Genomic Variation in Dairy Cattle

Identification and Use

Promotor:	**Prof. Dr. ir. J. A. M. van Arendonk** Hoogleraar Fokkerij en Genetica Wageningen Universiteit
Co-promotor:	**Dr. ir. H. Bovenhuis** Universitair docent, Leerstoelgroep Fokkerij en Genetica Wageningen Universiteit
Promotiecommissie:	**Dr. D. Boichard** (Institut National de la Recherche Agronomique, Jouy-en-Josas, France) **Prof. Dr. M. A. J. Georges** (University of Liege, Liege, Belgium) **Prof. Dr. J. A. M. Leunissen** (Wageningen Universiteit) **Dr. ir. J. K. Oldenbroek** (Animal Sciences Group, Wageningen UR)

Genomic Variation in Dairy Cattle

Identification and Use

Christiaan Schrooten

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit
prof. dr. ir. L. Speelman
in het openbaar te verdedigen
op vrijdag 27 februari 2004
des namiddags te half twee in de Aula.

Genomic Variation in Dairy Cattle - Identification and Use

Doctoral thesis

C. Schrooten, 2004

Animal Breeding and Genetics Group, Wageningen University, Wageningen and

Holland Genetics BV, CR Delta VRV Holding BV, Arnhem

ISBN: 90-8504-009-4

Abstract

The development of molecular techniques has offered possibilities to identify quantitative trait loci (QTL). Studies in dairy cattle have mainly focused on milk production traits. This thesis first gives an overview of the main identified QTL for milk production traits. Subsequently, a study to detect QTL affecting 27 conformation traits and functional traits was performed. A granddaughter design consisting of 20 Holstein-Friesian grandsires and 833 sons was analyzed by multi-marker regression. This across-family analysis suggested the presence of 61 QTL. Ten of these QTL exceeded the genomewise significance threshold. These were mainly QTL affecting body size traits and udder conformation traits.

When QTL-information is used to select for a certain trait, genetic progress in other traits may be influenced as well, due to pleiotropic effects of QTL, or due to closely linked QTL. A method was developed to identify regions affecting multiple traits. The method is based on the covariance between marker contrasts from single-trait regression analysis for different traits. Application of this method to data on fifteen traits (milk production, udder conformation, udder health and fertility traits) in our granddaughter design resulted in 59 multiple trait quantitative trait regions (MQR). Most MQR were found on BTA 6, 13, 14, 19, 22, 23 and 25.

QTL-information can be used in breeding schemes (marker-assisted selection, MAS) to increase the rate of genetic improvement. A number of multi-stage dairy cattle breeding schemes was evaluated, studying the impact of increased preselection using QTL-information. Response in multi-stage MAS-schemes was 4.5% to 31.3% higher than response in corresponding schemes without QTL-information. In some of the MAS-schemes with a reduced number of progeny tested bulls, genetic progress was identical to or higher than genetic progress in the original schemes. The results indicate opportunities to improve current breeding schemes. The gains depend on the amount of genomic variation explained by QTL.

Currently available pedigrees and methods offer excellent opportunities to identify more QTL, thus increasing the fraction of the genomic variation explained by QTL. New initiatives, like sequencing the bovine genome, will further facilitate the identification of genomic variation, and its use.

Table of Contents

Chapter		Page
1	General introduction	1
2	Quantitative trait loci for milk production traits in dairy cattle	7
3	Whole genome scan to detect quantitative trait loci for conformation and functional traits in dairy cattle	21
4	Detection of pleiotropic effects of quantitative trait loci in outbred populations using regression analysis	47
5	Whole genome scan to detect chromosomal regions affecting multiple traits in dairy cattle	69
6	Genetic progress in multi-stage dairy cattle breeding schemes using genetic markers	93
7	General discussion	115
	Summary	141
	Samenvatting	147
	Nawoord	155
	Curriculum Vitae	157

CHAPTER 1

General Introduction

Background

The aim of breeding programs is to improve the genetic level for traits of economic interest. The breeding goal defines which traits need to be improved, and describes the relative weights of each of the traits. In dairy cattle, genetic progress is achieved by selecting animals that rank highest on the selection criterion or selection index as parents for the next generation. The selection index usually consists of pedigree information and phenotypes measured on selection candidates as well as relatives. Phenotypic information on breeding goal traits as well as correlated traits may be included.

A number of factors determine the annual genetic progress:

- selection intensity, which is related to the fraction of animals selected as parents
- accuracy of the selection criterion, i.e. the correlation between the selection criterion and the breeding goal
- variance of the breeding goal
- generation interval, or the average age of parents when their offspring are born

Breeding programs are continuously trying to achieve maximum response, given certain restrictions with regard to budget, increase of inbreeding etc. Over the years, new techniques have been developed, like reproduction techniques and molecular techniques. These techniques have offered opportunities to increase the intensity and accuracy of selection, and decrease the generation interval, resulting in increased genetic progress.

Technological Developments

Reproduction techniques. The introduction of artificial insemination (AI), approximately 50 years ago, was the first major improvement to dairy cattle breeding programs. Through AI, males can produce large numbers of offspring, so fewer males need to be selected to produce the next generation. This has resulted in a considerable increase of the selection intensity. Because males do not have phenotypes for most of the important traits in dairy cattle, information on large numbers of offspring could now be used to obtain highly accurate estimates of the genetic level of males. Therefore, dairy cattle breeding programs were set up around progeny test programs and annual genetic progress increased considerably. Nowadays, progeny testing is still the key element of dairy cattle breeding programs.

