were produced assuming a finite locus model for which all the allelic effects were strictly additive. The genetic model was based on the general approach of Mackinnon and Georges (1998). The parameters used to simulate the base situation are

in Table 5.1. Several comparisons were made for certain scenarios generated with alternative parameters. In the base situation, all the genetic variability in the trait of interest was explained by 20 loci, each with 5 alleles. Mackinnon and Georges (1998) also considered QTL with multiple alleles. Multiple alleles have been reported for several genes, including blood groups, milk protein genes, and loci in the major histocompatability complex. Different QTL effects and positions were randomly assigned in each replicate. Allelic effects were simulated using a double exponential distribution. Compared to the normal distribution, the double exponential distribution is more sharply peaked, yielding relatively fewer intermediate allelic effects and slightly more large ones. The initial QTL allele frequencies were uniformly distributed and standardized to sum to 1.0. The phenotypic variance was 400. In the base scenario, heritability was 0.30. The QTL were randomly distributed across the genome, with no upper limit placed on the number of QTL per chromosome.

Thirty chromosomes of 90 cM each were simulated. There were 300 co-dominant markers, with 6 alleles each, distributed across the 30 chromosomes, evenly spaced and separated by a recombination rate (θ) of 0.10.

Marker assisted selection

A daughter design was used, every generation, to determine which marker alleles were associated with the favourable QTL alleles. Each sire had 100 daughters from the commercial population, with a lactation record and complete marker genotypes. Marker contrasts were calculated for all heterozygous markers within sire. All daughters that were informative at a given locus were used to calculate the contrast, ignoring the

genotype of the dams. The contrast was simply the difference in the mean of EBV of daughters carrying the alternative sire alleles. These EBV were based on one lactation record. Contrasts were then divided by their standard errors to account for differences in the number of informative daughters across loci.

Marker assisted selection was applied to select the best sons from among full-sib families of three males. Selection was based on one or more of the markers with the highest contrasts. If all the sons were uninformative or carried the same sire allele then selection was random.

The efficiency of QTL detection experiments and MAS varies depending on the genetic variance (Lande and Thompson, 1990; Ruane and Colleau, 1996; Meuwissen and Goddard, 1996), so two alternate levels of heritability were considered (0.10 and 0.50, Table 5.1). Also, the effect of the map density was examined. In addition to the base scenario, responses to selection were examined when using a relatively FINE ($\theta = 0.05$) or COARSE ($\theta = 0.15$) genetic map of markers. For these alternative maps, the number of markers was changed and the length of the genome remained constant.

Selection strategies

Five different selection strategies were examined (Table 5.1). In the first approach (BEST*1*), which served as the basis for most comparisons, the selection was based on the single best marker contrast for each sire. The selected marker could be different for each sire. The second strategy selected the young bull based on the best n marker loci, where n was equal to 3, 5 and 10 (BEST*3*, BEST*5* and BEST*10*, respectively). For each son an index was calculated upon which to base the selection:

$$I_{ij} = \sum_{k=1}^{n} (-1)^m * C_{ik} \qquad [2]$$

where i is the sire number, j is the son number, k is the locus number, n is 3, 5 or 10, C_{ik} is the contrast for sire i at locus k, and m is and indicator variable with value 0 or 1 if the first or second sire allele was inherited, respectively. This index differed for each son depending on which sire alleles were inherited.

The third strategy was a variation of the second. This approach first identified the marker with the highest contrast for each chromosome and then selected bulls based on an index (equation 2) for the best n of these markers. The limit of a single marker per chromosome was designed to prevent selection on multiple markers linked to the same QTL. Two levels of n were used, 3 and 5 markers (LIM*3* and LIM*5*, respectively).

The fourth strategy (THRES) involved considering all the markers for which the standardized contrast exceeded a given threshold. For the base situation, the threshold was set at 1.96, to reflect an α = 5% significance test (THRES*5*). For comparison, the simulation was repeated by using lower (THRES*10*) and higher (THRES*1*) to reflect α = 10% and α = 1%, respectively.

The fifth strategy, LIMT, combined aspects of LIM and THRES. First, the marker with the greatest contrast was identified for each chromosome. Then selection was based on all of these 30 markers with contrasts that exceeded the threshold 1.96.

The final strategy (DD) evaluated the feasibility of applying the results of the daughter design in the first generation to making selection decisions in subsequent generations without repeating the genome scan. Markers were traced from the initial sires of sons to their offspring and MAS was applied to choose sons of informative sires.

No specific effort was made to locate and characterize the QTL, but the QTL were assumed to be nearest to the markers with the greatest contrast. Because selection was based on an index of contrasts [Equation 2], the markers most likely to be near QTL implicitly received the greatest weight in the index.

Generational and cumulative responses to MAS were compared to response to conventional selection (random within-family selection). The extra gain per generation from MAS was the difference between the means of TBV of bulls selected using markers and bulls selected randomly. This quantity was expressed as a percentage by dividing by the gain achieved through random selection.

Many of the past theoretical studies on MAS have assumed a single QTL (Ruane and Colleau, 1996; Spelman and Garrick, 1998), clusters of QTL (Meuwissen and Van Arendonk, 1992; Meuwissen and Goddard, 1996) or few QTL (Spelman and Bovenhuis, 1998; Schulman et al., 1999), with previous knowledge of the position of the QTL relative to flanking markers. Thus, in these studies, the QTL being selected was predetermined. In contrast, the present study assumed no previous knowledge of QTL location. Therefore, several analyses were performed to investigate which markers were used in selection decisions, their location relative to the QTL, and the size of the QTL, in terms of genetic variance. To do this, the marker upon which selection was applied was monitored for each sire. The location of the marker relative to the nearest QTL was observed. The proportion of times that a QTL was adjacent to the marker, within three marker intervals or simply on the same chromosome was recorded. In addition, the genetic variance contributed by the marked QTL was recorded and all QTL were ranked in terms of their genetic variance.

Some previous studies (e.g. Schulman et al., 1999) have shown that MAS induces negative linkage disequilibrium among QTL greater than that expected to occur as a result of conventional selection. Negative linkage disequilibrium occurs when a negative covariance exists between the selected locus and the remaining loci. This disequilibrium has the potential to reduce genetic gain from MAS and possibly diminish the power of experiments designed to detect QTL. To test for evidence of increased disequilibrium (manifested in decreased power) due to MAS, we compared the mean value of contrasts following 5 generations of MAS versus 5 generations of conventional selection. However, changes in the relative magnitude in the contrasts between MAS and conventional selection could also be the result of differences in allelic frequencies and decreased genetic variance due to increased selection response from MAS. To account for differences in the contrasts due to changes in allelic frequencies, the allelic frequencies for both MAS and conventional schemes were recorded after 5 generations (GEN5) and then used as initial gene frequencies to establish a new, unrelated base population (NEW). Then, marker contrasts were re-estimated in this new base population. The difference between maximum marker contrasts from the GEN5 and NEW were then compared for MAS and conventional selection. A significant difference between the magnitude of contrasts from GEN5 and NEW populations was assumed to indicate a difference in disequilibrium.

To compare long term responses for marker assisted and conventional selection, the simulation was allowed to continue for many generations until the genetic variance was nearly exhausted (< 0.1% of the original variance) and cumulative response to MAS was compared to conventional selection. Much uncertainty and controversy exists

regarding the true number of genes affecting quantitative traits (Lande, 1981; Zeng, 1992). Therefore, for this comparison, a situation in which the trait was controlled by 50 QTL with 5 alleles each (Hayes and Goddard, 2000)was also simulated.

5.4 Results and Discussion

The following results pertain to the base set of parameters outlined in Table 5.1 unless stated otherwise.

Under the genetic parameters of the model with 20 QTL, on average (across replicates) 13.1% of the individual QTL accounted for at least 10% of the total genetic variance. With 20 QTL, a single randomly selected QTL was expected to account for 5% of the total genetic variance. Also, 2.3% of the QTL accounted for at least 20% of the variance, and 0.5% and 0.1% of the individual QTL accounted for at least 30 and 40%, respectively, in the base population. On average, the maximum difference between the highest and lowest allelic effect for any QTL was 9.90, which corresponds to approximately 0.9 genetic standard deviations or 0.5 phenotypic standard deviations. Previous analyses of field data (Ashwell et al., 1997; Georges et al., 1995) have uncovered QTL with estimated effects of this size or greater.

Table 5.2 shows the percentage increase above random selection in the average TBV of the chosen progeny test bulls for each of the five MAS approaches. This value was calculated as previously explained in Chapter 4.

The percentage increase ranged from 6.73 for the BEST*1* to 11.98 for THRES*5*, in the first generation. Gains were roughly halved by generation 5. Cumulative gains in

the mean TBV of the commercial population are also presented (Table 5.2) and they followed the same pattern. Cumulative gains in the commercial population (ranging from 1.68% to 2.77%) were less than the average advantage per generation for young bulls because selection of young bulls is only one of the factors that affect genetic progress in the commercial population.

One can consider BEST*1* and BEST*3* as applications of THRES with much higher thresholds. On average, 19 marker locations exceeded the 1.96 critical value for THRES*5* (compared to 1 and 3 for BEST*1* and BEST*3*). Clearly, some of the potential gain from MAS was lost when a too stringent critical value was applied. Basing selection decisions on a fixed threshold rather than on a fixed number of loci allows many potential loci on several chromosomes to be considered, if statistical evidence supports the segregation of numerous loci. In addition, the threshold method easily allows for selection based on an interval of markers surrounding the same QTL. This helps to account for imprecision in the location of the QTL with respect to the markers (on the contrary BEST*1* assumes that the QTL is in the interval adjacent to the marker), therefore increasing response to MAS (see Chapter 4, Appendix 1). Another potential advantage for the threshold was that it could allow for variability among the sires in the numbers of marker contrasts used in selection. With the THRES*5*, sires ranged from 4 to 39 (sd = 6.7) in the number of marker contrasts that exceeded the threshold, which indicates variability among sires in the number of segregating QTL with widely different allelic effects. When BEST*10* was applied, cumulative gain was 2.80%, which exceeded (although not significantly) the gain from THRES*5*. Therefore, the advantage of THRES*5* over the other approaches was more a function of selection on more loci than of

allowing variability among sires in the number of markers that contributed to the selection criterion.

Some efficacy is lost by applying a very liberal threshold because of excessive type I errors. For this simulated population, the optimal threshold was within the range corresponding to a comparison wise α of between 1 and 5%. When applying the higher threshold ($\alpha = 1\%$), the cumulative genetic gain over 5 generations was 2.84%, which was not significantly greater than the response of 2.77% achieved when the threshold was $\alpha = 5\%$. When the lower threshold ($\alpha = 10\%$) was applied, cumulative genetic gain was significantly ($P < 0.01$) decreased, at 2.52%.

Response to MAS was decreased, relative to THRES*5*, when applying the LIM*3*, LIM*5* and LIMT approaches, which restricted selection to a single marker per chromosome. Although these methods outperformed BEST*1* because they allow for selection on multiple QTL, by restricting selection to a single marker per chromosome, some precision was lost when the underlying QTL was not in an interval adjacent to the selected marker. A selected son could share the sire's haplotype at the selected markers but not the desired QTL allele, if recombination occurred. No significant differences in selection response were observed when considering 3 versus 5 loci (LIM*3* versus LIM*5*).

For all approaches, the advantages of MAS decreased with each generation. These decreases occurred because selection decreased the amount of genetic variability in the population and, therefore, the expected difference between sire alleles at the QTL. This loss of variability not only decreased the potential gain that could be achieved, but also decreased the accuracy of the estimates of associations between markers and QTL. The number of markers exceeding the given thresholds also decreased each generation.

The risk associated with the various strategies can be evaluated by comparing the mean advantage in TBV of selected bulls with the standard deviation (SD) of this advantage (Table 5.2). For example, in Generation 1, the advantage of the BEST1 strategy was 6.73% with a SD of 2.94%. Thus, the TBV of bulls selected by BEST1 exceeded that of the randomly selected bulls by an average of 2.28 SD. Assuming that the advantage was normally distributed, the randomly selected young bulls were expected to be equal or superior to the MAS bulls in only about 1% of the replicates. Not surprisingly, the more successful strategies were even less risky, as the means in the advantages of these strategies in TBV for Generation 1 were all in the range of 3.0 SD, meaning that the TBV of the randomly selected bulls would be expected to meet or exceed the TBV of MAS bulls in fewer than 0.15% of replicates.

As expected, the risk increased as more generations of MAS were practiced and the advantage of the MAS bulls decreased. In Generation 5, the TBV of THRES5 bulls was only 1.70 SD greater than randomly selected bulls on the average, so the randomly selected bulls were superior to the MAS bulls about 4.5% of the time. This level of risk was still rather low.

The cumulative response in the commercial population after five generations followed similar patterns. When BEST1 was practiced, the mean TBV of the commercial population was approximately 2.15 SD above the mean when only conventional selection was practiced (1.58% risk). By practicing THRES5, this risk was <0.05%.

Table 5.3 gives, for each generation, the frequency for which the marker with the highest contrast was located near a QTL, for the BEST*1* strategy. Clearly, the probability that a QTL existed nearby the marker with the highest contrast decreased over

generations. For example, in the first generation there was a 60% probability that a QTL was located in the interval between the selected marker and the next marker on the chromosome. This dropped by almost half (to 33.8%) in generation 5. Nevertheless, MAS was still effective in the fifth generation because this percentage was still 3 times greater than probability expected by choosing a marker at random. In addition, about 55% of the time a QTL was within 30 cM from the marker with the best contrast and 69% of the time on the same chromosome. Approximately 12% of the time no QTL was found on the same chromosome as the selected marker, even in the first generation. This result stresses the importance of considering multiple markers by using one of the other methods such as THRES.

Although the marker with the largest contrast was usually near a QTL, this QTL was rarely the most important locus as measured by the amount of genetic variance (in the commercial population) accounted for by the QTL (Table 5.4). For each replicate and generation, each QTL was ranked based on its genetic variance in the commercial population. For each sire, we determined the QTL that was being selected upon by the marker with the greatest contrast. Only 10.10 % of the times the selected QTL had the most variance.

On average, selection was for the seventh most variable QTL in generation 1 and this value decreased with increasing generations (Table 5.4). When using an additive model, the variance accounted for by a QTL is a function of allelic frequencies and substitution effects (Falconer and McKay, 1996). With large substitution effects, conventional selection would favor animals that are homozygous for the most favorable locus. Therefore, even if a QTL had the greatest genetic variance in the commercial

population, many of the nucleus sires may have been homozygous for the most favorable allele (Table 5.4). If so, then markers near this QTL would not show a large contrast. Spelman and Garrick (1998) demonstrated, under similar conditions of selection intensity, but using a single QTL, that the average contrast decreased by 25% in two generations of selection. For each QTL, the correlation for that QTL was calculated between the genetic variance contributed by that QTL and the proportion of sires that were homozygous for that QTL. We also recorded the proportion of sires that were homozygous at the QTL that contributed the most genetic variance. The correlation between QTL variance and proportion of homozygous sires was positive and significant ($P<0.0001$), ranging from 0.23 in generation 1 to >0.5 after generation 3 (Table 5.4). The proportion of sires homozygous at the most variable QTL was 0.3 in first generation (expected = 0.2 in the absence of selection) and more than 0.6 in generation 5.

Although QTL detection experiments have been shown to be most powerful when applied across family (Spelman et al., 1999), these results underscore the importance of considering results within family when subsequently applying MAS. Not only must transmission studies within family be used to determine the phase of the association between marker and QTL (Weller et al., 1990), but also to determine whether each family is actually segregating. These results indicate that as QTL variance (and thus the power for detection) increases, the probability that the QTL will be segregating within highly selected animals is decreased.

No difference was observed in the relative effectiveness of MAS with respect to the true heritability of the trait. For all three heritability levels (0.10, 0.30, 0.50), the average TBV for young bulls chosen with MAS in the first generation was 6.73% greater

than of those bulls selected randomly. These percentages varied slightly in later generations, but not significantly. With a lower heritability, the marker contrasts were estimated less accurately, but this factor was balanced by the fact that sires were less likely to be homozygous for the most variable QTL. These results differed from other studies (Chapter 4; Ruane and Colleau, 1996), but those studies considered only a single QTL. Moreau et al. (1998) showed that the response to MAS decreased for very low heritabilities (< 0.15). They estimated the existence of an optimal heritability around 0.15- 0.20, varying slightly depending on the percentage of genetic variance associated with the markers.

The effect of marker density on efficiency of MAS was also examined (Table 5.5). As expected, finer mapping yielded a superior group of progeny test bulls. However, the marginal advantage of increasing the marker density decreased as the density increased. For, example, when the COARSE mapping strategy was applied in generation 1, the average TBV of selected young bulls exceeded the EBV of randomly selected bulls by 6.28% (Table 5.5). This gain was increased by 7.2% (to 6.73% for BEST*I*) by increasing the marker density by 67% (from 6 to 10 markers per chromosome). Increasing the number of markers per chromosome by an additional 80% (from 10 to 18 with FINE) resulted in an increase of only 3.4% (to 6.96%). In fact, when cumulative gain was considered over 5 generations, the difference between using the base (BEST*I*) and FINE map was not significant ($P > 0.10$). The reason to place several markers on a chromosome is to use recombination events between markers to help locate the QTL more precisely. Spelman and Bovenhuis (1998) showed that, for this reason, having smaller marker brackets increased the response to MAS. In our study,

diminishing returns with finer maps occurred because, as maps become more dense, the number of daughters (or sons) with a recombination between two particular markers decreases.

Some breeding companies may be interested in trying to apply the DD approach because the QTL detection is only performed once and, therefore, costs decrease greatly. However, the benefits of applying MAS also decreased greatly. In the second generation of MAS, the average TBV of young bulls was only 1.27 % greater than random within-family selection and the advantage was nearly negligible in later generations. If within-family MAS is effective, nearly all surviving members of a given sire family will have inherited the best alleles at the loci upon which MAS was applied.

Means of the maximum contrasts (absolute value) following five generations of conventional selection and MAS are in Table 5.6. The means of maximum contrasts when the resulting allelic frequencies were used to establish new populations (NEW) are also presented. The maximum contrast following MAS (3.18) was significantly ($P<0.0001$) but only slightly lower than after the same period of random within-family selection of progeny test bulls (3.26). No difference was observed in the maximum contrasts for NEW populations established with the resulting allelic frequencies, so the difference between GEN5 and NEW contrasts was also greater ($P<0.0001$) for MAS than conventional selection. This difference was likely due to increased negative linkage disequilibrium among QTL when MAS was applied. Contrasts were greater in NEW versus GEN5 populations for both selection schemes. Increased genetic relationships among animals in the commercial population after 5 generations of selection were likely responsible for much of this difference, but should not have greatly affected the

difference between MAS and conventional selection in GEN5 because MAS was applied only within families.

