Improved Marker-Assisted Selection in Dairy Cattle Breeding Schemes:

Selective Genotyping/Phenotyping

Ph.D. thesis

Saeid Ansari-Mahyari

2007

Department of Large Animal Sciences
Faculty of Life Sciences
University of Copenhagen

Department of Genetics and Biotechnology
Faculty of Agricultural Sciences
University of Aarhus

In the Name of Allah ,
the Most Compassionate and the Most Merciful

Preface

All praise is to Allah, to whom all will return and whose knowledge is infinite, eternal.

This thesis was financed by Agriculture Research and Education Organization of Iranian Ministry of Jahad-e-Agriculture. The research was carried out at the Department of Genetics and Biotechnology (GBI-Foulum), Faculty of Agricultural Sciences, University of Aarhus and it is submitted to the Faculty of Life Sciences, University of Copenhagen.

First of all I like to use this preface to thank all employees, scientists and students in GBI-Foulum for making a great positive and relaxed atmosphere to complete this research.

I would like to thank Peer Berg for his excellent guidance, helpful comments, inspirations and scientific idea. His advice and patience on this study improved my understanding in genetic evaluation and animal breeding strategies. My sincere thanks to Lars Gjøl Christensen for all the official help and assistance he has provided during my PhD period.

I am also grateful to Mogens Sandø Lund for his supervision in my thesis and his fruitful discussions and useful suggestions about QTL analysis approaches and the practical solutions. Thanks to Anders Christian Sørensen for great assistances in the last part of my study and discussions about dairy breeding schemes. His skills in programming provided a considerable influence to improve the simulations.

I would like to express my grateful thanks to Bernt Guldbrandtsen for providing me with valuable scientific discussions and helps in Linux system.

Great thanks to Peter Løvendahl for filling up the last months of my PhD period with friendly and fruitful scientific discussions.

I am thankful to all other people and friends in GBI-Foulum for lots of scientific and non-scientific discussions: Mohammad Shariati, Hauke Thomsen, Peter Sørensen, Per Madsen, Louise Pedersen, Elise Norberg, Goutam Sahana, Bart Buitenhuis, Trine Villumsen, and Johanna Hoglund. I have enjoyed the working environment where I always felt welcome at my office.

Lastly but most sincerely, thanks to my wife and the best friend in my life ***Masoomeh Kazemi*** for encouragements and being wonderful and patient with me whilst my study in Denmark, and my lovely daughter ***Elmira***, the great joy of my life.

Saeid Ansari-Mahyari
July 2007

Summary

Use of detected QTL (quantitative trait loci) to increase genetic gain in dairy cattle breeding has been argued as a new tool during the last decade. In some traits, the cost of genotyping is higher compared to phenotyping which is routinely recorded e.g. milk yield and milk contents. In these traits, selecting the most informative individuals for genotyping is cost-effective and can effectively detect QTL while reducing genotyping effort and cost. Therefore, instead of genotyping all animals to detect and estimate the effect of QTL, under selective genotyping only a fraction of the population could be genotyped. There are problems in phenotypic recording of some other traits e.g. quality traits or disease resistance traits which are typically expensive to record. In such situations the costs of phenotyping might exceed or equal the costs of genotyping and therefore, dynamic selective phenotyping strategies are needed to only phenotype the most informative individuals. The general objectives of this thesis are to develop genotyping and phenotyping strategies and to test these strategies in QTL detection and also using selective genotyping in marker-assisted selection (MAS) in a dairy cattle breeding scheme. It is the objective to develop genotyping and phenotyping strategies to select animals where recording has the largest marginal effect in the two applications, thus maximizing the power of QTL detection given constraints in recording phenotypes or genotypes.

The objectives of chapter 1 are to give an overview and introducing the current selective approaches in QTL studies, MAS and also the criteria in selective genotyping and phenotyping. In addition, this chapter motivates the use and development of selective methods used for detection of QTL and application in a breeding scheme.

The following three chapters present the results from different simulation studies. The first paper, chapter 2, evaluates two selective genotyping strategies based on combined phenotypic and genotypic information. These methods were compared to selective genotyping based on only phenotypic information (conventional selective genotyping), and random sampling animals for genotyping in QTL detection experiments. Power to detect QTL was significantly higher for the combined methods compared to conventional selective genotyping and random genotyping methods, at all genotyping levels (10, 20 and 30%). Combining phenotypic and genotypic information can be considered as an alternative approach in order to: decrease genotyping costs, unbiased QTL effects, decrease sampling variance of the QTL variance component and increase power of QTL detection.

The second paper, chapter 3, compares the strategies in selective phenotyping in order to sample more informative individuals to be phenotyped based on linkage information within the half-sib families or linkage disequilibrium information across families in the population. Selecting progeny that are genetically similar to paternal haplotypes provided a better contrast in detecting QTL than random phenotyping. However, to estimate unbiased QTL parameters using this criterion for selective phenotyping requires large family sizes or prior information on QTL position to detect and position the QTL. Selective phenotyping across families can improve sensitivity of QTL position compared to random selective phenotyping. In addition, by increasing the proportion of phenotyping (30% to 50%) using linkage disequilibrium information, the estimated QTL effect approaches the true value quicker than using criteria based on linkage information.

The third paper, chapter 4, examines the advantage of using selective genotyping strategies in MAS, to decrease the amount of genotyping and devise a method for genotyping the animals contributing the most information. Initially, a sex-limited trait was simulated with $h^2 = 0.04$ and QTL effect equal to $\frac{1}{4}\sigma^2_G$. Two stages of selection among the male progenies and also females were carried out for 25 years. One genotyping strategy was based on the sum of QTL and polygenic estimated effects in each candidate. Another strategy used an index based on both total genetic effects and Mendelian sampling variance of the parents due to QTL. Random genotyping was also considered to compare with the strategies. Totally, MAS schemes increased genetic response due to a detected QTL and also total genetic gain and decreased inbreeding rate compared to traditional BLUP. The selective genotyping strategies showed similar genetic gain over selection within genotyping level. The results showed that it would be possible to achieve minimal loss of response at a detected QTL using fractional genotyping compared to complete genotyping and less increasing in inbreeding compared to traditional BLUP in a practical dairy cattle breeding.

Finally, the thesis concludes with a discussion and conclusion of the results from the three papers. The derived selective strategies illustrated the possibility to select more informative candidates to decrease genotyping or phenotyping costs. Additionally, the developed genotyping/phenotyping strategies can enhance the precision of QTL parameters in QTL experiments in comparison to random methods, and increase the genetic gains using MAS in dairy cattle breeding schemes compared to current selection methods.

Contents

Preface **i**

Summary **iii**

Chapter 1

General Introduction **1**

Introduction **3**

Objectives **15**

Chapter 2

Combined use of phenotypic and genotypic information in sampling animals for genotyping in detection of quantitative trait loci **25**

Chapter 3

Fine mapping QTL under selective phenotyping strategies based on linkage and linkage disequilibrium criteria **49**

Chapter 4

Across-family marker-assisted selection using selective genotyping strategies in dairy cattle breeding schemes **75**

Discussion and conclusion **103**

Chapter 1

General Introduction

Introduction

1. Background

Over the last decades, conventional animal breeding successfully utilized phenotypic and pedigree information to best linear unbiased predictors (BLUP) of breeding values using an animal model. The dairy cattle breeding schemes are designed such that genetic gain is maximized by increasing the accuracy of selection (e.g. progeny testing schemes in dairy cattle) or by decreasing the generation interval. Including genetic markers which are linked to quantitative trait locus (QTL) can increase the selection accuracy of young bulls and bull dams during the young ages. The extra genetic response achieved by marker information was proved to be limited when QTL information was used to augment conventional progeny testing programs (Meuwissen and van Arendonk, 1992; Spelman and Garrick, 1997; Spelman and Garrick, 1998). In situations where the classical phenotype selection (based on BLUP evaluation) has several limitations, for example a trait occur as sex-limited, expressed late in life, difficult to record (costly or measuring state) and low heritability, including QTL in the genetic evaluation is an efficient method to increase the genetic gains (Lande and Thompson, 1990; Meuwissen and Goddard, 1996).

The use of linkage disequilibrium (LD) to locate genes which affect quantitative traits has increased recently. Linkage disequilibrium occurs when the QTL and genetic markers are transferred together more frequently than what is expected by chance and therefore, QTL can be detected using the identified haplotypes along a chromosome (Meuwissen and Goddard 2004). QTL (fine) mapping can also be achieved using combined linkage disequilibrium and linkage analysis techniques in animal breeding schemes.

Mapping of quantitative traits is an important topic in the field of animal science with a lot of projects (Misztal, 2006). The present genotyping cost is generally high and is still considered as one of the major problem in the QTL studies. However, there are limited information on cost effectiveness of QTL detection and utilization in animal breeding plans. Because of high genotyping costs, it is important to find more efficient ways in the QTL/gene identification, utilization and verification. In order to decrease the costs of genotype identification, one alternative is to reduce the number of animals for genotyping. In this case, it is necessary to choose the most informative individuals for genotyping. The focused investigations on the selection of individuals for genotyping have concluded that it is not necessary to genotype the whole population but by a design using the selectively genotyped individuals high power in QTL detection could be achieved (Darvasi and Soller, 1992; Kinghorn, 1999).

There are some traits which are difficult or very expensive to measure and therefore, a dynamic strategy to identify informative individuals for phenotyping is necessary to improve these traits. The size of an experiment to detect QTL is also limited by the rearing costs of individuals for recording the phenotype. Selective phenotyping (SP) is a sampling strategy which is proposed for these traits in a phenotypic recording program. In compared to selection of informative individuals for genotyping (as explained in the previous paragraph), it is assumed in SP that all individuals are easily genotyped but not phenotyped. With complete genotyping, one alternative is to practice SP for the target trait or correlated traits in a population. Casu *et al.* (2003) indicated that after verifying a genomic region, selective phenotyping of the 50% of the population which showing the extreme values in the correlated traits, was good enough in QTL detection of target trait.

2. Selective genotyping approaches

2.1. QTL detection

The first step in marker-assisted selection (MAS) is to detect genes of interest. Sometimes the evidences from QTL experiments are not conclusive. In order to apply the QTLs in selection schemes, they need to be verified in the next experiments to confirm the results (Bovenhuis and Spelman, 2000), through for example fine mapping approaches.

The cost of genotyping an entire population is generally prohibitive. Selecting individuals for genotyping was firstly introduced by Lander and Botstein (1989). Darvasi and Soller (1992) introduced selective genotyping for a single marker linked to a QTL. With selective genotyping, only individuals from the high and low phenotypic extremes are genotyped. Therefore, this method considers the phenotype in a population and only a part of the population is chosen for genotyping. Due to selecting a part of the population for genotyping, the selective genotyping could cause a bias in QTL estimations and it is necessary to use a correction factor for the means to do the contrast of the extreme means (Darvasi and Soller, 1992). In addition, genetic variance needs to be corrected. The general principle exploited in selective genotyping is that most linkage information can be inferred from individuals showing extreme phenotype values. For a given number of individuals genotyped in this approach, the power to detect QTL is increased.

Kinghorn (1997) presented an index to indicate the information content of genotype probabilities (GPI), which had been derived from the results of segregation analysis. The GPI identifies which individuals (or loci) should be genotyped in order to maximize the benefit of genotyping in the population (Kinghorn, 1999). The genotyping models could be used to identify and genotype through individual-by-individual (Kinghorn, 1997 and 1999) or group-by-group (Macrossan, 2004). In other words, these methods could help to facilitate detection of major genes affecting quantitative traits and also in the description of genetic distance between populations, to maximize the benefit/cost ratio of genotyping operations. However, this approach is based on the genotype probabilities and there is not any report to indicate comparison with the conventional selective genotyping methods. In addition, GPI method becomes more complicated when a large haplotype is considered over several generations.

Current selective methods for genotyping have used either extreme phenotypes (EP) or genotypic information (by GPI), even if a proportion of animals already have been genotyped. A combination of both genotypic and phenotypic information to find the most informative animals for genotyping can be expected to be more powerful in QTL mapping studies. Therefore using all available genotypic and phenotypic records could give more accurate estimation of the QTL parameters and ultimately, decreasing number of animals for genotyping.

2.2. Marker-assisted selection

After detecting a genomic region which contains a segregating QTL, incorporating molecular information in the breeding schemes with marker-assisted selection (MAS), can be used to accelerate the genetic gain achievable through standard methods. Fernando and Grossman (1989) showed a method to incorporate information on a single marker-linked QTL into the mixed model equations to estimate the best linear unbiased predictions of breeding values (BLUP-EBV). According to their gametic model, breeding values are partitioned as additive effects of the linked QTL allele from sire and dam plus remaining effects of unmarked trait loci.

There are several simulation studies on the use of MAS (Meuwissen and Goddard, 1996; Spelman and Bovenhuis, 1998). Using marker information in selection schemes indicated that the genetic merit can enhance through increasing the accuracy of genetic evaluation in outbred populations (Lande and Thompson, 1990; Goddard and Hayes, 2002; Villanueva *et al.*, 2005), where the accuracy of selection is low when using conventional selection schemes (*e.g.* fertility traits). In most situations, there is not complete genotype information in the population. However it is possible in some cases to find a genotype probability of an individual without genotype information, using stochastic methods or using deterministic methods (Pong-Wong *et al*., 2001). Those probabilities can be used to estimate the gametic IBD (identical by descent) probabilities. A stochastic method is fairly flexible since it uses Monte Carlo Markov Chain approaches to handle pedigrees in any scope and structure with missing marker information. One major problem in this method is the computational time required. Deterministic methods are faster (but less accurate), and known as an attractive alternative to stochastic methods. Nagamine and colleagues (2002) introduced the simple deterministic method to estimate IBD coefficients for a population with a simple two generation pedigree.

After detection and verification of QTLs by different strategies, the QTL-linked markers can be used in MAS schemes. In the best situation when the marker is completely linked to a specific QTL, it is possible to detect the alleles with the best and the worst performers. These identified markers upon confirmation, will be highly useful for the selection plans. Equally important, they will be useful for checking the success of traditional selective breeding programs (Liu, 2001). In MAS, genetic information of the markers has been used as a criterion of indirect selection for genetic improvement of a given quantitative trait. When MAS is used in a population, the frequency of the favorable QTL allele is quickly increased during the first generations compared to conventional selection based on BLUP evaluation, especially if the detected QTL has a large effect compared to the total genetic effect of the trait. In order to assist selection program with detected QTL in the long-term selection, new QTL-linked markers should continuously be discovered and used in evaluation.

The increased efficiency of MAS, however, is accompanied by the increased cost involved in sample collection for genotyping. Therefore, the described strategies in this study may reduce the costs of using MAS while maintain consistent extra genetic gain. Selection of an animal for genotyping should be related to the linkage of marker loci and the QTL, which is based on the marginal effect of genotyping of that particular individual on the selection response. In order to use detected QTL, it is necessary to identify the impact of using informative individuals either in detection or selection. However, as in this project only the most informative animals will be selected for genotyping, it is important to study the effect of incomplete genotyping information in the selection schemes.

2.3. Observed criteria

2.3.1. Uncorrected criteria: It is proposed that the highest and lowest phenotypic values are most informative for mapping the QTLs (Taylor *et al.*, 1994) and with only genotyping the extremes, the costs of detection would be reduced. This is an advantage of genotyping the extreme phenotypes when the traits are easily and routinely collected. Casu *et al.* (2003) used a daughter design (DD) combined with the extreme phenotypes (EP) and showed that the number of genotyped individuals was lower in DD when combined with EP (50 and 25% genotyping of the population) with considerable power for intermediate QTL effects than complete genotyping (with power of 68, 83 and 87% for 25, 50 and 100% genotyping, respectively). The choice of what fraction of the extremes to be genotyped depends on the relative costs of phenotyping and genotyping (Darvasi, 1997). Extreme phenotype method is often considered to have a major limitation if experiments are aimed at studying many traits. It is also important to notice that EP may cause biased estimates of QTL effects for traits correlated with the selected trait. This bias could be reduced and the power of analysis increased if single trait analysis is replaced by multi-trait analysis (Henshall and Goddard, 1999). Another disadvantage of EP is that linear model estimates of the QTL effects are conditional on the individuals with genotype information from the analysis and this might cause a bias. One alternative to remove this problem is to use a mixture model approach which was presented by Jansen *et al.* (1998) for the mapping of

QTLs in outbred populations. Johnson *et al.* (1999) have simulated this approach in a half-sib family to demonstrate that estimates of the allelic effects of a QTL influencing the trait are unbiased not only for the main trait used to select individuals but also for a correlated trait when both traits were jointly analyzed in a bivariate model. In this case, Markov chain Monte Carlo methods are appropriate for sampling missing data (for individuals without genotype information) and then all phenotype records in the population could be used for the mapping of QTL. In addition, Ronin *et al.* (1998) showed that it is possible to estimate unbiased parameters if all phenotype records for the trait under selection are included in the model analysis. It has been demonstrated that in EP, the power of QTL detection was at least as great as the random genotyping (Ronin *et al.*, 1998; Bovenhuis and Spelman, 2000). Stella and Boettcher (2004) used ten different strategies in genotyping and concluded that all strategies approximately gave the same precision of estimates of the QTL position but they were better than random sampling from the population.

2.3.2. Corrected criteria (extremes of EBV/DYD): Instead of using the individuals located in the tails of the phenotypic distribution, Ajmone-Marsan *et al.* (2000) used the extremes of the estimated breeding values (EBV) distribution in Italian Friesian cattle population from two half-sib families. From a practical point of view, it would be better to use daughter yield deviations (DYD) of the sires in dairy cattle breeding plans in instead of raw observations (Hoeschele and VanRaden, 1993; Bennewitz *et al.*, 2004), because DYD have been adjusted for systematic environmental effects and also corrected for the additive genetic values of the daughter's dams (VanRaden and Wiggans, 1991). Israel and Weller (1998) used an animal model with a QTL effect and a fraction of the population genotyped and showed that using EBV as dependent variable gives biased estimates of QTL effect compared to DYD. However, they did not report the difference between using DYD and EBV in the power of detection. Thomsen *et al.* (2001) showed low power of detection in using EBV compared to DYD or deregressed breeding values.

There is not any investigation in order to compare phenotypic records and EBV under selective genotyping methods and mostly phenotypes are used as dependent variables in selective genotyping. Thus, this study used the phenotypes as a dependent variable under a daughter design to investigate different selective genotyping strategies.

2.3.3. Genotype probability index (GPI): In contrast to previously described criteria, GPI is conditional on genotypic information. The GPI is an estimate of the information content from genotype probabilities of segregation analysis and does not use phenotypes. It is zero when there is not any genetic information as well as Hardy-Weinberg equilibrium, and 100 when the genotype is clear and is known. Kinghorn (1999) developed an iterative genotyping strategy based on genotype probabilities from segregation analysis to use in the index and determine the most informative individuals for each genotyping cycle. The objective of this method was to maximize the utility of information across the entire population (utility is defined as the mean of GPI in live individuals after each individual genotyping, iteratively). On the other hand, as the cost of genotyping for single loci is generally high, first the probable genotypes of individuals are inferred by segregation analysis, and then through an iterative process, the most informative individuals are identified for genotyping. In the GPI method, individuals with low utility have to be genotyped because the information from the segregation analysis could not be helpful to make a decision on their genotype.

Macrossan and Kinghorn (2003) proposed different criteria for individual ranking based on GPI and concluded that at 50% genotyping, the indices based on GPI approach (Kinghorn, 1999) showed significant superiority, especially at higher frequencies of the favorable QTL allele, over a linear regression approach. For logistic reasons, animals may have to be genotyped in groups rather than individually (Macrossan, 2004), and the group of animals chosen is expected to differ from the same sized group chosen when animals are genotyped separately, since relationships between animals resulting from the pedigree means that genotyping one animal, may change the value of genotyping another individual. Kinghorn (1999) showed that when genotyping in groups of the top one, 10, 20 and 50 individuals chosen on a ranking index based on the segregation analysis, the cumulated utility (average GPI of live individuals) decreases as group size increases. Several scenarios were investigated by Macrossan (2004) for genotyping in groups. The main disadvantage in GPI methods is that it only uses pedigree and the results from segregation analysis to find the candidates for genotyping. However, the animal with the largest uncertainty on their genotype content information is not necessarily the most informative candidate for QTL detection studies. Moreover, GPI can not use any phenotypic information to consider linkage between the genotypes and the phenotypes. Recently GPI was developed for multiple-alleles (Percy and Kinghorn, 2005), but GPI computations become more complicated when a large haplotype is considered over several generations.

3. Selective phenotyping approaches

3.1. QTL detection

Some traits are costly and/or difficult to collect, e.g. post-slaughter trait, disease resistance related traits and physiological profiles. For such traits the costs of phenotyping might exceed or equal the costs of genotyping. Instead of considering the phenotype information to identify informative individuals in genotyping, selective phenotyping (SP) in the sampling strategies are necessary for the mentioned traits. Medugorac and Soller (2001) and Casu *et al.* (2003) used SP approach in a main trait based on the correlation with an easily-recordable trait. Two plant experiments investigated an index for selective phenotyping in the crosses based on information from the main trait (Jin *et al.*, 2004; Jannink, 2005). They used several criteria based on linkage information to maximize genetic dissimilarity in selection of candidates for phenotyping which were recombinants. On the other hand, after detecting the QTL in the main trait based on an initial genome scan experiment, these linkage criteria can be used for further mapping accuracy in QTL fine mapping. Selecting the individuals for phenotyping can be based on the entire genome where there is a little information about the genetic contents in the population. In the situations that previous studies have proposed the certain genomic regions may be important, selective phenotyping can be considered through the identified regions. This approach uses information from only a few markers (haplotypes) in the identified genomic regions (Jin *et al.*, 2004). Indeed the underlying basis of the proposed method is based on the correlation between QTL and the markers across the identified chromosome segment, and then phenotyping only the recombinants to increase mapping accuracy for a detected QTL.

Linkage disequilibrium (LD) between the markers and a linked QTL is central to QTL detection. Using historical recombination based on LD allows to fine map in outbred populations *e.g.* dairy cattle, instead of experimental populations for QTL studies.

Combining the methods of LD and linkage analysis is becoming popular for QTL fine mapping because creation of populations such as advanced intercrossed lines is nearly impossible in livestock due to time demands, financial constraints and inbreeding depression. Therefore, an alternative to the current linkage criteria in selective phenotyping is to use LD information in order to identify informative individuals for phenotyping. In fact, LD information may be particularly useful to fine map or confirm association of candidate regions that have already been shown to include QTLs.

3.2. Observed criteria

3.2.1. Correlated traits: There are some traits which are expensive to collect or difficult to measure and, therefore, it is not possible to have complete phenotype information, such as: sex limited traits, carcass quality traits, age limited traits (expressed late in life *e.g.* longevity, mastitis). Therefore for these traits, identification of informative individuals would be helpful for phenotyping. When a correlated trait is available for which the measurement is less time-consuming, less costly and more convenient, selecting individuals for phenotyping on trait of interest from the extremes of correlated traits is a useful strategy (Medugorac and Soller, 2001). It has been shown that use of an auxiliary trait could improve accuracy of prediction (Narain, 2003). If a genomic region affecting the expression of the trait was identified, phenotyping the 50% of the population showing the extremes of the correlated traits would be a proper strategy to increase the power of QTL detection on the main trait (Casu and Carta, 2001). In this case, analysis of the data for QTL effect estimation should take into account methods able to correct bias induced by selection on the completely genotyped trait and in order to reduce the risk of false detection. Bovenhuis and Spelman (2000) presented several formulas to make it possible to obtain unbiased estimates of QTL effects for correlated traits with relatively simple statistical techniques (an adjusted method in regression coefficients). Their formulas were valid for a range of QTL effects ($0.2\sigma_P \sim 0.4\sigma_P$).

3.2.2. Genetic dissimilarity: Instead of considering the correlated trait for phenotyping, Jin *et al.* (2004) developed an index for selective phenotyping which is based on the minimum uniformity of the genetic structure in the animals selected for phenotyping.

They used several criteria based on linkage information to maximize genetic dissimilarity in the phenotyped individuals given paternal genotype information. After detecting the QTL based on an initial genome scan, these linkage criteria can be used for further mapping accuracy in QTL fine mapping (Darvasi, 1998; Jannink, 2005).