Approximately 30 years ago, techniques that increased female reproduction capacity were introduced. The use of multiple ovulation and embryo transfer (MOET) (Rowson, 1971) and ovum pick up (OPU) in combination with in-vitro fertilisation (IVF) (Kruip et al., 1991) provided

opportunities to increase selection intensity in females and to reduce the generation interval (Meuwissen, 1991).

Molecular techniques. Variation between animals with regard to their genetic level for a certain trait is caused by differences in DNA-composition of these animals. Selection of animals with the best DNA-composition results in genetic progress. Until recently, selection was based only on phenotypic information of animals. Due to the influence of environmental factors on the expression of traits, phenotypes can only provide part of the information on the genetic level of animals. The use of molecular techniques that have been developed in the past decades can increase the amount of information on the genetic level of animals. Genetic markers like microsatellites provide opportunities to follow the inheritance of DNA-segments from parents to offspring. This can be used, for example, to categorise sons of a sire into two groups, based on which variant (allele) of the marker they received. Differences between these groups with regard to relevant traits suggest that this particular DNA-segment (or quantitative trait locus, QTL) is involved in the expression of these traits. Weller et al. (1990) showed that this strategy has considerable power in pedigree structures already available in dairy cattle. Because genetic markers are available early in life, the accuracy of selection at young age can be increased and the generation interval can be reduced. Soller and Beckman (1982) were the first to address the potential benefit of selection for genetic markers (marker assisted selection, MAS) for genetic response in dairy cattle breeding programs.

QTL-Detection and Marker-Assisted Selection

The availability of molecular techniques and their potential benefit for genetic progress has led to numerous experiments to detect QTL. In dairy cattle, these studies have mainly focused on milk production traits. Georges et al. (1995) were the first to report QTL affecting these traits, followed by numerous other studies (e.g. Spelman et al., 1996). However, milk production traits constitute only part of the breeding goal. In recent years, traits like fertility and health have become increasingly important. Compared to QTL for production traits, QTL for fertility and health may be relatively more important, for example due to low heritability of these traits, resulting in low accuracy of the current selection criterion, or due to difficulties to obtain phenotypes. Therefore, information on QTL for these traits is essential for efficient selection.

Once a QTL that affects a particular trait has been detected, it can be used in selection. However, selection for this QTL may also influence other traits, due to pleiotropy (i.e. the QTL affects multiple traits), or because the QTL is linked to QTL affecting other traits. To avoid

negative side effects of selection for certain chromosomal regions or to exploit positive effects of MAS on other traits of interest, insight into the pleiotropic effects of a QTL, or the effect of a chromosomal region on multiple traits, is needed. Until now, little attention has been paid to this topic.

To increase accuracy of selection, males are currently progeny tested and females are tested for performance, after preselection on pedigree information. Accuracy of preselection can be increased by including molecular information, provided markers or genes explain a substantial fraction of the genetic variance. Selection intensity in the pre-selection stage can be increased, by combining reproduction technology (increased embryo production) with molecular technology (MAS of embryos). The opportunities to change the design of the breeding program for the various applications of pre-selection MAS, need to be explored.

Aim

The aim of this thesis is to identify quantitative trait loci (QTL) in dairy cattle, and to study the benefits from selection on QTL-information. Emphasis is on 1) QTL affecting conformation and functional traits, 2) QTL or chromosomal regions affecting multiple traits, and 3) potential consequences of selection on QTL-information for genetic response and design of the breeding program.

Outline

Initially, QTL-detection studies in dairy cattle focused on detection of QTL for milk production traits. Chapter 2 of this thesis gives an overview of QTL affecting milk production traits. Due to the growing importance of non-production traits in dairy cattle breeding, there is also a need for QTL for these traits. Chapter 3 reports on a whole genome scan to detect QTL affecting functional and conformation traits.

When information on a QTL is used in selection, other traits may be influenced as well, due to pleiotropy or close linkage between QTL affecting different traits. Chapter 4 describes a method that was developed to detect pleiotropic QTL or closely linked QTL in an outbred population. A simulation study was used to verify the method. Chapter 5 describes the application of this method to scan the genome for chromosomal regions affecting multiple traits. This study involves milk production traits, udder conformation traits, udder health traits and fertility traits.

Genetic markers provide information that can be used early in life. Chapter 6 addresses general aspects of multi-stage selection and also looks at the possibilities to reduce the number of progeny

tested young bulls in a MAS-scheme. Besides, genetic progress in MAS-breeding schemes with increased embryo production and genotyping of embryos is studied.

Chapter 7, the general discussion, reports on current MAS-applications in dairy cattle breeding programs and addresses the possibilities to detect a sufficiently large fraction of the genetic variance, necessary to start MAS. The precision of the QTL-location largely determines how MAS can be applied. Advantages and disadvantages of alternative MAS-types, and the need for fine-mapping and/or gene-detection, are discussed. The final part of this chapter deals with the impact of new developments, in particular sequencing the bovine genome, on QTL detection and the use of molecular information in dairy cattle breeding.

REFERENCES

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

Kruip, T. A., M. C. Pieterse, T. H. van Beneden, P. L. Vos, Y. A. Wurth, and M. A. Taverne, 1991. A new method for bovine embryo production: a potential alternative to superovulation. Vet. Rec. 128: 208-210.

Meuwissen, T. H. E., 1991. The use of increased female reproduction rates in dairy cattle breeding schemes. Anim. Prod. 52: 21-31.

Rowson, L. E., 1971. Egg transfer in domestic animals. Nature 233: 379-381.