Previous studies have demonstrated inferior long-term genetic response to MAS when compared to conventional selection (Gibson, 1994; Villanueva et al., 1999). That phenomenon was not observed in our study for either 20 or 50 QTL models. Cumulative response with marker assisted selection was consistently greater than with random selection within full-sib family, even after selection had continued to the point where genetic variance was reduced to less then 0.01% of the initial variance. Our study differed from the previous, however, in several features. First, selection strategies differed. In our study, MAS was applied only to choose progeny test bulls from within families. In contrast Gibson (1994) and Villanueva et al. (1999) practiced population wide selection based on an index of the marked gene and the remaining polygenes. Also, previous studies simulated continued MAS on a single locus, rather than a varying array of loci as was the case in the present study. Applying MAS on many loci may be a more robust approach that helps prevent drift from fixing some loci at non-favorable alleles.

5.5 Conclusions

Applying full genome scans in each generation allowed for selection on different QTL across time. Within a given family, repeated selection for the same marker locus over multiple generations was of little value because of loss of information and decreased genetic variability from previous selection. Maximum progress was achieved by selecting in each generation upon several sets of markers neighboring multiple potential QTL,

rather than only the most likely QTL. The markers selected in each family should be based on a significant threshold rather than on arbitrary limitations on number of markers per chromosome or family.

The cost of the simulated MAS program would likely be prohibitive with current genotyping technology. Therefore, perhaps some combination of repeated scans on targeted areas of the genome and application of previously obtained results may be more efficient for breeding companies. Future technologies may, however, greatly decrease the cost of repeated full genome scans.

Within family MAS was rarely applied to the most important QTL in the commercial population because sires of sons were often homozygous for the best alleles at these loci. When MAS was practiced on different QTL in different generations, cumulative response with MAS maintained an advantage over response from conventional selection.

Table 5.1 **Simulation parameters.**

Parameter	Base scenario	Alternatives
Heritability	0.30	0.10, 0.50
Number of QTL	20	50
Recombination rate between markers	0.10	0.05, 0.15
MAS strategy	BEST*1*	BEST*n*, LIM*n*, THRES*n*, LIMT, DD

Table 5.2 **Average (SD) increase[1] in mean TBV of the selected group of young bulls when MAS is applied within family and cumulative gain (relative to conventional selection) in the commercial populations after five generations of selection.**

	MAS method				
Generation	**BEST1**	**BEST3**	**THRES5**	**LIM3**	**LIMT**
	----------%----------				
1	6.73^{a} (2.94)	10.14^{b} (3.34)	11.98^{c} (3.35)	9.72^{d} (3.33)	9.55^{d} (2.99)
2	5.34^{a} (2.94)	8.32^{b} (3.46)	9.85^{c} (3.47)	7.91^{d} (3.45)	7.89^{d} (3.23)
3	4.82^{a} (2.85)	7.44^{b} (3.33)	8.62^{c} (3.32)	6.97^{d} (3.48)	7.05^{d} (3.07)
4	4.16^{a} (2.95)	6.25^{b} (3.30)	7.04^{c} (3.40)	5.94^{bd} (3.30)	5.69^{d} (2.91)
5	3.43^{a} (2.81)	5.20^{b} (3.41)	5.79^{c} (3.41)	4.81^{b} (3.20)	4.63^{b} (2.80)
Cumulative[2]	1.68^{a} (0.78)	2.31^{b} (0.82)	2.77^{c} (0.82)	2.29^{b} (0.81)	2.34^{b} (0.80)

[1] Expressed as a percentage of the genetic response with conventional selection.

[a,b,c,d] Values in the same row with different superscripts are significantly different ($p < 0.0001$)

[2] Cumulative gain in the commercial population

Table 5.3 **Frequency of finding the largest marker contrast near a QTL and the expected values for a marker selected at random.**

Generation	Position of the best contrast		
	Within ± 10cM	Within ± 30cM	Whole Chromosome
	----------%----------		
1	56.9	78.7	87.2
2	47.5	69.9	80.6
3	43.2	65.6	76.9
4	37.9	59.6	72.9
5	33.8	54.9	69.0
Random	12.5	29.9	49.2

Table 5.4 **Average rank in genetic variance[1] of the QTL selected upon, the average percent of sires that were homozygous for the QTL that contributed the greatest genetic variance[1] and the correlation (r) of genetic variance[1] per QTL with the proportion of sires (r) that were homozygous.**

Generation	Rank[2]	Homozygous sires	r
1	7.16	30.6%	0.23
2	9.06	37.7%	0.38
3	10.05	44.2%	0.47
4	11.00	52.3%	0.51
5	11.91	60.3%	0.53

[1]In to the commercial population.
[2]Out of 20.

Table 5.5 **Effect of marker density on the increase[1] in mean TBV of the selected group of young bulls when MAS is applied within family.**

Generation	Marker Density[2]		
	Coarse (θ=0.15)	BEST1 (θ=0.10)	Fine (θ=0.05)
1	6.28	6.73	6.96
2	4.85	5.34	5.60
3	4.25	4.82	5.05
4	3.63	4.16	4.37
5	3.08	3.43	3.58

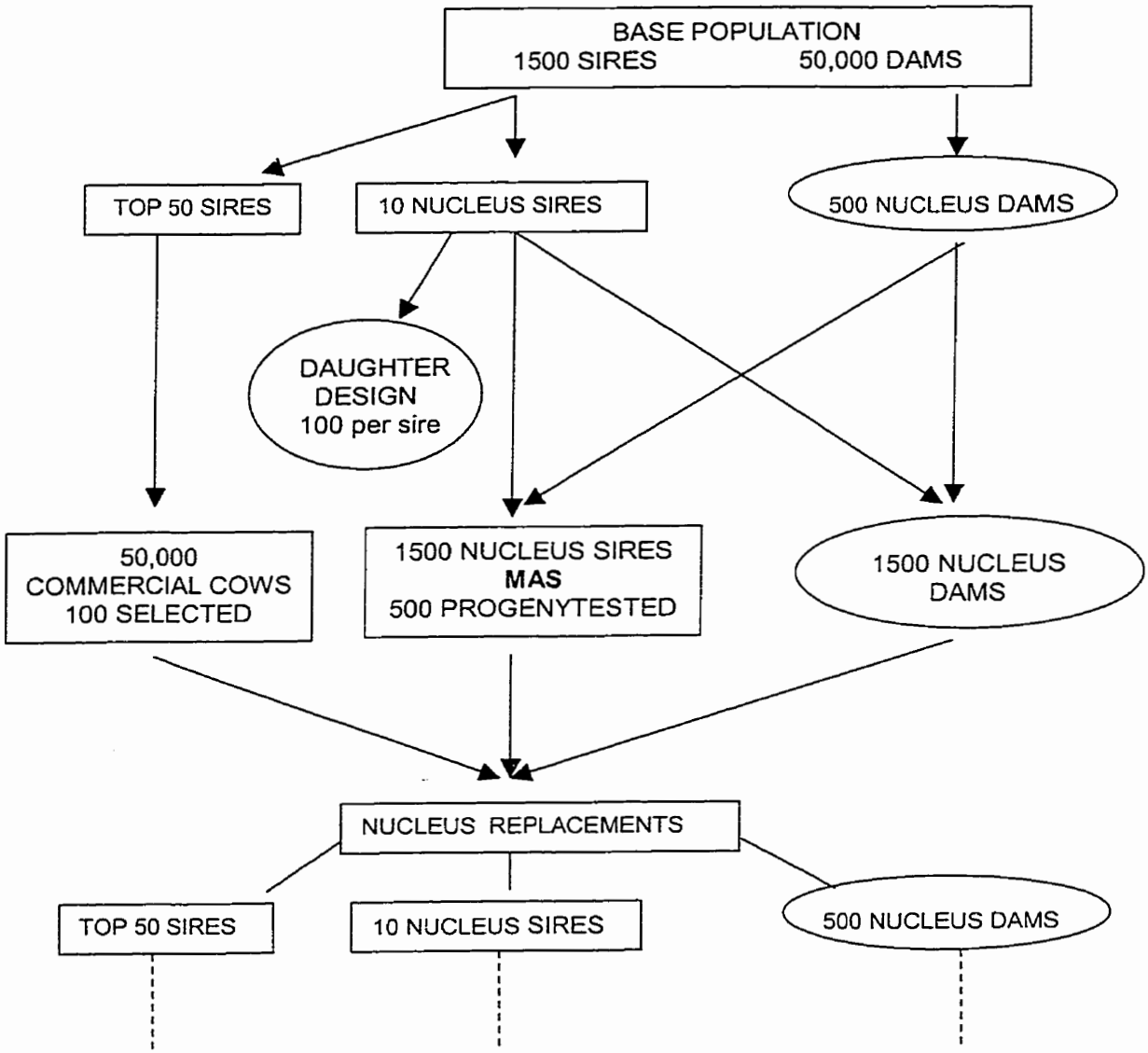
[1]Expressed as a percentage of the genetic response with conventional selection.
[2]Recomination rate = 0.15, 0.10, and 0.05, for Coarse, Best1, and Fine, respectively.
θ = recombination rate between adjacent markers.

Table 5.6 **Mean of maximum contrasts[1] following five generations (GEN5) of conventional and marker assisted selection (MAS) and in unrelated populations with the same allelic frequencies (NEW).**

Population	Selection scheme	
	Conventional	MAS
GEN5	3.26[a]	3.18[b]
NEW	3.46[a]	3.46[a]
Difference[2]	-0.20[a]	-0.28[b]

[1]Standardized
[2]Difference between GEN5 and NEW
[a,b] Values with different superscripts in each row are significantly different ($p < 0.0001$)

Figure 5.1. **Structure of the population.**



6. MAS for a multiple trait objective in an open nucleus herd.

6.1 Summary

The objective of this study was to examine the efficacy of applying MAS when the selection goal included two traits. A nucleus herd was simulated for five generations and MAS was applied within full sib families to choose bulls for progeny testing. A finite locus model was assumed with 20 QTL on 30 chromosomes. Three hundred markers were distributed evenly across the genome and a daughter design was used in each generation to determine which markers were associated with QTL. The selection goal included two traits. Trait 1 had an economic weight three times greater than trait 2. Heritabilities were 0.30 and 0.10 for traits 1 and 2, respectively, and the genetic correlation was either –0.25 or 0.25. Multiple trait MAS was applied in two ways, 1) by calculating separate contrasts for the two traits and combining the results together, or 2) calculating a single contrast using an index of the two traits. Multiple trait MAS was compared to applying MAS for trait 1 only and conventional selection alone. Situations where trait 2 directly affected fitness were also simulated. Multiple MAS decreased response for trait 1 relative to both single trait MAS or conventional selection. However, response for trait 2 increased to a greater degree and, therefore, response for the index was greatest with multiple trait MAS. This result was consistent whether the traits were positively or negatively correlated. No significant differences were observed between the two approaches for multiple trait MAS.

6.2 Introduction

Many studies have demonstrated that genetic response can be increased by applying marker assisted selection to increase selection accuracy (e.g. Lande and Thompson, 1990; Spelman and Garrick, 1997; Mackinnon and Georges, 1998). These studies have generally been applied for the genetic improvement of a single important trait. Many traits are recorded for dairy cattle and experiments to detect relationships between genetic markers and underlying QTL have been applied for most of these traits (e.g. Georges et al., 1995; Spelman et al., 1996; Ashwell et al., 1997) and several approaches to multitrait analysis for QTL mapping have been proposed (Ronin et al., 1995; Weller et al., 1996; Knott and Haley, 2000). However, with the exception of a few studies (DeKoning and Weller, 1994; Xie and Xu, 1998; Bernardo, 1999), work on the application of QTL information in breeding programs has generally focused on improvement for response to single trait selection objectives.

Single trait MAS is unlikely to be the most beneficial approach for application of information about QTL in selection programs. Although milk production is of primary concern for dairy cattle, other traits are economically important. For example, in Canada, many dairy traits are recorded and evaluated genetically. Commercially available sires can be selected based upon indexes that account for type traits, longevity, and udder health, in addition to production (Dekkers, 1995). Also, many of the non-production traits such as mastitis resistance (Emanuelson et al., 1988; Poso and Mantysaari, 1996) and reproductive performance (Roth et al., 1999) have unfavorable genetic correlations with production (Pryce et al., 1997), which may affect the efficacy of MAS. Also, phenotypic effects of reduced fertility due to high production (Roth et al., 1999) could

affect the number of offspring produced, decreasing selection intensity and genetic progress.

The objective of this study was to examine the efficacy of MAS when the selection goal included two traits. The comparisons were made for situations where the traits were negatively and positively correlated. In addition, MAS was applied when the traits were negatively correlated and one of the traits represented fertility.

6.3 Methods

The simulated population was originally described in Chapter 5. Selection for sex limited traits was practiced within an open nucleus and a commercial herd (Figure 6.1). The nucleus was formed by selecting, from a base population of 1500 males and 50,000 females, the best ten sires and 500 dams according to an index of EBV (Spelman and Garrick, 1998) for two traits. The EBV of the sires were assumed to have accuracies corresponding to a progeny test with 100 daughters and the accuracy of EBV for dams corresponded to one lactation record.

In subsequent generations, each of the ten selected sires was mated randomly to 50 nucleus dams to generate full-sib families of three males and three females. Within each family, the best male was selected based on his genotype, considering results of a daughter design that was previously conducted on each of the ten sires. All of the females were eligible for selection on their index. The top 400 nucleus females were selected for the next generation of dams along with 100 dams from a commercial population. The commercial population was comprised of 50,000 cows that were sired by the top 50 bulls in the previous generation. Five generations were simulated. To decrease variability

across comparisons, the same seed value was used to initiate the random number generator for all simulations.

Genetic model

Two traits were included in the selection goal. Trait 1 was designed to represent production, whereas Trait 2 had a lower heritability and was of secondary importance, representing a trait such as conformation, longevity, or a measure of health or functionality. Trait 1 received a standardized selection emphasis three times greater than did trait 2. For comparison, in Canada, production receives 2.5 times more weight than does longevity and >6 times more weight than does udder health (Dekkers, 1995). Trait 1 had a heritability of 0.30 and trait 2 had a heritability of 0.10. The phenotypic variance was arbitrarily set to 400 for both traits. To evaluate how selection response varied as a function of the relationship between the traits, separate situations were simulated in which the traits had genetic correlations of –0.25 and 0.25, respectively. The environmental correlation between the traits was assumed to be zero.]

Two additional situations were simulated where trait 2 directly affected fitness or reproductive rate. The genetic correlation between traits 1 and 2 was –0.25, based on estimates from the literature of genetic correlations between milk production and fertility (e.g. Roth et al., 1999). In the first of these two situations, trait 2 was a threshold trait affecting female fertility rate. For each female, an underlying variate for fertility was simulated by adding a random normal deviate ($\sigma_e^2 = 360$) to the TBV of each cow. Cows with variate $\geq$1.0 standard deviation below the mean of the base population were allowed to have only two offspring of each sex, rather than three. Cows with phenotypes $\geq$2.0

and $\geq$2.5 standard deviations below the mean had only one and zero offspring, respectively. The other situation was similar to having a lethal recessive that was unfavorably associated with trait 1. Numeric effects were assigned to all alleles by using the procedure described with a correlation of –0.25 between the traits. The allele with the greatest detrimental effect on trait 2 was declared the deleterious allele. Individuals (both males and females) that were homozygous for this unfavorable allele were unable to produce any offspring. The effect on the TBV and phenotypes of carriers of this allele was equal to the numeric effect.

The TBV were produced assuming a finite locus model for which all the allelic effects were strictly additive (Mackinnon and Georges, 1998). All genetic variability in the traits of interest was explained by 20 loci, each with 5 alleles. The QTL position was randomly assigned in each replicate. The initial allele frequencies of allele effects were uniformly distributed and standardized to sum to 1.0.

A pleiotropic effect was simulated for each QTL so that all QTL had some effect on both traits. The simulation of allelic effects was a three-step process. In each replicate, allelic effects for trait 1 were first drawn from a double exponential distribution with unit variance. Then, each of the effects for trait 1 was multiplied by the genetic correlation, r_g, and summed to a second random deviate from a double exponential distribution with variance $[1- r_g^2]$ to generate the effects for trait 2. Finally, the allelic effects for each trait were scaled to give the desired genetic variance given the initial allele frequencies. This procedure resulted in a few QTL with large effects on both traits, some QTL with large effects on only one trait, and many QTL with little effect on either trait.

The QTL were randomly distributed across the genome, with no upper limit placed on the number of QTL per chromosome. Thirty chromosomes of 90 cM each were simulated. There were 300 co-dominant markers, with 6 alleles each, distributed across the 30 chromosomes, evenly spaced and separated by a recombination rate (θ) of 0.10.

Marker assisted selection

A daughter design was used, every generation, to determine which marker alleles were favorably associated with the QTL alleles. Each sire had 100 daughters from the commercial population, with a lactation record and complete marker genotypes. Marker contrasts were calculated for all heterozygous markers within sire. All daughters that were informative at a given locus were used to calculate the contrast. The genotype of the dams was ignored in the calculation of the contrast. The contrast was simply the difference in the mean of EBV of daughters carrying the alternative alleles. These EBV were assumed to have the precision of EBV based on one lactation record. Contrasts were then divided by their respective standard errors to account for differences in the number of informative daughters across loci.

Marker assisted selection was then applied to select the best sons from among full-sib families of three males. Selection was based on the THRES approach of Chapter 5, because this approach was superior among different strategies compared in that study. This approach identified all marker alleles at which the marker contrast exceeded a given threshold. Then, for each son, contrast values were summed across all significant markers to form an index for selection that implicitly gave the most weight to the markers with the

greatest contrasts. The son with the highest sum was selected. A threshold of 2.65 standard units, corresponding to 5% comparison wise type I error rate, was used to determine which contrasts contributed to the sum.