4. Population designs in QTL studies

In outbred populations QTL is detected using half-sib designs which are based on the paternal half-sib family information. The two common half-sib experimental designs used in dairy cattle are daughter design (DD) and grand-daughter design (GDD) (Weller *et al.*, 1990). These designs are applied in the segregating populations in which heterozygous markers and QTL genotypes exist. Linkage disequilibrium information can be combined with linkage analysis in DD in order to increase power of detection and estimate unbiased QTL parameters.

4.1. Daughter design: In this design sire and his daughters are genotyped (Figure 1), which increase number of individuals for genotyping (sons and their daughters). The aim is to find significant different means of the daughters (in each sire group) that inherited **M** or **m** allele of the sire that has genotype **Mm**, and therefore, the marker alleles linked to QTL. On the other hand, since sires and their daughters in this design should be genotyped, genotyping costs will inevitably be higher than in GDD. Several strategies will be considered to identify the most informative individuals for genotyping in this study. These approaches could be compared to other designs for genotyping like grand-daughter design.

4.2. Grand-daughter design: This design uses linkage information within paternal half-sib families (Figure 2). DNA markers are genotyped only from a proven sire and his sons and then daughters of the sons are phenotyped. Therefore, it is logistically easier to collect genetic sample from AI sons (in GDD) compared to cows (in DD), which are scattered over a much larger number of herds.

The benefits of GDD compared to DD are: less required genotyping individuals, proven sires and his sons are in AI centers (test stations) therefore DNA samples can be collected more easily, and also higher statistical power per genotyped individual (since

GDD considers three generations and uses more phenotypic information). However, GDD needs to have a specific designed population *i.e.* proven sires and each of them with many progeny-tested sons. The daughter design is simpler than GDD and can be used where the herds have several sires and each sire with hundreds of daughters. Additionally, it should be possible to detect QTL with smaller magnitude in DD (with many genotyped daughters) compared to GDD (Ron *et al.*, 2004). However, if few progeny-tested sires are available it is necessary to consider DD to detect QTLs with small effects and reduce the confidence interval for QTL position to the length for which positional cloning becomes possible (Ron *et al.*, 2004). In order to use DD in QTL studies, *e.g.* genome scan experiments, it is necessary to investigate cost-effective methods for genotyping.

Both GDD and DD are better than crossing designs because of the saving in time and also the costs of rearing which are important in large animals such as cattle with long generation interval and uniparous reproduction. The basic idea behind these breeding designs (DD or GDD) is to exploit linkage disequilibrium between markers and QTL within a half-sib family. Disadvantages of this segregating population is described as the loss in statistical power compared to crossing inbred lines in QTL detection studies, and also the increase in uncertainties of the underlying genetic factors influencing the phenotypes (Da, 2003).

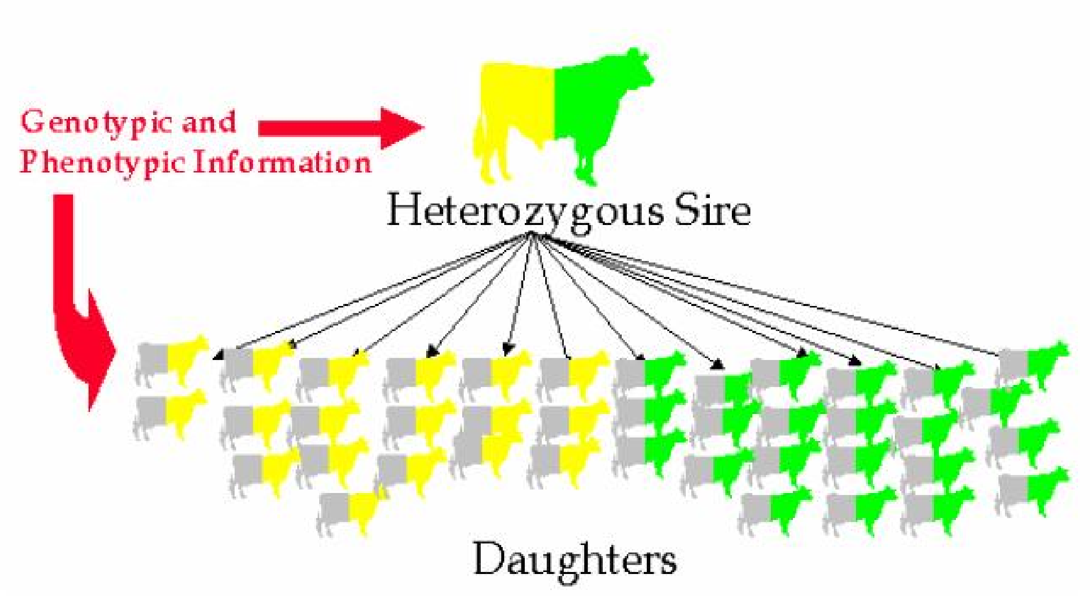

Figure 1: An example of daughter design. The heterozygous sire transfers either **M** allele (showed with yellow colour) or **m** allele (showed with green colour). In this design, phenotypic and genotypic information is collected and used to analyses from the sire and his daughters (Muncie, 2005).

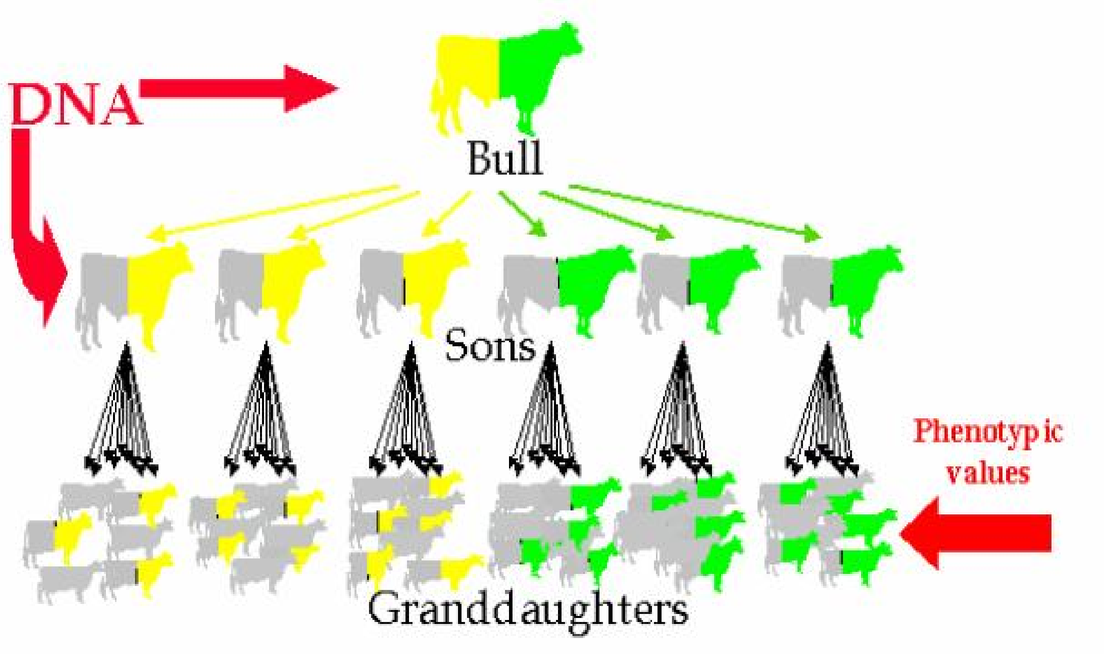

Figure 2: An example of grand-daughter design. The heterozygous proven bull transfers either **M** allele (showed with yellow colour) or **m** allele (showed with green colour) to his sons. The grand-daughters in generation three are evaluated for the phenotypes (Muncie, 2005).

Objectives

1. Outline of the objectives

The main objectives of the thesis are threefold and each of them with several specific aims.

1.1. Selective genotyping in QTL detection

The objective of this part of the thesis is to develop and compare the strategies to infer informative individuals for genotyping in QTL detection. In fact it would be achieved through reducing the number of individuals for genotyping instead of whole population genotyping. This is more efficient when individuals with the largest marginal effect in detection are genotyped. Methods, which have been derived to find informative individuals, will be used and compared for several scenarios to find the appropriate strategy which could have lower genotyping costs, as below:

> 1.1.1. Finding new approaches to combine genotypic and phenotypic information for ranking animals for genotype identification in several genotyping levels.
>
> 1.1.2. Comparison of these approaches with existing criteria for ranking animals *e.g.* extreme methods with respect to power of detecting a QTL in simulated data.

1.2. Selective phenotyping in QTL detection

When phenotyping the individuals are expensive or phenotype of a trait is difficult to collect, one needs to use a proper approach to identify QTL of a quantitative trait. One solution to cover this problem is by reducing the number of phenotyped individuals. Therefore, the general objective of this part in this thesis is to develop criteria for ranking animals for phenotyping based on genotype information. Developed criteria aim to sample more informative individuals for phenotyping based on linkage information within the half-sib families or linkage disequilibrium information across the families in a population. One major problem in selective phenotyping is the biased estimates of QTL due to the incomplete phenotype information in the trait of interest. Briefly the following objectives will be considered:

1.2.1. Construct criteria for ranking animals based on linkage and linkage disequilibrium information for phenotyping in three genotyping levels.
1.2.2. Comparing the developed criteria in a set of simulated data based on half-sib families *e.g.* dairy cattle.
1.3.3. Different proportions of the phenotyped animals will be studied for QTL detection in comparison to a random phenotyping approach.

1.3. Selective genotyping with MAS

MAS applied in a simulated population based on different selective genotyping strategies in several generations. The true breeding values were computed assuming a mixture model and additive allelic QTL effects. Use of linkage disequilibrium information to identify animals for genotyping could be considered as more powerful tools in animal breeding schemes compared to using linkage information in the top-down approach. Two proportions of genotyped animals were compared with respect to genetic gain and accuracy of genetic evaluation in a dairy cattle selection scheme. The following objectives are studied under different selective genotyping strategies:

1.3.1. Define dynamic criteria for genotyping to decrease number of genotyped animals in (young bulls and bull dams), but maintain at large response to selection under constrained genotyping costs.

1.3.2. Compare selective genotyping strategies with conventional BLUP using stochastic simulation with respect to genetic gain and selection accuracy.

1.3.3. Evaluation of the developed methods with random genotyping in a practical dairy cattle breeding scheme.

2. Evaluation of the objectives

Use of gene (marker) information of quantitative traits to increase genetic gain in breeding schemes has been argued as an alternative in dairy cattle selection. The cost of genotyping is generally high but phenotype of some economic traits are easily recorded in dairy cattle e.g. milk records and milk contents. So a combination of the observable records and using available genotypes has a potential to be an efficient and low cost scheme. The purpose is to find informative individuals for detection of QTL and the selection for quantitative traits. In contrast with selective genotyping contexts in QTL detection, however, selective phenotyping, which is used to identify more informative individuals for phenotyping, is particularly useful if the trait is difficult or expensive to measure.

2.1. QTL detection

The optimization problem in QTL detection depends on two important items: position and effect. In general, QTL studies employing traditional experimental designs and large sample sizes will readily identify QTL that are of significant effect. The limitations of QTL analyses to identify the position and effect of underlying QTL typically reside in the limitation test populations of finite size. For the purpose of this discussion, this study will seek to develop a method to identify a fraction of individuals for genotyping/phenotyping based on genotyped/phenotyped animals and also all available records. The strategies will be compared with the current approaches and also random selective genotyping (phenotyping).

2.2. Marker-assisted selection

Use of information from the detected QTL in the selection schemes requires developing selection criteria to join this molecular information with phenotypic information. The optimum selection should identify outstanding individuals as the parents of the next generation. Two important concepts to achieve an optimum selection are inbreeding and genetic trend. In the ideal case, selection increases the average of breeding values while keeping inbreeding rate[1] at an acceptable level. The second aim is not the objective in this

[1] Inbreeding rate = ["inbreeding in present generation" – "inbreeding in previous generation"] / [1 – "inbreeding in present generation"] (Falconer and Mackay 1996).

study. The breeding values are estimated for major genes (QTLs) and polygenic effects and deviation of the predicted from the true breeding values should be decreased. For traits regulated by a QTL with large effects ($>0.2\sigma_P$), and for which phenotypic selection is expensive, MAS can be efficiently used. However, use of MAS requires linkage disequilibrium which could be used in dairy cattle as within family MAS. One problem of MAS within family is the large number of offspring required from each half-sib family in order to estimate unbiased effects. The described strategies using the detected QTL in this study are compared with current selection schemes in dairy cattle breeding.

References

Ajmone-Marsan, P., E. Milanesi, R. Negrini, C. Gorni, F. Miglior, A. Samore, I. Cappuccio, M.C. Savarese and A. Valentini. 2000. Use of molecular markers, selective genotyping and DNA pooling for the identification of QTL for milk protein percentage in the Italian Friesian cattle breed. Zoot. Nutr. Anim., XXVI, 3:161-167.

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

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

Casu, S. and A. Carta. 2001. Power of QTL mapping experiments based on selection of individuals on phenotypic values of correlated traits. In: Proc. ASPA XIV Congress, Firenze (Italy), pp. 43-45.

Casu, S., A. Carta and J.M. Elsen. 2003. Strategies to optimize QTL detection designs in dairy sheep populations: the example of the Sarda breed. Options Mediterraneennes, Vol A55: 19-24.

Da, Y. 2003. Statistical methods and experimental designs for mapping genes of complex traits in domestic animals. Acta Genetica Sinica, 30:1183-1192.

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. 1997. The effect of selective genotyping on QTL mapping accuracy. Mammalian Genome, 8:67- 68.

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

Falconer, D.S. and T.F.C. Mackay. 1996. Introduction to quantitative genetics. 4th edn. Longman, New York, 464 pp.

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

Goddard, M.E. and B.J. Hayes. 2002. Optimisation of response using molecular data. CD-ROM Communication No. 22-01 in Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France.

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

Hoeschele, I. and P.M. VanRaden. 1993. Bayesian analysis of linkage between genetic markers and quantitative trait loci. II. Combining prior knowledge with experimental evidence. Theor. Appl. Genet., 85:946–952.

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

Jannink, J.L. 2005. Selective phenotyping to accurately map quantitative trait loci. Crop Sci., 45: 901–908.

Jansen, R.C., D.L. Johnson and J.A.M. van Arendonk. 1998. A mixture model approach to the mapping of quantitative trait loci in complex populations with an application to multiple cattle families. Genetics, 148: 391–399.

Jin, C., L. Lan, A.D. Attie, G.A. Churchill, D. Bulutuglo and B.S. Yandell. 2004. Selective phenotyping for increased efficiency in genetic mapping studies. Genetics, 168: 2285–2293.

Johnson, D.L., R.C. Jansen and J.A.M. van Arendonk. 1999. Mapping quantitative trait loci in a selectively genotyped outbred population using a mixture model approach. Genet. Res. Camb., 73:75–83.

Kinghorn, B.P. 1997. An index of information content for genotype probabilities derived from segregation analysis. Genetics, 145: 478-483.

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

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

Lander, E. and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics, 121: 185-199.

Liu, Z. 2001. Gene mapping, marker-assisted selection, gene cloning, genetic engineering and integrated genetic improvement programs at Auburn University, p. 109-118. In: M.V. Gupta and B.O. Acosta (eds.) Fish genetics research in member countries and institutions of the International Network on Genetics in Aquaculture. ICLARM Conf. Proc. 64, pp179.

Macrossan, P.E. 2004. Strategies to Minimise DNA Testing Costs for Research and Development Programs Involving Pedigreed Populations. PhD Thesis, University of New England, Australia.

Macrossan, P.E. and B.P. Kinghorn. 2003. A genetic algorithm to investigate genotyping in groups. Proc. Assoc. Advmt. Anim. Breed. Genet., 15: 43–46.

Medugorac, I. and M. Soller, 2001. Selective genotyping with a main trait and a correlated trait. J. Anim. Breed. Genet., 118: 285–295.

Meuwissen, T.H.E. and M.E. Goddard. 1996. The Use of Marker Haplotypes in Animal Breeding Schemes. Genet. Sel. Evol., 28:161-176.

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

Muncie, S.A. 2005. Fine Mapping Quantitative Trait Loci Affecting Health and Reproduction in US Holstein Cattle on Chromosome 18. MSc Thesis, North Carolina State University.

Misztal, I. 2006. Challenges of application of marker assisted selection - a review. Animal Science Papers and Reports, 24 (1): 5-10.

Nagamine, Y., S.A. Knott, P.M. Visscher and C.S. Haley. 2002. Simple deterministic identity-by-descents coefficients and estimation of QTL allelic effects in full and half sibs. Genet. Res., 80:237–243.

Narain, P. 2003. Accuracy of marker-assisted selection with auxiliary traits. J. Biosci., 28(5) :569-579.

Percy, A. and B.P. Kinghorn. 2005. A genotype probability index for multiple alleles and haplotypes. J Anim Breed Genet., 122:387-392.

Pong-Wong, R., A.W. George, J.A. Wooliams and C.S. Haley. 2001 A simple and rapid method for calculating identity-by-descent matrices using multiple markers. Genet. Sel. Evol., 33: 435–471.

Ron, M., E. Feldmesser, M. Golik *et al*. 2004. A complete genome scan of the Israeli Holstein population for quantitative trait loci by a daughter design. J Dairy Sci., 87: 476–90.

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.

Soller, M. and J.M. Reecy. 2004. QTL mapping and cloning in beef cattle. AgBioTechNet Proceedings 004 paper, 3:1-8.

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.

Stella, A. and P.J. Boettcher. 2004. Optimal designs for linkage disequilibrium mapping and candidate gene association tests in livestock populations. Genetics, 166: 341-350.

Taylor, B.A., A. Navin and S.J. Phillips. 1994. PCR-amplification of simple sequence repeat variants from pooled DNA samples for rapidly mapping new mutations of the mouse. Genomics, 21:626–632.

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

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

Villanueva, B., R. Pong-Wong, J. Fernández and M.A. Toro. 2005. Benefits from marker-assisted selection under an additive polygenic genetic model. J. Anim. Sci., 83:1747-1752.

Weller, J.I., Y. Kashi and M. Soller. 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J. Dairy Sci., 74:2525-2537.

Chapter 2

Paper I

Combined use of phenotypic and genotypic information in sampling animals for genotyping in detection of quantitative trait loci

Saeid Ansari-Mahyari, Peer Berg

Submitted to

Journal of Animal Breeding and Genetics

Combined use of phenotypic and genotypic information in sampling animals for genotyping in detection of quantitative trait loci

Saeid Ansari-Mahyari[1,2,3], Peer Berg[1]

[1]Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, University of Aarhus, Denmark

[2]Agriculture Research and Education Organization, Isfahan Agricultural & Natural Resources Research Center, Isfahan, Iran

[3]Department of Large Animal Sciences, Faculty of Life Sciences, Copenhagen University, Denmark

Summary

Conventional selective genotyping which is using the extreme phenotypes (EP) was compared with alternative criteria to find the most informative animals for genotyping with respects to mapping quantitative trait loci (QTL). Alternative sampling strategies were based on minimizing the sampling error of the estimated QTL effect (MinERR) and maximizing likelihood ratio test (MaxLRT) to use both phenotypic and genotypic information. Random selection of animals either across or within families was also tested. One hundred data sets were simulated each with 30 half-sib families and 120 daughters per family. The strategies were compared in these datasets with respect to estimated effect and position of a QTL within a previously defined genomic region at 10, 20 and 30% genotyping levels. Combined linkage disequilibrium linkage analysis (LALD) was applied in a variance component approach. Power to detect QTL was significantly higher for both MinERR and MaxLRT compared to EP and random genotyping methods (either across or within family), at all genotyping levels. Power to detect significant QTL (α=0.01) with 20% genotyping for MinERR and MaxLRT were 80 and 75% of that obtained with complete genotyping compared to 70 and 38% genotyping for EP within and across families, respectively. With 30% genotyping the powers were 78, 83, 78 and 58%, respectively. The estimated variance components showed that variance components were unbiased in EP strategies (within and across family) in a LD-population, only through at least 30%

genotyping. However, in order to decrease number of individuals for genotyping, either MinERR or MaxLRT could be considered. With 20% genotyping in MinERR, the estimated QTL variance components were not significant compared to complete genotype information but all studied strategies in this study at 20% genotyping significantly overestimated the QTL effect, compared to the simulated QTL effects. Results showed that combining the phenotypic and genotypic information in selective genotyping (*e.g.* MinERR and MaxLRT) is better than only using the extreme phenotypes and the combined methods can be considered as alternative approaches in order to decrease genotyping costs, with unbiased QTL effects, decreased sampling variance of the QTL variance component and also increased the power of QTL detection.

Introduction

Use of gene (marker) information of quantitative traits to increase genetic gain in breeding schemes has been argued as an alternative in dairy cattle selection. The cost of genotyping is generally high but phenotypes of some economically important traits are routinely recorded in dairy cattle e.g. milk yield and milk contents. Selecting individuals for genotyping is an attempt to remedy this problem. Selective genotyping was firstly introduced by Lander and Botstein (1989) to increase power of detecting QTL with a small effect. This approach selects portion of individuals for genotyping, on the basis of the individuals' phenotypes (generally those with extremely high or low phenotypic values). The advantage of this is that fewer individuals need to be genotyped for a given probability of identifying putative QTLs. Selective genotyping yields significant savings since genotyping is expensive. Darvasi & Soller (1992) practiced this method for a single marker linked to a QTL to genotype only the potentially most informative observations. The extreme phenotype method (EP) considers individuals with the most extreme phenotypes, without using available pedigree or marker genotype information. The general principle exploited is that most linkage information can be inferred by individuals with extreme phenotype values and for a given number of individuals genotyped, this increases power to detect a QTL compared to random genotyping. However, due to selecting a subset of animals for genotyping, EP could cause a bias in QTL parameter estimations (Lander & Botstein 1989; Darvasi & Soller 1992).

Instead of phenotype information, Kinghorn (1997) presented an index to indicate the information content of genotype probabilities (GPI) derived from a segregation analysis. According to this index, individuals with the least accurately known genotypes can be identified sequentially. The genotyping models could be used for identifying individual-by-individual (Kinghorn 1997 and 1999) or group-by-group (Macrossan 2004). However, this method becomes more complicated when a large haplotype is considered.

Simulation studies proposed that selective methods (e.g. extreme phenotypes) are using most informative individuals for detecting the QTLs (Van Gestel *et al.* 2000; Martinez *et al.* 1998; Muranty & Goffinet 1997) compared to random genotyping and the costs would be reduced. This is an advantage of genotyping the extreme phenotypes when the traits are easily and routinely collected in human and animal genetics studies in large populations (Casas *et al.* 2000). Casu *et al.* (2003) used a daughter design (DD) combined with EP and showed that number of genotyped individuals are lower in DD when combined with EP (50 and 25% of the population) with a reasonable power for intermediate QTL effects than using combined EP with grand daughter design. The choice of what fraction to genotype depends on the relative cost of phenotyping and genotyping (Darvasi 1997). It has been demonstrated that in EP, the power of QTL detection was at least as great as random genotyping (Ronin *et al.* 1998; Bovenhuis & Spelman, 2000). Stella & Boettcher (2004) used ten different strategies in genotyping and concluded that all strategies approximately had the same precise estimates of the QTL position but they were better than random sampling from the population.

One disadvantage of EP is that linear model estimates of the QTL effects are conditional on the individuals with genotype information from the analysis and this might cause a bias. One solution to solve this problem is to use a mixture model approach which was presented by Jansen *et al.* (1998) for the mapping of QTLs in an outbred population. Johnson *et al.* (1999) have simulated this approach in a half-sib family to demonstrate that estimates of the allelic effects of a QTL are unbiased not only for the main trait used to select individuals but also for a correlated trait when both traits were jointly analyzed in a bivariate model. In this case, Markov chain Monte Carlo methods are appropriate for sampling missing data (for individuals without genotype information) and then all phenotype records in the population could be used for the mapping of QTL. In addition, Ronin *et al.* (1998) showed that it is possible to estimate unbiased parameters if all phenotype records for the trait under selection are included in the analysis.