Soller, M., and J. S. Beckmann, 1982. Restriction fragment length polymorphisms and genetic improvement. Proc. 2nd World Congr. Genet. Appl. Livest. Prod., Madrid, Spain VI: 396-404.

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

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

CHAPTER 2

Quantitative Trait Loci for Milk Production Traits in Dairy Cattle

H. Bovenhuis and C. Schrooten

Animal Breeding and Genetics Group, Wageningen University, P.O. Box 338, 6700 AH, Wageningen, The Netherlands

Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France 31: 27-34 (2002)

ABSTRACT

Most of the QTL mapping studies that are presently being carried out use the granddaughter design. Even the largest designs that are being carried out at this moment are expected to detect only a small fraction of the QTL underlying milk production traits. Therefore, methods to increase the power of QTL mapping designs need to be explored. A survey of the results from QTL mapping studies reveals that especially for protein content significant QTL have been reported. Linkage has been confirmed for QTL located on chromosomes 3, 6, 14 and 20. The gene located on chromosome 14 has a major effect on fat content and has been cloned. Chromosomes 3, 6 and 20 have effects on protein content where chromosome 6 is likely to contain more than one QTL. Chromosomal regions showing effects on milk production traits also seem to have effects on other traits. Multiple trait analyses should be used to determine whether pleiotropic gene effects play a role.

INTRODUCTION

Since the first complete genome scan in dairy cattle by Georges et al. (1995) many QTL mapping studies have been initiated. It is expected that information from these experiments will increase our knowledge on the genetic and biological backgrounds of milk production traits. Furthermore, information on the location and effect of these genes can be used in breeding. The aim of this paper is to summarise the main results of the QTL mapping studies that have been carried out in dairy cattle.

EXPERIMENTAL DESIGNS

Most of the QTL mapping experiments that have been carried out in dairy cattle have used a granddaughter design (Boichard, 1998). The largest granddaughter design consists of 14 grandsires and approximately 1800 sons, i.e. the first large scale QTL mapping experiment in dairy cattle, which was initiated by Georges et al. (1995) and extended in later years (Zhang et al., 1998). Experiments that have been carried out in France, Germany and The Netherlands/New Zealand consist of 1000–1500 sons, whereas the designs in Scandinavian countries have around 400 progeny tested sons (Boichard, 1998). Lipkin et al. (1998) and Mosig et al. (2001) used selective DNA pooling. The power of this design is considerably higher than the power of the granddaughter designs that are currently being carried out. However, the full power of detecting QTL only applies to one trait, i.e. protein%. Most of the studies focussed on the Holstein breed. The world-wide

exchange of semen in this breed implies that most of these studies are likely to have common ancestors. In some situations the same grandsires might be included in different studies. This overlap makes it reasonable to assume that a large fraction of the segregating QTL are in fact identical.

QTL detection in a granddaughter design is limited to routinely collected traits. In all designs these comprise the milk production traits and conformation traits. Furthermore, in several countries information regarding reproduction (e.g. calving ease and non-return rate), health (somatic cell score) and workability (milking speed and temperament) is routinely collected, which enables detection of QTL for these traits. In addition, in Scandinavian countries veterinary records of individual cows are available.

Based on results from QTL mapping experiments in dairy cattle, Hayes and Goddard (2001) predicted the distribution of QTL effects for an "average" quantitative trait. Combining these results with power calculations makes it possible to construct the distribution of significantly detected QTL. Figure 1 shows that in a medium sized granddaughter design (15 grandsires with 70 sons per grandsire) around 5.4% of the QTL will be detected. For a large granddaughter design (20 grandsires with 75 sons per grandsire) this figure is approximately 7%. Hayes and Goddard (2001) concluded that, in order to explain 90% of the genetic variance, QTL as small as $0.1\sigma_p$ have to be detected. At present, even the largest granddaughter designs do not meet this requirement. Using selective DNA pooling, Mosig et al. (2001) report that as much as 90% of the QTL affecting protein% have been detected.

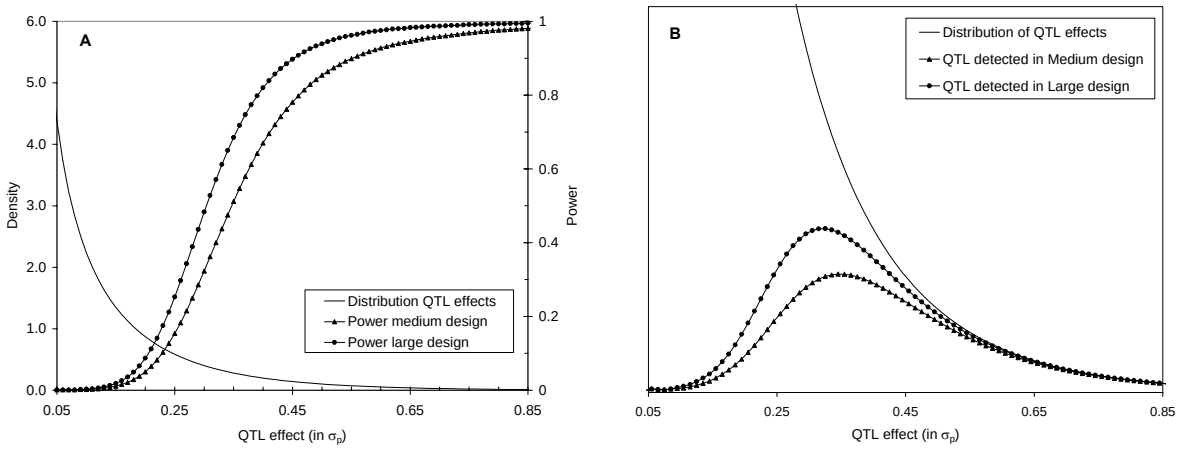

Figure 1. Distribution of QTL effects for milk production traits in dairy cattle according to Hayes and Goddard (2001) combined with the power of a medium sized and a large granddaughter design (A) and the expected distribution of detected QTL for both designs (B).