Three approaches to MAS were developed. In the first situation (SEPARATE), standardized marker contrasts were calculated separately for each trait. Then the two sums were added, weighted according to the relative economic importance of each trait. For the second approach (INDEX), an index of the two traits was calculated for each daughter by multiplying the single trait EBVs for Trait 1 and 2 by selection index weights which accounted for the economic weights, the genetic correlation and the amount of information the EBVs were based on. Then sons were selected based on marker contrasts for this index. The final situation (SINGLE) mimicked MAS as it may be applied in the current breeding industries, in which the underlying selection objective includes several traits but selection is formally practiced on one or a few traits. In SINGLE, both conventional selection and MAS were practiced only on trait 1.

Several comparisons were made to evaluate the relative effectiveness of conventional selection (RANDOM), in which young sires were selected randomly within-family, and the three MAS approaches. First, in each generation the average TBV of the sons chosen by MAS was compared to average TBV of a group of sons chosen randomly following the same approach as in Chapter 4 (equation [6]). This comparison was made for both traits 1 and 2 and an index of the two traits. Second, the genetic response in the commercial population was also compared. Finally, because the goal of a multiple trait selection program is to increase the frequencies of alleles that favorably affect both traits, the difference between generation one and five in the frequency of the

best allele was monitored within the commercial population. To determine the best allele, an index was calculated for each allele at each QTL,

$$I_{ij} = 3*a_{ij1}+1*a_{ij2},$$

where a_{ij1} and a_{ij2} are the effects of allele j of QTL i on traits 1 and 2 (standardized for genetic standard devation), respectively. The "best" allele was the allele for which I was greatest.

Similar comparisons were made for the situations that examined responses to MAS when differences in fertility were accounted for. However, rather than monitoring the frequency of the best allele, the frequency of the lethal recessive was recorded within the commercial population. Changes in the frequencies of the recessive allele were compared across selection strategies to determine which strategy most quickly decreased its frequency.

6.4 Results

Considering both traits while applying MAS within families resulted in significantly ($P > 0.05$) greater response for the selection objective than did either considering only the trait of primary importance (trait 1) or selecting young bulls via RANDOM. These advantages were obtained by improving the efficacy of selection for the second trait. The TBV of bulls selected by MAS were greater for trait 2 than were the TBV of randomly selected bulls. The relative differences between the TBV of bulls chosen by MAS versus randomly are given in Table 6.1 for multiple trait MAS strategies (INDEX and SEPARATE) and for selection on trait 1 only (SINGLE). Results in Table 6.1 are for the first generation of the simulation, but similar advantages were observed in

all generations. Differences in TBV are expressed as a percentage of the TBV of bulls selected randomly.

When the traits were negatively correlated, the bulls selected by multiple trait MAS had lower mean TBV for trait 1 than did the bulls selected randomly ($P < 0.05$). This difference occurred because multiple trait MAS placed relatively more emphasis on trait 2. Table 6.1 has values of >100% for this set of circumstances (heritability, genetic correlation, and selection weights) because essentially no improvement was obtained in trait 2 when the bulls for progeny testing were selected randomly. For example, the average TBV for trait 2 in the commercial population was –0.056. The average TBV of bulls when selecting randomly was –0.209 (lower, but not significantly different, $P > 0.01$). Bulls selected by the SEPARATE and INDEX approach had mean TBV for trait 2 of 0.291 and 0.502, respectively, both of which were significantly different from the means of both the commercial population and the bulls selected randomly ($P < 0.05$). However, the mean TBV for SEPARATE and INDEX were not significantly different from each other. Given that multiple trait MAS was advantageous (relative to random within-family) selection for trait 2 but not for trait 1, the final advantage in the index was relatively small but still significant (Table 6.1). Again, the SEPARATE and INDEX methods were not significantly different from each other. The increased emphasis on trait 2 resulting from MAS may be due to the fact that MAS, in general, increases selection accuracy and this increase was relatively more important for trait 2, because of its lower heritability.

Multiple trait MAS was more effective when the genetic correlation between the traits was positive. When the traits were positively correlated ($r_g = 0.25$) bulls selected

by multiple trait MAS had greater TBV for both traits (Table 6.1), than did randomly selected bulls. As was the case when traits were negatively correlated, the advantage was greater for trait 2. For example, advantages over random selection for trait 2 were 16.90% and 19.22% for SEPARATE and INDEX, respectively, versus 2.70% and 3.02% for trait 1. Because trait 1 received more weight in the selection index than did trait 2, the final advantage for the index was 4.62% and 5.91% for SEPARATE and INDEX, respectively, which were closer to the results for trait 1 than for trait 2.

All of the individual results, for both positive and negative genetic correlation slightly favored the INDEX approach over SEPARATE, but the differences were not statistically significant.

Applying SINGLE trait selection (both conventional and marker assisted) for trait 1 yielded expected results (Table 6.1). The selected young bulls were much superior for trait 1, having mean TBV 12.19% greater than randomly selected bulls. This advantage was exactly the same for both positive and negative correlations between traits, because the same random seed was used for the simulation and correlation had no effect on selection for trait 1 only. Superiority for trait 1 came at the expense of trait 2, as the mean TBV for trait 2 for selected bulls was decreased by greater than 10% in both instances. Because trait 1 was more important and received more weight in the index, SINGLE MAS for trait 1 still yielded slightly greater advantages in the index than did random selection. For the individual traits, differences between SINGLE and SEPARATE and INDEX were significant ($P < 0.05$), with differences being more pronounced for trait 2. Again, because of the importance of trait 1, differences in the overall index were less than between individual traits.

The differences between multiple trait MAS and random selection of young bulls tended to decrease in subsequent generations (results not shown). For example, in generation five when the correlation between traits was positive, the young sires selected by SEPARATE were only 2.24% greater for the index than were randomly selected sires versus 5.91% in generation one. When the traits were negatively correlated and INDEX was applied, the difference between MAS and randomly selected bulls was 1.64% in generation 5 versus 1.87% in generation one. In Chapter 5 we observed a similar result when single trait MAS was applied for within family selection of progeny test bulls and attributed it to effects of previous selection decreasing genetic variance at QTL with large effects. Also, in the initial generation the SINGLE approach was superior to random selection of young bulls by 1.50% (Table 6.1), mostly due to high response in trait 1 that compensated for decreases for trait 2. In later generations, as response to trait 1 decreased and the deficit in trait 2 increased, bulls selected with SINGLE had poorer values for the index than did randomly selected bulls.

Similar trends were observed when differences between multiple trait MAS and single trait MAS or random selection of young bulls were evaluated in terms of genetic response in the commercial population after five generations (Table 6.2). Specifically, when the traits were negatively correlated, SINGLE MAS yielded the most response for trait 1 and the least for trait 2. Random selection of young bulls yielded slightly (but not significantly) greater response for trait 1 than did the SEPARATE and INDEX approaches, but the SEPARATE and INDEX approaches were significantly superior ($P < 0.05$) for response in trait 2. Table 6.2 also gives responses after 5 generations to the overall index, expressed relative to a value of 100 for response to random selection of

young bulls. Responses in the index to SEPARATE and INDEX approaches were greatest, followed by RANDOM and SINGLE. The advantages of SEPARATE and INDEX relative to RANDOM in cumulative response after five generations of 0.8% and 1.4%, respectively (Table 6.2) were less than the relative advantages in young bull TBV of 1.42% and 1.87% (Table 6.1). This decrease was due to two reasons. First, other pathways of selection, in addition to the selection of bulls to progeny test, have effects on response in the commercial population. Second, results in Table 6.2 were after five generations of selection and, as mentioned earlier, the relative advantages of MAS for selection of progeny test bulls decreased over time. Trends of response were similar when the traits were positively correlated, except that, as expected, responses for trait 2 were greater (Table 6.2) when it was positively correlated with trait 1. Even when the traits were positively correlated, the benefits of multiple trait MAS were minor and much less than observed for single trait MAS in the previous two chapters. This result may be due to the fact that extreme values for contrasts were less common in the multiple trait scenarios. Few alleles would be expected to have extremely favorable effects for both traits. If so, this result may even be worse for situations with more traits or traits with similar weights and heritabilities.

Differences between generation 1 and 5 in the frequencies of the most favorable alleles for the two traits and the index are in Table 6.3. When the traits were negatively correlated, the greatest increase in frequency of the best allele for trait 1 was obtained by SINGLE trait selection (0.51). RANDOM selection gave the second greatest increase (0.48) which was only slightly, but significantly ($P < 0.05$), different from multiple trait MAS (0.47). The opposite pattern was observed for trait 2, with multiple trait MAS

being the most effective approach. Increases in the frequencies of the best allele for trait 1 were much greater than for trait 2, regardless of the strategy. This difference was due to the greater selection weight and higher heritability for trait 1. As a consequence of the greater importance of trait 1 in the index, the greatest response for the best allele for the index was achieved by SINGLE trait selection, 0.51 versus 0.48 for the other methods. When the traits were positively correlated, trends were similar, but the multiple trait approaches were more effective than when traits were negatively correlated. No difference was observed between these methods and RANDOM for the frequency of the best allele for trait 1 and the differences in frequencies between methods were greater for trait 2 than they were when the traits were negatively correlated.

The different approaches to selection also significantly affected the underlying genetic correlation between the two traits. When the traits were negatively correlated, multiple trait selection increased the magnitude of the correlation from –0.25 to –0.29. SINGLE trait selection reduced it to –0.21. When the traits were positively correlated both approaches decreased the genetic correlation, but the decrease was greater (to 0.12 versus 0.20) with multiple trait selection.

Changes in selection responses in the commercial population due to the effect of fertility and the presence of a deleterious recessive were much greater for multiple trait MAS (INDEX) than for SINGLE trait MAS (Table 6.4). When effects of fertility were accounted for, response for trait 1, after five generations of selection, was reduced by 1.6%, from 39.66 to 39.03. However, response to trait 2 was increased by 27%, from 1.08 to 1.37. In contrast, no significant effects on response were observed for SINGLE trait selection. Differences were more pronounced when a deleterious recessive was

simulated. Response for trait 1 was reduced by 11.3%, from 39.66 to 35.16 when the deleterious recessive was present and multiple trait MAS was practiced, whereas response to trait 2 was increased by 63.9%, to 1.77. Effects of the deleterious recessive on response to SINGLE trait MAS were less than with INDEX selection, but were significant ($P < 0.05$). Response to trait 1 was decreased by 2.8% and response to trait 2 was increased by 4.0%.

The use of a multiple trait INDEX for MAS also decreased the frequency of the deleterious gene more than did SINGLE trait MAS. After 5 generations of INDEX MAS, the frequency of the deleterious allele was decreased by 0.057, despite the unfavorable correlation with trait 1. In contrast, the frequency of the deleterious allele was increased by 0.015 when SINGLE trait MAS was applied. Although the deleterious allele had little effect on response to SINGLE trait selection in the short term, long term response may be limited as the frequency of the deleterious allele continues to increase at a similar rate. This result suggests that multiple trait MAS may be a more robust approach for the long term than is single trait selection.

6.5 Discussion

Marker assisted selection for the multiple trait selection objective was more effective than was conventional selection, when applied to the choice of young bulls prior to progeny testing. However, the benefits of multiple trait MAS were less than previously observed for single trait MAS. Considering the scenarios simulated, the average TBV (for the index) of MAS young bulls were up to 6% greater than the average TBV of randomly selected young bulls. This advantage tended to be less than the

response of approximately 12% observed for single trait selection in the most directly comparable study (see Chapter 5). Under similar assumptions and approaches regarding the application of MAS, Spelman and Garrick (1999) also observed advantages ranging up to 10%. Other studies (e.g., Kashi et al., 1990; Mackinon and Georges, 1998; Schulman and Dentine, 1998) reported gains that varied from less than five percent to greater than 20%, but the underlying genetic models differed greatly. Averaging of the effects of the multiple trait resulted in having fewer alleles with large favorable effects on both traits.

In this study, response to MAS for the primary trait (Trait 1) was significantly decreased compared to both single trait MAS and conventional selection for that trait. However, gains in the multiple trait selection objective resulted from large increases in response for the secondary trait.

In contrast to the results reported here, De Koning and Weller (1994) observed relatively greater gains from MAS for a multiple trait objective than for a single trait objective. This discrepancy between the current study may be explained by differences in the underlying genetic models used for the simulations. First, De Koning and Weller (1994) assumed that QTL genotype was known without error, while in the present study the QTL position was unknown. Moreover, they simulated QTL for which the individual allelic effects for each trait were correlated by either 1 or -1, whereas each QTL was simulated with a different covariance in this study. The QTL simulated by De Koning and Weller (1994) were also biallelic, rather than multi-allelic. All of these factors would be expected to contribute to greater response to multiple trait MAS than for the conditions that we simulated. Finally, the previous study applied the same economic

weight for the two traits. This assumption may also have favored multiple trait MAS relative to our study. The major gains were for the second trait, which had a much lower economic weight (and thus contributed less to overall response to MAS) than did the first trait.

The advantages achieved from multiple trait MAS were considerably greater in our study when the traits were positively correlated than when negatively correlated. De Koning and Weller (1994) observed the opposite result. This difference was most likely due to the fact that they assumed precise knowledge of the genotypes for the QTL. In our study, the negative correlations between the traits likely decreased the accuracy with which markers were statistically associated with QTL.

When the secondary trait (2) in our study affected fitness and was negatively correlated with the primary trait (1), multiple trait MAS led to increased response in trait 2 (relative to the situation where reproductive rate was unaffected). Although this phenomenon may need to be corroborated and confirmed in repeated studies, it may provide an additional reason why MAS may be especially beneficial for health and fitness traits, in addition to reasons given by previous authors (e.g. Haley and Visscher, 1998; Davis and DeNise, 1998).

Additional research in multiple trait MAS is definitely needed. The increased use of MAS and the addition of more traits to breeding goals and selection indexes for dairy cattle have been predicted for the future (Boettcher, 2000; Cassell, 2000; Welper, 2000). One specific topic to address is the mathematical approaches for the design of indexes that incorporate MAS and conventional selection on EBV. No differences were observed between the two approaches used in this study for the application of multiple trait MAS,

but the selection criteria used were rather simple. More research is needed on approaches to combine information about marked QTL and EBV for the remainder of the genome, possibly by using variations of approaches developed for single QTL and single traits (Dekkers and Van Arendonk, 1998). Efficient approaches for detecting QTL with effects on several traits are also needed, perhaps by expanding on previous work by (Ronin et al., 1998; Henshall and Goddard, 1999; Bovenhuis and Spelman, 2000).

Table 6.1 **Percent difference at generation one in TBV of young bulls selected through MAS and randomly, when two traits were negatively and positively correlated and three different selection strategies were applied.**

MAS Strategy	Genetic correlation					
	Negative			Positive		
	Trait 1	Trait 2	Index	Trait 1	Trait 2	Index
	----------%----------					
SEPARATE	-1.34[a]	>100.00[a]	1.42[a]	2.70[a]	16.90[a]	4.62[a]
INDEX	-2.11[a]	>100.00[a]	1.87[a]	3.02[a]	19.22[a]	5.91[a]
SINGLE	12.19[b]	-10.26[b]	1.50[a]	12.19[b]	-12.67[b]	2.11[b]

[a,b] Values in the same column with different superscripts are significantly different ($P < 0.05$)

Table 6.2 **Responses in the commercial population to five generations of selection using four different strategies of multiple trait MAS, when two traits were negatively and positively correlated.**

MAS Strategy	Genetic correlation					
	Negative			Positive		
	Trait 1	Trait 2	Index	Trait 1	Trait 2	Index
RANDOM	40.10[a]	0.09[a]	100.0[a]	40.90[a]	12.29[a]	100.0[a]
SEPARATE	39.75[a]	1.22[b]	100.8[b]	40.94[a]	12.97[b]	100.9[b]
INDEX	39.66[a]	1.80[b]	101.4[b]	40.97[a]	13.19[b]	101.2[b]
SINGLE	43.98[b]	-7.83[c]	98.5[c]	43.78[b]	4.52[c]	96.6[c]

[a,b,c] Values in the same column with different superscripts are significantly different ($P < 0.05$)

Table 6.3 **Changes, after five generations of selection, in the frequencies of the most favorable alleles when different strategies of multiple trait MAS were applied and the two traits were negatively and positively correlated**

MAS Strategy	Genetic correlation					
	Negative			Positive		
	Trait 1	Trait 2	Index	Trait 1	Trait 2	Index
RANDOM	0.48[a]	0.05[a]	0.48[a]	0.48[a]	0.16[a]	0.49[a]
SEPARATE	0.47[b]	0.06[b]	0.48[a]	0.48[a]	0.18[b]	0.49[a]
INDEX	0.47[b]	0.06[b]	0.48[a]	0.48[a]	0.19[b]	0.49[a]
SINGLE	0.51[c]	-0.04[c]	0.51[b]	0.51[b]	0.05[c]	0.51[b]

[a,b,c] Values in the same column with different superscripts are significantly different ($P < 0.05$)