All current traditional selective methods for QTL mapping experiments have used either phenotypic or genotypic information, even if a proportion of animals already have been genotyped. Therefore, QTL mapping studies is expected to be more powerful by utilizing both genotype and phenotype information to finding the most informative animals for genotyping.

The objective of the current study was to compare strategies for selective genotyping with respect to power of detecting a QTL in simulated data. New criteria based on both phenotypic and genotypic information is contrasted with traditional selective genotyping approaches.

Material and Methods

Data Simulation:

Selective genotyping approaches were compared on simulated data and precision of the QTL mapping and power of detection related to the selective genotyping criteria were studied in a daughter design *e.g.* dairy cattle population.

Population structure and genetic model: The simulation of the parents (base population) was based on the method of Meuwissen and Goddard (2000), with discrete generation assumption to generate linkage disequilibrium. This method assumed that variation in a QTL is due to a mutation that occurred 100 generations ago due to random drift, given a specific effective population size. Therefore, according to this method and based on the genetic model described below, 100 generations of completely random mating was simulated with effective population size of 200. In generation 1, the genotypes were simulated for 6 marker loci equally spaced in a 50 cM chromosome segment and also an associated QTL placed in the midpoint between markers 3 and 4. A simple bi-allelic QTL and markers with five alleles were assumed. In the first generation, unique alleles were sampled and the frequencies of alleles in each marker were equal to $\frac{1}{5}$. Initially in generation 1, the frequencies in the QTL alleles were the same and equal to 0.50. Marker and QTL genotypes (each animal with two unique alleles) were assigned according to Mendelian inheritance and allowed for recombination within the region. Comparisons of the strategies were based on the final generation (generation 101). In the final generation, genotypes of the progenies (generation 102) were sampled from the parental haplotypes allowing for recombination. A Poisson distribution, given the distance between markers or QTL, was used to determine the probability of an uneven number of crossovers. In the last generation (generation 100) a mutant QTL allele was sampled at random, with the requirement that the frequency ($\boldsymbol{P_Q}$) was between 0.45 and 0.55, otherwise if $\boldsymbol{P_Q}$ was out of this range then a new LD population was simulated.

Phenotype: One quantitative trait was simulated, as the sum of QTL, polygenic and residual effects, with moderate inheritance e.g. milk yield with assumed heritability of 0.25 and total phenotypic variance scaled to one. The phenotype records were assigned to all the daughters in the analysis and not sires. The variance due to the QTL was calculated as $Vqtl=2\times\boldsymbol{P_Q}\times(1-\boldsymbol{P_Q})\times\alpha^2$, where $\boldsymbol{P_Q}$ is the frequency of the favorable QTL allele in the (final) offspring generation and α is the gene substitution effect. Proportion of QTL variance relative to total phenotypic variance was between 0.0620 and 0.0627 and therefore, allele substitution effect was 0.354. Polygenic effects were drawn from a Normal Distribution (ND), ND($\frac{1}{2}a_s+\frac{1}{2}a_d$, $\frac{1}{2}\sigma^2_a$) where a_s (a_d) is the polygenic breeding value of the sire (dam) and σ^2_a is the polygenic variance (0.1874). For base animals polygenic breeding values are sampled from ND(0, σ^2_a). Residual effects are sampled from ND(0,σ^2_e), where σ^2_e is the residual variance (0.75).

Genotyping Selection Strategies

Four strategies for selective genotyping in a population which contains paternal half-sib family groups were defined based on a daughter design. Each strategy was used to select daughters for genotyping until 10, 20 and 30% of the 3600 daughters in the population were genotyped. All genotyping strategies were used for the daughters and it was assumed the sire’s genotypes was available.

1-Random Genotyping: Random selection of daughters for genotyping was used. The random strategy was practiced either randomly across family (RAN_A) or randomly within family (RAN_W).

2-Extreme Phenotype Genotyping: In this strategy, the daughters with extreme phenotypes (EP) were identified for genotyping either across sire families (EP_A) or within sire families (EP_W). EP_A strategy is known as usual conventional selective genotyping. The daughters with phenotypes in the highest and lowest percentages were selected for genotyping with equal proportions in each extreme.

3-Minimum Error of QTL variance component (MinERR): This approach uses all available phenotype and genotype information in selecting daughters for genotyping. Based on a mixed inheritance model with QTL as random effect (as it will be described later in **Analysis**), and via restricted maximum likelihood (REML), the asymptotic standard error of the estimated QTL variance component effect was computed, conditional on the QTL position. This standard error was derived from the second derivative of the likelihood function equation (3) and computed over all sire groups.

For the first level of genotyping (10%), genotype information from 2% of the extremes (4% in total) in each sire family (within family) was genotypes as a start point from which to begin genotyping cycles. Then the next daughters for genotyping at each iteration cycle were chosen on the largest reduction in standard error of the QTL variance component until 10%, 20% or 30% genotyping were achieved.

In each iteration, and given all phenotypic and a fraction of genotypic information, first maximum likelihood tests were computed for the positions in the middle of the marker intervals (5, 15, 25, 35 and 45 cM), in order to identify the most likely position of the QTL. Then the standard error of the QTL variance component was computed through solving the mixed model equations conditional on the identified QTL position for all candidates. The candidates were the potential daughters for genotyping in each step of genotyping. Because identical by descent (IBD) computation is time consuming in solving the equations (2½ hrs for 2% genotyping and about 30hrs for each replicate), only the extremes of each sire family was considered as candidates (10 offspring per sire family). Therefore, depending on available daughters per sire, up to 300 candidates were chosen each iteration. To assign putative genotypes for the candidates, recombination was assumed in the sires with randomly assigned as paternal haplotype part, and for maternal part, 100 genotype samples were assigned based on marker allele frequencies in the dam population. Parents were assumed to be unrelated, and no previous generations were taken into account.

The IBD matrix was computed as a function of available marker data and the position of a putative QTL on the chromosome. This process was repeated for all candidates, one at a time. In each genotyping step, 72 offspring from 300 candidates (24% of the candidates) with the largest reduction in standard error of QTL variance component were genotyped. Therefore in each step, 2% of the daughters were genotyped. This approach was repeated to find further daughters (next 2 percent) for genotyping until 10, 20 and 30% genotyping of the population.

4-Maximizing Likelihood Ratio Test (MaxLRT): Another criterion to use all available phenotypic and genotypic information in selecting the daughters for genotyping was based on maximum increase in likelihood ratio test (MaxLRT). The daughters were selected for genotyping based on the largest LRT in potentially selected daughters for genotyping. Thus, given the most likely QTL position, all phenotype information, available genotype information and genotype of these daughters based on sire's genotype (for paternal haplotype part) and marker allele frequencies in dam population (for maternal haplotype part), this criterion was computed through a mixed model approach. In each iteration of genotyping cycles, 2% of the candidates with the highest LRT were selected for genotyping. Again, this criterion considered QTL positions in the middle of the marker intervals (5, 15, 25, 35 and 45 cM) and maximum chosen as the QTL location. This process of selective genotyping was continued until 10, 20 and 30% of the population was genotyped.

Analysis

In total 100 data sets were simulated and each strategy used the simulated data as described above. After selection of the daughters to genotype, a mixed inheritance model with QTL as random effect (as below), was fitted and the variance component of QTL effect was computed, conditional on the QTL position on the identified segment of the chromosome:

$$\mathbf{y} = \mathrm{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Wq} + \mathbf{e} \qquad \textbf{(1)}$$

where: **y** is an (N × 1) vector of phenotypes of the daughters and N is number of daughter (genotyped and non-genotyped animals) phenotypic records, X is an incidence matrix for **β**, which reduces to a vector on N ones, **β** is a vector of fixed effects, which reduces to the overall mean here, Z is an (N × *p*) incidence matrix relating phenotypes to random sire effects, **u** is a (*p*×1) vector of additive polygenic effect for each sire (polygenic effect), W is an (N × q) incidence matrix relating phenotypes to random QTL effects**, q** is a (q×1) vector of random additive effects due to the QTL (haplotype effect) and **e** (N × 1) are residuals. The phenotypic variance of the observations is:

$$\mathbf{V = Z\,A\,Z'\,\sigma^2_u + W\,G_pW'\,\sigma^2_q + R} \qquad \textbf{(2)}$$

where: **A** is the numerator relationship matrix based on additive genetic relationships, which reduces to an identity matrix here, $\mathbf{G_p}$ is the matrix containing the IBD probabilities of a putative QTL at location (five positions considered between six markers) and **R** = $\mathbf{I\,\sigma^2_e}$ (**I** is an identity matrix).

Gametic relationship matrix at each putative QTL position (G_p) was derived with assumption about historical population structure and clustering approach for LD+LA analysis across sire family groups using the method described by Meuwissen and Goddard (2000, 2001). Here it is assumed that paternal and offspring only (a fraction of offspring in the strategies) marker data is available and that their phases are known. The linkage information is based on the IBD probability between a parental and an offspring haplotype given that both animals are genotyped. Analyses of LD+LA and comparisons of genotype strategies were based on the final generation (generation 101) and their daughters in a daughter design approach.

The residual log likelihood under multivariate normality in the model equation (1) was:

$$\begin{aligned}&\log L(G_p,\sigma_u^2,\sigma_q^2,\sigma_e^2)\propto\\&-\frac{1}{2}(\log|R|+\log|A\sigma_u^2|+\log|G_p\sigma_q^2|+\log|C|+y'R^{-1}y-\hat{\beta}'X'R^{-1}y-\hat{u}'Z'R^{-1}y-\hat{q}'W'R^{-1}y)\end{aligned} \qquad \textbf{(3)}$$

where: C is the coefficient matrix of the mixed model equations from model (1) and the rest are as explained in equations (1) and (2). Given IBD matrix in a putative QTL position, the log *L* was maximized using the Newton-Raphson algorithm to obtain variance components of random effects in model equation (1).

The DMU program package (Madsen and Jensen, 2002) was used for estimation of variance components and obtains MLE. This program uses average information restricted maximum likelihood (AI-REML) algorithm for estimation of (co)variance components in mixed models and the restricted likelihood is maximized with respect to variance components associated to the random effects (Sørensen *et al*., 2003). A likelihood ratio test (**LRT**) was calculated as follows:

$$\textbf{LRT} = -2 * (\textbf{log likelihood } (H_0) - \textbf{log likelihood } (H_1)) \qquad \textbf{(4),}$$

where: log likelihood (H_1) is the likelihood for a model with a QTL-effect and it is calculated for each bracket. Log likelihood (H_0) is based on a model excluding the QTL-effect(s).The LRT-statistic has a Chi-square distribution with one degree of freedom (because of one QTL). The LRT statistic does not take into account that multiple tests are performed along the chromosome, but our simulations with no-QTL effect ($\mathbf{Phen_{no\_QTL}}$) will provide an estimate of chromosome-wise false positive rate. The LRT statistic thresholds for significant ($P < 0.05$), and highly significant ($P < 0.01$) QTL effects were calculated for each data set using an approximate quick method according to the method of Piepho (2001). Power of detection was counted as the number of QTL detected at a given genomewide significance level (α), based on LRT. Sensitivity of position was the number of data sets where the estimated QTL position was in the bracket containing the QTL.

Results

Power in QTL Detection: The number of simulations where the test statistic based on LRT was significant (p-value <0.01 and 0.05) is presented in Table 1 and partly in Figures 1 and 2. Substantial differences can be observed between random genotyping strategy and other strategies for selective genotyping. The power of detection for EP_A is much smaller than what was obtained for EP_W. Number of significant QTL at a significance level of 0.01 with 20% genotyping is found for MinERR to be 80% of that obtained with complete genotyping, and in MaxLRT and EP_W relative power was 75% and 70%, respectively. With 30% genotyping these figures changed to 78, 83 and 78 percent, respectively. Similar differences in power of detection were achieved at a significance level of 0.05 (Figure 1) with MinERR, MaxLRT and EP_W being superior to EP_W and Random genotyping.

Table 1. Power of QTL detection and sensitivity of position† in different selective genotyping strategies with two levels of error type-I

Strategy	Genotype (%)	Detection (α=1%)	Detection (α=5%)	Position (α=1%)	Position (α=5%)
All*	100	88	95	86	95
RAN_W	10	2	7	0	5
RAN_W	20	3	19	1	10
RAN_W	30	14	31	10	24
RAN_A	10	0	7	0	5
RAN_A	20	4	17	3	14
RAN_A	30	11	29	8	22
EP_W	10	38	51	33	47
EP_W	20	62	72	54	67
EP_W	30	69	78	64	75
EP_A	10	8	18	7	18
EP_A	20	33	43	26	38
EP_A	30	51	58	43	53
START**	4	8	22	7	14
MinERR	10	46	63	35	55
MinERR	20	70	80	63	75
MinERR	30	69	82	67	77
MaxLRT	10	45	57	37	48
MaxLRT	20	66	74	62	70
MaxLRT	30	73	83	69	80

† Sensitivity of QTL position shows number of datasets out of 100 where the estimated QTL was in the marker bracket containing the QTL

* All genotypes from the offspring were included

** First 4 percent in the strategies MinERR and MaxLRT is extreme phenotyping as start point

The percentage of simulations with a significant QTL (p-values < 0.05 or 0.01) for no-QTL effect (**Phen$_{\mathbf{no\_QTL}}$**), *i.e.* false positives, is given in Table 2. At a significance level of 0.05, five out of 100 simulations are expected to show a significant QTL, and one out of 100 simulations are expected to show a significant QTL at a significance level of 0.01. In almost all situations explored, expected number of false positive (1% or 5%) closed to the observed number in Table 2. With low level of genotyping in EP methods, (especially in the start point for MaxLRT and MinERR which is based on EP within family), false numbers were exceeded. However, this effect disappeared with increasing the genotyped proportion. Fewer false positives were found for the 20 and 30 percent of genotyping in MinERR, when compared to MaxLRT and the EP strategies, and were close to the all genotyped progeny strategy.

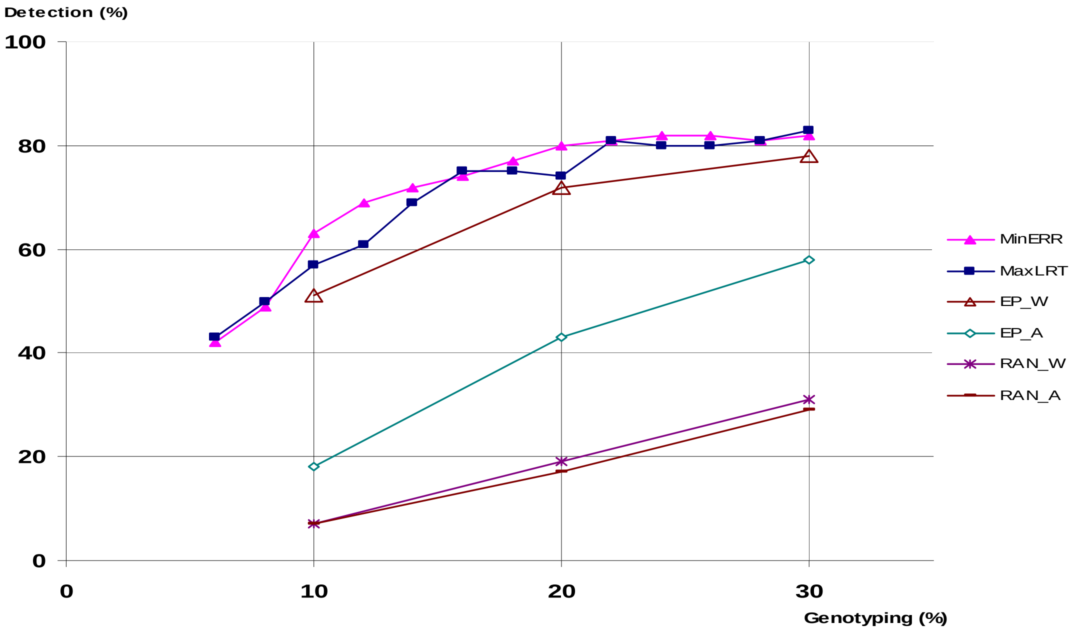

Figure 1. Power of QTL detection under different selective genotyping criteria with a significant QTL (p<0.05)

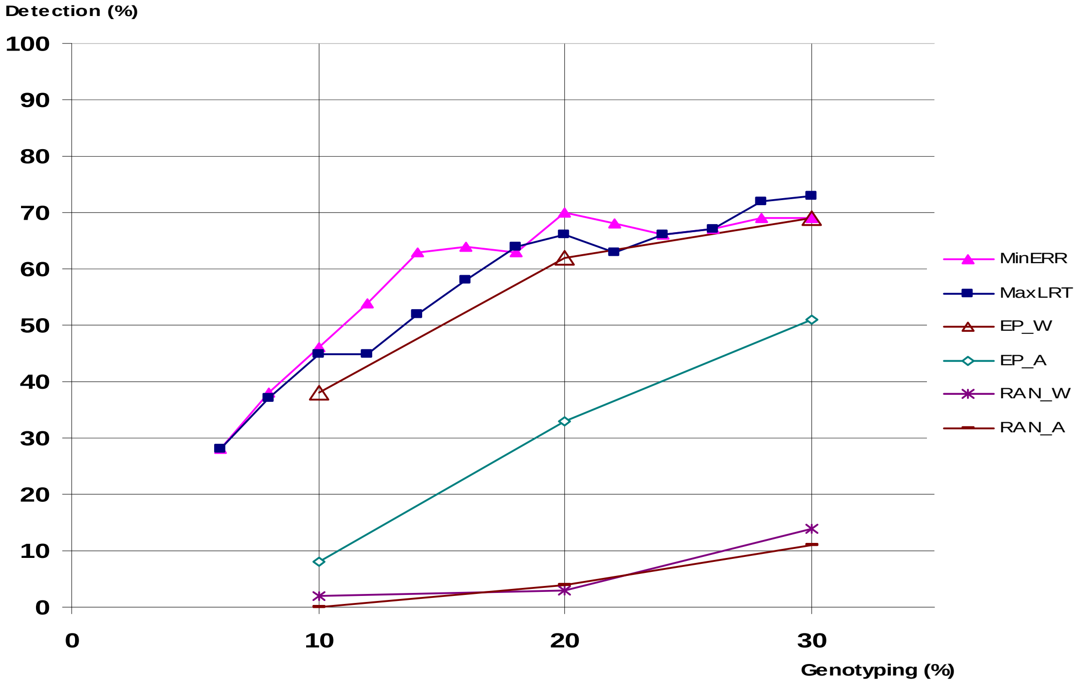

Figure 2. Power of QTL detection under different selective genotyping criteria with a highly significant QTL (p<0.01)

Table 2. Frequency of false positives (out of 100) in QTL detection in different selective genotyping strategies with two levels of error type-I

Strategy	Genotype (%)	Detection (α=1%)	Detection (α=5%)
All*	100	1	1
RAN_W	10	0	0
RAN_W	20	0	0
RAN_W	30	0	1
RAN_A	10	2	4
RAN_A	20	0	1
RAN_A	30	0	1
EP_W	10	3	4
EP_W	20	1	4
EP_W	30	0	4
EP_A	10	1	5
EP_A	20	1	2
EP_A	30	2	4
START**	4	5	7
MinERR	10	4	8
MinERR	20	0	1
MinERR	30	0	1
MaxLRT	10	1	5
MaxLRT	20	0	5
MaxLRT	30	0	5

* All genotypes from the offspring were included

** First 4 percent in the strategies MinERR and MaxLRT is extreme phenotyping as start point

Sensitivity of QTL Position: The estimation of the correct location of the QTL is very important for the design of subsequent fine mapping experiments. In Table 1, the number of datasets (of 100) where the estimated QTL position was in the bracket containing the QTL (25cM) is given. In total, the percentage datasets with a significant QTL after 30% genotyping at correct location (simulated QTL position) ranged from 8 to 69, with α = 0.01, and 22 to 80 with α = 0.05. The lowest sensitivity in detecting the correct location was with random genotyping (both across and within sire family) and EP_A, in contrast with the other genotyping strategies. MaxLRT, MinERR and EP_W was obtained a higher sensitivity of simulated QTL location (both at type I error levels of 1% and 5%) after 20% genotyping. Using a stringent threshold, *e.g.* 0.01, the correct position was found less frequently compared to the 0.05 threshold. Comparing the EP_W and EP_A (Figure 2) showed a large effect of sampling equally from all families when selecting extreme phenotypes, whereas there were only a minor effect for random genotyping at all levels of selective genotyping. The means and standard error of the QTL position estimates (in cM) and statistic test criterion can be found in Table 3. The mean of position estimates over one hundred replicates are approximately similar for the strategies used in this study and close to the true value position (±1 to ±2.5 cM), except RAN_A at 10% genotyping. With regard to the standard error of the position estimates, MaxLRT, MinERR and EP_W are slightly superior at 20 and 30% genotyping (less than 1 cM). Likewise the MinERR, MaxLRT and EP_W strategies have a higher LRT statistic compared to the others.

Table 3 Means (±SE) of QTL position estimates in cM, and the statistic test criterion (likelihood ratio), based on 100 replicates in different selective genotyping strategies

Strategy	Genotype(%)	Position	Likelihood Ratio Test
True*		25.0	
All**	100	25.5(±0.64)	24.7
RAN_W	10	23.4(±1.46)	2.4
RAN_W	20	23.6(±1.37)	3.7
RAN_W	30	26.0(±1.23)	5.7
RAN_A	10	21.0(±1.42)	2.3
RAN_A	20	23.0(±1.42)	4.0
RAN_A	30	25.7(±1.15)	5.7
EP_W	10	24.9(±1.10)	10.5
EP_W	20	24.4(±0.97)	16.1
EP_W	30	24.7(±0.88)	19.3
EP_A	10	23.8(±1.35)	4.7
EP_A	20	23.4(±1.02)	8.3
EP_A	30	22.8(±1.04)	12.0
START	4	22.5(±1.44)	5.1
MinERR	10	24.3(±1.17)	12.6
MinERR	20	24.2(±0.97)	17.5
MinERR	30	24.3(±0.86)	19.0
MaxLRT	10	23.3(±1.12)	12.0
MaxLRT	20	23.9(±0.90)	18.3
MaxLRT	30	25.0(±0.84)	19.7

* True value of the position used in simulation ** All genotypes from the offspring were included

Variance Components: A sire model (equation 1) was applied to the simulated data to partition total variance into QTL, polygenic, and residual components. Estimation of these parameters required an iterative solving of more than 7280 equations (depend on the strategy) at each QTL position (five points), and also for each candidate (300 candidates per each 2% of animals to be genotyped, as described in **Material and Methods**). Thus, it was computationally intensive and time consuming and therefore, given all available genotyped individuals in each iteration of genotyping, the most likely position of QTL with the highest likelihood ratio value was used to estimate the variance components and their standard errors. The resulting estimates are shown in Table 4, as the average over 100 data sets. Gametic QTL variance was estimated as half of the total QTL variance (biallelic QTL was assumed). Polygenic variance (from sire model) is a quarter of the total additive genetic variance used in the simulation. Therefore, 3/4 of the additive genetic effect is a part of the estimated residual variance components.