More recently, projects using crosses between cattle breeds have been started up. At Roslin Institute (Edinburgh) a cross between the Holstein and Charolais breeds has been produced (Williams and Wooliams, 1998) while in New Zealand a cross between Jerseys and Holsteins was established (Spelman, 1998). Specifically setting up these crosses has the advantage that traits not collected routinely can be included in the experiment. On the other hand, the large generation interval and the costs of housing animals make these experiments very costly. Crosses between breeds enable the detection of QTL which explain the difference between breeds and it is therefore uncertain if QTL detected in these experiments are also segregating in purebred populations.

RESULTS OF QTL MAPPING EXPERIMENTS

Table 1 shows an overview of significant and suggestive QTL affecting milk production traits. Summarising results causes some difficulties as studies use different criteria to conclude significance. In order to make results comparable, false discovery rates used by Mosig et al. (2001) and Thomsen et al. (2001) were converted.

Table 1 shows that for protein% more significant and suggestive QTL were reported than for other production traits, even when the study by Mosig et al. (2001), reporting many QTL for protein%, is not considered. This seems surprising considering the higher heritability for protein percentage which is expected to have a negative effect on the statistical power of detecting QTL in a granddaughter design: for a medium sized granddaughter design the power of detecting a QTL with an effect of $0.4\sigma_P$ is 67% when the heritability is 0.3 and 31% when the heritability is 0.6. Therefore, this suggests that for protein% more QTL with large effects are segregating than for other milk production traits. Also for milk yield a considerable number of significant QTL are reported (Table 1). However, this can be attributed to the large number of QTL reported by De Koning et al. (2001). The use of significant QTL as cofactors by De Koning et al. (2001) is expected to increase the power. However, the size of this design (~500 young bulls) suggests that the large number of QTL reported in this study might also be related to the population under study, i.e. the Finnish Ayrshire population.

Table 1 shows that especially for BTA14, 6, 20 and 3 significant QTL effects have been reported. The significant results in Table 1 do not necessarily represent the same QTL and therefore results for those 4 chromosomes will be presented in more detail (Figure 2). Distances between genetic markers in Figure 2 were based on the MARC map. Not all of the results presented in Figure 2 reached genomewide or suggestive significance levels.

Table 1. Main results from QTL mapping studies in dairy cattle[1].

BTA	Milk Yield	Fat Yield	Protein Yield	Fat %	Protein %
1	**[J]				*[I]
2					***[I]
3			*[E]	**[E]	***[D], *[E] **[H]
5	**[J]				
6	***[D] **[J]			***[D]	***[A], ***[D], *[F], **[G], **[H], ***[I]
7			**[G]		***[I]
8					***[I]
9		*[D]			
10					***[I]
11					*[I]
12	**[J]				
13					***[I]
14	***[C], **[G] ***[K]	*[E], **[G], **[H], ***[K]	***[K]	**[E], **[G], **[H], ***[C]	***[C], **[G], **[H], ***[I]
16	*[K]	**[K]	**[K]		
17	*[D]				
20	**[J]			***[D]	***[B], ***[D], **[G], **[H], *[I]
21	**[J]				***[I]
22					*[I]
23	**[J]		*[K]		***[I]
26		**[G]	**[G]	*[D]	
27					*[I]
28				*[D]	*[D], **[H]
29	**[J]				*[I]

[1] ***=Experimentwise, **=Genomewide, *=suggestive linkage

[A]Spelman et al. (1996); [B]Arranz et al. (1998); [C]Coppieters et al. (1998); [D]Zhang et al. (1998); [E]Heyen et al. (1999); [F]Velmala et al. (1999); [G]Boichard et al. (2000) [H]Ashwell et al. (2001); [I]Mosig et al. (2001); [J]De Koning et al. (2001); [K]Thomsen et al. (2001).

BTA14. Coppieters et al. (1998) report experimentwise evidence for the presence of a QTL affecting fat%, protein% and milk yield on the centromeric end of BTA14, i.e. at marker *CSSM66*. At the same location Boichard et al. (2000) found significant evidence for effects on milk yield, fat yield, fat% and protein %. Heyen et al. (1999) found effects on fat% and fat yield in the same

chromosomal region but closer to marker *ILSTS039*. In the region bracketed by *ILST039* and *CSSM66*, Looft et al. (2001) detected significant effects on milk, fat and protein yield and found substantial linkage disequilibrium between marker *KIEL_E8* and milk production traits. Ashwell et al. (2001) report a significant effect for fat% and fat yield at marker *BMS1678*. At the other end of BTA14, close to marker *BM6425,* Mosig et al. (2001) report a QTL affecting protein%. This is likely to be a different QTL from the ones reported in other studies.