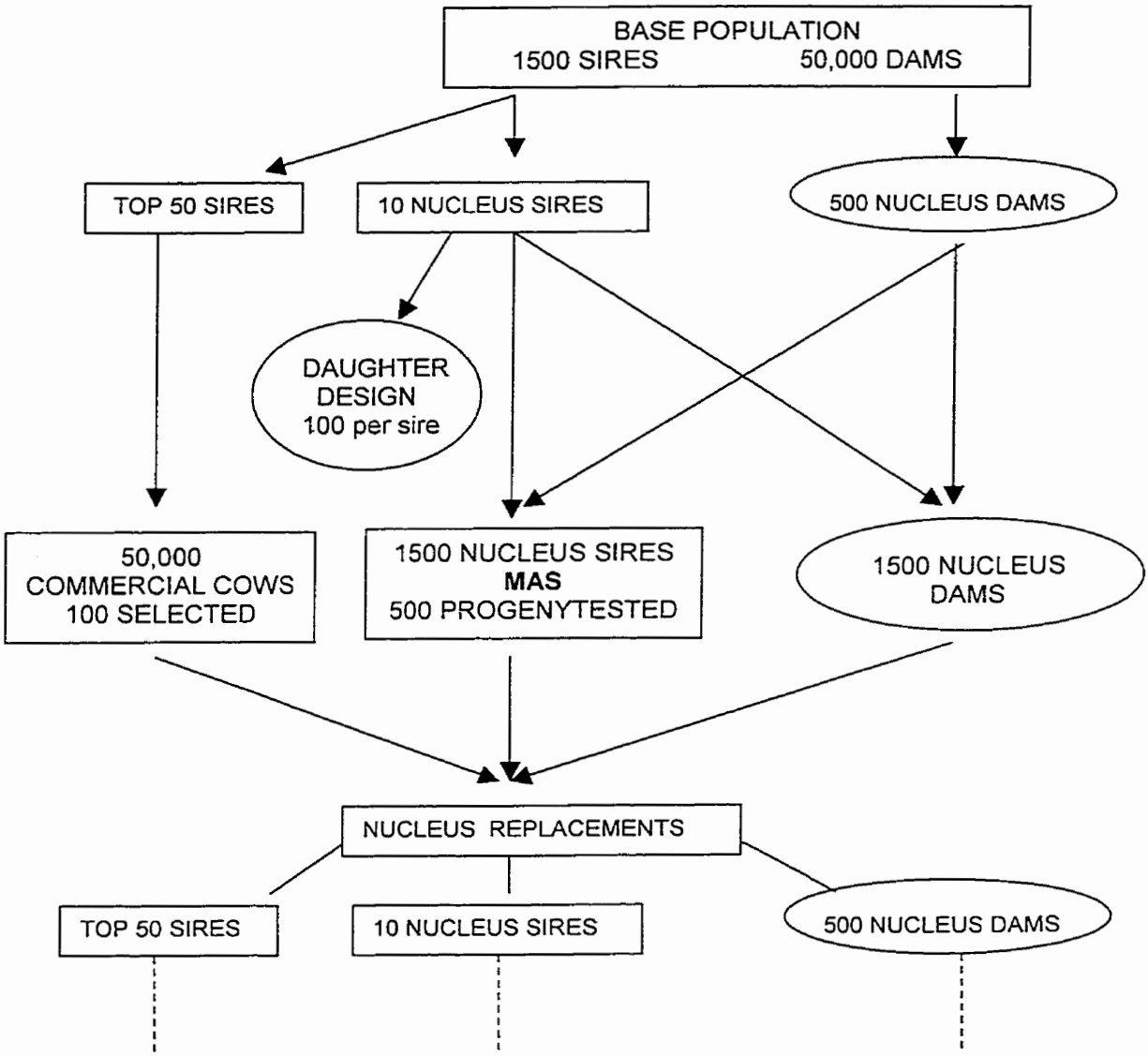
Table 6.4 **Responses[1] to five generations of INDEX and SINGLE MAS, when two traits were negatively correlated, when trait 2 was fertility and when one allele was a lethal recessive.**

Strategy	Trait 1	Trait 2
INDEX		
Negative	39.66[a]	1.08[a]
Fertility	39.03[b]	1.37[b]
Recessive	35.16[c]	1.77[c]
SINGLE		
Negative	43.38[d]	-6.26[d]
Fertility	43.31[d]	-6.36[d]
Recessive	42.18[e]	-6.02[e]

[1] Response in the commercial population.

[a,b,c,d,e] Values in the same column with different superscripts are significantly different ($P < 0.05$)

Figure 6.1 **Structure of the population.**



7. Models for genetic effects in simulations of marker assisted selection in dairy cattle

7.1 Summary

The objective of this study was to determine how different assumptions about the underlying genetic model for a population affected the distribution of allelic effects and the long-term response to marker assisted selection (MAS) in dairy cattle. An open nucleus and commercial population was simulated, with MAS being used within family to select young bulls to enter a progeny-testing program. A finite genetic model with up to 200 loci was assumed. Effects of mutation were included to maintain genetic variance. One hundred generations were simulated. The genetic models differed in the distribution of allelic effects. Gamma, double exponential and normal distributions were used to generate allelic effects. Models also differed with respect to mutation rate, number of segregating loci, and the distribution of initial allelic frequencies. Genetic variances in the commercial population were monitored. Genetic variances changed over time, but eventually stabilized. Based on inference from the final generation, after the genetic variance became stable, a gamma distribution with long tails seemed to best describe the allelic effects. The 3 largest QTL explained approximately 40% of the genetic variance, versus <20% for the normal model. The 20 largest QTL accounted for >99% of the variance. The distribution of allelic effects was clearly U-shaped, with most alleles having very high or very low frequencies. To test the efficacy of MAS, the breeding values of selected bulls were compared to randomly chosen bulls from the same families. The superiority of MAS was maintained throughout the course of the simulation,

regardless of the model used. Over time, the trend in the advantage of MAS mirrored the trend for genetic variance in the commercial population.

7.2 Introduction

Different simulations of MAS have used different underlying genetic models. This greatly affected the results on the relative advantage of MAS over conventional selection (for review, Spelman, 1998). One major difference is the use of a mixed inheritance model with a polygenic background and one or more QTL (e.g. Chapter 4; Meuwissen and Goddard, 1996; Spelman and Garrick, 1998) versus a finite locus model (e.g. Chapter 5; Mackinnon and Georges, 1998). When stochastic simulations have considered only short or medium term prediction, the results have tended to be similar for the two previously mentioned models. However, the finite locus model is unable to predict a long-term response to selection, because allele frequencies are rapidly driven to fixation (Gibson, 1999). Recently, QTL of moderate effect have been detected in selected populations (e.g. Coppieters et al., 1998). Hayes and Goddard (2000) performed a meta-analysis of QTL mapping experiment results from the literature. They estimated the number of genes affecting quantitative traits in dairy cattle and swine to be between 50 and 100. These results highlight the need to consider the finite model in the prediction of selection response, since, for these analyses, the infinitesimal model is flawed.

Mutation makes a substantial contribution to genetic variation (Hill, 1982). Along with migration and non-additive genetic effects (e.g. dominance and epistasis), mutation is likely one of the major factors in the maintenance of variation (and long term response)

in populations under selection (Lynch and Walsh, 1998). To be able to reproduce the long term selection response that has been observed in commercial livestock populations (e.g. Cassell, 2000), mutation should be considered in modeling MAS breeding schemes.

The first objective of this study was to determine how different assumptions about the underlying genetic model for a population affected response to selection and the distribution of allelic effects following many generations of selection. A second objective of the study was to determine the superiority of MAS under a genetic model designed to sustain long term selection response.

7.3 Methods

Populations were simulated by following the basic desig outlined in chapter 5. Selection for a sex-limited trait was applied within an open nucleus and a commercial herd (Figure 7.1). The nucleus was formed by selecting, according to EBV (Spelman and Garrick, 1998), the best ten sires and 500 dams from a base population of 1500 males and 50,000 females. The EBV of the sires were assumed to have accuracies corresponding to a progeny test with 100 daughters. The accuracy of EBV for dams corresponded to one-lactation record.

In the first and following generations, each of the 10 sires was mated randomly to 50 nucleus dams to generate full-sib families of three males and three females. Within each family, the single best male was selected to enter a progeny test, based on favourable marker-QTL associations determined in a daughter design. All nucleus born females were eligible for selection and the top 400 females were selected for the next

generation of nucleus dams, along with 100 cows from the commercial population. The commercial population was comprised of 50,000 cows that were sired by the top 50 bulls in the previous generation. One-hundred generations were simulated in each of 200 replicates.

Genetic model

For each simulation, a different underlying genetic model was assumed. All of the simulations had several characteristics of the genetic model in common. Then, one aspect of the genetic model was modified for each situation, in order to evaluate the effects on maintenance of genetic variability and response to MAS.

Common aspects across simulations. For all scenarios, the TBV of each animal was determined by the sum of additive effects for 200 QTL in a finite locus model. Each QTL had five alleles. The QTL were randomly distributed across the genome. Thirty chromosomes of 90 cM each were simulated and no upper limit was placed on the number of QTL per chromosome.

The phenotypic variance for the simulated trait was 400. Initial genetic variance was targeted at 120 (heritability = 0.30), but varied slightly depending on the actual allelic effects and frequencies obtained in a replicate. For MAS, 180 co-dominant markers, each with 6 alleles, were distributed across the 30 chromosomes, evenly spaced (i.e. 6 per chromosome) and separated by a recombination rate (θ) of 0.18.

Variable aspects across simulations.

Three alternatives were used to describe the initial distribution of allelic effects:

1. normal distribution,

2. double exponential distribution,

3. gamma distribution.

Figure 7.2 shows the upper half of each of these distributions. The normal distribution has been used by many previous authors when simulating underlying allelic effects (e.g. Chapter 2; Meuwissen and Goddard, 1996; Spelman and Garrick, 1998) and is the proposed distribution of QTL for simultaneous genetic evaluation of identified QTL and remaining polygenes as random effects in a linear model (Fernando and Grossman, 1989). To obtain the desired variance, allelic effects were drawn from a normal distribution with $\mu = 0$ and $\sigma = 2.0$.

The double exponential distribution is more peaked and has longer tails than does the normal distribution (Figure 7.2), resulting in fewer alleles with intermediate effects and more with small or large effects. Mackinnon and Georges (1998) assumed that allelic effects were distributed as a double exponential. The double exponential distribution is defined by a scale parameter. The scale parameter used to achieve the desired genetic variances in these simulations was 1.35.

The gamma distribution is very general (in fact, the normal and exponential are specific types of gamma distributions) and its shape varies greatly depending upon the values of shape and scale parameters used to define the distribution. The distribution used in the simulation had shape and scale parameters of 0.5 and 0.55, respectively, resulting in the distribution shown in Figure 7.2. In the simulation, allelic effects were randomly drawn from symmetric distributions with both positive and negative values. This gamma distribution results in relatively more large and small allelic effects and fewer intermediate effects than either the normal or double exponential distributions.

Hayes and Goddard (2000) proposed a gamma distribution for allelic effects based on meta-analysis of previous QTL detection studies.

In addition to the differences in the initial distribution of allelic effects, the following aspects of the genetic model were also examined:

1) mutation rate,
2) number of segregating loci,
3) initial allelic frequencies.

Each of these differences were applied to populations simulated with the gamma distribution as the initial distribution for allelic effects.

Mutation rate. If allelic effects are fixed in a population over time and no new alleles enter the population, selection and genetic drift will eventually exhaust most of the genetic variation, fixing each QTL at a given allele. Such a phenomenon has not been observed in selected experimental or commercial populations, suggesting that new genetic variation is continually introduced through mutation (Lynch, 1988; Keightley, 1998).

Simulation of mutation was the most difficult aspect of this study and required the most liberal assumptions. In this study two different kinds of mutation were simulated. First, mutation in segregating QTL that changed one allele to another one of the five initial alleles was generated. The second type of mutation created variability in new QTL. Of the 200 simulated QTL most (165) were initially fixed at one randomly selected allele. These were the QTL at which new mutations would occur in later generations of the simulation. The value of 200 QTL was chosen as a result of computing resources. The genes that affect quantitative traits are likely to include thousands, or even millions,

of base pairs at which a polymorphism could alter performance. Therefore, with unlimited resources, one would probably want to design a simulation with a much higher number of potential loci at which future mutation could introduce variability.

For this study, we generated 35 alleles at which variability existed in the first generation. The actual number of segregating QTL contributing genetic variability of a production trait is obviously unknown but 35 was chosen because this value was the mid-point between 20 by Mackinnon and Georges (1998) and 50 by Hayes and Goddard (2000). Therefore, in early generations of the simulation, the ratio of probabilities of mutation in a new versus segregating QTL was approximately 165 : 35.

To help circumvent the restriction to 200 total QTL, in addition to the initial 165 potential new QTL, the opportunity for further new QTL alleles was also simulated. This situation was simulated by monitoring each segregating allele until fixation was reached. Fixation was declared for a given QTL when the frequency of one allele exceeded 0.995 in the commercial population. When fixation was reached, new values were simulated for the other 4 alleles at that locus, each with an initial frequency of 0.0. Thus, additional new mutations could be generated continuously and were not limited to the original 165.

A standard mutation rate was used that corresponded with a genome wide rate of 10^{-5} (Lynch and Walsh, 1998). At this rate, mutation was expected to occur in approximately 12 animals in the nucleus each generation (2 alleles * 200 QTL * 3000 nucleus animals * 10^{-5}). To observe the effects of an increased mutation rate, one set of simulations was performed with a higher mutation rate of 2×10^{-5}.

For simplicity, mutation was not simulated at marker loci. Although this is clearly a rather unrealistic assumption, it was assumed that the primary effect of such

mutations on MAS would be to increase the number of apparent genotyping errors and uninformative individuals and thus decrease the advantages of MAS at a rate proportional to the frequency of mutation. The analysis of this interesting relationship was not among the objectives of the study.

Number of segregating loci. As mentioned previously, for most situations, the initial number of segregating QTL was set at 35. An additional situation was simulated with only 20 segregating QTL in the first generation of the simulation. With only 20 segregating QTL, each locus contributed more genetic variance than when 35 QTL were segregating. Therefore, allelic effects for this model were drawn from a gamma distribution with scale parameter of 0.42, rather than 0.55. Preliminary studies with a greater number of segregating QTL (50) indicated little difference from a starting value of 35 QTL. Thus, populations with more than 35 QTL were not examined further.

Initial allelic frequencies. In the standard model, allelic frequencies for segregating QTL were generated following the approach of Mackinnon and Georges (1998). For each QTL, frequencies for the five alleles were initially drawn from a uniform distribution bounded by 0.0 and 1.0. The five values were then standardized to sum to 1.0.

Some authors have suggested (Crow and Kimura, 1970) that the distribution of allelic frequencies is more likely to be somewhat U-shaped, rather than uniform. With a U-shaped distribution, for most QTL, one allele has a much greater frequency than do the alternative alleles. For example, in selected populations, the most favorable allele is expected to have the highest frequency at most QTL.

For comparison to the standard situation with uniform allelic frequencies, one set of simulations was generated with a U-shaped distribution for alleles. To simulate the U-shaped distribution, the allelic frequency (p_H) for one randomly chosen allele was drawn from a uniform distribution bounded by 0.9 and 1.0. The frequencies for the other 4 alleles were then drawn from another uniform distribution, but standardized to sum to (1 – p_H). For all simulations, marker alleles were assumed to have equal initial frequencies.

Marker assisted selection

A daughter design was used, every generation, to determine which marker alleles were favorably associated with the unknown QTL alleles. Each sire had 100 daughters from the commercial population with a lactation record and complete set of marker genotypes. Marker contrasts were calculated for all heterozygous markers within sire. All daughters that were informative at a given locus were used to calculate the contrast. The genotype of the dams was ignored. The contrast was simply the difference in the mean of EBV of daughters carrying the alternative alleles. These EBV were based on one lactation record. Contrasts were then divided by their standard errors to account for differences in the number of informative daughters across marker loci.

Marker assisted selection was applied to select the best sons from among full-sib families of three males. Selection was based on the THRES approach of Chapter 5, because this approach was superior among different strategies compared in that study. This approach first identified all marker alleles at which the marker contrast exceed a given threshold. Then, for each son, contrast values were summed across all significant

markers and the son with the highest sum was selected. A threshold of 2.65 standard units was used. This form of MAS was performed in every generation of the simulation.

Analyses performed

Two-hundred replicates were generated for each genetic model. For each model, the advantage of within family MAS over random selection was calculated in each generation. This quantity was determined by calculating the difference between the average TBV of bulls selected by MAS and a randomly selected set of bulls. In addition, the genetic variance and frequencies of each QTL allele within the commercial population was also monitored. From this information, the shape of the distribution of allelic frequencies that occurred following many generations of selection could be determined and the number of segregating QTL and the amount of genetic variance contributed by each could be calculated.

Based on these results, a genetic model for which genetic variability and selection response could be maintained at a relatively steady state for many generations could be proposed. The effectiveness of MAS was tested in this population.

7.4 Results and Discussion

Distribution of allelic effects

Genetic variance was calculated in the commercial population in every generation of every simulation. Figure 7.3 shows the trend in genetic variances for models that assumed that allelic effects were from normal, double exponential, and gamma

distributions. For all models, the genetic variance changed greatly across generations. All three models showed the same pattern. Initially the genetic variance decreased until about generation 5. Then for 10 to 15 generations the variance increased to reach maxima at approximately 1.7 to 1.9 times the original variance. Variances for all three models then decreased steadily until around generation 60 where variance for the gamma model became stable. Variance for the other two models continued to decrease, but at a slower rate, beyond this point. Both normal and double exponential models stabilized at around generation 90.

The initial decrease in variance occurred due to selection prior to the point where new mutations started to exert an effect. The decline in variance was likely a consequence of both the changes in frequencies of some alleles and disequilibrium among loci (Bulmer, 1971). This decrease was greatest for the gamma model, which reached a minimum of 105 versus 110 for the other two models ($P < 0.05$). The variance for the gamma model decreased the most because effects of individual alleles were greatest for this distribution (Figure 7.2) and, therefore, selection moved these loci more quickly toward fixation.

The subsequent upsurge in variance indicated the period where the new mutations in the 165 loci that were initially fixed started to occur. These QTL were initially fixed at a randomly selected allele. Therefore, the best allele at most of these loci (4 out of 5) was at a very low frequency during the early generations of the simulation. Through selection the frequency of these alleles increased, thereby increasing the total genetic variance during this second phase of the simulation. The upward trend in variance occurred most quickly, but reached the lowest peak, for the gamma model because of the previously

mentioned larger effects of the extreme alleles, upon which selection was most effective. For this reason the trend for the double exponential distribution was intermediate.

The stage of declining genetic variances that was observed between generations 25 to 60 occurred because mutation at most of the 165 initially fixed QTL took place in previous generations. The frequencies of the best alleles for these QTL passed 0.50 and further selection decreased genetic variance rather than increasing it. The genetic variance tended to eventually reach a steady state for all three models. This state occurred much earlier with the gamma distribution and at a greater magnitude than with the normal or double exponential models. The variance for the gamma model stabilized at approximately 40 versus approximately 20 for the other two models (Figure 7.3). The variance was likely maintained at a higher level for the gamma model because the extreme effects of new mutations were greater and more easily retained in the immediately subsequent generations. Genetic drift was seemingly either more likely to eliminate the new mutations in the normal and double exponential models or these mutations had smaller effects than the extreme mutations with the gamma model, and thus each contributed less variance.