Table 4 Averaged QTL, polygenic (sire) and residual variance components (±SE) over all QTL positions for different levels of genotyping and different strategies

Strategy	Genotype(%)	QTL effect†	Sire effect	Residual
True*		0.0313	0.0469	0.8901
All**	100	0.0335(±0.00134)	0.0444(±0.00176)	0.9052(±0.0024)
RAN_W	10	0.0494(±0.00407)	0.0370(±0.00225)	0.8890(±0.0046)
RAN_W	20	0.0426(±0.00286)	0.0398(±0.00198)	0.8960(±0.0035)
RAN_W	30	0.0383(±0.00228)	0.0414(±0.00187)	0.9003(±0.0031)
RAN_A	10	0.0492(±0.00399)	0.0366(±0.00222)	0.8893(±0.0045)
RAN_A	20	0.0423(±0.00286)	0.0399(±0.00199)	0.8961(±0.0035)
RAN_A	30	0.0379(±0.00227)	0.0420(±0.00190)	0.9006(±0.0031)
EP_W	10	0.0480(±0.00242)	0.0389(±0.00194)	0.8879(±0.0032)
EP_W	20	0.0389(±0.00165)	0.0426(±0.00185)	0.8960(±0.0027)
EP_W	30	0.0335(±0.00134)	0.0444(±0.00180)	0.9010(±0.0025)
EP_A	10	0.0404(±0.00295)	0.0413(±0.00209)	0.8973(±0.0036)
EP_A	20	0.0384(±0.00215)	0.0419(±0.00190)	0.8988(±0.0029)
EP_A	30	0.0374(±0.001801)	0.0427(±0.00183)	0.8995(±0.0027)
START	4	0.0587(±0.00400)	0.0340(±0.00226)	0.8785(±0.0045)
MinERR	10	0.0473(±0.00222)	0.0389(±0.00190)	0.8882(±0.0031)
MinERR	20	0.0352(±0.00149)	0.0437(±0.00182)	0.8994(±0.0026)
MinERR	30	0.0288(±0.00121)	0.0461(±0.00180)	0.9057(±0.0024)
MaxLRT	10	0.0458(±0.00225)	0.0395(±0.00192)	0.8898(±0.0031)
MaxLRT	20	0.0364(±0.00154)	0.0435(±0.00184)	0.8980(±0.0026)
MaxLRT	30	0.0306(±0.00128)	0.0453(±0.00180)	0.9038(±0.0025)

† Variance component of the QTL is gametic QTL effect

* True value of the position used in simulation

** All genotypes from the offspring were included

Corresponding QTL variance component estimates for random genotyping (RAN_W and RAN_A) and extreme genotyping across family were significantly higher than the true QTL variance and estimates obtained from complete genotyping. With increasing genotype information (from 10% to 30%) in different putative QTL positions, the estimated QTL variance component decreased but the estimated polygenic variance increased. In contrast with genetic effects, the residual component remained relatively stable at all levels of genotyping regardless of the strategy used. However the variance components due to the QTL by RAN_A, RAN_W and EP_A still overestimated after 30% genotyping.

All genetic variance components were biased with EP_A, MinERR and MaxLRT at 10% genotyping. As it was expected, by increasing the proportion of genotyping in these strategies, the estimated QTL effect approaches the true value. In contrast to the QTL effects, the polygenic variances underestimated at the low levels of genotyping. This quantifies overestimation of QTL effects induced by selection of the extremes or random samples. With 20% genotyping, the QTL variance component was not significant between the MinERR strategy and complete genotype information but MaxLRT and random genotyping methods_were significantly different than complete genotyping. Variance component estimates in the MaxLRT strategy has been closed to the estimated values of MinERR. Standard error of the components showed a similar trend. The standard error of the QTL variance component was always lowest in MinERR strategy and highest in random genotyping.

Discussion

The present study proposes selective methods to identify more informative individuals for genotyping given all available marker, pedigree and phenotype information in a daughter design half-sib family, *e.g.* dairy cows. Whereas detection and positioning of QTL utilize the linkage between marker and phenotype information, the EP method (currently used in selective genotyping) only consider phenotypes. MinERR and MaxLRT in this study can be used to map QTL more accurately compared to EP, and decrease genotyping costs in QTL mapping experiments. The major restriction in QTL mapping and detection is the costs of collecting and typing of marker data and therefore, sampling approaches for genotyping have been designed to provide considerable cost saving, particularly when the phenotype is routinely collected. Various strategies for selective genotyping have been proposed for human and animal QTL detection experiments using kinships (Cardon & Fulker 1994, Stella & Boettcher 2004). The principle underlying these methods is that the difference in phenotypes between pairs of sibs (within family) becomes larger as they share a decreasing number of alleles at a QTL identical by descent from their parents (Chatziplis *et al.* 2001), and therefore in such families a QTL is more likely to be segregating. The current study showed combining phenotype information with marker information would provide more accurate detection of QTL compared to only using the extreme phenotypes. Additionally, in order to decrease number of individuals in genotyping, MinERR (or MaxLRT) could be considered. MinERR was an alternative approach to decrease genotyping costs and needed to genotype only 20-25% of animals to achieve unbiased parameters compared to complete genotyping. The power with 20% genotyping for MinERR and MaxLRT were 80% and 75% of that obtained with complete genotyping compared to 70% and 38% with EP within and across families, respectively. Higher power to detect QTL in the strategies based on both phenotypic and genotypic information (MinERR and MaxLRT) compared to only considering the phenotypic information (EP methods) and also random approach, at the same level of genotyping is due to selection of more informative individuals for genotyping. In MinERR and MaxLRT, daughters from all sires were potentially candidates to genotype but some daughter groups were never included in genotyping. Therefore, increased power is also due to selection of daughters for genotyping from segregating sires. Ultimately this causes more genotyped daughters from segregated sired to contribute information to contrasts between marker haplotypes. However in EP_W strategy, power was not usually significantly different compared to MinERR and MaxLRT. In this study, it was assumed that frequency of the mutant QTL allele was between 0.45-0.55. Therefore, the sires were more likely to be heterozygous compared to lower frequencies for favorable QTL allele. Using lower frequency could affect the within family genotyping strategies and decrease the power of detection as heterozygous sires (informative sires) will be rarer than in the current study. Besides in MinERR and MaxLRT, the genotype of the candidates were as a group (2% by 2% according) to get more realistic strategies. Therefore if it can be possible to genotype the candidates one by one, it is expected that the power of MinERR and MaxLRT would be even larger than those power obtained here.

QTL parameters indicated that application of EP selection strategies was superior to random genotyping of individuals, as also shown in previous experiments (*e.g.* Stella & Boettcher, 2004). With increasing level of genotyping, both EP across and EP within family, the extreme phenotype approach has significantly increased power of QTL detection and sensitivity of QTL position (Meuwissen *et al.* 2005). This shows that EP methods in the current study discriminate between the presence and absence of a segregating QTL in the families. Sampling induces a correlation between estimated residual and QTL effects. This correlation increased when the extreme daughters of all sires were selected for example in EP_A. It indicates that in the extremely selective genotyping approaches there is a real probability that an erroneous conclusion can be drawn about the location. In the extreme phenotype methods, estimated effects are biased upwards when genotyping the extremes within the population or at low levels of genotyping (less than 20%), even if all available phenotype information was used. This phenomenon might be due to the positive correlation between residual effects and the QTL effect in the sampled individuals for genotyping that magnifies the allelic effect. MinERR and MaxLRT obtain a power of detecting a QTL of app. 70% at 20 ~ 30 percent genotyping compared to a power of 88% with complete genotyping. These two criteria showed significantly increased power compared to random genotyping and genotyping the extremes across families, and marginally increased power compared to genotyping extremes within family. Power of a selective genotyping approach to detect a QTL can be affected by several factors such as the number of genotyped animals, the effect of segregating QTL (α) and heritability of the trait. Stella & Boettcher (2004) showed EP strategy was more precise than random genotyping when the heritability and QTL variance of the trait were low. However, lower power is expected in EP genotyping because in these situations (low heritability and QTL variance), the phenotypes provide relatively little information about the genotypes and EP method is based on the phenotype records. The combined strategies in this study could also use genotypic information and therefore, they might be more useful for selective genotyping of a low heritable trait.

Estimated variance component from the MaxLRT and MinERR strategy differed only slightly. The QTL position from the MaxLRT strategy was more accurate compared to the MinERR method and therefore MaxLRT could used to fine map a QTL. However, with using the EP strategy within families at 30% genotyping, estimated variance components (QTL and polygenic effects) were not significantly different compared to those obtained with complete genotype information. Besides, as proportion genotyped increased (10 to 30%) in both MinERR and MaxLRT, standard error of the estimated QTL variance component decreased, compared to other selective strategies in this study.

The genetic variance explained by QTL in this experiment ($\frac{1}{4}\sigma^2_G$) was in the range given by Druet *et al.* (2006). They reported that the proportion of genetic variance explained by a QTL was up to 36.0% for dairy traits in a Holstein population. The variance of QTL is a function of allele frequencies and the QTL substitution effect. The proportion of segregating sires (heterozygous sires) for QTL also is a function of QTL allele frequencies and therefore directly influences the QTL variance. Decreasing the variance of QTL by using lower favorable QTL frequency (*e.g.* $\boldsymbol{P_Q}<0.3$) will affect the efficiency of assessing which sires are heterozygous. If some sires are not informative, it is not possible to determine whether they are heterozygous. Therefore, the estimated proportion of heterozygous sires will be underestimated.

MinERR and MaxLRT methods do not come without disadvantages. These strategies have higher computational requirements than EP and random genotyping. The most time-consuming part was the likelihood maximization, because MinERR and MaxLRT strategies are based on standard error of estimated variance component of QTL effect and the maximum likelihood function of the model, respectively. However, if there is not possibility to use these strategies, genotyping based on the phenotypes and using the extremes within family can be suggested in order to obtain a powerful approach in QTL detection with unbiased estimates with at least 30% genotyping.

The current study considered a single trait and one QTL on a specific chromosome segment. If either more than one correlated trait or QTL is considered, some decrease in the selection intensity of the samples may secure sufficient power of detection for all traits. The reverse way can be possible when traits are uncorrelated. This could be an interesting area for further study.

Conclusion

MinERR and MaxLRT, can be used as approaches in selecting individuals for genotyping and the results from simulation in this study showed that unbiased gene substitution effects could be estimated with only 20% of the animals genotyped. QTL position parameters from MaxLRT strategy were more accurate compared to MinERR method and therefore it should be preferred. MinERR method decreased the level of genotyping (up to 22% compare to 30% in extreme phenotyping) to increase the power and estimate true value of QTL parameters.

Acknowledgements

This research was supported by a grant from Agriculture Research and Education Organization of the Ministry of Jahad-e-Agriculture, Iran. The authors are grateful to Peter Sørensen and Hauke Thomsen for using the IBD program and colleagues in Department of Genetics and Bioinformatics, Foulum, Denmark for their helpful comments in this study.

References:

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

Cardon L. R., Fulker D.W. (1994) The power of interval mapping of quantitative trait loci, using selected sib pairs. *Am. J. Hum. Genet.,* **55**, 825–833.

Casas E., Shackelford S.D., Keele J.W., Stone R.T., Kappes S.M., Koohmaraie M. (2000) Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate forms of myostatin. *J. Anim. Sci.,* **78**, 560–569

Casu S., Carta A., Elsen J.M. (2003) .Strategies to optimize QTL detection designs in dairy sheep populations: The example of the Sarda breed. In: Gabina D. (ed.), Sanna S. (ed.), *Breeding programmes for improving the quality and safety of products. New traits, tools, rules and organization? Zaragoza: CIHEAM-IAMZ,* pp. 19-23.

Chatziplis D.G., Hamann H., Haley C.S. (2001) Selection and subsequent analysis of sib pair data for QTL detection. *Genet. Res.,* **78,** 177–186.

Darvasi A. (1997) The effect of selective genotyping on QTL mapping accuracy. *Mammalian Genome,* **8,** 67-68.

Darvasi A., Soller M. (1992) Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theor. Appl. Genet.,* **85**, 353 - 359.

Druet, T., Fritz S., Boichard D., Colleau J.J. (2006) Estimation of genetic parameters for quantitative trait loci for dairy traits in the French Holstein population. *J.Dairy Sci.,* **89**:4070–4076

Jansen R.C., Johnson D.L., Arendonk J.A.M.van. (1998) Mixture model approach to the mapping of quantitative trait loci in complex populations with an application to multiple cattle families. *Genetics,* **148**, 391-399.

Johnson D.L., Jansen R.C., Arendonk J.A.M.van. (1999) Mapping quantitative trait loci in a selectively genotyped outbred population using a mixture model approach. *Genetical Research,* **73**, 75-83.

Kinghorn B.P. (1997) An index of information content for genotype probabilities derived from segregation analysis. *Genetics,* **145**, 479–483.

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

Lander E.S., Botstein D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics,* **121**, 185-199.

Macrossan P.E. (2004) Strategies to Minimise DNA Testing Costs for Research and Development Programs Involving Pedigreed Populations. PhD Thesis, University of New England, Australia.

Madsen P., Jensen J. (2002) A user’s guide to DMU. A package for analysing multivariate mixed models. Version 6, release 4.4, Danish Institute of Agricultural Sciences, Tjele, Denmark.

Martinez M.L., Vukasinovic N., Freeman A.E., Fernando R.L. (1998) Mapping QTL in outbred populations using selected samples. *Genet. Sel. Evol.,* **30**, 453-468.

Meuwissen T.H.E., Goddard M.E. (2000) Fine scale mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics,* **155**, 421-430.

Meuwissen T.H.E., Goddard M. E. (2001) Prediction of identity by descent probabilities from marker-haplotypes. *Gen. Sel. Evol.,* **33**, 605-634.

Meuwissen T.H.E., Janss L.L.G., Bovenhuis H. (2005) QTL detection and fine mapping in complex pedigree. Animal Breeding and Genetics Group, Wageningen University, The Netherlands.

Muranty H., Goffinet B. (1997) Selective genotyping for location and estimation of the effect of a quantitative trait locus. *Biometrics,* **53**, 629-643.

Piepho H.P. (2001) A quick method for computing approximate thresholds for quantitative trait loci detection. *Genetics*, **157***,* 425–432.

Ronin Y.I., Korol A.B., Weller J.I. (1998) Selective genotyping to detect quantitative trait loci affecting multiple traits: interval mapping analysis. *Theor. Appl. Genet***., 97**, 1169-1178.

Stella A., Boettcher P.J. (2004) Optimal designs for linkage disequilibrium mapping and candidate gene association tests in livestock populations. *Genetics*, **166***,* 341-350.

Sørensen P., Lund M.S., Guldbrandtsen B., Jensen J., Sorensen D. (2003) A comparison of bivariate and univariate QTL mapping in livestock populations. *Genet. Sel. Evol.,* **35**, 605–622

Van Gestel S., Houwing-Duistermaat J.J., Adolfsson R., van Duijn C.M., Van Broeckhoven C. (2000) Power of selective genotyping in genetic association analyses of quantitative traits. *Behavior Genetics,* **30**, 141-146.

Chapter 3

Paper II

Fine mapping QTL under selective phenotyping strategies based on linkage and linkage disequilibrium criteria

Saeid Ansari-Mahyari, Peer Berg , Mogens Sandø Lund

Submitted to

Genetics Selection Evolution

Fine mapping QTL under selective phenotyping strategies based on linkage and linkage disequilibrium criteria

Saeid Ansari-Mahyari [1, 2, 3], Peer Berg [1], Mogens Sandø Lund [1]

[1] Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, University of Aarhus, Denmark

[2] Agriculture Research and Education Organization, Isfahan Agricultural & Natural Resources Research Center, Iran

[3] Department of Large Animal Sciences, Faculty of Life Sciences, Copenhagen University, Denmark

Abstract: In fine mapping experiments where phenotypes are very expensive, difficult to collect or time-demanding, selective phenotyping (SP) could be used to phenotype the most informative individuals and reduce phenotyping costs. Linkage-based (LA) and linkage disequilibrium-based (LD) criteria for mapping quantitative trait loci (QTL) were investigated and compared to random phenotyping. Several selective phenotyping strategies based on LA and LD information were compared using stochastic simulations with three levels of phenotyping (30, 40 and 50%). The LA criteria were applied within sire groups and QTL were detected using a linkage method. The LD criteria were used to select animals for phenotyping across sire families and QTL were detected with a method combining linkage disequilibrium and linkage information. Selecting individuals with similar haplotypes to the paternal haplotypes rather than selecting them randomly, increased the power to detect QTL. In order to use LA criterion to estimate unbiased QTL parameters, a large half-sib family size or prior information on QTL position required. The SP based on LD criteria used more information across families and improved the accuracy to estimate the QTL position compared to random selective phenotyping with the same sample size. Additionally, LD criteria increased power of detection when only 30% of all individuals were phenotyped. By increasing the proportion of phenotyping when applying LD strategies (30% to 50%), the estimated QTL effect approaches the true value quicker than when LA strategies were applied. The results showed that the developed LD criteria were better than the LA criteria to select individuals for phenotyping in QTL (fine) mapping studies.

Keywords: Linkage criteria/ linkage disequilibrium criteria/ phenotyping costs

1. Introduction

In detecting and mapping loci affecting quantitative traits (QTL), the association between the phenotype and markers spanning across the genome is assessed in a genetically segregating population. Due to recent developments in high-throughput technologies, genotyping costs are generally less limiting to the sample size in QTL mapping experiments. The genotyping facilities have drastically reduced the cost of genotyping [13]. This is particularly true for single nucleotide polymorphism (SNP) markers [1]. In contrast to these developments for genotyping, there are some situations where recording the phenotypes of traits are difficult or very expensive. This includes the phenotypes on complex physiological or behavioral traits [12]. In addition to these phenotypic records, gene mapping studies based on microarray expression data as phenotype is now emerging and will be soon commonplace [8]. However, from cost-power perspective in microarray experiments, the limits of sample size need to be considered carefully through an efficient sampling strategy to reduce the number of microarrays.

Darvasi and Soller [6] introduced a genotyping sampling method for identifying the individuals which called selective genotyping (SG). This method selected the most informative individuals (half of the population) and the power of detection retained as compared to when the whole population was genotyped. The SG may be preferred if expenses of rearing are fairly low and the trait is routinely evaluated. Instead of considering the phenotype to identify informative individuals in genotyping, selective phenotyping (SP) are necessary for situations where costs of phenotyping are larger than costs of genotyping or the trait can not easily collected. Medugorac and Soller [15] practiced selective phenotyping when a main trait is highly correlated with an indicator trait which could be easily recorded. Casu *et al.* [4] showed SP through a correlated trait in a daughter design can be suggested in QTL detection wherever the main trait is difficult or expensive to record. These studies indirectly used the SP based on information from the correlated traits.

Jin *et al.* [12] investigated an index in selective phenotyping methods for crosses based on information from the main trait. Their criteria aimed to maximize genetic dissimilarity in the candidates for phenotyping. After detecting the QTLs based on an initial genome scan, these linkage criteria can be used to obtain further accuracy in QTL fine mapping [5, 11].

Linkage analysis approaches are used to detect QTL through an initial genome scan experiment. A proper experimental design is required in linkage studies to identify recombinant individuals. However, utilizing historical recombination based on linkage disequilibrium information (LD) allows to finemap using current outbred populations *e.g.* dairy cattle, instead of experimental populations for QTL studies. The LD information are becoming more popular for QTL experiments because creation of populations such as advanced intercrossed lines is nearly impossible in livestock due to time demands, financial constraints and inbreeding depression. If linkage disequilibrium exists in the population, one possibility is to use non-random association of alleles at marker brackets for fine mapping *e.g.* [3, 26]. Different factors can generate linkage disequilibrium in a population such as number of founders, population admixture, selection, and mutation. Modern livestock populations are usually a combination of other smaller populations and therefore, LD could be generated due to difference in the allele frequencies [19, 24]. Selective phenotyping using criteria based on LD across a population can utilize information on historical recombinations of alleles at different loci in an identified chromosome segment. These methods can be used to select and genotype the animals for fine map QTLs. For example with half-sib family design in outbred populations [10] or grand daughter designs as in Olsen *et al.* [20] who utilized the combined linkage and LD mapping in dairy cattle based on IBD method of Meuwissen *et al.* [16]. Moreover, LD information may be particularly useful to fine map or confirm association of candidate regions that have already been shown to include QTLs.

The objective of this study was to develop criteria for ranking animals for phenotyping based on genotype information for QTL detection purposes. The criteria were constructed from linkage and LD information in a set of half-sib families using simulated data.

Different proportions of the phenotyped animals studied for detection in comparison to a random phenotyping approach.

2. Material and methods

2.1 Outline of the simulation

A stochastic simulation was used to generate data to compare several SP criteria. Linkage disequilibrium between markers and QTL was simulated over 100 generations of random mating in a historical population without recording the pedigree and with constant population size of 100 males and 100 females in each generation (see Meuwissen and Goddard [17]). The genotypes for all individuals were simulated for 21 marker loci equally spaced in a 100 cM chromosome segment and also an associated QTL placed in the midpoint between markers 5 and 6. A simple bi-allelic QTL and markers with five alleles were assumed. In the first generation, unique alleles were sampled and the frequencies of alleles in each marker were equal to 0.2. Marker and QTL genotypes were sampled according to rules of Mendelian inheritance and allowed for recombination within the segment. Additionally, it was assumed that recombinants could be identified unambiguously, with known genotype phase. In the last generation a mutant QTL allele was sampled at random, with the requirement that the frequency of positive QTL ($\boldsymbol{P_Q}$) was between 0.45 and 0.55.

The generation 101 was assumed as base population with unknown parents (founders). Marker genotypes were assumed known for parents (generation 101) and marker genotypes, pedigree and phenotypic values were known in the progenies (generation 102) in order to use for the strategies in this study. A quantitative trait with moderate inheritance (h^2 = 0.25) was modeled with total phenotypic variance of one. Proportion of QTL variance relative to the total phenotypic variance was between 0.062 and 0.063 [$\sigma^2_{QTL} = 2 \times \boldsymbol{P_Q} \times (\boldsymbol{P_Q} - 1) \times \alpha^2$], where α is additive allele substitution effect ($\alpha = \mu_{QQ} - \mu_{qq}$) and assumed 0.354. Polygenic and residual effects were drawn from N(0, $\mathbf{A}\sigma^2_a$) and N(0, $\mathbf{I}\sigma^2_e$), where $\mathbf{A}$ is the numerator relationship matrix which ignored the ancestral relationship beyond the known pedigree, $\mathbf{I}$ is an identity matrix, and σ^2_a and σ^2_e are polygenic and residual variance components, respectively. Polygenic effects (true breeding values) were equal to means of the paternal and maternal values plus a Mendelian sampling component N(0, 1/2σ^2_a). After generating an observation, the QTL effect was added. In this study, 30 unrelated sires were simulated with 80 daughters per sire. The progenies were paternal half-sibs, i.e. all dams were assumed unrelated.

2.2 Phenotyping Criteria

In order to choose the most informative individuals, two sources of information in a daughter design were considered, based on linkage information within each sire family (LA criteria) and linkage disequilibrium across half-sib families (LD criteria). Markers on both sides of a putative QTL position were considered as a group which called “haplotype”. The size of the haplotype was between 6 to 10 markers, depending on the QTL position. Twenty marker intervals between were studied for QTL position. If putative QTL position was in the intervals 5 to 16, then the haplotype size was 10 (*e.g.* MMMMMQMMMMM) and otherwise less than 10.

In order to evaluate QTL parameters in this study, the following strategies were used to select daughters for phenotyping.

2.2.1 Random phenotyping

With this method, individuals were randomly phenotyped either within sire families or across the families. In 3 levels of phenotyping (30, 40 and 50%), either an equal number of randomly chosen daughters within sire families or a random sample across sire families were considered. The reason for two approaches in random phenotyping was to compare randomly phenotyped individuals in both LA and LD criteria.

2.2.2 LA Criteria

The main objective in this approach was to use linkage information by following the recombination events within each half-sib family. Therefore, available marker scores in the parents and progenies were considered in the LA strategies.

Maximum Recombinants (MaxRec): This strategy was considered to maximize recombination across the haplotype and sample the most recombinant offspring given parental genotypes with three levels of phenotyping (30, 40, and 50%). For each progeny (*j)* and sire (*k*)*,* defined R$_{lijk}$ = 1 if the progeny was recombinant in marker interval *i* in the paternally (*l*=1) or maternally (*l*=2) inherited genotype, and R$_{lijk}$ = 0 otherwise. The number of recombinant marker intervals for progeny *j* within sire *k* as:

$$R_{\circ\circ jk} = \sum_{\ell=1}^{2} \sum_{i=1}^{Npos} R_{\ell,i,j,k} \qquad \text{(Eq. 1)}$$

The *N* progeny with most recombinations were selected for phenotyping within sire family groups.

Maximum-Uniform Recombinants (MaxUniRec): While the previous strategy maximized the overall linkage for QTL mapping which is available among the selected progeny in each sire group, the uniform number of recombination used to sample individuals with many recombinations, such that the recombinations across sampled individuals were uniformly distributed over the identified chromosomal region. The rationale for this strategy followed from the assumption that the QTL position was unknown in the identified segment and therefore, it was desirable to have mapping information evenly distributed. Thus, this method proposed as a simple way to meet the two objectives of maximizing total information and also its uniformity.