In a follow-up of the study by Coppieters et al. (1998), Riquet et al. (1999) describe the fine mapping of a QTL on BTA14 by developing a high density marker map and searching for identity-by-descent regions. Recently, Grisart et al. (2002) reported the positional cloning of this QTL. This is the first QTL in dairy cattle that has been successfully identified. The proposed candidate gene is *DGAT1* and is located close to *ILSTS039*. The gene catalyses the final step in triglyceride synthesis and as 98% of the milk fat consists of triglycerides this is a likely candidate. Grisart et al. (2002) hypothesise that the functionality of the enzyme has changed due to the identified mutation which would explain the effect on fat%. In the Dutch population 51% of the variation in fat% could be explained by the mutation indicating that this is a gene with a major effect. For the New Zealand population 31% of the variation in fat% could be explained by the mutation. Allele frequencies were 0.63 in the Dutch population and 0.30 in the New Zealand population. The reason a QTL with such large effects is still segregating in both populations is probably that the gene has a negligible effect on the net-merit index used in both countries.

In addition to the QTL affecting milk production traits a number of studies reported QTL with effects on non-production traits. The effects on somatic cell score by Zhang et al. (1998) and front teat placement and fore udder attachment by Ashwell et al. (2001) might be the result of pleiotropic effects of one gene.

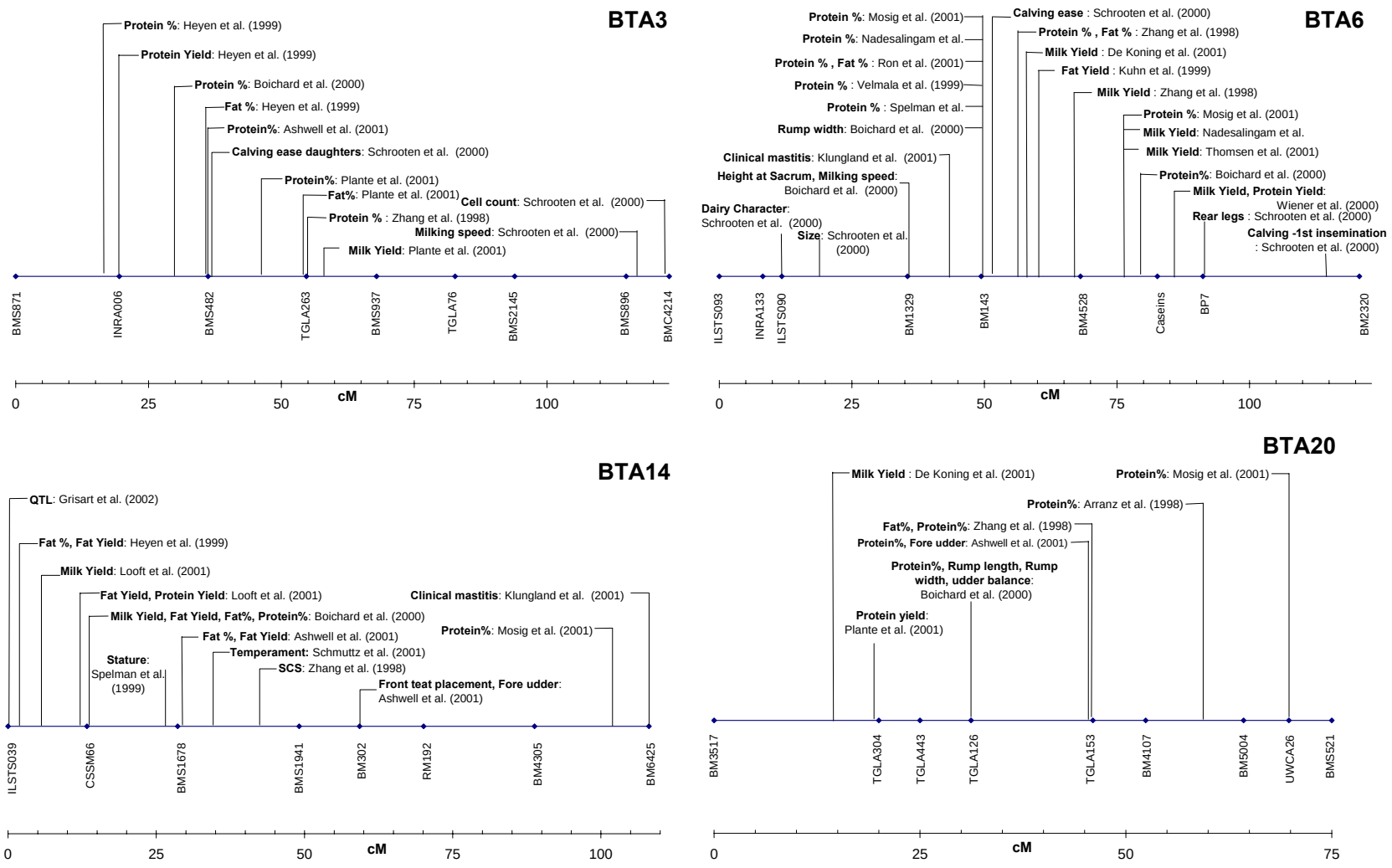

Figure 2. The most likely location of QTL on BTA3, BTA6, BTA14 and BTA20 affecting milk production traits and non-production traits in dairy cattle.

BTA6. The QTL detected by Georges et al. (1995) on BTA6 in combination with the presence of the casein gene cluster encouraged many researchers to study this chromosome. Figure 2 shows the location of reported QTL. There is a remarkable consistency between different studies on the location of a QTL affecting protein %. Most of the studies locate the QTL at or close to marker *BM143* and Ron et al. (2001) report a 95% confidence interval of 4 cM around this marker. In addition to the effect on protein%, Zhang et al. (1998) and Ron et al. (2001) also report a QTL affecting fat% in the same region. The estimated locations suggest that this is the same gene. Note that Klungland et al. (2001) found significant evidence for a QTL affecting clinical mastitis close to marker *BM143*. Boichard et al. (2000) and Mosig et al. (2001) report a QTL affecting protein% that is located closer to the casein gene cluster. In the same region several studies reported a QTL with an effect on milk yield.