The relative impact of individual QTL from the different models can be evaluated by examining variance accounted for by each QTL. Figure 7.4 shows the proportion of total variance accounted for by each of the 200 QTL for the gamma and normal models. The QTL are ordered from left to right in terms of total variance. The results shown are means from generation 100, after variability had stabilized for both models. Results from the double exponential model were more similar to the normal and, therefore, are not shown. As can be clearly seen, the extremely large QTL for the gamma model accounted

for much greater proportions of total genetic variance than for the normal model. On average, the largest QTL for the gamma model accounted for >20% of the total genetic variance, versus only about 6% for the normal model. In fact, both the second (>10%) and third (~7%) most variable QTL for the gamma model accounted for more variance than did the most variable QTL from the normal model. Overall, the 10 largest ranked QTL for the gamma model accounted for more of the total genetic variance than did the correspondingly ranked QTL from the normal models.

Although extension of results from simulation to real life situations can be difficult and must be done with caution, these results regarding the variance accounted for by individual QTL may suggest that the "true" distribution of QTL effects in real livestock populations may more closely resemble the gamma distribution than either the normal or double exponential. Previous QTL detection studies (Georges et al., 1995; Knott et al., 1998; Coppieters et al., 1998) have attributed to individual QTL proportions of total genetic variance that were consistent with the values for the largest alleles from the gamma distribution and much greater than the largest alleles from the simulated normal or double exponential distributions.

Other factors

Assumptions about mutation rate, number of segregating QTL, and the distribution initial allelic frequencies had noticeable effects on the genetic variance within the commercial population, particularly in the early generations. Figure 7.5 shows the genetic variation in the commercial population by generation for the models with increased mutation rate, 20 rather than 35 initially segregating QTL, and a U-shaped

distribution for the initial allelic frequencies. Because all models except the twenty-QTL model eventually converged to approximately the same variance (not significantly different) after approximately 50 generations, only the first fifty generations are shown.

All three modifications caused genetic variance to increase in the early generations relative to the reference gamma model (Figure 7.3). The greatest difference was for the model with only 20 QTL controlling genetic variance. This factor led to greater variance because with fewer QTL, effects of each QTL were larger. This increase in effect applied to both initially segregating and new QTL. Given that the QTL formed by new mutations tended to be larger than with the other models, genetic variance rose faster and peaked at a higher level than when the other models were used. The variance was also maintained at a greater level in the long term.

The genetic variance for the model with increased mutation rate also had higher variance in the short term than did the reference gamma. This gain in variance occurred because, in the short term, relatively more new segregating QTL were created, due to increased rates of mutation. In the long term, this model began to converge toward the standard gamma because there was no difference between the models in the size of the allelic effect and number of new loci and these variables controlled the genetic variance in later generations.

Differences between the U-shape model and the standard gamma with uniform allelic frequencies were only observed in the very short term. The initial drop in genetic variance due to selection was not as severe because, for a greater number of QTL, the frequency of the best allele was increasing toward 0.50. None of the six initial models maintained a stable genetic variance throughout the term of the simulation, but the trends

observed provided information about how to alter the simulation model to help maintain a more consistent trend in variance. The early drops, sharp upswings, and subsequent steep declines observed for genetic variance highlighted potential flaws in the design of the initial simulation. The initial drop in variance probably occurred because the approach to assign original allelic frequencies was not entirely adequate for a simulation with many generations. Beyond five generations, the genetic variability would have continued to decline quickly, had not mutation been included in the model.

The simulation of a U-shaped distribution for initial gene frequencies would likely have prevented the initial drop in variance that was observed. A U-shaped distribution of allelic frequencies implies that some of the best alleles at a few of the QTL have very low frequencies (less than 0.01, for example). Selection would be expected to increase the genetic variance contributed by these QTL (until the frequency of the best allele passed 0.50) while decreasing the variance only slightly at the alleles where the frequency of the best allele was already high. Therefore, the total genetic variance would have remained more stable overall.

The proposal for a U-shaped distribution of allelic frequencies was supported by observation of the allele frequencies that resulted in generation 100, after stability was reached in the models (Figure 7.6). Figure 7.6 shows for the gamma model a histogram of frequencies for all the alleles of the QTL that were segregating in generation 100. (Only results for the gamma model are shown, because the distribution of allelic frequencies was of a similar shape for all models.) Clearly, the distribution of allelic frequencies at the point of stable genetic variance was U-shaped. Most of the alleles had frequencies <0.05 or >0.95.

The dramatic rise in genetic variance indicated another potential problem in the simulation model, suggesting that the rate of new mutations was too high in these generations. The simulated mutation rate was designed to correspond with a genome wide rate within the range of predictions from literature and thus, may have been adequate. The observed result was probably not a result of an excessive mutation rate, alternatively, the distribution of allelic effects for new mutations may not have been realistic. In the simulation, five allelic effects were initially simulated for the new QTL and the genome was fixed randomly at one of the five alleles. To most closely simulate reality, most of the QTL probably should have been fixed at the best allele. Long term forces of selection (both natural and artificial) have created organisms for which most genes encode proteins that are of a nearly optimal structure for their particular purpose. Thus, most mutations to change this structure are much more likely to be detrimental rather than beneficial (Hartl, 1999). This concept also helps to explain why the DNA sequences for the functional parts of many proteins are highly conserved, even across species (Lyn et al., 1995; Bemark et al., 1998; Thaller et al., 1998). Years of evolution have selected for the ideal protein at these sequences, which is the same regardless of species.

The eventual stabilization of genetic variances at levels less than the initial variances could be an indication that the size (variability) of allelic effects was too low. The genetic variance with the 20 QTL model stabilized at a higher level than did the variances with the other models, and this model had greater variance for the individual allelic effects.

New Simulation

In an effort to improve the simulation model to obtain a more stable trend in genetic variance, a number of modifications were made. First, a U-shaped distribution for initial gene frequencies was simulated to help prevent the early decreases in variance due to selection. To generate the U-shaped distribution, the procedure followed was similar to that described previously (sampling one allelic frequency of one allele from U(0.90,1.00)), except that an additional condition was added. That is, for most (75%) of the QTL, the most favourable allele was most common. This modification was designed to simulate the fact that past selection would have increased the frequency of the favourable alleles. The 75% was chosen based on the final allelic frequencies (generation 100) observed in the previous simulations. The rate of new and beneficial mutations was decreased to an expected value of three per generation to help decrease the large peak in the genetic variance observed in generations 10 to 15 with the previous models (Figures 3 and 4). The choice of three new mutations per generation was chosen following a grid search and corresponded to approximately 1.2×10^{-6} per locus per meiosis. When fewer mutations were simulated, genetic variance continued to decrease; when more mutations were simulated, variance continued to reach excessively high levels. Finally, the size and variability of allelic effects for new mutations was also increased, in order to maintain long-term variance at a level more consistent with the starting variance. These effects were generated with a gamma distribution with scale parameter of 0.40.

The results of the modifications can be seen in Figure 7.7, which shows the genetic variance for each generation, based on 200 replicates of the new simulation model. Clearly, the changes helped maintain a higher and more stable genetic variance

throughout that stabilized at a greater value in later generations than did the previous simulations. However, some trends remained that were difficult to avoid due to conflicting forces between the forces affecting genetic variance. Use of a U-shaped distribution for allelic frequencies maintained the genetic variance at or above the original level for a few generations, but effects of selection decreased the variance significantly below the starting level by generation 10. The effect of new mutations, being fewer in frequency than in previous simulations, did not begin to override the decreases due to selection until about generation 20. The new mutations eventually increased genetic variance above the original level, peaking at around 125% of the original value at generation 60. The variance then declined until reaching an equilibrium of approximately 100% of original variance in generations 90 and greater.

Advantage of MAS

Table 7 also shows the advantage obtained by MAS throughout the simulation, expressed as percent of superiority in MAS young bulls relative to randomly selected bulls. Although some variability was observed, the superiority remained significant stabilizing at around 9% (+/- 0.5%) in the later generations.

As discussed previously, the other simulation models differed in a variety of aspects, but regardless of the genetic model simulated, the TBV of bulls selected by MAS was greater than the TBV of bulls selected randomly from the same families. Figure 7.8 shows for all models the average percentage difference in TBV of MAS and randomly selected bulls for each generation. The trends of this advantage in TBV for MAS bulls follow the same pattern as observed for genetic variance in the commercial population

(Figures 3 and 5). As variance increased in the commercial population, so did the advantage due to MAS. The different trends each increased to a peak in early generations when many QTL with new mutations were segregating in the population. The peak was greatest (up to 20%) for the model that started with 20 segregating QTL, which was also the model with the highest initial increase in variance.

The use of MAS in early generations was particularly beneficial in these simulations because at many loci, the best alleles were newly formed by mutation and were initially at relatively low frequencies. Selection quickly began to increase the frequencies of these beneficial alleles toward fixation, however, such that the benefits of MAS (and genetic variance, as shown in Figures 3 and 5) decreased in generations 20 to 50. Eventually, the benefits of MAS stabilized in the later generations at lower levels than observed with the "improved" model (Figure 7.7) but TBV of bulls selected by MAS remained at a level significantly ($P < 0.05$) greater than for random selection for all models. In these later generations, the influx of new and beneficial mutations was less than in early years, but still high enough to maintain some genetic variation.

Gibson (1994) and Muir and Stick (1998) both used simulation to demonstrate that long-term gains with MAS were less than with conventional selection. However, their models differed from ours, primarily because they considered MAS for a single locus and did not consider the possibility of mutation, which could have created new favourable alleles at the locus. In their studies, as that locus approached fixation for the selected allele, continuing to emphasize its selection compromised genetic improvement at QTL in the rest of the genome. Gibson (1994) and Muir and Stick (1998) also simulated individual selection based on an index of QTL and remaining genetic

information, which differed from the approach applied in this study of selecting the animal with the best QTL genotype within family. In our study, all of the directly competing candidates (full-sibs) for MAS had equal EBV for the remaining (non MAS) portion of the genome, so no danger existed in over-emphasizing the MAS portion of the genome.

Recently, Dekkers and Van Arendonk (1998) demonstrated that a more complicated index that optimised weights on QTL and remaining genetic information could be devised so that response to MAS could meet or exceed that for conventional selection for any given planning horizon. Based on their results, and the observations made in this study, a seemingly logical conclusion is that for most situations, MAS programs can be designed to increase response over conventional selection methods and maintain such an advantage for many generations. The precise magnitude of the advantage obtainable by MAS is likely to be highly variable and based strongly on the rates of mutations and effects of the new alleles created.

7.5 Conclusions

In this study, due to limits on computing resources and knowledge of true rates and sizes of mutations, many simplifying assumptions were made. The greatest assumptions were made about the mutation effects, including the rate and size of mutations, and the number of alleles affecting the quantitative traits. The validity of these assumptions can not be tested until more knowledge about the genome underlying quantitative traits is obtained, but a range of values were used and compared in this study to help account for the lack of prior knowledge.

Figure 7.1 **Structure of the population**

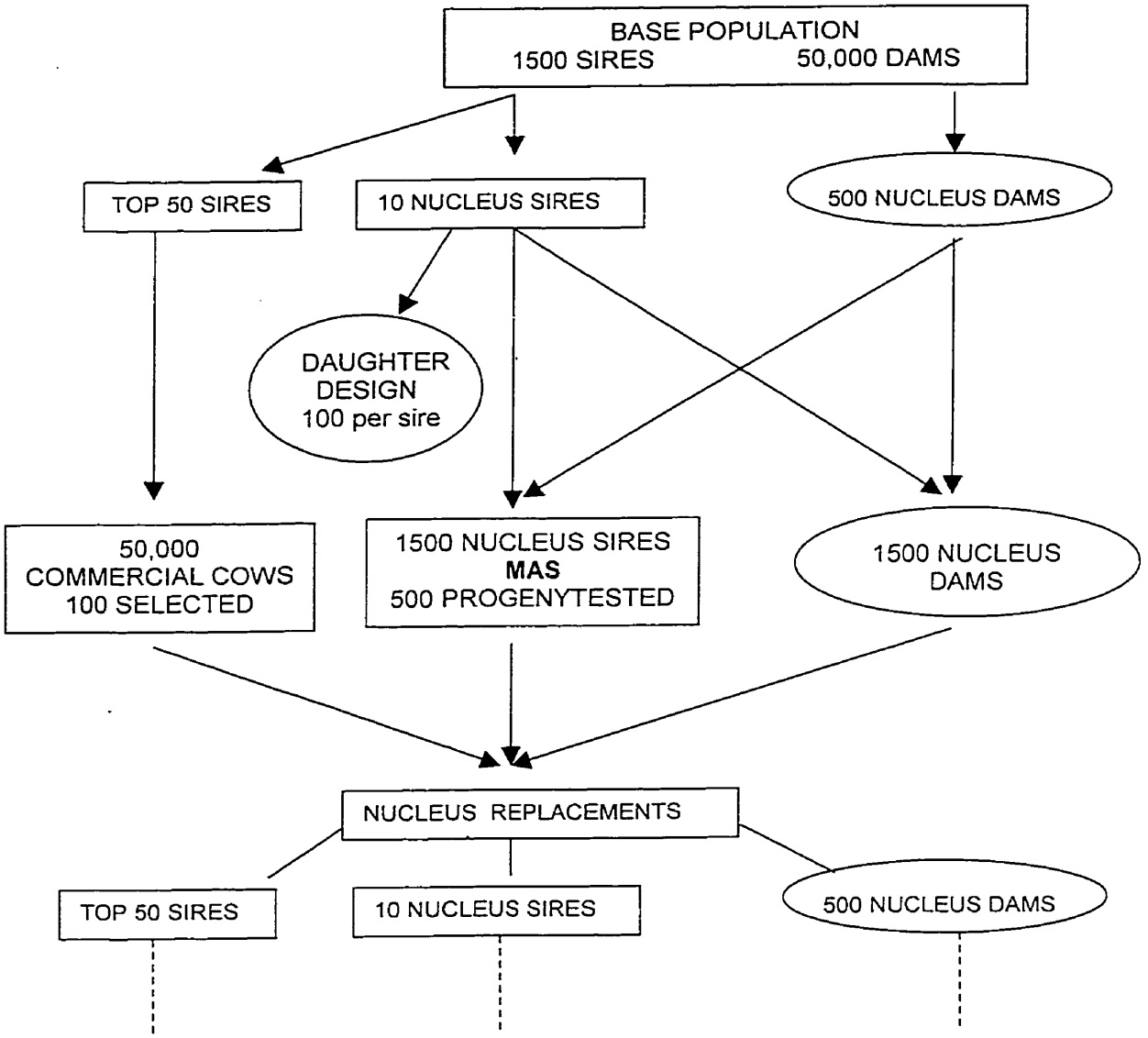


Figure 7.2 **The upper halves of distribution of allelic effects from the normal, double exponential and gamma distributions.**

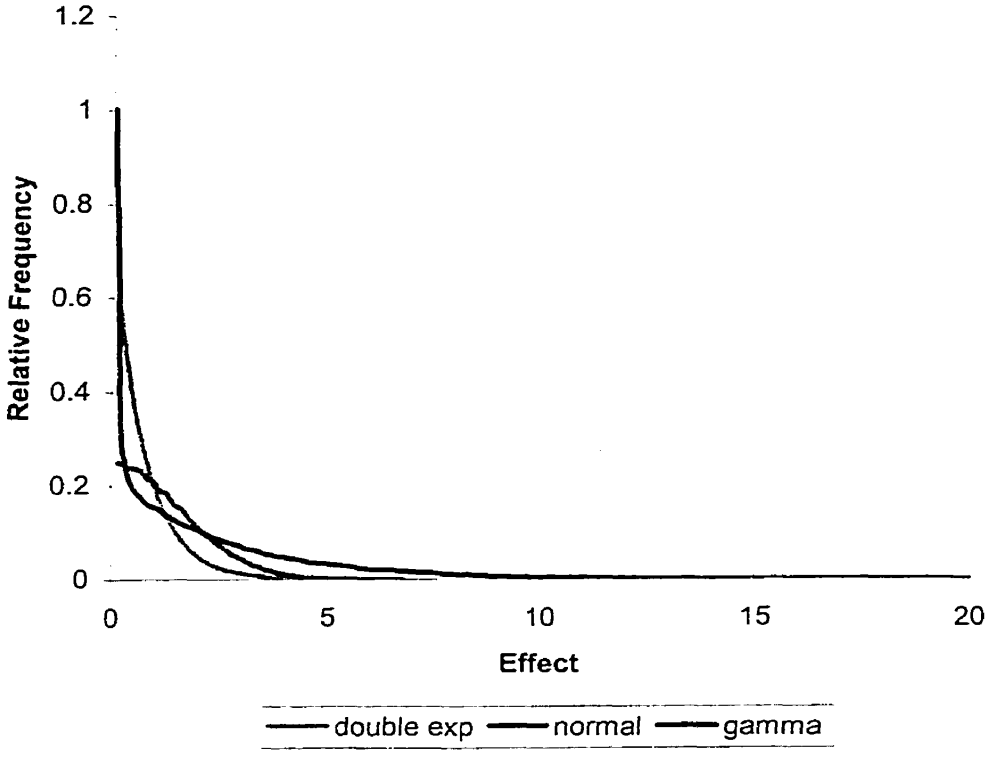


Figure 7.3 **Trends for genetic variance in the commercial populations when allelic effects had normal, double exponential and gamma distributions.**