Three levels of phenotyping (30, 40, and 50%) were considered for which the sampling was performed in two stages. In the first stage, half the candidates were selected based on the maximum recombinations (MaxRec) leading to 12, 16 and 20 of the highest recombinant progenies within each sire group being phenotyped. Therefore in total, 360 progenies in 30% phenotyping; 480 progenies in 40% phenotyping; and 600 progenies in 50% phenotyping were selected in the first stage based on maximum recombination. All selected individuals were assigned to the set of *S*. In the second stage, the following steps were iterated until *N* progeny were selected (720, 960, and 1200 selected for the levels of 30, 40 and 50%, respectively):

1) The rest of progenies within each sire family were added to *S*, one by one, and recombination events were summed for the 20 intervals (the intervals of 21 markers). Then, standard deviation (*STD*) of recombinations across the intervals was calculated as:

$STD=\sqrt{\frac{1}{N_{\text{int}}-1}\sum_{i=1}^{N_{\text{int}}}(x_i-\bar{x})^2}$; where, N_{int} is the number of the marker intervals across identified region (N_{int} =20), x_i is the total number of recombinations in interval *i*, and $\bar{x}$ is the mean of all intervals. This approach was done for all sire groups to achieve the changes in *STD* of recombination when one offspring is added to *S*.

2) In the next step, within each sire group, the progeny resulting in the lowest *STD* when added to *S* is retained in *S*. If several progeny have the same standard deviation, one of them was picked at random.

3) Start again at step 1 to determine the next progeny to phenotype.

Minimum Recombination (MinRec): This strategy minimized recombination events across the intervals only in the paternal haplotype of half-sib progenies. The score (modified from Equation 1) is the sum of the recombinant marker intervals for progeny *j* within sire *k*:

$$R_{0\,jk} = \sum_{i=1}^{Npos} R_{i,j,k} \qquad \text{(Eq. 2)}$$

The *N* progeny with lowest recombination score within the sires were selected for phenotyping. The same numbers of progenies were phenotyped in each sire family. Therefore, this strategy primarily samples non-recombinant offspring for phenotyping. From each sire group 30, 40 and 50% of the offspring were selected for phenotyping.

2.2.3 LD Criteria

In order to use LD information within an outcross population *e.g.* half-sib families, several criteria were applied. The main objective in these criteria was to phenotype informative individuals across the sire families for QTL analysis based on the historical generations.

The parental haplotypes were clustered with maximum 5 markers on the left and 5 markers on the right sides of the putative QTL location (as detailed by Meuwissen and Goddard [18]).

Maximum Frequency (MaxFre): This approach aimed to maximize frequencies of the clustered haplotype in the founders (parents) while minimizing recombinations in offspring. First in each putative QTL position, all unique haplotypes in the founders were identified. Then after clustering the haplotypes in parents, the frequency of clustered haplotypes were assessed. In a particular interval the cluster group of the paternal and maternal haplotypes in the offspring could be identified. Over all putative QTL positions, the frequencies of paternal and maternal haplotypes were summed in the offspring (**HapFre**) as below. Equation 3 was used when a recombination is observed in the marker bracket (interval that assumed there is a QTL), and equation 4 if not:

$$HapFre_{individual} = \sum_{ipos=1}^{Npos} \left[(\frac{1}{2}P_1 + \frac{1}{2}P_2) + (\frac{1}{2}M_1 + \frac{1}{2}M_2) \right] \qquad \text{(Eq. 3)}$$

$$HapFre_{individual} = \sum_{ipos=1}^{Npos} [(P + M)] \qquad \text{(Eq. 4)}$$

where, P_1 and P_2 are frequencies of paternal and maternal clustered haplotypes in the sire, and M_1 and M_2 in the dam. P is paternal and M is paternal and maternal clustered haplotype frequencies of sire and dam, respectively. Progenies with the highest HapFre across sire families were chosen for phenotyping. Three levels of phenotyping (30, 40 and 50%) were used.

Equal Highly Frequent haplotypes (EqHigh): This strategy was used to decrease the variance of most frequent clustered haplotypes among the phenotyped animals. Therefore, the objective was to increase the uniformity in the most frequent clusters in each putative QTL position and sample a balanced set of frequent clustered haplotypes. To start this strategy, first the optimum number of clusters was found by minimizing the standard error of the variance component in a one way random model [22]. Preliminary study showed that using 50 of the most frequent clustered haplotype codes, was optimal. Moreover, two preliminary steps need to begin this strategy. In the step one, based on clustering in the QTL positions, the 50 most frequent haplotypes were found. In the step two, 10% of the population was phenotyped, such that clustered haplotypes of their parents should have maximum similarity with the most frequent clustered haplotypes (which were identified in the step one).

Then after two initial steps, the next individuals were iteratively chosen for phenotyping based on the following index until 30, 40 and 50% of the population was sampled:

$$Index = \sum_{ipos=1}^{Npos=20} \left(\sum_{ihap=1}^{Nhap=50} (N_{ipos,ihap} - \frac{2 \times N_{animal}}{50})^2 + N_{unsel}^2 \right) \qquad \text{(Eq. 5)}$$

where,N_{pos} is the number of putative QTL positions across the identified chromosomal segment, $N_{ipos,ihap}$ is number of the most frequent clustered haplotypes in each position, N_{animal} is the number of offspring to be phenotyped in each level of phenotyping, N_{unsel} is the sum of the cluster(s) of each individual which did not exist in the list of most frequent haplotypes. If the haplotype codes of the candidates existed in the codes derived in the first initial step, then N^2_{unsel} assigned was zero. Therefore, N^2_{unsel} can be different for each candidate, according to their founder haplotypes.

Maximum Similar Highly Frequented Clusters (Max): In this strategy, individuals were selected for phenotyping if their parental haplotypes have high similarity with highly frequent haplotypes in the population. First, all putative QTL positions were identified and then according to the founder genotypes, the most frequent clustered haplotypes were identified. If the parents of candidates have maximum similarity with the highly frequent clustered haplotypes over all the positions, then the candidate was chosen to be phenotyped. This process continued until three levels of phenotyping (30, 40 and 50%). Two sizes of the most frequent haplotype clusters were used in order to evaluate the size of the most frequent clustered haplotypes and also investigate robustness of this criterion. In **Max10**, only ten of the most frequent clusters in each position were assigned and in **Max50**, fifty of the most frequent clustered haplotype codes.

2.3 Analyses of the simulated data

In total 100 data sets were simulated and the criteria and levels of phenotyping were applied in each data set. In each scenario data from daughters selected for phenotyping were analysed with a mixed inheritance model with QTL as random effect. Variance components of QTL effect were estimated for each putative QTL position using REML.

Relationship (IBD) matrix between QTL alleles of any two founder haplotypes was computed by two procedures:

1) Linkage association analysis within sire group using a recursive algorithm [25],

2) LA+LD analysis (combination analysis) across sire family groups [18]. In clustering of the haplotypes a window of maximum five markers on each side of the putative QTL position was used over all strategies.

QTL analysis was carried out using the following linear mixed model:

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Zu} + \mathbf{Wq} + \mathbf{e}$$

where **y** is a vector of phenotypes for selected daughters, **μ** is overall mean, **Z** and **W** are known incidence matrices relating the phenotypes to its polygenic and QTL effects, **u** is a vector of random additive genetic effect due to the sires, **q** is a vector of random additive genetic effects due to the QTL effect and **e** is random residual effect. The random variables assumed to be normally distributed and mutually independent. Specifically the parameters used characterized as: **u** from N(0, $\mathbf{A}\sigma^2_u$), **q** from N(0, $\mathbf{H}_{\boldsymbol{p}}\sigma^2_q$), and **e** from N(0, $\mathbf{I}\sigma^2_e$), where **A** is the additive relationship matrix, $\mathbf{H}_p$ is the identity by descent (IBD) matrix that contains the IBD probabilities for the QTL at position *p*, and **I** is an identity matrix. As sires were assumed to be unrelated, **A** matrix simplifies to an identity matrix. The IBD probabilities for the QTL based on the haplotype were calculated using the analytical method of Meuwissen and Goddard [18]. Depended on the putative QTL position, the size of the haplotype was between 6 to 10 markers from both sides.

The DMU program package [14] was used for estimation of variance components and also computation of the maximum likelihood equations (MLE). This program uses average information restricted maximum likelihood (AI-REML) algorithm for estimation of (co)variance components in mixed models and the restricted likelihood is maximized with respect to variance components associated to the random effects [23]. The parameters were estimated at the marker brackets (mid points of the each marker interval) along the identified chromosome.

Hypothesis tests to detect QTL were based on the asymptotic distribution of likelihood ratio test (**LRT)= –2* (log likelihood$_{\text{(reduced)}}$ – log likelihood$_{\text{(full)}}$)** where, log likelihood$_{\text{(reduced)}}$ is from a model excluding the QTL-effect and log likelihood$_{\text{(full)}}$ is the likelihood for a model with a QTL-effect and it is calculated for each bracket. The LRT-statistic has a Chi-square distribution with one degree of freedom. The LRT-statistic thresholds for significant (P<0.05) and highly significant (P<0.01) QTL effect were calculated using Piepho approach [21], using LRT in each putative QTL location, number of chromosomes/segments, degree of freedom (which is difference in number of parameters between reduced and full models), and chromosome-wise type-I (α). Power of detection was assessed as the proportion of the 100 replicates in which a QTL was detected at a given genomewide significance level (α). Accuracy of position was assessed as the proportion of the 100 replicates in which a QTL was detected in the simulated position ± 5 cM.

3. Results

3.1 LA strategies

When progenies were selected based on the number of recombinations along the identified segment of chromosome using MaxRec and MaxUniRec strategies, no improvement were observed in power to detect QTL and accuracy of position compared to random selection (Table I) with same level of phenotyping. However, using the MinRec strategy which aimed to minimize recombination in paternal haplotypes of the progenies with almost balanced sampling of alternative haplotypes showed selective phenotyping was able to increase power of detection over random phenotyping (Fig. 1). Using this strategy, progenies were sampled with maximum similarity in the paternal haplotypes. Results in Table I indicated that sampling a set of non recombinant offspring has minimized the standard error of the contrast between the sire haplotypes and thereby increased the power of detection in MinRec strategy compared to sampling of recombinant offspring in other LA strategies. However, MinRec strategy was not able to estimate QTL position compared to other LA strategies. Averages of the likelihood ratios over hundred iterations in indicated that all LA strategies, except MinRec, have the highest values at the simulated QTL position (22.5 cM), and the peak of the expected curves have been reached in this point with increasing proportion phenotyped (Fig. 1).

Table I. Power of QTL detection and sensitivity of position[†] in All, Random and different LA strategies with two level of error type-I within sire family by linkage analysis

Strategy	Phenotype (%)	Detection (α=1%)	Detection (α=5%)	Position (α=1%)	Position (α=5%)
All [‡]	100	54	68	54	68
Random	30	0	6	0	5
	40	3	13	2	10
	50	12	20	9	19
MaxRec	30	0	7	0	5
	40	4	11	3	10
	50	10	17	10	17
MaxUniRec	30	2	5	2	5
	40	3	9	3	6
	50	5	17	3	14
MinRec	30	8	17	6	15
	40	9	26	7	17
	50	16	31	11	23

[†] Sensitivity of QTL position shows number of finding around an interval QTL position in simulation (22.5±5.0 cM) out of 100 iterations

[‡] All phenotypes from the offspring were included

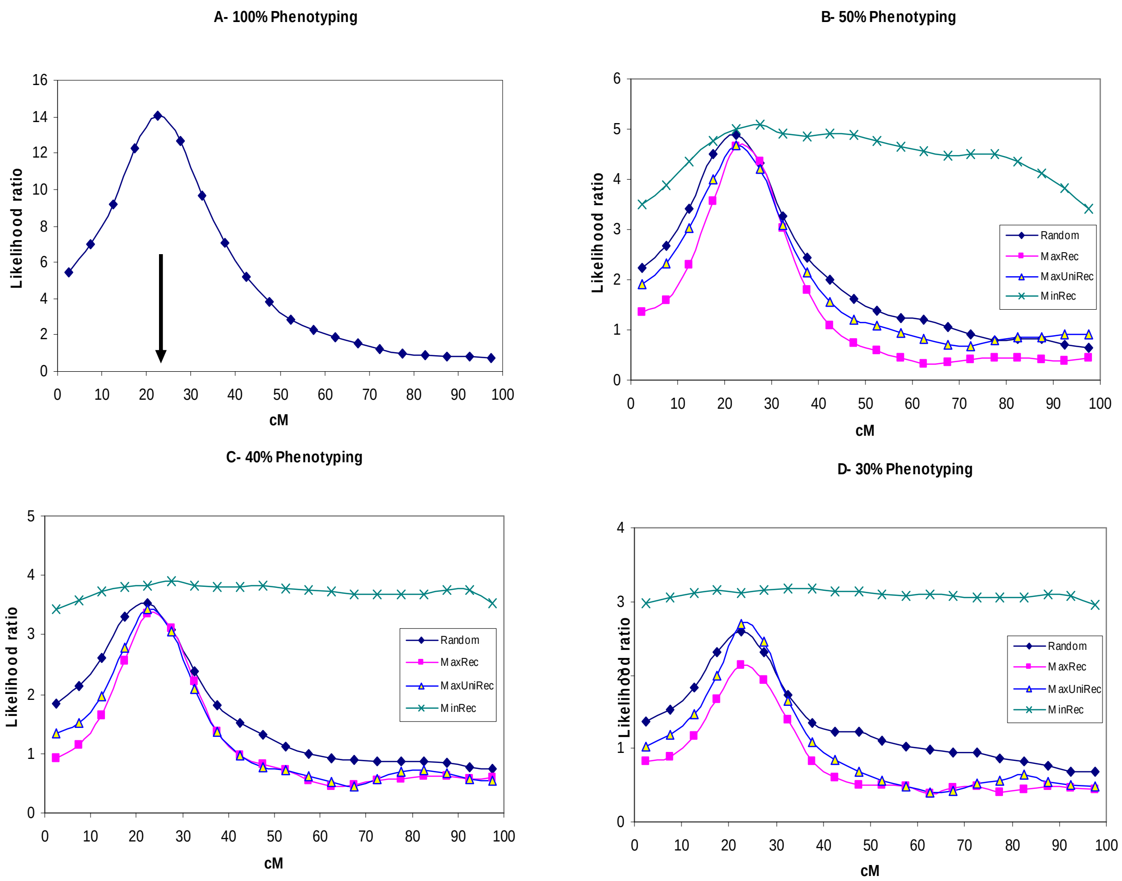

Figure 1. Map comparisons of different selective phenotyping methods using LA strategies. (A) all phenotyped individuals with simulated QTL position (22.5cM); (B, C and D) levels of phenotyping.

Estimated QTL position with MaxRec and MaxUniRec were not further away from the simulated QTL position in 50% phenotyping compared to MinRec (Table II). As might be expected, the sets of progenies selected by the MaxUniRec strategy had fewer recombinations than those selected by the MaxRec strategy. With 50% phenotyping, the average of recombinations was 3.085±0.103 in MaxRec and 2.417±0.165 in MaxUniRec, which is not a great difference. The MaxUniRec strategy captured 75 to 80% of the increase in recombinations achieved by the MaxRec strategy. Also, the two methods differed in term of their effects on the variance of recombinations across the marker intervals. With 50% phenotyping, coefficient of variations was 33.25% in MaxRec and 68.50% in MaxUniRec.

Table II. True and means (±SE) of QTL position estimates in cM; gametic QTL, polygenic and residual variance components based on 100 replicates in different selective LA strategies within sire family

Methods	Phenotype(%)	Position(cM)	$\hat{\sigma}^2_{QTL} \pm SE$	$\hat{\sigma}^2_{Sire} \pm SE$	$\hat{\sigma}^2_{error} \pm SE$
True[†]		22.5	0.0313	0.0469	0.8901
All [‡]	100	24.40±1.24	0.0343±0.00137	0.0449±0.00216	0.8225±0.0045
Random	30	36.30±2.65	0.0511±0.00257	0.0364±0.00277	0.7647±0.0092
	40	34.40±2.74	0.0423±0.00209	0.0413±0.00260	0.7933±0.0074
	50	28.95±2.00	0.0381±0.00195	0.0421±0.00248	0.8053±0.0066
MaxRec	30	38.65±2.82	0.0520±0.00252	0.0366±0.00291	0.7701±0.0089
	40	33.55±2.56	0.0455±0.00220	0.0399±0.00260	0.7883±0.0076
	50	31.15±2.31	0.0400±0.00197	0.0434±0.00253	0.8056±0.0069
MaxUniRec	30	35.20±2.70	0.0537±0.00244	0.0378±0.00267	0.7606±0.0087
	40	32.50±2.58	0.0447±0.00212	0.0408±0.00242	0.7904±0.0075
	50	31.25±2.43	0.0400±0.00174	0.0421±0.00245	0.8070±0.0058
MinRec	30	44.35±3.69	0.0420±0.00314	0.0450±0.00294	0.7882±0.0108
	40	45.65±3.57	0.0389±0.00231	0.0445±0.00252	0.8027±0.0079
	50	43.65±3.08	0.0370±0.00200	0.0442±0.00236	0.8092±0.0071

[†] True value of the parameters used in simulation

** All phenotypes from the offspring were included

No considerable difference was observed between LA criteria and random phenotyping in the estimated gametic QTL effect and also standard error of the variance component estimation of QTL (Table II), with the same number of phenotyped progenies. Nevertheless, polygenic effect in MinRec was always unbiased and similar to the analysis of data sets which were including all phenotype information.

3.2 LD strategies

These criteria were based on information across families in the population. All available phenotypic and genotypic information during the phenotyping process have been analyzed to detect QTL using a combined linkage disequilibrium and linkage analysis (LDLA) method. Table III presents the power of detecting QTL and accuracy of QTL location for the LD criteria. The precision is given as the number of replicates in which the estimated position is within ±5.0 cM from the true position (22.5 cM). Values that are higher than in the random scenario for a given percent phenotyped are given in bold. The results are presented for two levels of alpha (1% and 5%). The LDLA approach was more powerful in detecting the QTL than LA method when all progenies were phenotyped (Tables I and III).

Table III. Power of QTL detection and sensitivity of position[†] in All, Random and different LD strategies with two level of error type-I across sire families by LDLA analysis

Strategy	Phenotype (%)	Detection (α=1%)	Detection (α=5%)	Position (α=1%)	Position (α=5%)
All [‡]	100	100	100	100	100
Random	30	7	28	5	20
	40	30	47	24	43
	50	55	72	52	68
MaxFre	30	***29***	***52***	***22***	***46***
	40	***47***	***69***	***45***	***62***
	50	***63***	***80***	***59***	***75***
EqHigh	30	***17***	***40***	***13***	***33***
	40	***33***	***48***	***30***	42
	50	50	67	47	63
Max10	30	***26***	***45***	***21***	***38***
	40	***41***	***71***	***39***	***59***
	50	***70***	***82***	***60***	***78***
Max50	30	***27***	***51***	***23***	***45***
	40	***46***	***68***	***45***	***64***
	50	***63***	***80***	***59***	***80***

[†] Sensitivity of QTL position shows number of finding around an interval QTL position in simulation (22.5±5.0 cM) out of 100 iterations

[‡] All phenotypes from the offspring were included

Compared to the LA criteria, using LD criteria improved power of QTL detection when the percentage of phenotyped individuals increased. Precision of QTL location also increased with increasing level of phenotyping when the LD criteria were used. Generally, the likelihood ratio peaks were higher for the different LD criteria when compared to random phenotyping in the same level of phenotyping (30, 40 and 50%). The only exception was EqHigh for which the curves were similar to random at 50% phenotyping (Fig. 2). Both power to detect QTL and accuracy of estimating QTL position in the strategies with maximum similarity among the most frequent clusters (Max10 and Max50) were almost identical (Table III). Although precision of QTL location in low levels of phenotyping (till 40%) was higher with using more frequent clusters (Max50).

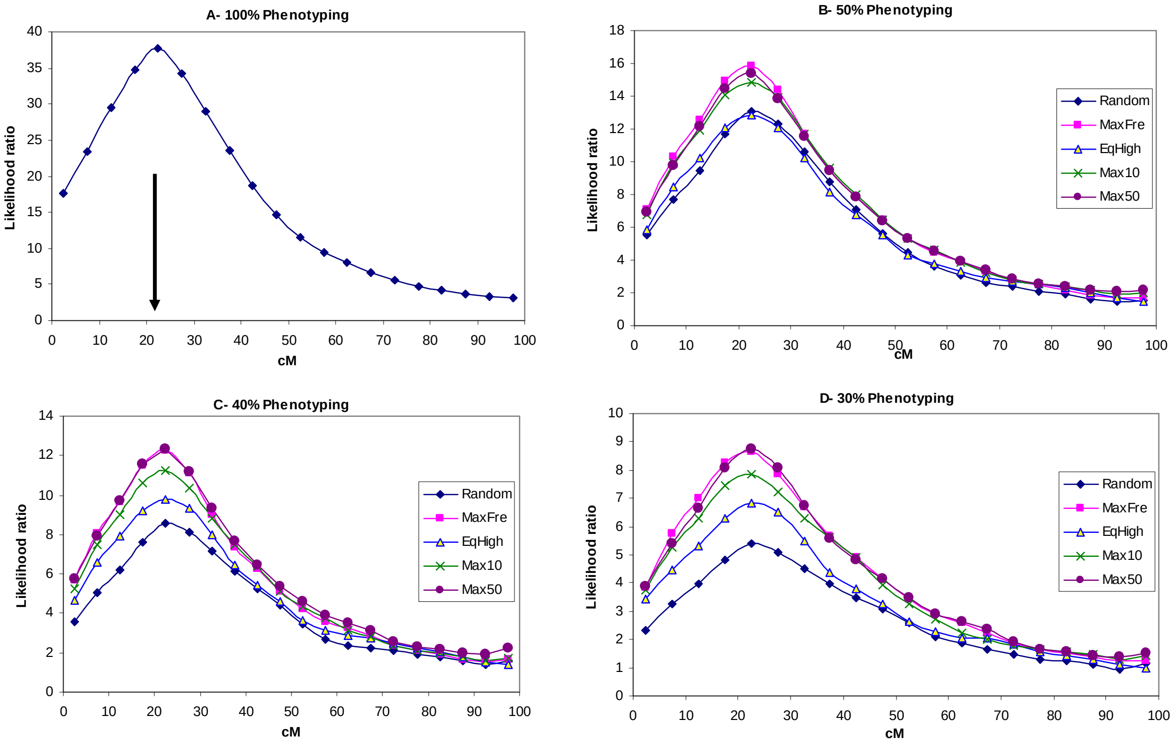

Figure 2. Map comparisons of different selective phenotyping methods using LD strategies. (A) all phenotyped individuals with simulated QTL position (22.5cM); (B, C and D) levels of phenotyping.

Precision of QTL positioning improved as the level of phenotyping increased with the same criterion (Table IV). Generally LD criteria in this study were increased QTL precision compared to random phenotyping (Fig. 2). Especially with low percents of phenotyping (30 or 40%), they were considerably closer to the simulated QTL position compared to the random phenotyping (Table IV). However, the peaks of likelihood ratios in Fig. 2 showed that all LD criteria and also random phenotyping closely estimated QTL position to the true location (22.5 cM), regardless of increasing the amount of phenotype records. Moreover, results in Table III showed that LD criteria were robust and more powerful in QTL studies and it is possible to reach the same power in detection and position with less phenotyped progenies (until 25%) if LD criteria be used compared to random phenotyping. For example, in 30% phenotyping with the strategy Max50, 90-108% of power for detection and 96-105% for position were achieved compare to 40% random phenotyping.

Estimations of the QTL, polygenic and residual variance components based on LD criteria are shown in Table IV. The estimations of variance components using all phenotypic information through LDLA QTL-analysis were considerably closer to the true values compared to LA QTL-analysis (Tables II and IV). When only 30% of the individuals were phenotyped the estimates of QTL variance was biased in all LD strategies as well as with random sampling. However, when the percentage of phenotyped individuals increased to for example 50%, the QTL variance was still biased for random sampling, but unbiased estimates were reached using both Max50 and EqHigh. As shown in Table IV, by increasing the proportion of phenotyping in each LD strategy, the estimated QTL effect approached the true value. In contrast to the QTL variance component, residual variance estimate was underestimated at the low levels of phenotyping when LD criteria were used. Polygenic variance was also underestimated through MaxFre and EqHigh but unbiased with all levels of phenotyping in Max10 and Max50. The estimates of polygenic and residual variances using random phenotyping were not significantly different from estimates when all phenotypes were used.