Spelman et al. (1996), Zhang et al. (1998) and Velmala et al. (1999) fitted a 2-QTL model to BTA6. Velmala et al. (1999) found evidence for the presence of 2 QTL affecting protein%, milk yield and fat yield. Their analyses suggest the presence of one QTL close to *BM143* and another QTL located around the casein cluster. Zhang et al. (1998) found significant evidence for 2 QTL affecting fat%. These two loci were only 12 cM apart: one locus was located close to *BM1329* and the other at *TGLA37*. Zhang et al. (1998) indicated that in all the cases where there was evidence in favour of a 2 QTL model, the two linked loci were in repulsion phase in those families where both of them were segregating. Selection might have induced negative covariances between loci. Closely linked QTL in repulsion phase are expected to remain undetected in single QTL analyses, as was illustrated by Velmala et al. (1999) for the QTL affecting fat yield on BTA6.

BTA20. On BTA20 effects on protein% were reported by Arranz et al. (1998), Zhang et al. (1998), Boichard et al. (2000), Ashwell et al. (2001), and Mosig et al. (2001). All these effects were located in a chromosomal region of approximately 40 cM and therefore might be due to a single QTL. In the same chromosomal region, Zhang et al. (1998), also showed significant evidence for a QTL affecting fat%. Boichard et al. (2000) and Ashwell et al. (2001) reported effects on udder characteristics at the same chromosomal region as where effects on protein% were found.

BTA3. On BTA3 QTL for protein% were reported by Zhang et al. (1998), Heyen et al. (1999), Boichard et al. (2000) and Ashwell et al. (2001). These effects were located in an area of about 40 cM. In addition, in the same region significant effects were found for protein yield and fat% (Heyen et al., 1999). At the other end of the chromosome Schrooten et al. (2000) detected effects on milking speed and somatic cell count.

DISCUSSION

It was concluded that even the largest QTL mapping studies carried out to date are expected to detect only a limited fraction of the QTL underlying milk production traits. More QTL can be detected by increasing the size of the design or by using alternative designs, e.g. selective DNA pooling. However, in most studies the information present has not been fully exploited. Most of the studies used regression methods for QTL detection. Regression methods have been proven to be robust, relatively simple to apply and computationally not very demanding which allows the use of permutation tests for calculation of significance thresholds. However, regression methods assume unrelated families and can handle only two generations of genotyped individuals. In dairy cattle pedigrees, usually additional relations exist, e.g. due to relationships between grandsires or due to dams with multiple genotyped sons. Bink and Van Arendonk (1999) and Bolard and Boichard (2001) showed that including additional relationships increases the power of the experiment. These methods have not been fully exploited in most experiments. An alternative strategy for increasing the power of existing designs was demonstrated by De Koning et al. (2001). By fitting cofactors De Koning et al. (2001) significantly detected 8 QTL affecting milk yield. Without the use of cofactors 5 suggestive QTL were detected.

Genes influencing more than one trait are believed to be the main source for genetic correlations between traits. Little attention has been paid to potential pleiotropic effects of chromosomal regions. However, before MAS for certain chromosomal regions can be implemented, effects of chromosomal regions on multiple traits need to be studied. Figure 2 demonstrates that chromosomal regions affecting milk production also have effects on other traits. So far attention has focussed on single trait analyses. Multiple trait analyses could be used to determine whether pleiotropic gene effects play a role.

While results from QTL mapping experiments accumulate, focus will turn towards fine mapping, cloning genes and utilising this information in breeding. So far, the first gene affecting milk production has been identified on BTA14 (Grisart et al., 2002). Linkage has been confirmed on three other chromosomes; BTA3, BTA6 and BTA20 where BTA6 is likely to contain more than one QTL. It is expected that in the coming years more chromosomal regions affecting milk production traits will be detected. However, these QTL are likely to have smaller effects, which will complicate positional cloning.

REFERENCES

Arranz, J.-J., W. Coppieters, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L. Moreau, C. Mezer, J. Riquet, P. Simon, P. Vanmanshoven, D. Wagenaar, and M. Georges, 1998. A QTL affecting milk yield and composition maps to bovine chromosome 20: a confirmation. Anim. Genet. 29: 107-115.

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

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

Boichard, D., 1998. QTL detection with genetic markers in dairy cattle: an overview. Paper presented at 49th Annu. Meeting of the European Assoc. of Animal Production, August 24-27 1998, Warsaw, Poland.

Boichard, D., C. Grohs, F. Bourgeois, F. Cerqueira, R. Faugeras, A. Neau, R. Rupp, Y. Amigues, M. Y. Boscher, and H. Levéziel, 2000. Detection of genes influencing economic traits in three French dairy cattle breeds. Paper presented at 51st Annu. Meeting of the European Assoc. of Animal Production, August 21-24 2000, the Hague, The Netherlands.

Bolard, M., and D. Boichard, 2001. Use of maternal information for QTL detection in a (grand)daughter design. Paper presented at 52nd Annu. Meeting of the European Assoc. of Animal Production, August 26-29 2001, Budapest, Hungary.