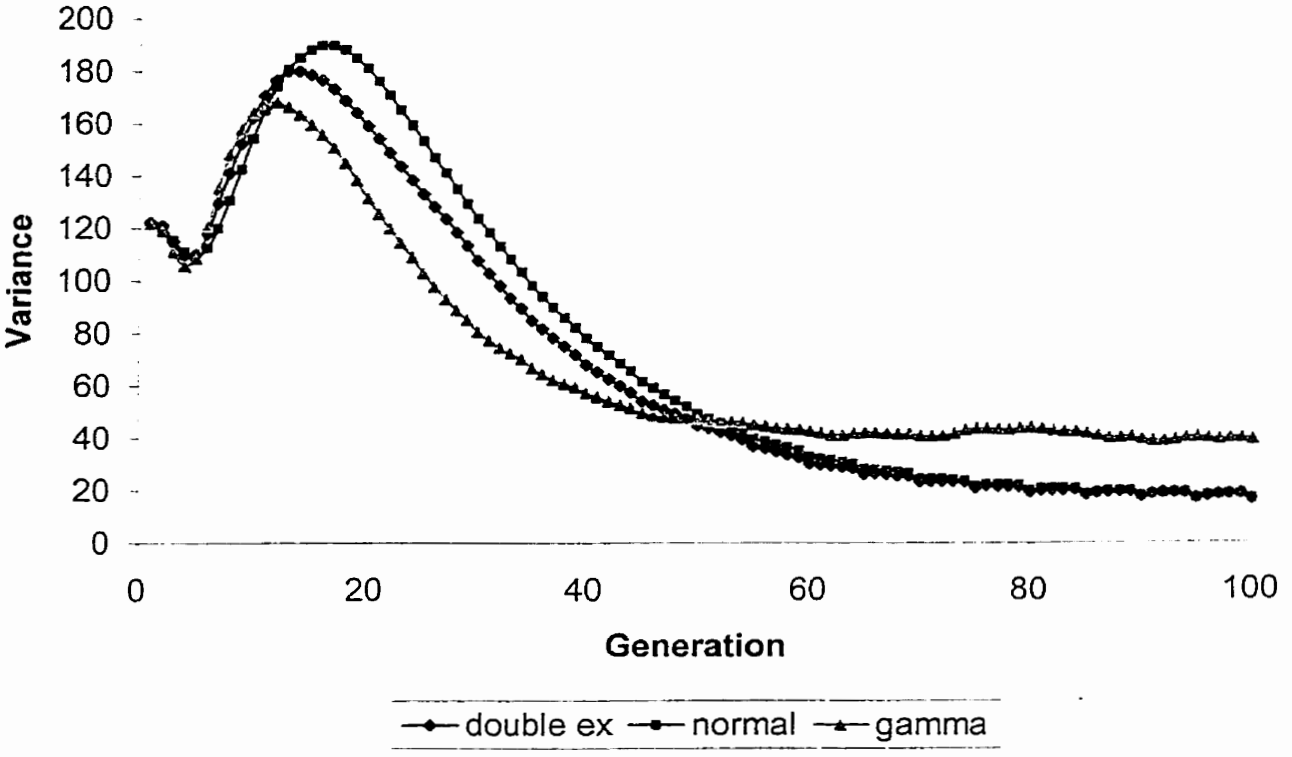


Figure 7.4 **Percent of the total genetic variance contributed by each QTL QTL after 100 generations of selection.**

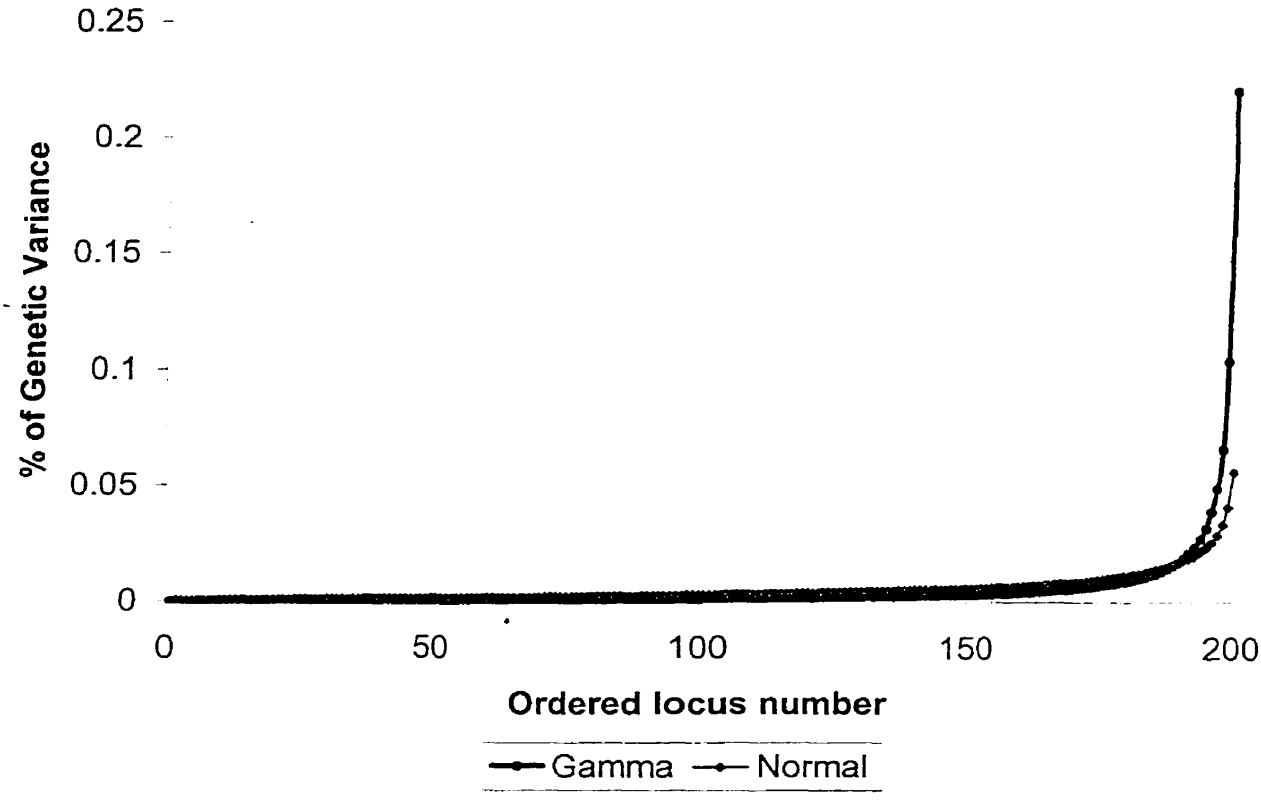


Figure 7.5 **Genetic variance for each generation in the commercial population for the models with high mutation, 20 QTL and U-shaped distribution for initial frequencies.**

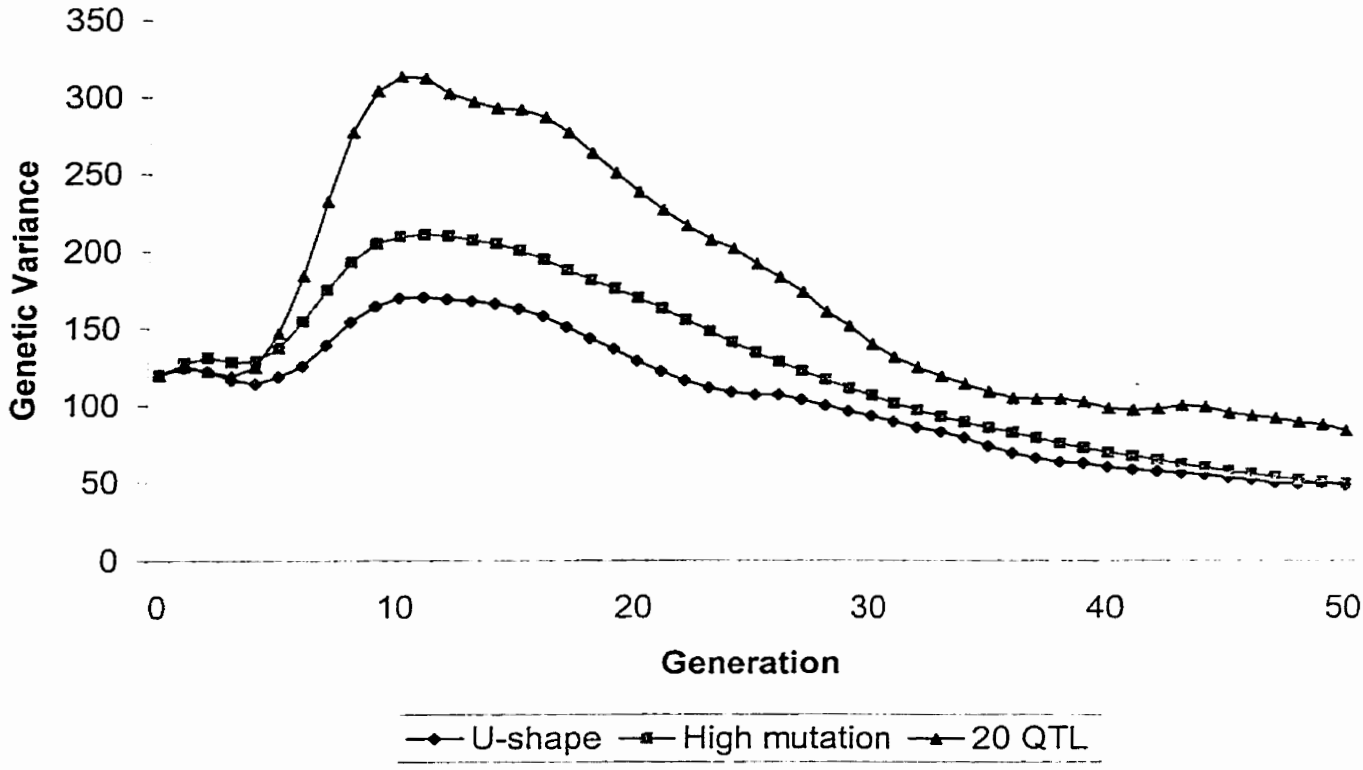


Figure 7.6 **Distribution of allelic frequencies in generation 100.**

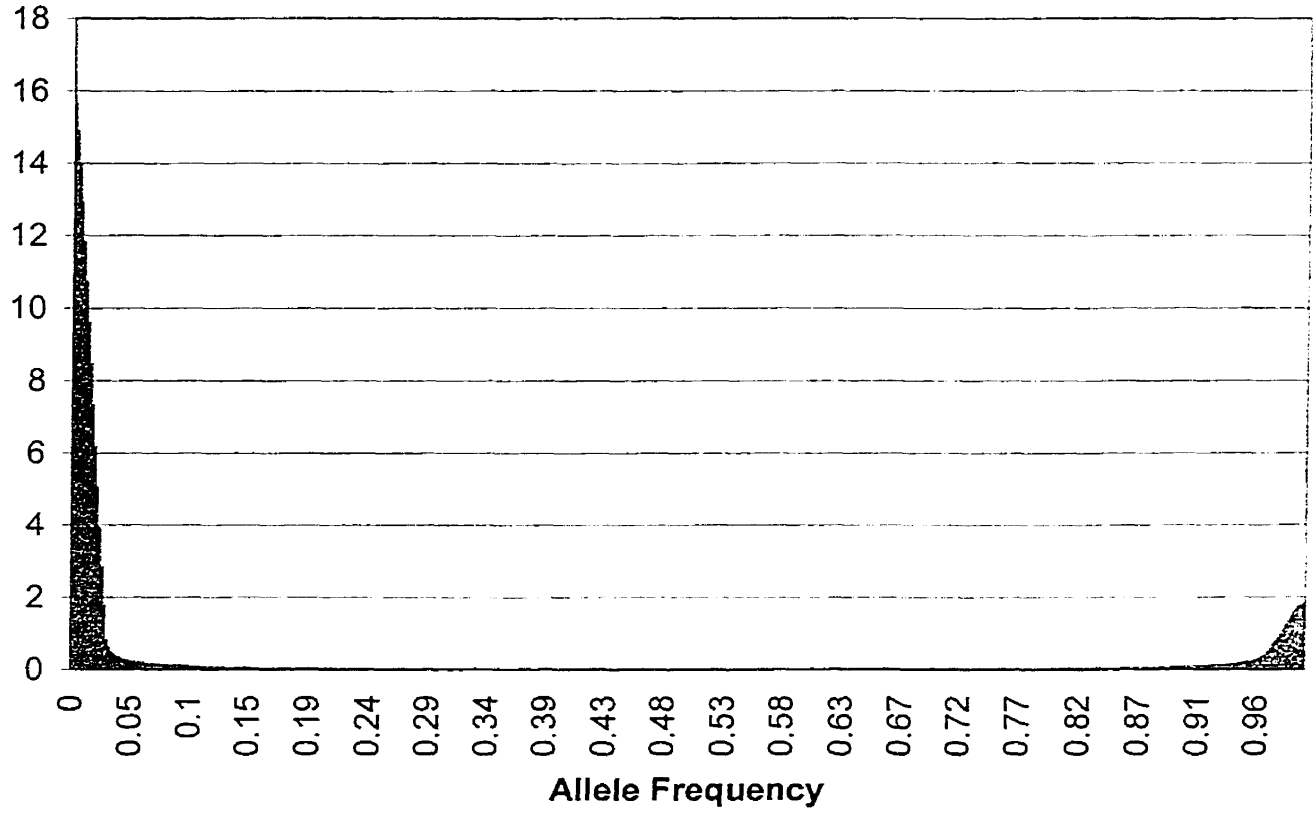


Figure 7.7 **Means of genetic variance in the commercial population and percentage increase in TBV of bulls chosen by MAS across 100 generations from the "improved" simulation.**

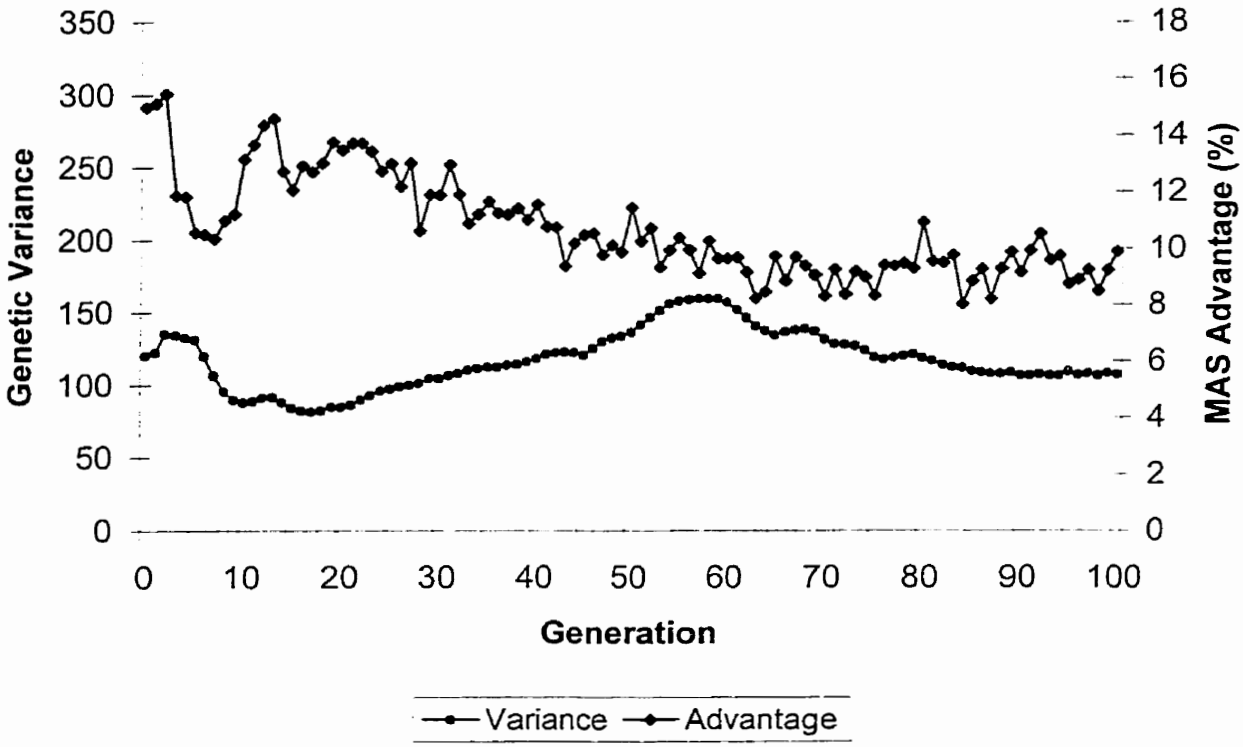
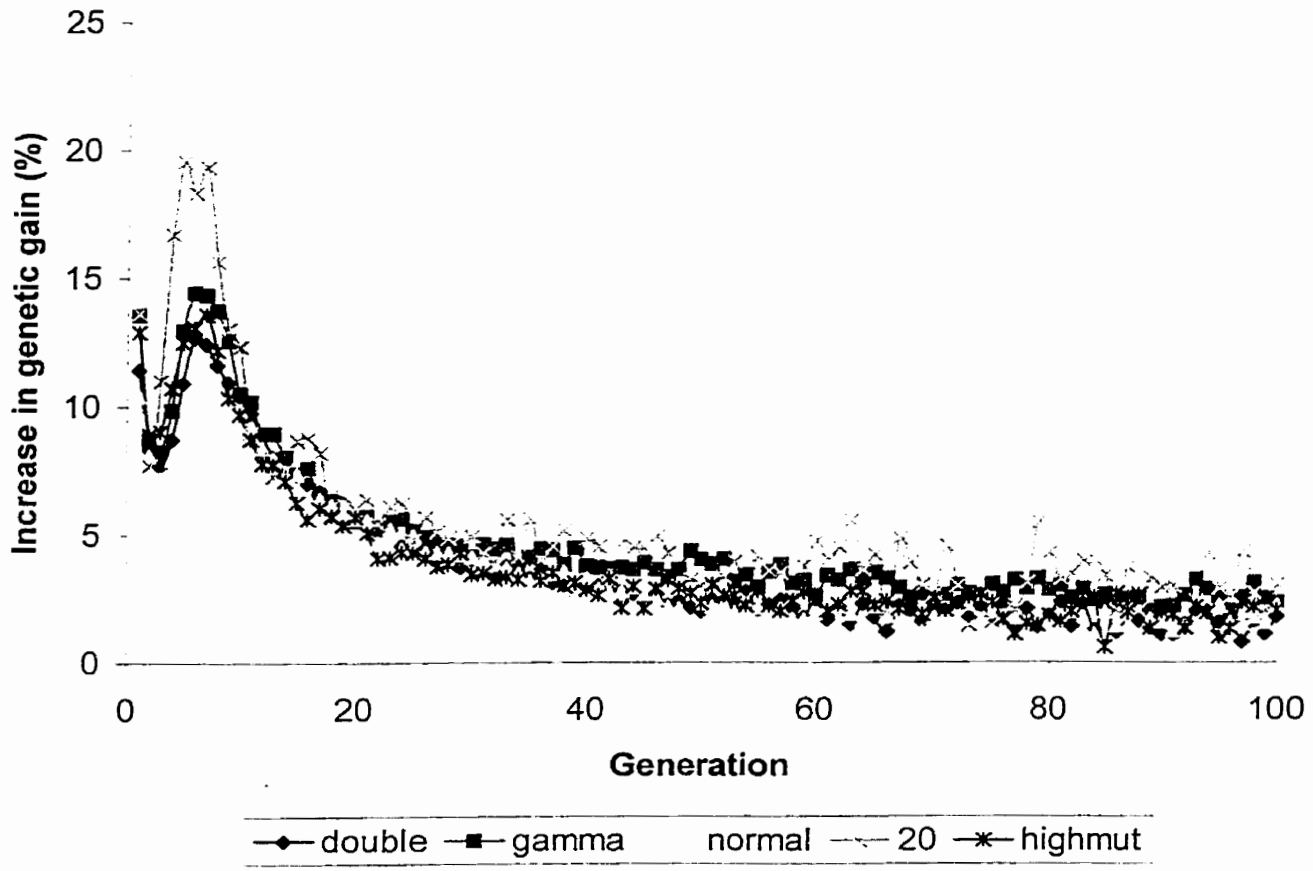


Figure 7.8 **Means across all generations for each simulation model of the percent increase in TBV of young bulls chosen by marker assisted selection.**



8. General Discussion

Computer simulation is a powerful tool with an almost infinite number of uses. The average member of the public is probably most familiar with the use of simulation in computer games or training tools for airline pilots, but simulation is perhaps most valuable when used for scientific study. Simulation is used in some way in nearly all branches of science, from medicine (Mihalas, 1998) to nuclear physics (Laedermann and Décombaz, 2000) and, of course, animal production (Korver and Van Arendonk, 1988).