Table IV. True and means (±SE) of QTL position estimates in cM; gametic QTL, polygenic and residual variance components based on 100 replicates in different selective LD strategies across sire families

Methods	Phenotype(%)	Position(cM)	$\hat{\sigma}^2_{QTL} \pm SE$	$\hat{\sigma}^2_{Sire} \pm SE$	$\hat{\sigma}^2_{error} \pm SE$
True[†]		22.5	0.0313	0.0469	0.8901
All [‡]	100	22.30±0.45	0.0328±0.00064	0.0461±0.00175	0.8901±0.0026
Random	30	32.25±2.17	0.0378±0.00152	0.0452±0.00280	0.8891±0.0052
	40	29.95±1.95	0.0348±0.00124	0.0455±0.00241	0.8959±0.0044
	50	27.60±1.49	0.0350±0.00116	0.0455±0.00208	0.8917±0.0040
MaxFre	30	30.30±1.90	0.0389±0.00155	0.0455±0.00343	0.8772±0.0049
	40	28.00±1.72	0.0358±0.00138	0.0459±0.00262	0.8841±0.0038
	50	24.85±1.20	0.0340±0.00117	0.0453±0.00227	0.8851±0.0034
EqHigh	30	25.75±1.86	0.0394±0.00155	0.0414±0.00264	0.8768±0.0052
	40	27.70±1.58	0.0358±0.00137	0.0426±0.00246	0.8862±0.0044
	50	26.30±1.28	0.0336±0.00111	0.0462±0.00234	0.8883±0.0039
Max10	30	28.70±1.85	0.0379±0.00166	0.0455±0.00299	0.8816±0.0050
	40	26.30±1.65	0.0360±0.00116	0.0466±0.00268	0.8845±0.0041
	50	26.60±1.50	0.0348±0.00106	0.0463±0.00221	0.8872±0.0037
Max50	30	28.55±1.70	0.0378±0.00148	0.0447±0.00305	0.8753±0.0043
	40	25.45±1.43	0.0353±0.00112	0.0442±0.00242	0.8823±0.0042
	50	24.30±1.04	0.0332±0.00094	0.0458±0.00237	0.8861±0.0038

[†] True value of the parameters used in simulation
[‡] All phenotypes from the offspring were included

4. Discussion

The current study showed that selective phenotyping based on linkage disequilibrium information can decrease the required number of individuals to phenotype compared to criteria based on linkage information within sire families or randomly phenotyped individuals. Using LD criteria could also improve power of detecting QTL with the same number of phenotyped individuals. One outbred population was considered in this study to identify the subset of phenotyped individuals and applied in detection and localization of a QTL. The methods can be easily extended to other experimental designs and different kinds of mating. Several studies have used LA information based on linkage information between the parents and their offspring [11, 12], where most information is provided by highly recombinant individuals to increase genetic dissimilarity in the sample sets. In fact, these methods proposed a set of linkage criteria which were based on prior knowledge of QTL position and maximum recombinations in the phenotyped individuals. Therefore, if little information about genetic architecture was available and the approximate region containing the QTL of interest was assumed to be known, shorter regions can be considered. In this situation MaxRec and MaxUni strategies could be more accurate to find highly dissimilar progeny based on markers around the putative QTL position. Therefore, these criteria could even give better results in confirming and localizing the QTL. In this study we used recombinations in the maternally inherited haplotypes as well as in the paternally inherited (Equation 1). Better estimation in QTL parameters in using MaxRec and/or MaxUniRec can probably be achieved by ignoring recombinations in the maternally inherited haplotypes, since this information is not used to map QTL in the LA analyses. Furthermore, with the LA criteria all markers were considered to find highly recombinant informative individuals. Using markers far away from the QTL position tended to select individuals that were not informative for the mapping. Moreover, in order to increase the power of detection and estimate QTL parameters with linkage information within sire family, especially when QTL effect is low, number of offspring per sire should be increased (unpublished results). However, Jannink [11] showed if high-density marker genotypes are available with large QTL effect to use in LA criterion, accuracy of QTL mapping may not improve by selective phenotyping. By selecting the individuals that have as genetically similar paternal genotype as possible with sire genotypes in each half sib group using MinRec, selected subsets of offspring provided a better contrast in detection than random sampling. MinRec assigned the individuals in two groups with maximum dissimilarity between different marker genotypes. Therefore, this criterion was optimal for detection or verification but yielded poor accuracy of position estimation. When the number of individuals for phenotyping is limited, LA criteria by selecting a small number of individuals in each sire family may be improved by selecting a higher fraction of offspring in fewer half-sib families. However, there was little evidence from this simulation study to suggest a considerable improvement in genome scan mapping when LA criteria were utilized in phenotyping.

LD criteria in this study increased both power of QTL detection and precision of position compared to random SP. These strategies used LD information in a population which showed it can increase accuracy of QTL studies [10]. Using LDLA analysis for QTL mapping indicated that fairly moderate QTL effect in simulations (explaining 6.20% of the phenotypic variance) can be detected and positioned compared to LA analysis. The LD information which is based on the historical ancestral recombinations contributed substantially both to the probability of detecting QTL and precision to estimate QTL position. This was not only with all individuals phenotyped, but also with incomplete phenotype information using LD criteria. Therefore, LD information perspective provided useful insight into considering phenotyping criteria. The results indicated that LD criteria which used non-random association of alleles at different loci in gametes across families can increase the power of detection and accuracy of QTL position compared to random phenotyping.

LA criteria required large family sizes or prior information around QTL position to identify informative progenies for phenotyping and to successfully detect and position QTL. Therefore, an alternative to LA criteria in linkage mapping is to use LD criteria and using both linkage disequilibrium and linkage analysis information simultaneously in QTL experiments [2, 9]. One benefit of LD criteria can be identified by limited number of genotyping and phenotyping the individuals, but without any specific family structure within population [7]. Additionally, using the LD criteria to find informative offspring for phenotyping could increase power to detect QTL. Additionally, low level of phenotypic information has led to overestimate QTL variance component even when selection is based on LD criteria. Therefore to estimate unbiased genetic (QTL and polygenic) and residual variance components using LD criteria, it is required to increase amount of phenotype information (depending on population size). It should be noticed that with increasing population size, the level of phenotyping could be decreased to 30 or even 20% to reach same power of detection using LD criteria (unpublished results of another simulations by authors). In practice LD occurs because mostly animals in the modern populations have inherited the same piece of chromosome from ancestor of breeding nucleuses who lived perhaps many generations ago [10]. Therefore, sampling methods based on LD information which could use these common regions (marker haplotypes) might be more reliable to select the informative individuals for phenotyping purposes.

5. Conclusion

The LD criteria can be used to select the most informative individuals to phenotype when phenotypes are very expensive or difficult to record in comparison to LA criteria and random phenotyping. Continued declines in the cost of genotyping progeny, expensive records and novel opportunities for future use of gene expression data as phenotypes are now opening new challenges in the realm of marker-assisted selection and QTL mapping. LD criteria involved historic recombinations and contributed substantially to the power of detection and precision of the parameter estimates by selecting informative individuals for QTL mapping using LD-based methods. Therefore, if a dense marker map is available in a region harboring QTL to be fine mapped or verified, the LD criteria can be used efficiently with larger difference in power than what was found in this study.

6. Acknowledgements

This research was supported by a grant from Agriculture Research and Education Organization of the Ministry of Jihad-e-Agriculture, Iran. The authors thank P. Sørensen and H. Thomsen for using IBD program.

7. References

[1] Bell P.A., Chaturvedi, Gelfand C.A., Huang C.Y., Kochersperger M., Kopla R., Modica F., Pohl M., Varde S., Zhao R., Zhao X., Boyce-Jacino M.T., SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery, BioTech.(Suppl) 32 (2002) S70-S77.

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

[3] Bodmer W.F., Human genetics: the molecular challenge, Cold Spring Harbor Symp. Quant. Biol. LI (1986) 1–13.

[4] Casu S., Carta A., Elsen J.M., Strategies to optimize QTL detection designs in dairy sheep populations: the example of the Sarda breed, Options Mediterraneennes A55 (2003) 19-24.

[5] Darvasi A., Experimental strategies for the genetic dissection of complex traits in animal models, Nat. Genet. 18 (1998) 19–24.

[6] Darvasi A., Soller M., Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus, Theor. Appl. Genet. 85 (1992) 353 - 359.

[7] Dekkers J.C., Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons, J. Anim. Sci. 82 (2004) 313–328.

[8] Doerge R.W., Mapping and analysis of quantitative trait loci in experimental populations, Nature Reviews Genetics 3 (2002) 43–52.

[9] Fan R., Spinka C., Jin L., Sun Jung J., Pedigree linkage disequilibrium mapping of quantitative trait loci, European J Hum Genet 13 (2005) 216–231.

[10] Hayes B.J., Gjuvsland A.B., Omholt S.W., Power of QTL mapping experiments in commercial Atlantic salmon populations, exploiting linkage and linkage disequilibrium and effect of limited recombination in males, Heredity 97 (2006) 19–26.

[11] Jannink J.L., Selective phenotyping to accurately map quantitative trait loci, Crop Sci. 45 (2005) 901–908.

[12] Jin C., Lan L., Attie A.D., Churchill G.A., Bulutuglo D., Yandell B.S., Selective phenotyping for increased efficiency in genetic mapping studies, Genetics 168 (2004) 2285–2293.

[13] Jobs M., Howell W.M., Stromqvist L., Mayr T., Brookes A.J., DASH-2: Flexible, low-cost, and high-throughput SNP genotyping by dynamic allele-specific hybridization on membrane arrays, Genome Research 13 (2003) 916-924.

[14] Madsen P., Jensen J., A user’s guide to DMU. A package for analysing multivariate mixed models. Version 6. release 4.4. Danish Institute of Agricultural Sciences, Tjele, Denmark (2002).

[15] Medugorac I., Soller M., Selective genotyping with a main trait and a correlated trait, J. Anim. Breed. Genet. 118 (2001) 285–295.

[16] Meuwissen T.H.E., Karlsen A., Lien S., Olsaker I., Goddard ME., Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping, Genetics 161 (2002) 373–379.

[17] Meuwissen T.H.E., Goddard M.E., Fine mapping of quantitative trait loci using linkage disequilibrium with closely linked marker loci, Genetics 155 (2000) 421–430.

[18] Meuwissen T.H.E., Goddard M.E., Prediction of identity by descent probabilities from marker-haplotypes, Genet. Sel. Evol. 33 (2001) 605-634.

[19] Nei M., Li W.H., Linkage disequilibrium in subdivided populations, Genetics 75 (1973) 213-219.

[20] Olsen H.G., Lien S., Svendsen M., Nilsen H., Roseth A., Aasland Opsal M., Meuwissen T.H.E., Fine mapping milk production QTL on BTA6 by combined linkage and linkage disequilibrium analysis, J. Dairy Sci. 87 (2004) 690–698

[21] Piepho H.P., A quick method for computing approximate thresholds for quantitative trait loci detection. Genetics 157 (2001) 425–432.

[22] Searle S.R., Casella G., McCulloch C.E.., Variance Components. New York: John Wiley and Sons, 1992.

[23] Sørensen P., Lund M.S., Guldbrandtsen B., Jensen J., Sorensen D., A comparison of bivariate and univariate QTL mapping in livestock populations, Genet. Sel. Evol. 35 (2003) 605–622

[24] Thomson G., Klitz W., Disequilibrium pattern analysis. I. Theory, Genetics 116 (1987) 623-632.

[25] Wang T., Fernando R.L., van der Beek S., Grossman M., Covariance between relatives for a marked quantitative trait locus, Genet. Sel. Evol. 27 (1995) 251-274.

[26] Xiong M., Guo S.W., Fine-scale genetic mapping based on linkage disequilibrium: theory and applications, American Journal of Human Genetics 60 (1997) 1513–1531.

Chapter 4

Paper III

Across-family marker-assisted selection using selective genotyping strategies in dairy cattle breeding schemes

Saeid Ansari-Mahyari, Anders Christian Sørensen, Mogens Sandø Lund, Peer Berg

Submitted to
Journal of Dairy Science

Across-family marker-assisted selection using selective genotyping strategies in dairy cattle breeding schemes

S. Ansari-Mahyari *†‡, A. C. Sørensen*, M. S. Lund* , P. Berg*

* Department of Genetics and Bioinformatics, Faculty of Agricultural Sciences, University of Aarhus, Denmark

† Agriculture Research and Education Organization, Isfahan Agricultural & Natural Resources Research Center, Isfahan, Iran

‡ Department of Large Animal Sciences, Faculty of Life Sciences, Copenhagen University, Denmark

ABSTRACT: This study investigated the potential loss expected from marker-assisted selection (MAS) when only a proportion of animals are genotyped using several selective genotyping strategies. A population resembling a commercial dairy cattle population were simulated and the potential breeding candidates (young-bulls and bull-dams) were selected across families to find the most informative individuals for genotyping over 25 years. Two strategies were used to identify the most informative animals. The first genotyping strategy was based on selecting individuals for genotyping based on their predicted total genetic effect (sum of the predicted QTL (quantitative trait locus) and polygenic effects) being close to the truncation point for selection. The second strategy used an index that extended the previous strategy to include the variance due to segregation of the QTL in the parents. The two strategies for selective genotyping were applied at two different genotyping levels and compared to random selection of candidates for genotyping and complete genotyping of the potential candidates. The cumulative genetic responses using marker information were significantly greater than the conventional selection (without incorporating QTL effect) until year 18. All selective genotyping strategies at the same proportion of genotyping showed similar cumulative genetic level. The frequency of the favorable QTL allele was increased faster with more animals genotyped. Extra response in total genetic effect (polygenic and QTL) was not significantly different between genotyping all candidates (100%) and 20% and 50% genotyping (except for year 13), but 20% genotyping resulted in significantly higher response than BLUP. With 50% (20%) genotyping of candidates for selection within a population, 95% (89%) of maximum cumulative QTL response was achieved in year 13. All MAS schemes resulted in a 19% reduction in rate of inbreeding compared to the BLUP scheme. Therefore, it is possible to use selective genotyping in practical dairy cattle breeding and decrease the genotyping costs with a minimal loss of response compared to complete genotyping of the potential candidates.

Key words: marker-assisted selection, dairy cattle, quantitative trait locus, selective genotyping

INTRODUCTION

Genetic improvements in quantitative traits during the last decades were achieved by selecting genetically superior parents using phenotypic and pedigree information through best linear unbiased prediction (BLUP). In dairy cattle breeding programs, proven sires are usually selected after progeny testing using evaluation with an animal model (*e.g.* Powell and Norman, 2006). In addition, selection of superior cows as bull–dams to improve the genetic gain is an important part in breeding programs. In a dairy breeding plan, young-bulls potentially can be selected for progeny testing based on information from quantitative trait loci (QTL) (Spelman and Garrick, 1998; Schrooten *et al*., 2005; Schulman and Dentine 2005). The reason to consider detected QTL (a detected region of chromosome that contains specific gene(s) affecting a quantitative trait) is to obtain a more accurate evaluation and increase the frequencies of favorable alleles in the population. Therefore, once a QTL is detected, incorporating molecular information in the breeding schemes with marker-assisted selection can be used to increase the genetic gain over that achievable through BLUP. Several studies have shown the benefits of using MAS in increasing the genetic merit due to higher accuracy of genetic evaluation in outbred populations (Lande and Thompson, 1990; Goddard and Hayes, 2002; Villanueva *et al*., 2005). Moreover, in situations where selection based on BLUP evaluations has several limitations, for example traits that are sex-limited, expressed late in life, costly to record and/or has a low heritability, MAS is expected to increase genetic gain compared to traditional breeding programs (Lande and Thompson, 1990; Meuwissen and Goddard, 1996). Several traits with these characteristics are included in dairy cattle breeding schemes such as fertility traits and disease resistance.

Before considering population-wide linkage disequilibrium (LD), the selection schemes using marker information for dairy cattle were based on within family information. The two main schemes to use linkage information within family in MAS are bottom-up (using daughter design) and top-down (using granddaughter design) approaches (Kashi *et al.*, 1990; Mackinnon and Georges, 1998). Genotyping of the bull-dams in these approaches is not widely common in dairy breeding schemes. Spelman and Garrick (1998) included dam genotypes in these schemes, which increased the amount of genotyping in genetic evaluation. Bottom-up and top-down approaches only consider linkage information within a family. Recently, use of LD information to locate QTL has increased (Pérez-Enciso 2003, Meuwissen and Goddard 2004, Uleberg and Meuwissen 2007). Therefore, the next step after fine-mapping QTL is to use them to increase information in prediction. There are several examples in dairy cattle to use LD markers for preselection of the candidates (*e.g.* Dekkers, 2004).

Fernando and Grossman (1989) initially presented the method to incorporate marker information in BLUP for evaluation of breeding animals resulting in marker-assisted BLUP (MA-BLUP). The cost of complete genotype information in MAS is the most important limitation in a breeding scheme. Genotyping the whole population is not practical for commercial dairy cattle populations. Several investigations in QTL experiments attempted to decrease genotyping costs by identifying the most informative individuals based on phenotypic information (Lander and Botstein, 1989; Darvasi and Soller, 1992), or segregation analysis (Kinghorn, 1999; Macrossan, 2004). Identification and genotyping the candidates in evaluation of young-bulls and bull-dams can improve the accuracy of selection.

Using DNA-information in a population with LD can enhance the accuracy to identify superior candidates. However, this approach would increase the genotyping costs in the MAS schemes. On the other hand, progeny testing in dairy cattle also has a high cost.

Therefore, using MAS to pre-screen potential young bulls and improve accuracy of selecting the best young bulls could increase genetic gain or potentially reduce the cost of progeny testing.

The main objective of this study was to compare strategies for selection the most informative individuals for genotyping in MAS, in a population with LD between markers and QTL. Two genotyping levels (20% and 50%) were considered and compared to genotyping all or no candidates, the latter resulting in selection being based on BLUP. Response to selection at the QTL and polygenes were compared in simulated dairy cattle breeding schemes for a low heritable trait (*e.g.* mastitis).

MATERIALS AND METHODS

A population resembling a commercial dairy cattle population was simulated. Parents were selected based on either BLUP or MA-BLUP predictions. Genotypes of both potential young-bulls and bull-dams were considered in MA-BLUP evaluations. Twenty-five years of selection was simulated and the following results were considered: cumulative genetic responses, annual genetic gain and accuracy of evaluation. In addition, the frequency of the positive QTL allele and rate of inbreeding were compared.

Genetic model

True breeding values (TBV) were simulated assuming a mixed inheritance model with an additive QTL effect. A sex-limited trait, like mastitis resistance, was considered with an initial heritability of 0.04 and the total phenotypic variance was unity. The polygenic effect of each animal (a_i) in the base population was sampled from N(0, σ_a^2), where σ_a^2 is variance of the additive polygenic component, which was 0.03. The polygenic effect of later generations was obtained from $a_i = \frac{1}{2}a_{sire} + \frac{1}{2}a_{dam} + m_i$, where m_i is the Mendelian sampling effect, which was sampled from N(0, ½(1- f) σ_a^2) and f is the average inbreeding coefficient of the sire and dam of i.

It was assumed that one biallelic QTL had been identified in earlier QTL mapping experiments. The variance due to the QTL was σ_{qtl}^2 = 2p × (1− p) × α^2 (Falconer and

Mackay, 1996), where *p* is the frequency of the favorable QTL allele in the base population and α is the gene substitution effect. Initially the frequency of the favorable QTL-allele was 0.10(±0.01), and σ^2_{qtl} = $\frac{1}{4}\sigma^2_G$ = 0.01, where σ^2_G is the variance of the additive genetic component of the trait ($\sigma^2_G = \sigma^2_a + \sigma^2_{qtl}$). Given that the QTL allele frequency changed, the total genetic variance changed over time. The gene substitution effect was constant and equal to 0.236. Four markers placed in an identified 4cM region of a chromosome and positioned at 0, 1.5, 3.5, 4 cM was used. The detected QTL was between the second and the third marker at position 2.5cM. Markers with five alleles were assumed. Marker and QTL genotypes were sampled according to the rules of Mendelian inheritance with recombination.

Population structure

The numbers of selected candidates in the scenarios are shown in Table 1. After simulating the base population with linkage disequilibrium, 40 herds of equal size (100 cows per herd) were generated. The population size was kept constant.

Selection of males: In the BLUP scenario, 200 young-bulls were selected based on the average predicted breeding values (EBV) of the parents and then progeny tested (100 daughters per candidate). As shown in Table 1, the best 10 progeny tested bulls were used as proven bulls. In the MA-BLUP scenarios, first bull calves were pre-selected as candidates for genotyping, according to the average of the parents EBV's. Depending on the selective genotyping strategy (described below), all or a fraction of them were genotyped. After realizing the genotypic information and re-doing the MA-BLUP evaluation, 200 young-bulls were selected based on EBV for progeny testing. The best 10 progeny tested bulls were selected as proven sires.

Selection of females: In the BLUP approach, cows were selected based on EBV's including own performance and offspring information and the best cows were selected as bull dams. In the MA-BLUP approach, potential bull dams were first selected for genotyping based on EBV's. Again, depending on the selective genotyping strategies, all or a fraction of them were genotyped and MA-BLUP evaluation were re-done including the new genotype information. Ultimately, the best cows were selected as bull dams. The order of events in a year of simulation is summarized in Figure 1.

Table 1. Population breeding scheme in BLUP and MA-BLUP evaluation during each year of simulation.

	Selection	Pre-selection		Selection	Age of
Candidates	Criteria	Genotyping	Progeny-testing		Reproduction(year)
Young bulls[†]					
	BLUP	--	200	10	1
	MA-BLUP	400	200	10	1
Bull dams[†]					
	BLUP	--	--	1000	1-5
	MA-BLUP	2000*	--	1000	1-5
Cow dams[††]					
	BLUP	--	--	3000	1-5
	MA-BLUP	--	--	3000	1-5

[†] The candidates selected across herds.

[††] The candidates selected within herds.

* This was the maximum number for genotyping in bull dams.

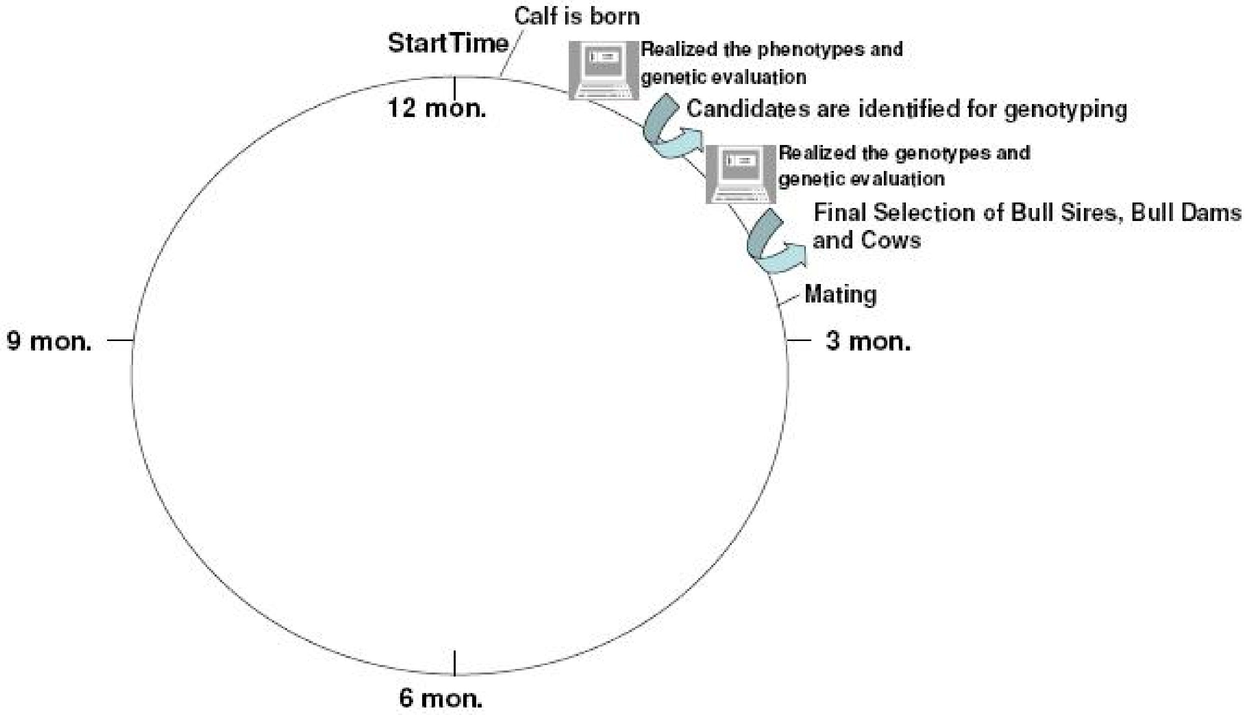
Calf is born
StartTime
Realized the phenotypes and genetic evaluation
12 mon.
Candidates are identified for genotyping
Realized the genotypes and genetic evaluation
Final Selection of Bull Sires, Bull Dams and Cows
Mating
9 mon.
3 mon.
6 mon.