Coppieters, W., J. Riquet, J.-J. Arranz, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L. Moreau, C. Nezer, P. Simon, P. Vanmanshoven, D. Wagenaar, and M. Georges, 1998. A QTL with major effect on milk yield and composition maps to bovine chromosome 14. Mamm. Genome 9: 540-544.

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

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

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

Hayes, B., and M. E. Goddard, 2001. The distribution of the effects of genes affecting quantitative traits in livestock. Genet. Sel. Evol. 33: 209-229.

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

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

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

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

Looft, C., N. Reinsch, C. Karall-Albrecht, S. Paul, M. Brink, H. Thomsen, G. Brockmann, C. Kühn, M. Schwerin, and E. Kalm, 2001. A mammary gland EST showing linkage disequilibrium to a milk production QTL on bovine chromosome 14. Mamm. Genome 12: 646-650.

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

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

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

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

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

Schmutz, S. M., J. M. Stookey, D. C. Winkelman-Sim, C. S. Waltz, Y. Plante, and F. C. Buchanan, 2001. A

QTL study of cattle behavioral traits in embryo transfer families. J. of Hered. 92: 290-292.

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

Spelman, R. J., 1998. Doctoral thesis, Wageningen Agricultural University, P. O. Box 338, 6700 AH Wageningen, The Netherlands.

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

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

Thomsen, H., N. Reinsch, N. Xu, C. Looft, S. Grupe, C. Kühn, G. A. Brockmann, M. Schwerin, B. Leyhe-Horn, S. Hiendleder, G. Erhardt, I. Medjugorac, I. Russ, M. Förster, B. Brenig, F. Reinhardt, R. Reents, J. Blümel, G. Averdunk, and E. 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.

Velmala, R. J., H. J. Vilkki, K. T. Elo, D. J. de Koning, and A. V. Mäki-Tanila, 1999. A search for quantitative trait loci for milk production traits on chromosome 6 in Finnish Ayrshire cattle. Animal Genetics 30: 136-143.

Wiener, P., I. Maclean, J. L. Williams, and J. A. Woolliams, 2000. Testing for the presence of previously identified QTL for milk production traits in new populations. Anim. Genet. 31: 385-395.

Williams, J., and J. Woolliams, 1998. Annual Report 97-98, Roslin Institute.

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

CHAPTER 3

Whole Genome Scan to Detect Quantitative Trait Loci for Conformation and Functional Traits in Dairy Cattle

C. Schrooten[1,2], H. Bovenhuis[1], W. Coppieters[3], and J. A. M. van Arendonk[1]

[1]Animal Breeding and Genetics Group, Wageningen University, PO Box 338, 6700 AH Wageningen, The Netherlands
[2]Holland Genetics, CR Delta, PO Box 5073, 6802 EB Arnhem, The Netherlands
[3]Department of Genetics, Faculty of Veterinary Medicine, University of Liège (B43), 20 Bd de Colonster, 4000-Liège, Belgium

Journal of Dairy Science 83: 795 – 806 (2000)

ABSTRACT

A granddaughter design was used to locate quantitative trait loci determining conformation and functional traits in dairy cattle. In this granddaughter design, consisting of 20 Holstein Friesian grandsires and 833 sons, genotypes were determined for 277 microsatellite markers covering the whole genome. Breeding values for 27 traits regarding conformation (18), fertility (2), birth (4), workability (2) and udder health (1), were evaluated in an across-family analysis using multimarker regression. Significance thresholds were determined using a permutation test.

The across-family analysis suggested the presence of 61 quantitative trait loci, when 27 (i.e. one for each trait) were expected by chance. The test statistic exceeded the genomewise significance threshold for the following traits and chromosomes: chest width on chromosome 2; gestation length on chromosome 4; stature, body capacity and size on chromosome 5; dairy character on chromosome 6; angularity on chromosome 12; fore udder attachment on chromosome 13; and fore udder attachment and front teat placement on chromosome 19. The quantitative trait loci for size traits on chromosomes 2, 5 and 6 may also have an effect on calving ease. The quantitative trait loci for udder traits on chromosomes 13 and 19 may also affect somatic cell score and mastitis resistance. If there are no negative effects on other economically important traits, marker assisted selection using markers associated with these quantitative trait loci can be applied.

(**Key words**: whole genome scan, quantitative trait loci, conformation, functional traits)

Abbreviation key: **MAS** = marker assisted selection, **QTL** = quantitative trait locus

INTRODUCTION

Most economically important traits in dairy cattle production are influenced by many genes as well as environmental factors. Breeding programs aim at selecting animals with the most favorable set of genes, to produce animals for the next generation. Selection in most of these breeding programs is for a combination of production, conformation and functional traits. Evaluation procedures like BLUP (Henderson, 1984) have been developed to estimate breeding values of animals. The nature of the underlying genes (quantitative trait loci, QTL) affecting traits, however, is still largely unknown.

Recently, efforts have been undertaken to locate genes affecting economically important traits in dairy cattle. Genetic markers associated with these genes can be used in marker assisted selection

(MAS), to increase genetic progress (Kashi et al., 1990). For dairy cattle, the focus was initially on milk production traits. The granddaughter design (Weller et al., 1990) was used to locate genes involved in milk, fat and protein production (Georges et al., 1995; Spelman et al., 1996; Arranz et al., 1998; Coppieters et al., 1998; Ron et al., 1998). Implementation of MAS in breeding strategies for production traits has been started (Spelman, 1998).