Simulation holds some distinct advantages over conventional experimentation. Possibly, the biggest advantage is cost. Time is also an important consideration. Dairy cattle take several years to turn over a generation, a process that can be done in milliseconds with today's computers. These huge savings in money and time allow simulated experiments to be large and replicated many times, which increases the statistical power. Another advantage is that the underlying parameters of the simulation are known, which allows for checking and verification of the simulation model. The parameters upon which the simulation is generated can be varied, which allows one to test results for sensitivity to changes in the underlying parameters. Because of these factors, relatively strong conclusions can often be drawn from simulation experiments.

Simulation experiments do have some disadvantages and potential pitfalls, however, which the investigator must consider in the design and especially in the interpretation of results from simulation. For a number of reasons, including imperfect knowledge, mathematical complexity, and insufficient computing resources, simulation programs are almost always subject to simplifying assumptions. In these studies, for example, discrete generations were simulated, although generations overlap in dairy cattle. This assumption was made to simplify the simulation and was not expected to greatly impact the general conclusions. In animal breeding, the infinitesimal (Falconer and Mackay, 1996) or mixed inheritance models of genetic effects are often used. Simulation studies of marker assisted selection (MAS) have often dealt with selection for one (e.g. Spelman and Garrick, 1998) or a few (e.g. Kashi et al., 1990) loci with the remainder of an animal's genetic value being represented by a normally distributed random variable. Because these simplifying assumptions are required, strong conclusions

can often not be made about the absolute numeric values obtained, but rather only about relative values and general trends in how results change as parameters of the simulation are varied over a range. Finally, with simulation, an investigator must often choose between adopting a very simple model about which only general and less precise conclusions can be made or a more specific model that may yield results that yield more precise conclusions that are applicable in a relatively specific context.

Many of the previous studies of the application of MAS have been quite general. The simulations presented in this thesis have taken a more specific approach. In all instances, MAS was simulated for choosing progeny test sires from among full sibs in a nucleus herd. From each of the experiments several firm conclusions can be made about MAS in this context and the general trends observed are likely to hold true in many approaches for MAS.

The primary objective of the first experiment was to investigate whether accounting for a confidence interval for locations for quantitative trait loci (QTL) when making selection decisions increased response relative to performing selection based only upon the genotype at the genomic location where the QTL most likely resided. The results indicated that, in fact, selection response was increased when the uncertainty in QTL location was taken into consideration and young bulls were selected according to an index of marker regression coefficients throughout a confidence interval of the marker location. The mean true breeding value (TBV) of bulls selected by considering the confidence interval exceeding that of randomly selected bulls by 2.60%, versus only 2.00% when only genotypes at the predicted QTL location was used. Accounting for this uncertainty was particularly beneficial in instances where the predicted QTL location was outside of the region bracketed by the markers adjacent to the QTL and recombination had occurred between the true and predicted locations of the QTL. These advantages increased as the size of full-sib families increased. General results for response to MAS were similar to reports by Spelman and Garrick (1998) who used a similar model.

An additional objective of the first study was to examine effects on selection response of using either bootstrapping or approximate LOD scores to construct the confidence interval. Although the bootstrapping approach tended to yield slightly wider confidence intervals than did the LOD approach, no significant difference in the average

TBV of selected young bulls was detected between these two methods. The index used to weight the different chromosomal locations within the confidence interval was relatively simple. A more complex approach that improves selection response could be derived, but this is somewhat doubtful, however, because the selected sons carried the desired allele with a high frequency (approximately 70%). Because of this latter factor, genetic variance at the QTL was quickly decreased throughout the three generations simulated, suggesting that the marker information obtained in the selected generation would have had little value had subsequent generations been simulated.

This final result was examined in more detail in the second experiment, for which MAS was simulated for five generations. A daughter design was used to determine the markers upon which to base selection with a finite locus model similar to Mackinnon and Georges (1997). As part of this experiment, MAS was practiced on the most favourable locus in the first generation within a given sire family and then again in following generations, if possible. As suggested by the first experiment, this locus was of relatively little statistical or practical significance in subsequent generations. For this reason, the primary objective of the second experiment was to implement and compare (in terms of selection response) several strategies for the use of full genome scans in each generation to apply MAS.

The most beneficial strategy was one that selected young bulls on their genotypes at all markers with allele contrasts exceeding a given threshold, whether or not those markers were associated with the same QTL. This approach combined the benefits of selecting on multiple QTL and included aspects of using a confidence interval for QTL location. The average TBV of bulls selected by this strategy exceeded the mean of randomly selected bulls by up to 12%, versus $\leq$7% when only the marker with the greatest contrast was used. An intermediate threshold was optimal. For this specific situation the optimum occurred at around 2.65 standard units, corresponding roughly to a nominal comparisonwise significance level of $\leq$1%, but this value is likely to differ for different situations. Of critical importance was to avoid setting the threshold too high and thus increasing the probability of ignoring some segregating genes. This factor was especially true for selection within a nucleus, because most of the sires were homozygous for the best allele of the most important (based on genetic variance) QTL in the

commercial population. As a result, the QTL for which these selected animals were heterozygous was often one of the QTL of secondary importance, which decreased the expected value of the contrast. Both genetic variance and the advantage of MAS decreased across generations.

The objective of the third experiment was to examine the benefits of MAS for a multiple-trait selection goal, a common characteristic of dairy cattle breeding schemes (Dekkers, 1995). Situations were simulated for an index that included a primary (production) and secondary (functional) trait. Both positive and negative correlations between the two traits were considered in separate simulations. Two approaches to multiple-trait MAS were considered. One was an index based on separate statistical analyses for each of the two traits and the other considered identifying markers that seemed to be linked to QTL that directly affected phenotypes for the index. No significant differences were found between these two approaches. In general, multiple-trait MAS was found to be effective, at least in terms of the total index. The selected young bulls were superior to randomly selected bulls for the secondary trait, but no better, or even somewhat inferior, for the trait of primary importance. The superiority of MAS decreased as the genetic correlation between traits increased. In general, MAS for a multiple-trait objective was less beneficial than was single-trait MAS. In the most favourable situation, the TBV of young bulls selected by MAS exceeded that of randomly selected bulls by about 6%. Many of the results conflicted with the findings of DeKoning and Weller (1994) who performed one of the few published experiments on multiple-trait MAS. However, their model was considerably different from the one used here, particularly because they simulated selection for QTL known without error.

The final experiment examined factors that could effect long-term maintenance of genetic variance and response to MAS and how these factors could be accounted for in simulation. Genetic variance (in the commercial population) was monitored as a function of changes in mutation rates, distributions of allelic effects, number of QTL, and distributions of allelic frequencies. The results suggested that allelic effects should be simulated with a gamma distribution, as proposed by Hayes and Goddard (2000). Allelic frequencies followed a U-shaped distribution, as suggested by Gibson (1999) and others.

In all models simulated, MAS was of benefit throughout the term of the experiment (100 generations).

As is common with most studies, this work answered a number of questions, but left many unanswered, including some that pertain to this work. One of the most important questions is how these results can be extended to real life applications. In some instances, more work may be required. For example, the final three studies were based on full genome scans in each generation. Although such an approach would probably yield the greatest selection response, it may not be optimal under today's cost structure. Some have predicted that rapid genotyping for many loci will soon be available at very low costs (e.g. Visscher and Haley, 1995), but that is not a reality today. Therefore, work may be needed on a strategy that targets the genome scans to certain areas of the genome, based on results of previous genome scans. Results across families could be combined. For example, perhaps an initial genome scan can be done for several families. Later genome scans can then ignore areas for which no indication of QTL was previously found. In the simulation used for the second and third experiments, 20 QTL were placed on 30 chromosomes, so at least 10 chromosomes had no QTL. Ideally, in actual selection programs, at least some of these chromosomes could be eliminated from future scans (although they should eventually be rescanned to detect possible mutations).

Also, a great deal of additional work is needed on the application of MAS to multiple-traits. The procedures used in the third study could possibly be made more elegant and effective. Several authors have suggested that MAS was most beneficial for the secondary traits, so any MAS upon such traits will have to simultaneously consider both the marked and unmarked genes influencing production. Weights for selection indexes with MAS must consider the additional accuracy obtained and the fact that this increase in accuracy may be greater for low heritability traits.

Although mutation is an important contributor to the maintenance of genetic variation, other factors also play a significant role. These factors include migration across populations and interactions among different loci and alleles, such as dominance and epistasis. The impact of these factors on short and long-term response to MAS should be examined.

Finally, additional work is required on uncovering the true genetic model for traits of interest in animal production. Achieving this goal will undoubtedly be aided by ongoing work on humans and model species. With increased knowledge on this broad subject, additional improvements can be made in the simulation of the genetics of livestock populations.

9. References.

Andersen, S. 1994. Calculation of response and variance reduction due to multi-stage and multiple trait selection. Anim. Prod. 58:1-9.

Andersson, L., C. S. Haley, H. Ellengren, S. A. Knott, M. Johansson, K. Andersson, L. Andersson-Eklund, I. Edfors-Lilja, M. Fredholm, I. Hansson, J. Hakansson, and K. Lundstrom. 1994. Genetic mapping of quantitative trait loci for growth and fatness in pigs. Science 263:1771-1774.

Andersson-Eklund, L., L. Marklund, K. Lundstrom, C. S. Haley, K. Andersson, I. Hansson, M. Moller M, and L. Andersson. 1998. Mapping quantitative trait loci for carcass and meat quality traits in a wild boar x Large White intercross. J Anim Sci, 76:694-700.

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

Barendse, W., S. M. Armitage, L. M. Kossarek, A. Shalom, B. W. Kirkpatrick, A. M. Ryan, D. Clayton, L. Li, H. L. Neibergs, N. Zhang, W. M. Grosse, J. Weiss, P. Creighton, F, McCarthy, M. Ron, A. J. Teale, R. Fries, R. A. McGraw, S. S. Moore, M. Georges, M. Soller, J. E. Womack, and D. J. S. Hetzel. 1994. A genetic linkage map of the bovine genome. Nat. Genet. 6:227-235.

Bemark, M., D. Liberg, and T. Leanderson. 1998. Conserved sequence elements in K promoters from mice and humans: implications for transcriptional regulation and repertoire expression. Immunogenetics 47:183-194.

Bernardo, R. 1999. Two-trait selection response with marker-based assortative mating. Theor. Appl. Genet. 98:551-556.

Bink, M. C. A. M. and J. A. M. Van Arendonk, and R. L. Quaas. 1998. Breeding value estimation with incomplete marker data. Genet. Sel. Evol. 30:45-58.

Bink, M. C. A. M. and J. A. M. Van Arendonk. 1994. Marker assisted prediction of breeding values in dairy cattle populations. Proc. 5th World Cong. Genet. Appl. Livest. Prod. 21:233-236.

Bink, M. C. A. M. and J. A. M. Van Arendonk. 1999. Detection of quantitative trait loci in outbred populations with incomplete marker data. Genetics 151:409-420.

Boettcher, P. J. 2000. 2020 Vision? The future of dairy cattle breeding from the academic perspective. Proc. 83rd Annu. Mtg. Am. Dairy Sci. Assoc., Baltimore, MD.

Bovenhuis, H. and I. J. M. DeBoer. 1994. The potential contribution of milk protein loci to improvement of dairy cattle. Proc. 5th World Cong. Genet. Appl. Livest. Prod. 19:311-318.

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. Genet. 137:267-280.

Bovenhuis, H. and R. J. Spelman. 2000. Selective genotyping to detect quantitative trait loci for multiple traits in outbred populations. J. Dairy Sci. 83:173-180.

Bulmer, M. G. 1971. The effect of selection on genetic variability. Am. Nat. 105:201-211.

Cassell, B. G. 2000. Optimal genetic improvement for the high producing cow. Proc. 83rd Annu. Mtg. Am. Dairy Sci. Assoc., Baltimore, MD.

Chakravarti, A. 1999. Population genetics – making sense out of sequence. Nat. Genet. 21:56-60.

Chevalet, C., M. Gillois and, J. VuTienKang. 1984. Conditional probabilities of identity of genes at a locus linked to a marker. Gent. Sel. Evol. 16:431-444.

Christensen, L. G. and T. Liboriussen. 1987. Embryo transfer in the genetic improvement of dairy cattle. Exploiting new technologies in animal breeding: genetic developments. Page 37.

Churchill, G. A. and R. W. Doerge. 1995. Empirical threshold values for quantitative trait mapping. Genetics 138: 963-971.

Cockett N. E., S. P. Jackson, T. L. Shay, F. Farnir, S. Berghmans, G. D. Snowder, D. M. Nielsen and, M. Georges. 1996. Polar overdominace at the ovine callipyge locus. Science 273:236-238.

Colleau, J.J. 1986. Genetic improvement by embryo transfer within an open selection nucleus in dairy cattle. Proc. 3rd World Cong. Genet. Appl. Livest. Prod. 9:127-130.

Coppieters, C., J. Riquet, J.-J. Arranz, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marq, P. Simon, P. Vanmanshoven, D. Wageenar, and M. Georges. 1998. A qtl with major effect on milk yield and composition maps to chromosome 14. Mammalian Genome 9:540-544.

Coppieters, C., Alexandre Kvasz, J.-J. Arranz, B. Grisart, J. Riquet, F. Farnir, M. Georges. 1999. The great-grand-daughter design: a simple strategy to increase the power of a grand-daughter design for QTL mapping. Genet. Res. 74:189-199.

Crow, J. F. and M Kimura. 1970. An introduction to population genetics theory. Harper and Row, NY, USA.

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.

Darvasi, A., A. Weinreb, 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 134:943-951.

Davis, G. P. and S. K. DeNise. 1998. The impact of genetic markers on selection. J. Anim. Sci. 76:2331-2339.

De Koning, D. J., L. L. G. Janss, A. P. Rattink, P. A. M. Van Oers, and B. J. De Vries et al., 1999 Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (Sus scrofa). Genetics 152:1679-1690.

De Koning, G. J. and J. I. Weller. 1994. Efficiency of direct selection on quantitative trait loci for a two-trait breeding objective. Theor. Appl. Genet. 88:669-677.

Dekkers, J. C. M. 1995. Genetic improvement of dairy cattle for profitability. Animal Science Research and Development: Moving Toward a New Century. M. Ivan (ed.). Agri-food Canada. Ottawa, Canada. pp. 307-327.

Dekkers J. C. M. and G. E. Shook. 1990. A semi-stochastic model for simulation of genetic progress in a large dairy cattle population. J. Anim. Breeding Genet. 107:321-333.

Dekkers, J. C. M. and M. R. Dentine. 1991. Quantitative genetic variance associated with chromosomal markers in segregating populations. Theor. Appl. Genet. 81:212-220.

Dekkers, J. C. M. and J. A. M. Van Arendonk. 1998. Optimizing selection for quantitative traits with information on an identified locus in outbred populations. Genet. Res. 71:257-276.

Dentine, M. R. and C. M. Cowan. 1990. An analytical model for the estimation of chromosome substitution effects in the offspring of individuals heterozygous at a segregating marker locus. Theor. Appl. Genet. 79:775-780.

Dupuis, J. and D. Siegmund. 1999. Statistical methods for mapping quantitative trait loci from a dense set of markers. Genetics. 151:373-386.

Emanuelson, U., B. Danell, and J. Philipsson. 1988. Genetic parameters for clinical mastitis, somatic cell counts, and milk production estimated by multiple-trait restricted maximum likelihood. J. Dairy Sci. 71:467-477.

Falconer, D. S. and T. F. C. Mackay. 1996. Introduction to quantitative genetics. Third edition. Longman, Essex, UK.

Farnir, F., W. Coppieters, J.-J. Arranz, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L.e Moreau, M. Mni, C. Nezer, P. Simon, P. Vanmanshoven, D. Wagenaar, and M. Georges. 2000. Extensive Genome-wide Linkage Disequilibrium in Cattle. Genome Res. 10:220-227.

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

Fujii, J., K. Otsu, F. Zorzato, S. de Leon, V. K. Khanna, J. Weller, P. J. O'Brien, and D. H. MacLennan. 1991. Identification of a mutation in a porcine ryanodine receptor that is associated with malignant Hyperthemia. Science 253:448-451.

Georges, M. and L. Andersson. 1996. Livestock genomics comes of age. Genome Research 6:907-921.

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

Georges, M., and W. Coppieters. 2000. QTL Markers and Mapping. CGIL Summer Course. June 5-9. Guelph, Ontario.

Gibson, J. P. 1994. Short-term gain at the expense of long-term response with selection of identified loci. . Proc. 5th World Cong. Genet. Appl. Livest. Prod. 21:201-204.

Gibson, J. P. 1999. Molecular and quantitative genetics: a useful flirtation. Proc. of From J. Lush to Genomics. May16-18, Iowa. pp:77-84.

Goddard, M. E. 1991. Mapping genes for quantitative traits using linkage disequilibrium. Genet. Sel. Evol. 25:131-134.

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

Gomez-Raya, L. and G. P. Gibson. Within-family selection at otherwise unselected locus in dairy cattle. Genome 36:433-439.

Grignola, F. E., I. Hoeschele, and B Tier. 1996. Mapping linked quantitative trait loci via residual maximum likelihood: I Methodology. Genet. Sel. Evol. 28:479-490.