Figure 1. Breeding scheme in each year based on MA-BLUP evaluation. In this approach genetic evaluations was performed twice a year, which was before genotyping the candidates and then after genotyping. In BLUP evaluation, the candidates were evaluated once a year.

Genotyping strategies

Several selective genotyping strategies were used under the MA-BLUP approach. In the first year of genotyping in this experiment (year 4), it was assumed that all founder sires which were selected for progeny testing had already been genotyped. Additionally from the fourth year onward, the sires of animals in evaluation were assumed to have marker information. The objective of the selective strategies was to maximize genetic gain given a reduced number of genotyped animals. Each strategy was replicated 50 times, from 50 different base populations. The same fifty base populations were used for all strategies. The criteria in the current study were:

1. Complete genotyping (ALL): All young-bulls and bull dams that can be potentially selected as candidates each year were genotyped. These individuals were identified based on an MA-BLUP evaluation.

2. Random genotyping (RAN): Based on random sampling, a proportion (20 and 50%) of pre-selected candidates was genotyped.

3. Average EBV's (AVE): A fraction of the candidates (young-bulls and bull-dams) were selected for genotyping based on the estimated effects of QTL plus polygenic across the families. The animals chosen for genotyping were those closest to the truncation point for selection. At this point, the candidates were selected for progeny testing from potential young-bull candidates and for use as bull-dams from potential bull-dam candidates for the next generation. Therefore, this strategy aimed to genotype the animals, which potentially could change from below to above the threshold or the opposite after redoing the genetic evaluations.

After an MA-BLUP evaluation, the overall haplotype effect computed in all pre-selected candidates for genotyping by: $\hat{Q}_{off} = \tfrac{1}{2}(\hat{Q}_{S1} + \hat{Q}_{S2}) + \tfrac{1}{2}(\hat{Q}_{D1} + \hat{Q}_{D2})$,

where $\hat{Q}_{off}$ is the total predicted QTL effect in the offspring, and $\hat{Q}_{S1}$ and $\hat{Q}_{S2}$ ($\hat{Q}_{D1}$ and $\hat{Q}_{D2}$) are the estimated QTL effects in the sire (dam). If one or both of the parents were not included in the evaluation, the average haplotype effects of the grand parents were used or in case they too were not in the evaluation, the expected effects of the parental haplotypes were computed by:

$$\hat{Q}_{parent} = 2 \times ((P_Q \times \alpha_Q) + (P_q \times \alpha_q)),$$

where $\hat{Q}_{parent}$ is the sum of the parental haplotype effects, P_Q (P_q) is allele frequency of positive (non-positive) QTL-allele in the birth year of the non-genotyped parent, and α_Q (α_q) is gene substitution effect of the positive (non-positive) QTL-allele. In this case, $\hat{Q}_{parent}$ was used for the sum of the haplotype effects if the parent was not included in the evaluation. The total genetic value in the offspring was computed as: $\hat{A}_{off} = \hat{Q}_{off} + \frac{1}{2}(\hat{a}_D + \hat{a}_S)$. Given the truncation point (*t*) among pre-selected animals, an index for each candidate was calculated as:

$$Index = \left| t - \hat{A}_{off} \right|$$

A proportion (20 and 50%) of candidates with the smallest index values was genotyped.

4. Combined average-EBV's and variance-QTL (COMBINE): While the previous strategy used the sum of estimated QTL and polygenic effects, this strategy used both the mean of the total genetic effects and the segregation variance of the detected QTL. The rationale for this strategy followed from the assumption that Mendelian sampling variance of the potential candidates depend on the parent's genotypes, which therefore affect the probability that genotyping may cause the total genotypic effect to cross the selection threshold. The expected breeding values ($\hat{A}_{off}$) was computed as in the previous strategy. The Mendelian sampling variance in the pre-selected candidates was calculated as follows. Each sire (dam) has two estimated haplotype effects as $\hat{Q}_{S1}$ and $\hat{Q}_{S2}$ ($\hat{Q}_{D1}$ and $\hat{Q}_{D2}$). There is four possible combinations of the QTL effects in offspring with equal probability as below:

$$\hat{Q}_1 = \hat{Q}_{S1} + \hat{Q}_{D1} \qquad \hat{Q}_2 = \hat{Q}_{S1} + \hat{Q}_{D2} \qquad \hat{Q}_3 = \hat{Q}_{S2} + \hat{Q}_{D1} \qquad \hat{Q}_4 = \hat{Q}_{S2} + \hat{Q}_{D2}$$

The Mendelian sampling variance was calculated for all pre-selected candidates by:

$$MenVar_{\hat{Q}_{off}} = \frac{1}{4}\left[\sum_{i=1}^{4} (\hat{Q}_i - \hat{Q}_{off})^2\right],$$

where $\hat{Q}_{off}$ is the total predicted QTL effect in the offspring. Again, if one of the parents or both were not genotyped, the mean of the haplotype effects was computed as previously described. However, in the situation that both parents were not evaluated, the average haplotype effects of the grand parents were used. If they were not in the evaluation, the effects in each parent were computed as $Q_1 = ½Q_{parent} + ½\delta$ for the first haplotype effect and $Q_2 = ½Q_{parent} - ½\delta$ for the second haplotype. The δ was computed as the mean of absolute differences of the estimated haplotype effects ($\hat{Q}_{1,i}$ and $\hat{Q}_{2,i}$) in all animals (N) which were included in MA-BLUP evaluation by: $\delta = \frac{\sum_{i=1}^{N} \left|\hat{Q}_{1,i} - \hat{Q}_{2,i}\right|}{N}$. The δ was used to compute more accurate Mendelian segregation variance for un-genotyped parents without affecting $\hat{Q}_{off}$. Finally, an index was calculated in order to identify candidates for genotyping as: $Index = \left|\Phi - 0.5\right|$, where $\Phi\,(\hat{A}_{off}$, $MenVar_{\hat{Q}_{off}}$, t) is the cumulative normal distribution function with mean $\hat{A}_{off}$ and variance $MenVar_{\hat{Q}_{off}}$, which is evaluated at the truncation point (t). Therefore, Φ gives the probability of not being selected. Either 20% or 50% of individuals with the lowest value of the index were genotyped. These were the animals most likely to move from one side of the truncation point to the other side when genotyped and reevaluated.

Breeding value evaluation

The MA-BLUP evaluation was used for selection based on phenotypic, pedigree, and marker information. For progeny tested bulls, daughter yield deviations (DYD) were sampled directly. Twice the DYD was used as phenotype of the sires in genetic evaluations. In all genotyping strategies during the first three years, animals were selected using BLUP evaluations without using genotypic information of the parents.

The MA-BLUP evaluation was based on the following mixed model:

$$\boldsymbol{y} = \mathrm{X}\boldsymbol{\beta} + \mathrm{Z}\boldsymbol{a} + \mathrm{W}\boldsymbol{q} + \boldsymbol{e}\,,$$

where $\boldsymbol{y}$ = vector of records for cows or double DYD for progeny tested bulls, X = incidence matrix relating phenotypes to fixed effects, $\boldsymbol{\beta}$ = vector of fixed Herd-Year-Season (HYS) effects, Z = incidence matrix relating phenotypes to animals, $\boldsymbol{a}$ = vector of additive polygenic effects for each animal, W = incidence matrix relating phenotypes to random QTL effects, $\boldsymbol{q}$ = vector of random additive effects due to the QTL (haplotype effect) and $\boldsymbol{e}$

= vector of residuals. The random vectors are assumed to be normally distributed and mutually independent. Specifically, $\boldsymbol{a}$ is N(0, $\mathbf{A}\sigma^2_u$), $\boldsymbol{q}$ is N(0, $\mathbf{H}_{\boldsymbol{p}}\sigma^2_q$), and $\mathbf{e}$ is N(0, $\mathbf{R}\sigma^2_e$),where $\mathbf{A}$ is the additive relationship matrix, $\mathbf{H}_{\boldsymbol{p}}$ is the identity by descent (IBD) matrix that contains the IBD probabilities for the QTL position *p* (2.5 cM in this study), and $\mathbf{R}$ is a diagonal matrix with the element of 1 for the cows and $\frac{1}{weight}$ for twice the DYD. The weight was calculated as the inverse of the variance of twice the DYD. Additionally, the variance component used for prediction of QTL effects was recalculated every year based on frequency of the QTL alleles.

For the BLUP evaluation, the QTL effect was excluded from the model and the evaluation was based on phenotypic and pedigree information. The total estimated breeding values for BLUP and MA-BLUP evaluations were $\hat{a}$ and $\hat{a}+\hat{q}$, respectively. Data used in genetic evaluation was constrained to the last 15 years prior to the date of evaluation.

The results of genetic responses, QTL frequency and rate of inbreeding from the strategies were tested using standard analysis of variance with genotyping strategy method and replicate as two fixed independent variables. The replicate was included in the model as the same 50 base populations were used.

RESULTS

Figure 2 shows that the cumulative genetic response due to the QTL in MAS schemes with two levels of genotyping (20% and 50%) was higher than BLUP . In year 7, the gain was slightly higher than ALL strategy. However, the cumulative genetic responses of three genotyping levels in years 5 to 8 were not significantly different (p<0.05). The responses were significantly higher in ALL strategy compared to 50% genotyping in years of 12 and 13 (Figure 2). At 20% genotyping, the proportional cumulative response was not significantly different from ALL strategy in year 18 and onward. Selection based on BLUP showed significantly lower QTL response until year 23 compared to all MAS schemes in this study. In the BLUP scheme, the genetic gain due to QTL was higher from year 14 and onward, since variance of QTL was higher in BLUP scheme compared to MAS schemes due to slower fixation of the QTL. Until year 20, the difference in cumulative genetic response between MAS and BLUP selection was significant (*p*<0.05). The difference was not significant only in year 24 and 25. Breeding values due to the QTL effects were accurately estimated in this study as the QTL variance was updated based on actual allele frequency at the QTL.

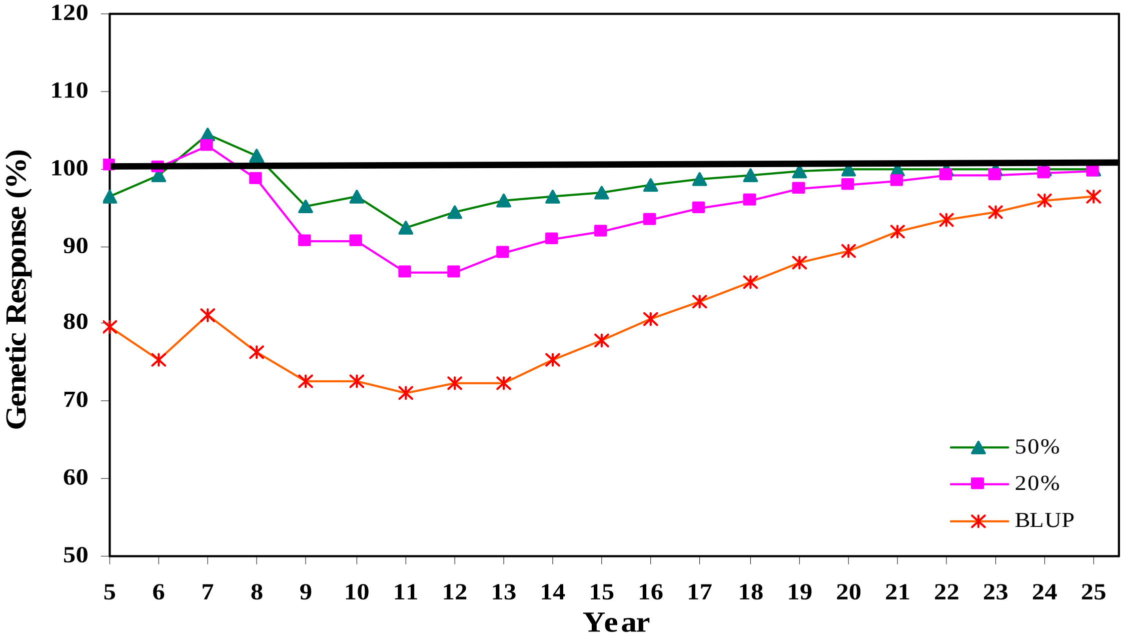

Figure 2. Percentage cumulative genetic response of detected QTL in comparison to "ALL" strategy (the straight line), using BLUP and MA-BLUP with two levels of genotyping (50% and 20% of the potential candidates) over selection.

As expected, frequencies of the favorable QTL allele with both BLUP and MA-BLUP selection schemes increased over time but faster with MAS (Figure 3a). All three selective genotyping strategies in this study (AVE, COMBINE and RAN) behaved similarly at each genotyping level and the genetic response was not significantly different between the strategies at any time (Figure 3b). However, the genetic gain of a detected QTL was faster at higher genotyping levels as expected. The maximum difference in QTL allele frequency between genotyping all potential candidates and 50% genotyping of the candidates was 7.6% at year 11. The scenario 'ALL' and the three selective genotyping strategies at 50%

genotyping showed frequencies close to fixation (>0.90) in year 14 and 15, respectively (Figure 3b).

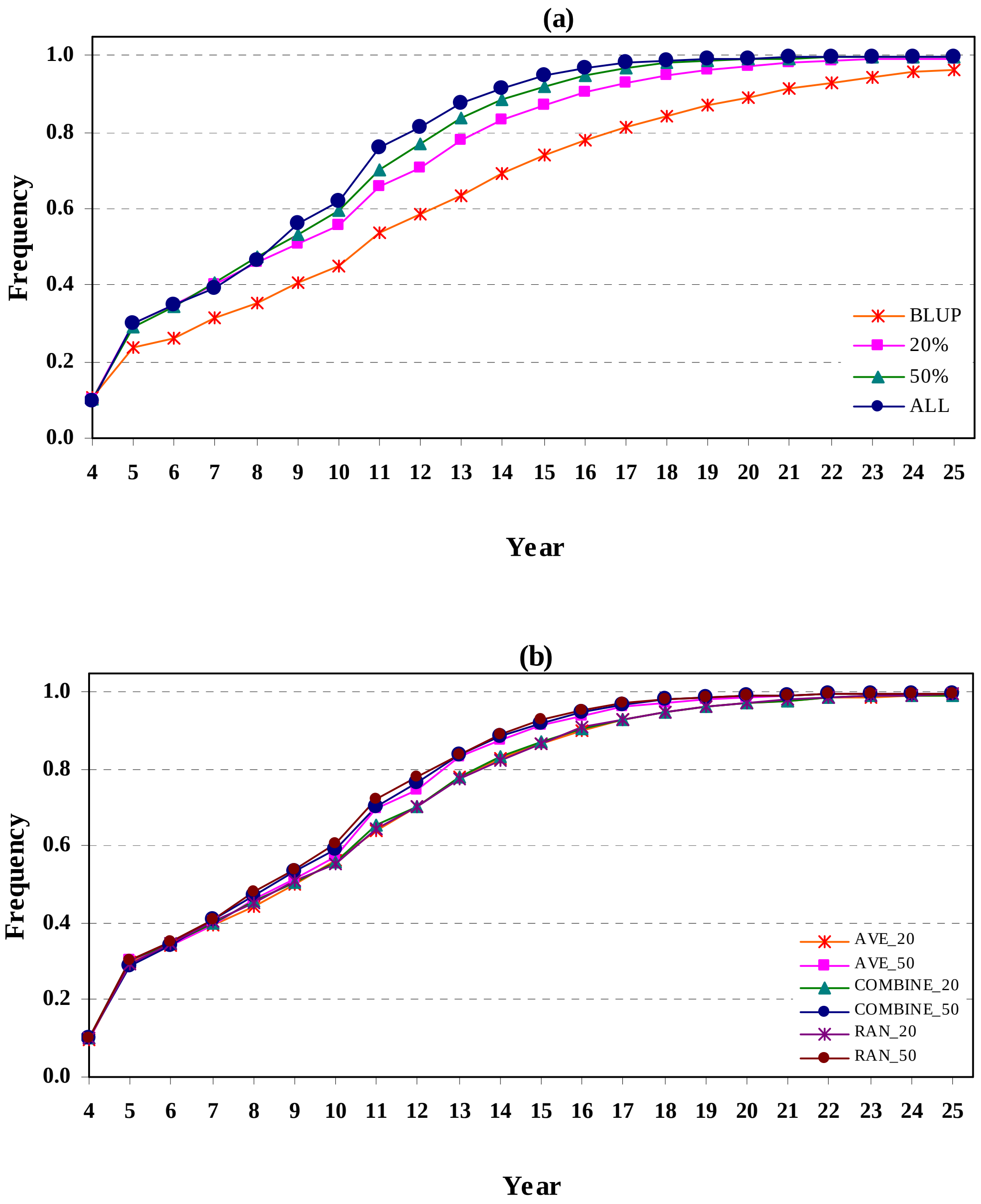

Figure 3. Frequency of the positive QTL allele: a) BLUP and MA-BLUP with three levels of genotyping (all, 50% and 20% of the potential candidates each year of selection), b) genotyping of the potential candidates under different selective genotyping criteria (AVE, COMBINE and RAN) using two levels of genotyping (50% and 20%).

The cumulative values of total genetic response and polygenic response relative to ALL strategy are shown in Figure 4. The total genetic responses (Figure 4a) using marker information with genotyping levels of 20% and 50% were higher in years 6 to 8 than ALL strategy, which was due to extra polygenic response compared to ALL strategy (Figure 4b). Selection with traditional BLUP until year 20 (Figure 4a) showed lower total genetic response than all MAS schemes. Total genetic response with 20% genotyping was not significantly different with 50% genotyping in year 18 and onward. All selective genotyping strategies showed similar total genetic responses with the same genotyping level. The BLUP procedure showed higher annual genetic response than MA-BLUP strategies from year 14 due to the QTL still not being fixed (results not shown). The maximum difference in cumulated total genetic response of MA-BLUP strategies over BLUP was achieved in year 13. Additionally, for the genotyping levels (20% and 50%) the annual genetic gain in most years were larger than for the BLUP scenario until year 20. However, there were only slight differences (and not significant) between the genotyping scenarios at each genotyping level (results not shown). The cumulative polygenic response using BLUP evaluations in comparison to ALL strategy were significantly higher than MA-BLUP schemes until year 18 (Figure 4b). The annual polygenic response (results not shown) in BLUP was lower than MA-BLUP from year 15, when the QTL was close to fixation with MA-BLUP, until the QTL was close to be fixed with BLUP in year 24 (Figures 3a and 4b). All selective genotyping strategies in different levels of genotyping led to very similar polygenic response (partly shown in Figure 4b).

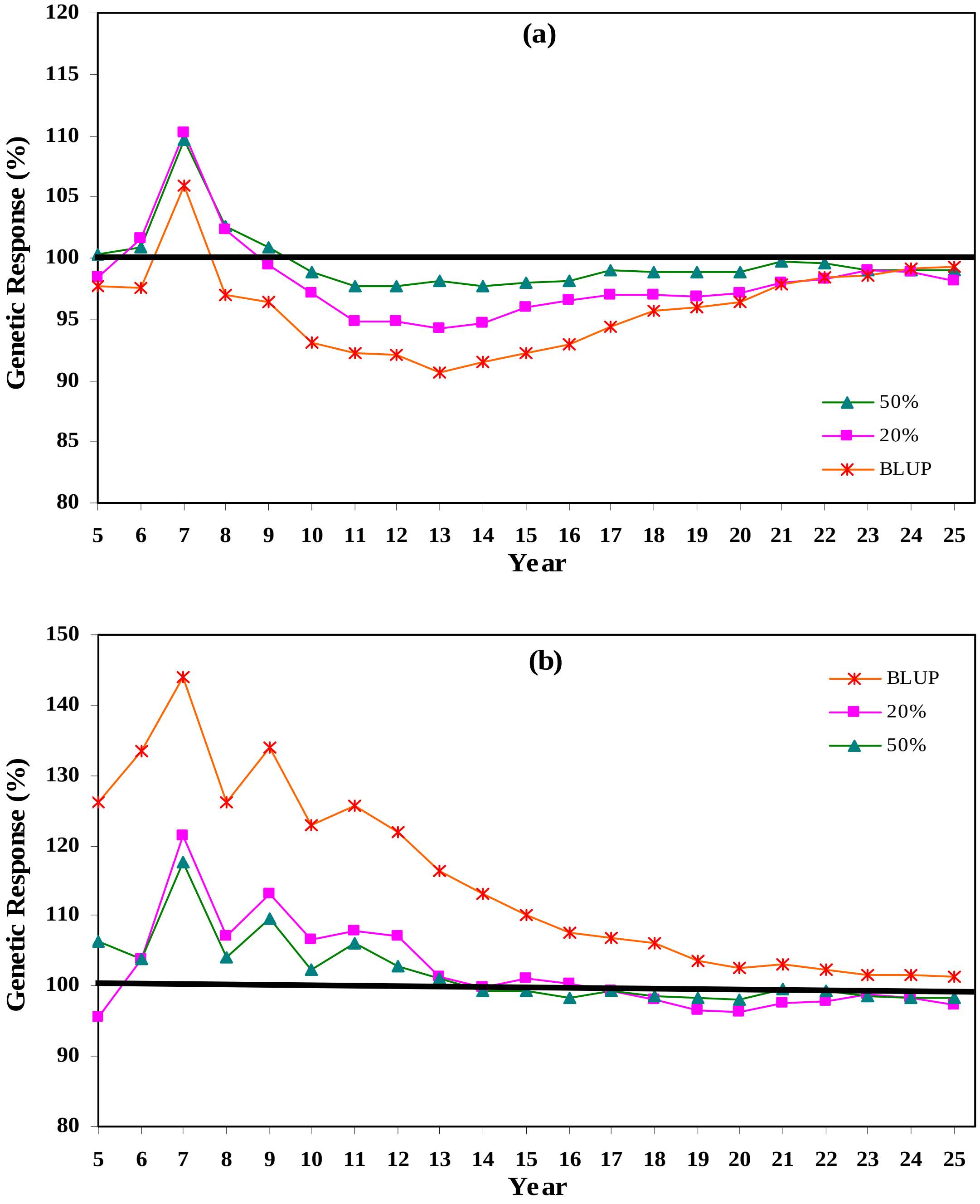

Figure 4. Percentage cumulative genetic response in comparison to "ALL" strategy (the straight line), using BLUP and MA-BLUP with two levels of genotyping (50% and 20% of the potential candidates) over selection:

a) total genetic response (polygenic + QTL), b) polygenic response.