Grobet, L , L. J. Martin, D. Poncelet, D. Pirottin, B. Brouwers, J. Riquet, A. Schoeberlein, S. Dunner, F. M'enissier, J. Massabanda, R. Fries, R. Hanset, and Georges M. 1997. A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. Nat. Genet. 17(1):71-4.

Groen A. F. and C. Smith. 1995. A stochastic simulation study of the efficiency of marker-assisted introgression in livestock. J. Anim. Breed. Genet. 92:161-170.

Hacia, J. 1999. Resequencing and mutational analysis using oligonucleotide microarrays. Nature Genetics 21:42-47.

Haley, C. 1999. Advances in quantitative trait locus mapping. Proc. of From J. Lush to genomics. May16-18, Iowa. pp:47-59.

Haley, C. S. and P. M. Visscher. 1998. Strategies to utilize marker-quantitative trait loci associations. J. Dairy Sci. 81:85-97.

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. and S. A. Knott. 1994. Interval mapping. Proc. 5th World Cong. Genet. Appl. Livest. Prod. 21:25-32.

Hartl, D. L. 1999. A Primer of Population Genetics. Third Edition. Sinauer Assoc. Sunderland, MA.

Hayes B. and M. E. Goddard. 2000. The distribution of the effects of genes affecting quantitative traits in livestock. Genet. Sel. Evol. (accepted).

Henderson, C. R. 1984. Prediction of random variable. Application of linear models to animal breeding. Univ. of Guelph, Guelph, Canada. pp. 33-68.

Henshall, J. M. and M. E. Goddard. 1999. Multiple-trait mapping of quantitative trait loci after selective genotyping using logistic regression. Genetics 151(2):885-894.

Hill, W. G. 1976. Order statistics of correlated variables and implications in genetic programmes. Biometrics. 32:889-902.

Hill, W. G. 1982. Prediction of response to artificial selection from new mutation. Genet. Res. 72:159-168.

Hoeschele, I, B. Tier and, H. U. Graser. 1995. Multiple-trait genetic evaluation for one polychotomous trait and several continuous traits with missing data and unequal models. J. Anim. Sci. 73:1609-27.

Hoeschele, I, P. Uimari, F. E. Grignola, Q. Zhang and K. M. Gage. 1997. Advances in statistical methods to map quantitative trait loci in outbred populations. Genetics 147:1445- 1457.

Hoeschele, I. 1994. Bayesian QTL mapping via the Gibbs sampler. Proc. 5th World Cong. Genet. Appl. Livest. Prod. 21:241-244.

Hoeschele, I. and P. M. Van Raden. 1993. Bayesian analysis of linkage between genetic markers and quantitative trait loci. I and II. Theor. Appl. Genet. 85:946-952, 953-960.

Hoeschele, I. and T. R. Meinert. 1990. Association of genetic defects with yield and type traits: The Weaver locus effect on yield. J. Dairy Sci. 73:2503-2515.

Jansen, R. C. 1992. A general mixture model for mapping quantitative trait loci using molecular markers. Theor. Appl. Genet. 85:252- 260.

Janss, L. L. G., R. Thompson, and J. A. M. Van Arendonk. 1995. Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations. Theor. Appl. Genet. 91:1137-1147.

Jensen, J. 1989. Estimation of recombination parameter between a quantitative trait locus (QTL) and two marker gene loci. Theor. Appl. Genet. 78:613-618.

Juga, J. and A. Maki-Tanila. 1987. Genetic change in a nucleus breeding dairy herd using embryo transfer. Acta Agric. Scand. 37:511-519.

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

Keightley, P. D. 1998. Genetic basis of response to 50 generations of selection on body weight in inbred mice. Genetics 148:1931-1939.

Kinghorn, B. P. and B. E. Clarke. 1997. Genetic evaluation at individual QTL. Animal Biotechnology 8: 63-68.

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

Knott, S. A., D. B. Neale, M. M. Sewell, and C. S. Haley. 1997. Multiple marker mapping of quantitative trait loci in an outbred population of Loblolly pine. Theor. Appl. Genet. 94:810–820.

Knott, S.A.; L. Marklund, C. S. Haley, K. Andersson, W. Davies, H. Ellegren, M. Fredholm, I. Hansson, B. Hoyheim, K. Lundstrom,M. Moller, and L. Andersson. 1998. Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and large white pigs. Genetics, 149:1069-80.

Knott, S. A. and C. S. Haley. 2000. Multitrait least squares for quantitative trait loci detection. Genetics 156:899-911.

Korver, S. and J. A. M. Van Arendonk. 1988. Modelling of Livestock Production Systems. Kluwer Academic Publ. London, UK.

Kuhn, C., G. Freyer, R. Weikard, T. Goldammer, and M. Schwerin. 1999. Detection of QTL for milk production traits in cattle by application of a specifically developed marker map of BTA6. Anim. Genet. 30:333-340.

Laedermann, J.P. and M. Décombaz. 2000. Simulation of nuclear decay. Appl. Radiation and Isotopes. 52:419-425.

Lande, R. 1981. The minimum number of genes contributing to quantitative variation between and within populations. Genetics 99:541-553.

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 L. Kruglyak. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genetics 11: 241-247.

Lipkin E , M. O. Mosig, A. Darvasi, E. Ezra, A. Shalom, A. Friedmann, and M. Soller. 1998. Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers: analysis of milk protein percentage. Genetics 149:1557-1567.

Liu, B.H. 1998. Statistical Genomics. CRC Press. New York, N.Y.

Lyn, D., N. L. Lystock, and M. Smulson. 1995. Conservation of sequences between human and gorilla lineages: ADP-ribosyltransferase (NAD^+) pseudogene 1 and neighboring retroposons. Genetics 155:241-250.

Lynch, M. 1988. The rate of polygenic mutation. Genet. Res. 51:137-148.

Lynch, M. and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer, MN, USA.

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

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

Malecot, G. 1984. Les mathematiques de l'heredite. Masson et Cie, Paris, France.

Mangin, B., B. Goffinet, and A. Rebai. 1994. Constructing confidence intervals for QTL location. Genetics 138:1301-1308.

Mangin, B. and B. Goffinet. 1997. Comparison of several confidence intervals for QTL location. Heredity 78:345-353.

McDaniel, B. T. and B. G. Cassell. 1991. Effects of embryo transfer on genetic change in dairy cattle. J. Dairy Sci. 64:2484-2494.

Meuwissen, T. H. E. 1990. The effect of the size of MOET nucleus dairy cattle breeding plans on the genetic gain and its variance. Proc. 4th World Cong. Genet. Appl. Livest. Prod. 14:271-274.

Meuwissen, T. H. E. and J. A. M. Van Arendonk. 1992. Potential improvement in rate of genetic gain from marker assisted selection in dairy cattle breeding schemes. J. Dairy Sci. 75:1651-1659.

Meuwissen, T. H. E. and M. E. Goddard. 1996. The use of marker haplotypes in animal breeding schemes. 28:161-176.

Mihalas, G.I. 1998. Modelling and simulation in medicine and life sciences. Medical Informatics. 23:93-96.

Moreau L., A. Charcosset, F. Hospital and A. Gallais. 1998. Marker assisted selection efficiency in populations of finite size. Genetics 148:1353-1365.

Muir, W. M. and D. A. Stick. 1998. Relative advantage of combining genes with a major effect in breeding programs: simulation results. Proc. 6th World Cong. Genet. Appl. Livest. Prod. 26:357-360.

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

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

Oliver, S. G. 1996. From DNA sequence to biological function. Nature 379:597-600.

Ollivier, L. 1998. The accuracy of marker assisted selection for quantitative traits within populations in linkage equilibrium. Genetics 148:1367-1372.

Pagnacco , G. and G. B. Jansen. 2000. Use of marker haplotypes to refine covariances among relatives for breeding value estimation. J. Anim. Breed. Genet. (accepted).

Poso, J. and E. A. Mantysaari. 1996. Relationships between clinical mastitis, somatic cell score, and production for the first three lactations of Finnish Ayrshire. J. Dairy Sci. 79:1284-1291.

Pryce, J. E., R. F. Veerkamp, R. Thompson, W. G. Hill, and G. Simm. 1997. Genetic aspects of common health disorders and measures of fertility in Holstein Friesian dairy cattle. Anim. Sci. 65:353-360.

Riquet, J., W. Coppieters, N. Cambisano, J. J. Arraz, P. Berzi, S. K. Davis, B. Grisart, F. Farnir, L. Karim, M. Mni, P. Simon, J. F. Taylor, P. Vanmanshoven, D. Wagenaar, J. E. Womack and, M. Georges. 1999. Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle. Proc. Natl. Acad. Sci. USA 96:9252-9257.

Rohrer, G. A. and J. W. Keele. 1998. Identification of quantitative trait loci affecting carcass composition in swine: I fat deposition traits. J. Anim. Sci 76:2247-2254.

Ronin, Y. I.; V. M. Kirzhner, and A. B. Korol. 1995. Linkage between loci of quantitative traits and marker loci: multitrait analysis with a single marker. . Theor. Appl. Genet. 90:776-786.

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. 97:1169-1178.

Roth, A., E. Strandberg, B. Bergland, U. Emanuelson, and J. Philipsson. 1999. Genetic correlations among female fertility traits in different parities in Swedish dairy cattle. pp. 177-181 *in* Occasional Publication No. 24, British Society of Animal Science.

Rothschild M. and M. Soller. 1997. Candidate gene analysis to detect genes controlling traits of economic importance in domestic livestock. Probe 8:13-20.

Rothschild, M., C. Jacobson, D. Vaske, C. Tuggle; L. Wang; T. Short, G. Eckardt, S. Sasaki, A. Vincent, D. McLaren, O. Southwood, H. van der Steen, A. Mileham, and G. Plastow. 1996. The estrogen receptor locus is associated with a major gene influencing litter size in pigs. Proc Natl Acad Sci U S A, 93(1):201-205.

Ruane, J. and J. J. Colleau. 1994. The value of marker assisted selection when one QTL is marked. 45th meeting EAAP. Edinburgh, U.K.

Ruane, J. and J. J. Colleau. 1995. Marker assisted selection for genetic improvement of animal populations when a single QTL is marked. Genet. Res. 66:71-78.

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.

Schulman, N. F., M. J. de Vries, and M. R. Dentine. 1999. Linkage disequilibrium in two stage marker-assisted selection. J. Anim. Breed. Genet. 116:99-110.

Shrimpton, A. E. and A. Robertson. 1988. The isolation of polygenic factors controlling bristle score in Drosophila Melanogaster. Genetics 118:445-459.

Schrooten, C. and J.A.M. Van Arendonk. 1992. Stochastic simulation of dairy cattle breeding schemes: Genetic evaluation of nucleus size and type. J. Anim. Breed. Genet. 109:1-15.

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

Smith, C. and S. P. Simpson. 1986. The use of genetic polymorphisms in livestock production. Anim. Prod. 20:1-10.

Smith, C. and D. B. Smith. 1993. The need for close linkages in marker assisted selection for economic merit in livestock. Anim. Breeding Abstr. 61:197-204.

Soller, M. and I. Medjugorac. 1999. Making the transition from quantitative trait locus mapping to marker-assisted selection. Proc. of From J. Lush to genomics. May16-18, Iowa. pp. 1-5

Spelman, R. J. 1998. Major factors in marker assisted selection genetic response in dairy cattle populations. Proc. 6th World Cong. Genet. Appl. Livest. Prod. 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. Genet. 144:1799-1808.

Spelman, R. J. and D. J. Garrick. 1997. Utilization of marker assisted selection in commercial dairy cow populations. Livest. Prod. Sci. 47:139-148.

Spelman, R. J. and J. A. M. Van Arendonk. 1997. Effect of inaccurate parameter on genetic response to marker assisted selection in outbred populations. J. Dairy Sci. 80:3399-3410.

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. 81:2942-2950.

Spelman, R. J. and H. Bovenhuis. 1998. Genetic responses 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., D. J. Garrick and, J. A. M. Van Arendonk. 1999. Utilization of genetic variation by marker assisted selection in commercial dairy cattle populations. Livest. Prod. Sci. 59:51-60.

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

Stam, P. 1987. Marker genes in selection: biochemical polymorphism as markers in selection for quantitative trait. Animal Genetics 18:97-99.

Thaller, M. C., S. Schippa, and G. M. Rossolini. 1998. Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase superfamily. Prot. Sci. 7:1647-1652.

Uimari, P., Q. Zhang, F. E. Grignola, I. Hoeschele and, G. Thaller. 1996. Analysis of QTL workshop. Granddaughter design data using least squares, residual maximum likelihood and Bayesian methods. J. Quant. Trait Loci 2:7.

Van Arendonk, J. A. M., B. Tier, and B. P. Kinghorn. 1994a. Use of multiple genetic markers in prediction of breeding values. Genetics 137:319-329.

Van Arendonk, J. A. M., H. Bovenhuis, S. Van Der Beek, and A. F. Groen. 1994b. Detection and exploitation of markers linked to quantitative traits in farm animals. Proc. 5th World Cong. Genet. Appl. Livest. Prod. 21:193-200.

Van Der Beek, S. and J. A. M. Van Arendonk. 1993. Criteria to optimize designs for detection and estimation of linkage between marker loci from segregating populations containing several families. Theor. Appl. Genet. 86:269-280.

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-sibs and half-sibs families. Theoret. Appl. Genet. 91:1115- 1124.

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

Velmala, R. J., H. J. Vilkki, K. T. Elo, D. J. De Koning, and A. V. Maki-Tanila. 1999. A search for quantitative trait loci for milk production traits on chromosome 6 in Finnish Ayrshire cattle. Anim. Genet. 30:136-43.

Vilkki, H. J., D. J. De Koning, K. Elo, R. Velmala, A. Maki-Tanila. 1999. Multiple marker mapping of quantitative trait loci of Finnish dairy cattle by regression. J Dairy Sci 80:198-204.

Villanueva, B., R. Pong Wong, B. Grundy and J. A. Woolliams. 1999. Potential benefits from using an identified major gene in BLUP evaluation with truncation and optimal selection. Genet. Sel. Evol. 31: 115-133

Visscher, P. M. and C. S. Haley. 1995. Utilizing genetic markers in pig breeding programs. Anim. Breed. Abstr. 63:1-8.

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

Wang, Y., Dekkers, J. C. M. and J. P. Gibson. 1994. Effect of data structure on statistical power of daughter and granddaughter designs for detecting marker-QTL associations. Proc. 5th World Cong. Genet. Appl. Livest. Prod. 21:252-255.

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. 1990. Experimental design for mapping quantitative trait loci in segregating populations. Proc. 4th World Cong. Genet. Appl. Livest. Prod. 13:113-116.

Weller, J. I., G. R. Wiggans, P. M. VanRaden, and M. Ron. 1996. Application of canonical trasformation to detection of quantitative trait loci with the aid of genetic markers in a multitrait experiment. Theor. Appl. Genet. 92:998-1002.

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. 73:2525-2537.

Welper, R. D. 2000. A futuristic look at the dairy cattle genetic industry. Proc. 83rd Annu. Mtg. Am. Dairy Sci. Assoc., Baltimore, MD.

Whittaker, J. C., R. N. Curnow, C. S. Haley, and R. Thompson. 1996. Using marker maps in marker assisted selection. Genet. Res. 66:255-265.

Woolliams, J. A. and C. Smith. 1988. The value of indicator traits in the genetic improvement of dairy cattle. Anim. Prod. 46:333-345.

Xie, C. and S. Xu. 1998. Efficiency of multistage marker-assisted selection in the improvement of multiple quantitative traits. Heredity 80:489-498.

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.

Zeng, Z. B. and C. C. Cockerham. 1993. Mutation models and quantitative genetic variation. Genetics 133:729-736.

Zhang, Q., D. Boichard, I. Hoeschele, C. Ernst, A. Eggen, B. Murkve, M. Pfister-Genskow, L. A. Witte, F. E. Grignola, P. Uimari, G. Thaller and, M. D. Bishop. 1998. Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. Genetics, 149:1959-73.

Zhang, W. and C. Smith. 1993. Computer simulation of marker assisted selection utilizing linkage disequilibrium. Theor. Appl. Genet. 83:813-820.

Appendix 1. Comparison of interval mapping and single marker contrasts for MAS

Interval mapping was one of the most time consuming aspects of the simulation in Chapter 4. In order to simplify simulations for the subsequent research objectives, it would be opportune to simplify MAS by using single marker contrasts. Thus, it was necessary to test how much advantage was really obtained from the more complex interval mapping approach.

The selection criteria used in Chapter 4 were based on dividing the chromosome into many loci separated each by one centiMorgan and calculating the probability of inheritance from grandsire to grandson at each location. Then the regression coefficients associated with each of these positions were multiplied by the transmission probability and summed across the confidence interval. However, markers were only available for a small proportion of these locations. Therefore, we hypothesized that considering in the selection criteria only those loci at which markers were present would yield results that were not significantly different from the approach that considers all loci.

To test this hypothesis, equation [3] from Chapter 4 was modified and the following equation was used for the selection criteria:

$$I = \sum_{i=1}^{m} p_i \beta_{ij}$$

where $i = 1,\ldots m$, refers to only those loci within the confidence interval where an informative marker was present, p_i is the probability that the son inherited the reference grandsire haplotype at cM i, and β_{ij} is the standardized regression coefficient for

grandsire j at cM i. The breeding values of the sons selected by this criterion were then compared to those of sons selected by equation [3] of Chapter 4 and randomly selected. Because no difference was observed between the two approaches for defining the confidence interval, this comparison was made for the LOD approach only. Forty offspring per dam were simulated.

As hypothesized, no significant difference was observed between the approach that considered all loci and the current approach, that considered only the marked loci. The average breeding value of bulls selected by considering all loci was 14.30 versus 14.28 when only marked loci were considered. Both were significantly superior to random selection (13.94).

These results can be explained by the fact that the genotypes at marker loci provided all of the available information about the transmission of the favourable QTL allele from grandsires to grandsons in intervening intervals. Thus, even though interval mapping provides more precise estimates of QTL location than do single marker contrasts, as the QTL location is not confounded with its effect (Liu 1998), for the purposes of MAS there is probably little difference between them. Precision with regard to QTL location is not highly critical for the purposes of MAS. Increased mapping precision has been shown to have little effect on the power of QTL detection (Darvasi et al., 1993: Dupuis and Siegmund, 1999) and, therefore, should have little effect on power and subsequent response to MAS. This finding justisfies the use of single marker contrasts in the studies in Chapters 5 to 7.