Two genotyping levels in the potential candidates (young-bulls and bull-dams) were studied to compare the genetic gain achieved relative to the maximum obtainable gain from genotyping all candidates. As explained in Materials and Methods above, in each year of selection regardless of the genotyping criteria, the number of genotyped candidates was based on the genotyping levels (20, 50 or 100%) and only non-genotyped animals were candidates for genotyping. Therefore, in the ALL strategy the numbers of genotyped candidates as potential bull-dams were mostly less than what was planned as some were already genotyped in a previous year. Figure 5 shows the percentage of cumulative numbers of genotyped animals (males, females and both) in genotyping levels of 20% and 50% of the potential candidates relative to ALL strategy (which is complete genotyping of potential candidates). In first year of MAS (year 4), the number of genotyped males was higher because of genotyping more eligible sires for progeny testing from previous years. On the other hand, the proportion of genotyping in females showed that relative to the ALL strategy, more than 20% and 50%, respectively of the candidates needed to be genotyped until year 8. This was due to less genotyping than planned in ALL strategy (Table 1). Consequently, on average in this study, the actual genotyping proportions relative to ALL strategy was 32% to 35% for 20% genotyping level and 72% to 75% for 50% genotyping level.

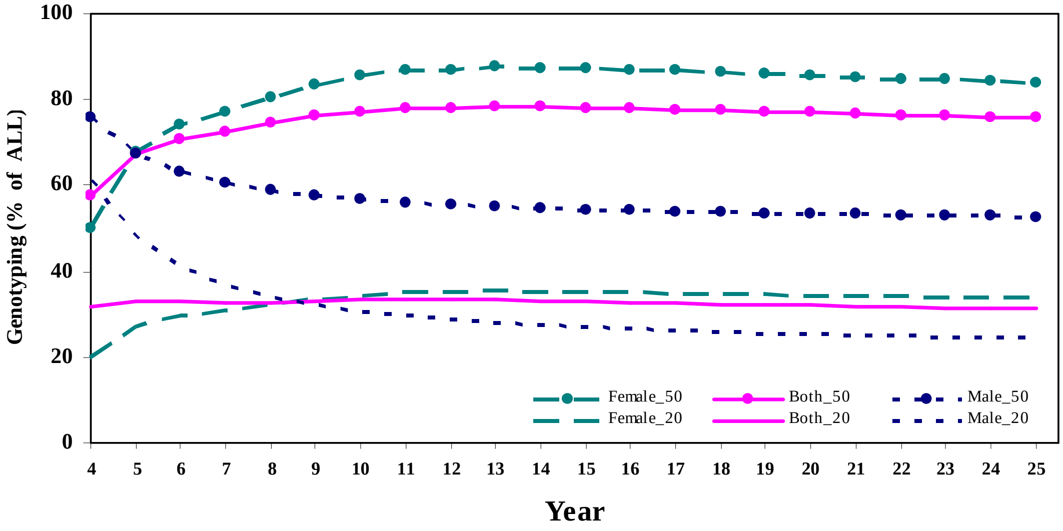

Figure 5. Percentage genotyped animals relative to the “ALL” strategy at two levels of genotyping (20 and 50%) versus the year of selection.

The inbreeding coefficients until year 7 were less than 0.006 for all selection schemes and not significantly different until year 10 (Figure 6). The rate of inbreeding (Δ*f*) was 0.72 percent/year in BLUP, while the MA-BLUP strategies (with 100% genotyping) resulted in a rate of inbreeding of 0.58 percent/year. Genotyping at 20% of the potential candidates showed slightly higher rate of inbreeding than 50% and 100% genotyping during the years of 11-19 (Figure 6). After fixation of the QTL (Figure 3), the rate of inbreeding with MA-BLUP increased to a value similar to the rate of inbreeding with BLUP.

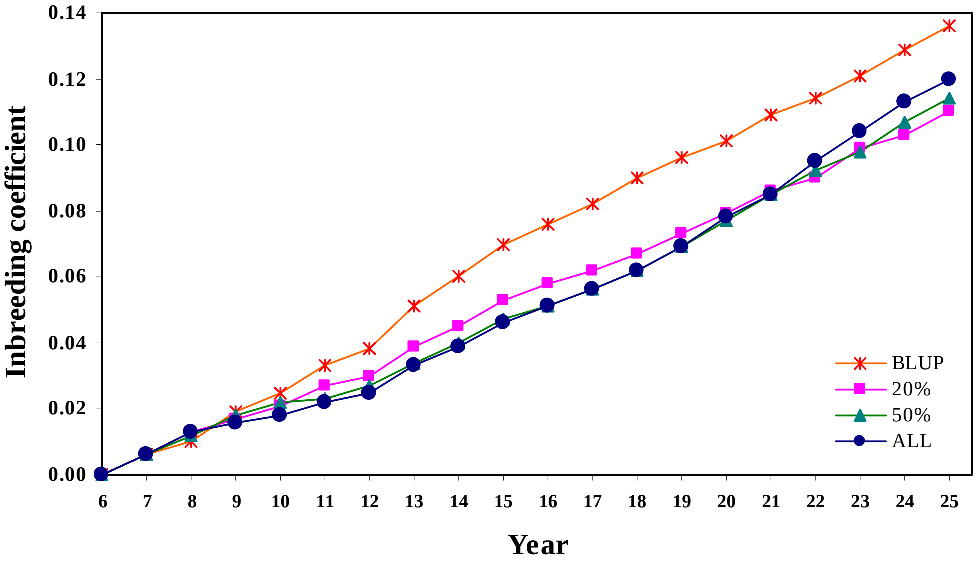

Figure 6. Inbreeding coefficient versus birth year resulting from BLUP and MAS using three levels of genotyping (ALL, 50% and 20% of the potential candidates).

DISCUSSION

In this study, several selective genotyping strategies were used to genotype the animals contributing most information when using MAS, in a population with population-wide LD between markers and QTL. The cumulative response of total genetic response using marker information was initially greater than the conventional selection until year 21. Allele frequency surpassed 0.9 in year 16 when 20% of the candidates were genotyped and in year

21 for BLUP selection (Figure 3.a). Higher initial response from MAS is mostly due to strong emphasis on marked genetic variance (Spelman and Garrick, 1997). However, in practical breeding schemes, more than one trait and/or QTL is used and therefore a slower fixation of the positive QTL allele will be expected. An increase in the accuracy of the estimated breeding values of QTL could be obtained using marker haplotypes with high-density genotyping (Meuwissen *et al.*, 2001). The markers in this study were assumed to be close to the QTL and in LD with the QTL. Additionally, the developed genotyping criteria used LD information to find the most informative animals for genotyping across the families. Using dense markers in LD with a QTL (as a haplotype), such that the marker–QTL associations persist across families in the population, makes the implementation of MAS much simpler and linkage phase for each sire family does not have to be established (Hayes and Goddard, 2003).

Three selective genotyping criteria were used to choose animals for genotyping from the potential candidates each year of selection. The AVE and COMBINE strategies sampled animals with the highest uncertainty on whether to be selected or not. In RAN strategy, the animals were randomly sampled from the pre-selected group. All proposed genotyping strategies showed similar cumulative genetic level each year, at similar proportion of genotyping (20% and 50%). The computational time demanding was not different between the genotyping strategies using the same genotyping level, and therefore any of them can be used in a MAS scheme based on selective genotyping program.

Cumulative polygenic response using BLUP evaluations were higher than MA-BLUP schemes. However, extra polygenic response in BLUP was considerably decreased compared to MA-BLUP from year 13, when the QTL was close to fixation with MA-BLUP selection. Other studies showed lower polygenic response to selection based on pedigree and marker information compared with using only pedigree information (*e.g.* Spelman and Garrick, 1997). Contrary to this, Sonesson (2007) used within-family MAS and showed that polygenic genetic gain was not significantly different when using marker information.

The base populations used in this study were simulated in LD. In the first years of selection based on marker information, the frequency of the favorable QTL allele was low (~0.10). More weight on the QTL during the first years of selection relative to the polygenic component affected the prediction and therefore, this could result in lower polygenic gain in MAS schemes. Ultimately, total gain was higher with MA-BLUP selection compared to BLUP in genetic evaluation.

Increased gain in MAS was due to more information in genetic evaluations compared to BLUP (*e.g.* Villanueva *et al.*, 2005), which results in increased accuracy of predicted breeding values. The accuracy of predicted QTL effects increased during the first years of selection since the increasing number of genotyped offspring gave more information per haplotype. Accuracies achieved when 50% of the potential candidates were genotyped were very close to the accuracies in the ALL strategy, while accuracies were lower with 20% of candidates genotyped. In both BLUP and MA-BLUP scenarios the accuracy decreased over time, because of decreasing genetic variance due to the Bulmer effect (Bulmer, 1971), inbreeding, and reduced heterozygosity at the QTL (after years 7-10).

Previous simulation studies investigated the benefit of using markers linked to known genes in dairy cattle breeding schemes (Meuwissen and van Arendonk, 1992; Stella *et al.*, 2002; Schrooten *et al.*, 2005), and showed higher genetic response. The progress was reported to increase with 5% to 64% compared to BLUP selection, depending on the trait being selected and the genetic models (Meuwissen and Goddard, 1996; Hayes and Goddard, 2003). The MAS studies assumed complete genotype information but genotyping the whole population or even only the potential candidates is expensive and not a practical approach for dairy cattle breeding (Dekkers and Van der Werf, 2007). For reasons mainly related to the genotyping costs, dairy cattle breeding programs, which are based on MAS, should identify and genotype the most informative breeding animals in multistage breeding programs. This simulation study attempted to clarify the usefulness of a proper genotyping criterion in a real situation of a dairy cattle breeding scheme.

Marshall *et al.* (2002) used several genotyping strategies and assumed an advanced breeding scheme (*e.g.* closed nucleus system) with complete information on pedigree and phenotypes. Based on the predicted marker genotypes with reasonable certainty (Kinghorn, 1999), Marshall *et al.* (2002) used an index based on genotype information with a single linked marker and EBV's of two traits to select a fraction of the population for genotyping. They indicated that the gain at a detected QTL with selective genotyping was close to the gain with complete genotyping. However, in more extensive breeding schemes .*e.g.* sheep and beef, there is incomplete recording in the pedigree. In addition, this approach becomes more complicated to compute the genotype probabilities when several markers or complicated pedigrees are considered. The method developed by Kinghorn (1999) to find the more informative individuals for genotyping does not take into account the phenotypic information and only uses the results from segregation analysis.

In dairy cattle breeding, it is necessary to develop more flexible selective genotyping methods for use in existing, outbred populations. This study considered the QTL as a random effect in the context of the mixed model terminology (Fernando and Grossman, 1989). The results showed that the selective genotyping approaches were useful to replace complete genotyping, with a minimal loss in genetic progress. In addition, the considered methods were quite simple and used the results from the previous MAS evaluation. With 20% genotyping, total cumulative genetic level per year was not significantly different from 50% genotyping, but 20% genotyping was significantly lower than complete genotyping in years 11-16. With 50% (20%) genotyping level, 95% (89%) of the maximum response at the QTL was observed in year 13.

The MAS schemes in dairy cattle in France and Germany, which are based on within family selection, have been described in the literature (Weller, 2007). These schemes use linkage information and need to genotype either all young-bulls and bull-dams or all sires and their daughters in the German and French breeding plans, respectively. The use of LD to locate genes which affect quantitative traits has increased recently (Meuwissen and Goddard, 2004; Uleberg and Meuwissen, 2007). The classical linkage and LD analysis are complementary and combined methods could simultaneously use linkage and LD information. Therefore, in order to increase the accuracy of the selection based on fine-mapped QTL and decrease genotyping costs in breeding programs, it is necessary to develop MAS across families in a population using selective genotyping methods. This study indicated that one method to decrease genotyping costs in a MAS scheme across families was genotyping the candidates close to the truncation point, when QTL effect estimation was based on both LD and linkage analysis information. Therefore, the markers those are close enough to the causative mutation and have consistent associations across families can be used in animal breeding programs (Dekkers, 2004). In addition, selection across families does not need any specific family structure (Dekkers, 2004). However, the benefits of implementing marker-linked QTL, using selective genotyping or in general for MAS, depend on the net result of extra costs, savings and amount of genetic progress of the chosen breeding program (Schrooten *et al*., 2005).

Another benefit of MAS schemes over BLUP in this study was a reduction in the rate of inbreeding, as using additional marker information allows selection of high-ranking animals within families rather than any restriction to choose only progeny from the highest-ranking families. In a practical breeding scheme, restricting Δ*f* can be achieved through selecting parents from more families *e.g.* optimum contribution selection for MAS (Villanueva *et al.,* 2002). Potentially the attempts to restrict Δ*f* in MAS is expected to increase genetic progress as has been shown for BLUP selection (Villanueva *et al.,* 2002).

CONCLUSIONS

Selective genotyping can be implemented by using any of the criteria in this study as a pre-selection step when selecting either young-bulls for progeny testing or bull-dams. Marker-assisted selection schemes using LD markers increased genetic response due to a detected QTL and also total genetic gain and decreased inbreeding rate compared to traditional BLUP. By genotyping the more informative breeding animals a reduction in genotyping costs can be achieved with a minimal loss of response compared to complete genotyping and low rate of inbreeding compared to traditional BLUP.

ACKNOWLEDGMENTS

This research was financed by a grant from Agriculture Research and Education Organization of Iranian Ministry of Jahad-e-Agriculture and “DNA based selection to improve disease resistance, fertility, calf survival and production in Danish dairy cattle” from Danish Directorate for Food, Fisheries and Agri Business j.no. 3401-04-00853. Authors thank Peter Sørensen and Hauke Thomsen for using IBD program.

REFERENCES

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

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.

Dekkers, J. C. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J. Anim. Sci. 82: (E-Suppl): E313–328.

Dekkers, J. C. M., and J. H. J. Van der Werf. 2007. Strategies, limitations, and opportunities for marker-assisted selection in livestock. *In*: Marker-Assisted Selection: Current Status and Future Perspectives in Crops, Livestock, Forestry and Fish. Edited by E. P. Guimarães, J. Ruane, B. D. Scherf, A. Sonnino, and J. D. Dargie. FAO, Rome, Italy.

Falconer, D.S., and T. F. C. Mackay. 1996. Introduction to Quantitative Genetics. 4th ed. Longman Group, Essex.

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

Goddard, M. E., and B. J. Hayes. 2002. Optimisation of response using molecular data. CD-ROM Communication No. 22-01 in Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France.

Hayes, B., and M. E. Goddard. 2003. Evaluation of MAS in pig enterprises. Livest. Prod. Sci. 81:197-211.

Kashi, Y., E. Hallerman, and M. Soller. 1990. Marker assisted selection of candidate bulls for progeny testing programs. Anim. Prod. 52:21-31.

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

Lande, R., and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124: 743–756.

Lander, E., and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199.

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

Macrossan, P. E. 2004. Strategies to minimise DNA testing costs for research and development programs involving pedigreed populations. PhD Thesis, University of New England, Australia.

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

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

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

Meuwissen, T. H. E., and M. E. Goddard. 2004. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multi-trait data. Genet. Sel. Evol. 36: 261-279.

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

Pérez-Enciso, M. 2003. Fine mapping of complex trait genes combining pedigree and linkage disequilibrium information: a Bayesian unified framework. Genetics 163**:**1497–15

Powell, R.L., and H. D. Norman. 2006. Major advances in genetic evaluation techniques. J. Dairy Sci. 89:1337-1348.

Schrooten, C., H. Bovenhuis, J. A. M. van Arendonk, and P. Bijma. 2005. Genetic progress in multistage dairy cattle breeding schemes using genetic markers. J. Dairy Sci. 88:1569-1581.

Schulman, N. F., and M. R. Dentine. 2005. Linkage disequilibrium and selection response in two-stage marker-assisted selection of dairy cattle over several generations. J. Anim. Breed. Genet. 122:110–116.

Sonesson, A. K. 2007. Within-family marker-assisted selection for aquaculture species. Genet. Sel. Evol. 39: 301-317.

Spelman, R. J., and D. J. Garrick. 1997. Utilisation of marker assisted selection in a commercial dairy cow population. Livest. Prod. Sci. 47:139–147.

Spelman, R. J., and D. J. Garrick. 1998. Genetic and economic responses for within-family marker-assisted selection in dairy cattle breeding schemes. J. Dairy Sci. 81: 2942–2950.

Stella, A., M. M. Lohuis, G. Pagnacco, and G. B. Jansen. 2002. Strategies for continual application of marker-assisted selection in an Open Nucleus Population. J. Dairy Sci. 85:2358–2367.

Uleberg, E., and T. H. E. Meuwissen. 2007. Fine mapping of multiple QTL using combined linkage and linkage disequilibrium mapping - A comparison of single QTL and multi QTL methods. Genet. Sel. Evol. 39:285-299.

Villanueva B., R. Pong-Wong, and J. A. Woolliams. 2002. Marker assisted selection with optimised contributions of the candidates to selection. Genet. Sel. Evol. 34: 679–703.

Villanueva, B., R. Pong-Wong, J. Fernández, and M. A. Toro. 2005. Benefits from marker-assisted selection under an additive polygenic genetic model. J. Anim. Sci. 83:1747-1752.

Weller, J. L. 2007. Marker-assisted selection in dairy cattle. *In*: Marker-Assisted Selection: Current Status and Future Perspectives in Crops, Livestock, Forestry and Fish. Edited by E. P. Guimarães, J. Ruane, B. D. Scherf, A. Sonnino, and J. D. Dargie. FAO, Rome, Italy.

DISCUSSION AND CONCLUSION

Detection and mapping of the QTL is a statistical inference of the alleles at the genome and their effects on the phenotypes in a population. Because quantitative traits are governed by more than one gene, studying the genotypes related to a trait gives hope of understanding the relationship between genotypes and phenotypes. Selective genotyping or phenotyping is a strategy to find a fraction of the population which is most informative and cost-effective to genotype or phenotype instead of the whole population. There are different approaches that allow selecting a proportion of the population for genotyping or phenotyping in QTL mapping and marker-assisted selection. These approaches could be improved by using all information. Additionally, similar situations arise in association studies based on linkage disequilibrium in QTL mapping and MAS. In this thesis, different sampling criteria are addressed to decrease the costs of genotyping or phenotyping.

In chapter 2 and 3, the consequences of sampling the informative animals in QTL experiments using daughter design were investigated. Simulation results established that selective genotyping with combining phenotypic and genotypic information to detect marker-QTL linkage for QTL which requires larger mapping population can decrease sampling variance of the QTL variance components. The criteria which used LD information in selective phenotyping could decrease the number of phenotype records until 25% with the same power of detection.

The results demonstrate potential consequences of increasing the sensitivity of QTL detection with selective criteria by identify the most informative individuals for genotyping based on accessible information (from paper-I). Selecting individuals for genotyping was firstly introduced by Lander and Botstein (1989) and then Darvasi and Soller (1992) used selectively genotyped animals (only the extremes for the recorded trait) for a single marker linked to a QTL. The focused investigations on the selection of individuals for genotyping have concluded that it is not necessary to genotype the whole population and high power in QTL detection could be achieved with genotyping the most informative individuals (Darvasi and Soller, 1992; Kinghorn, 1999). Instead of the current genotyping methods which are using only phenotypic or genotypic information, the combined methods presented in this thesis increased the power and closer QTL parameter estimations to the simulated values (with genotyping ~20% of the population using combined methods compared to 30% in extreme selective genotyping). Therefore, in a larger QTL mapping study, it seems important to use a selective genotyping method with including both genotype and phenotype data to decrease genotyping costs, with unbiased QTL parameters. It should be noticed that combined use of genotype and phenotype information does come at a cost of requiring intensive computations, depending on the size of population. In this situation, the extreme phenotypes within sire groups can be selected for genotyping as an alternative approach.

In chapter 3, several criteria are presented to identify and phenotype the most informative candidates in comparison to the selective genotyping approaches. The criteria based on LD information were more powerful to select the most informative individuals for phenotyping than the criteria based on linkage analysis (LA) information and also random phenotyping. The methodology was shown to substantially improve the power of QTL analyses in a half-sib design, especially regarding the LD criteria. In addition, by increasing the phenotyping level in LD strategies (30% to 50%), the estimated QTL effect approaches the simulated true values quicker than LA strategies. Other studies have used LA information based on linkage information between the parents and their offspring [Jin *et al.*, 2004; Jannink, 2005], where most information is provided by highly recombinant individuals to increase genetic dissimilarity in the sample sets. The LD criteria, which is based on the historical ancestral recombinations, contributed substantially both to the probability of detecting QTL and precision to estimate QTL position in a QTL (fine)mapping experiment. The LD usually exists in modern animal population, because of selection, closed nucleus breeding scheme and etc. Therefore, using a phenotyping criterion which can take into account this information might be more reliable to find informative animals for phenotyping, when phenotyping is very expensive or difficult to collect.

In chapter 4, the final experiment examined potential loss expected from using selective genotyping in a dairy cattle selection scheme based on marker-assisted selection. The genetic progress when genotyping a proportion of the animals were compared with complete genotyped candidates and non-genotyping strategy (BLUP). A random genotyping scenario plus two selective genotyping criteria were used based on either sum of QTL and polygenic estimated effects (EBV's) or an index of EBV's and Mendelian sampling variance due to QTL. Genotyping a fraction of the population in the MAS schemes increased genetic response due to a detected QTL and also cumulative total gain and decreased the rate of inbreeding compared to traditional BLUP. With genotyping more informative breeding animals, a minimal loss of response at a detected QTL was observed compared to complete genotyping and reduced rates of inbreeding than traditional BLUP. No significantly difference between the genotyping criteria at the same genotyping level indicated that all proposed selective genotyping methods can find the most informative candidates around the truncation point. According to this study, one alternative to decrease genotyping costs in a MAS scheme across families was genotyping the potential candidates close to the truncation point, where QTL effect estimation was based on both LD and linkage analysis information. Therefore, using the markers, which are close enough to the causative mutation and have consistent associations across families can be helped to use the methods in a selection breeding program (Dekkers, 2004). In addition, another benefit of selection across the families is that this approach does not need any specific family structure within population (Dekkers, 2004), e.g. bottom-up or top-down designs in pre-selecting of a dairy breeding program. However, the benefits of implementing marker-linked QTL, with using selective genotyping or in general for MAS, depend on the net result of extra costs, savings and amount of genetic progress of the chosen breeding program (Schrooten et al., 2005) and consequently, it is potentially possible to use selective genotyping around the truncation point in a practical dairy cattle breeding.

This work answered a number of questions relevant to improve selection of a sample of the population for typing (genotype or phenotype). One of the most important questions is how these genotyping/phenotyping strategies can be extended to real life applications. In some situation, more work may be needed to evaluate the proposed strategies. For example, the benefits of the criteria used in this experiment might be affected by several factors such as the levels of the LD in population, using in a trait with low/high heritability, and different QTL parameters (amount of the effect and frequency of favorable allele). The suggested genotyping methods for QTL mapping required intensive computations for a large population. The biggest problem is computing the gametic relationship matrix (IBD matrix) to be used in mixed-model equations. The same computational challenge is expected in using selective genotyping in MAS schemes in a large population. This problem could possibly be solved with developing new algorithms for computing the IBD matrix. The selective phenotyping methods have assumed that low-costly and easily genotyping for loci, which may not be under today's method. However, because of developments in genotyping technologies, the costs are continuing to decrease (Jin *et al.* 2004).

Also, a great deal of additional work is needed on the application of selective genotyping or phenotyping in QTL detection and MAS, simultaneously. Additionally, several authors have tried to use selective genotyping methods for multiple-traits (*e.g.* Bovenhuis and Spelman, 2000). Therefore, as most real-world practical breeding schemes are based on more than one trait/QTL, the criteria for selective genotyping and phenotyping for QTL (fine) mapping (or verification) and MAS need to be examined under realistic situations which are influencing the results from the selective genotyping/phenotyping criteria.

References

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

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.

Dekkers, J. C. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J. Anim. Sci., 82: (E-Suppl): E313–328.

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

Lander, E. and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics, 121: 185-199.

Jannink, J. L. 2005. Selective phenotyping to accurately map quantitative trait loci. Crop Sci., 45: 901–908.

Jin C., L. Lan, A.D. Attie, G.A. Churchill., D. Bulutuglo and B.S. Yandell. 2004. Selective phenotyping for increased efficiency in genetic mapping studies. Genetics, 168:2285–2293.

Schrooten, C., H. Bovenhuis, J. A. M. van Arendonk, and P. Bijma. 2005. Genetic progress in multistage dairy cattle breeding schemes using genetic markers. J. Dairy Sci., 88:1569-1581.

Notes