However, MAS is expected to be especially beneficial for traits that cannot be improved very efficiently by current breeding programs, for example, due to low heritability (Lande and Thompson, 1990). Important traits in dairy cattle breeding with low heritability include functional traits like fertility and health traits. Although some markers associated with functional traits were reported (Ashwell et al., 1996; Ashwell et al., 1997; Ashwell et al., 1998b; Zhang et al., 1998), these studies were limited to only a few traits or used a marker map that was not very dense.

In this study, a whole genome scan using a dense marker map was applied to a granddaughter design in the Dutch Holstein-Friesian population to locate QTL for conformation and functional traits.

MATERIALS AND METHODS

Family Structure and DNA Analysis

Semen samples from 949 young bulls, progeny-tested in the Netherlands from July 1987 to September 1993 and belonging to 22 half-sib families, were collected, and DNA was extracted from these samples. To avoid effects of selection within families (MacKinnon and Georges, 1992) young bulls in a sampling-region × period subclass were excluded from the analyses if only part of the young bulls in this subclass had DNA material available. After these edits, the data set consisted of 20 half-sib families with 833 sons. This granddaughter design was an extension of the design described previously by Spelman et al. (1996) and used to detect QTL for production traits. The number of families was identical, but some grandsires had additional sons in the analyses. These sons did not have information on breeding values for production traits when Spelman et al. (1996) performed their analyses. The number of young bulls per grandsire ranged from 12 to 147 with an average family size of 42. For each young bull and grandsire, genotypes for 277 microsatellite markers covering the 29 autosomes were determined as described by Georges et al. (1995) or by using the “four dye-one lane” technology on an ABI373 or ABI377 sequencer. A linkage map for the 29 autosomes was constructed using CRIMAP (Lander and Green, 1987) and ANIMAP (Georges et al., 1995). The Haldane mapping function was used to calculate length of the

chromosomes and the distance between markers on each chromosome (Table 1). Total length of the genome was almost 3200 cM.

Average marker interval per chromosome ranged from 4 to 29 cM. A graphical representation of marker distribution and marker density can be found in Figure 1.

Average heterozygosity for each chromosome was calculated as number of heterozygous marker loci across grandsires divided by number of marker loci across grandsires. Average heterozygosity per chromosome ranged from 45 to 73%. Average heterozygosity across chromosomes was 60% (Table 1).

Trait Data

Data on many traits of dairy cows are routinely collected in the Netherlands. Systematic environmental factors, such as herd, year, and season of calving, influence these traits. BLUP procedures (Henderson, 1984) are used to estimate breeding values of sires. These procedures adjust for systematic environmental factors and utilize all available data. The data used in this study were estimated breeding values for 27 of these traits (Table 2). These estimated breeding values were obtained in the national genetic evaluation by using a sire model or an animal model.

Data on calving ease and other birth traits are collected on offspring of young bulls. About 1000 inseminations per young bull are carried out, mainly on first lactation heifers. Calving survey cards to record birth traits on these heifers are supplied to farmers that have cows in calf to a young bull. Farmers score calving ease and birth weight and report to the national herdbook. Gestation length is derived from insemination date and birth date, which are in the national database.

Eighteen conformation traits are scored in the national herd classification system. All lactating heifers in herds participating in herd classification, i.e. 50% of all milk-recorded heifers (CR Delta, 1999), are classified once. Each herd is visited approximately twice a year by a professional classifier. The conformation traits can be subdivided into four general characteristics and 14 linearly scored traits. At the time of classification, farmers report a linear score for milking speed and temperament of the heifer during milking. For each of the traits, the scale for scoring is divided into 9 classes.

Two fertility traits, interval between calving and first insemination, and nonreturn at 56 d postinsemination are derived from calving and insemination data. The trait “nonreturn at 56 d postinsemination” is a measure for pregnancy rate. All cows not offered for AI within 56 d after insemination are considered pregnant.

Table 1. Number of markers per chromosome, chromosome length, average marker interval, heterozygosity, and information content

Chromosome	Markers (no.)	Length (cM)	Average marker interval (cM)	Average heterozygosity[1]	Average information content[2]
1	10	189	21.0	0.61	0.60
2	10	139	15.4	0.62	0.69
3	8	171	24.4	0.52	0.66
4	10	104	11.6	0.69	0.68
5	8	181	25.9	0.68	0.72
6	29	113	4.0	0.50	0.62
7	8	138	19.7	0.59	0.65
8	8	197	28.1	0.61	0.67
9	6	113	22.6	0.54	0.59
10	12	142	12.9	0.67	0.60
11	9	112	14.0	0.60	0.68
12	7	103	17.2	0.62	0.71
13	5	117	29.3	0.57	0.70
14	11	108	10.8	0.64	0.70
15	8	103	14.7	0.62	0.67
16	6	114	22.8	0.71	0.70
17	6	93	18.6	0.67	0.67
18	7	124	20.7	0.59	0.68
19	26	133	5.3	0.54	0.65
20	25	94	3.9	0.51	0.67
21	6	86	17.2	0.68	0.69
22	6	85	17.0	0.65	0.65
23	8	65	9.3	0.73	0.74
24	7	55	9.2	0.60	0.65
25	7	80	13.3	0.54	0.60
26	6	59	11.8	0.59	0.65
27	6	44	8.8	0.53	0.58
28	7	57	9.5	0.45	0.59
29	5	60	15.0	0.57	0.60
Total	277	3179	12.8	0.59	0.66

[1]Fraction of marker loci that are heterozygous, averaged across grandsires and all marker loci on each chromosome

[2]Calculated from variance of quantitative trait locus conditional probabilities at each centimorgan as a proportion of the variance when true descent is known and then averaged (Spelman et al., 1996)

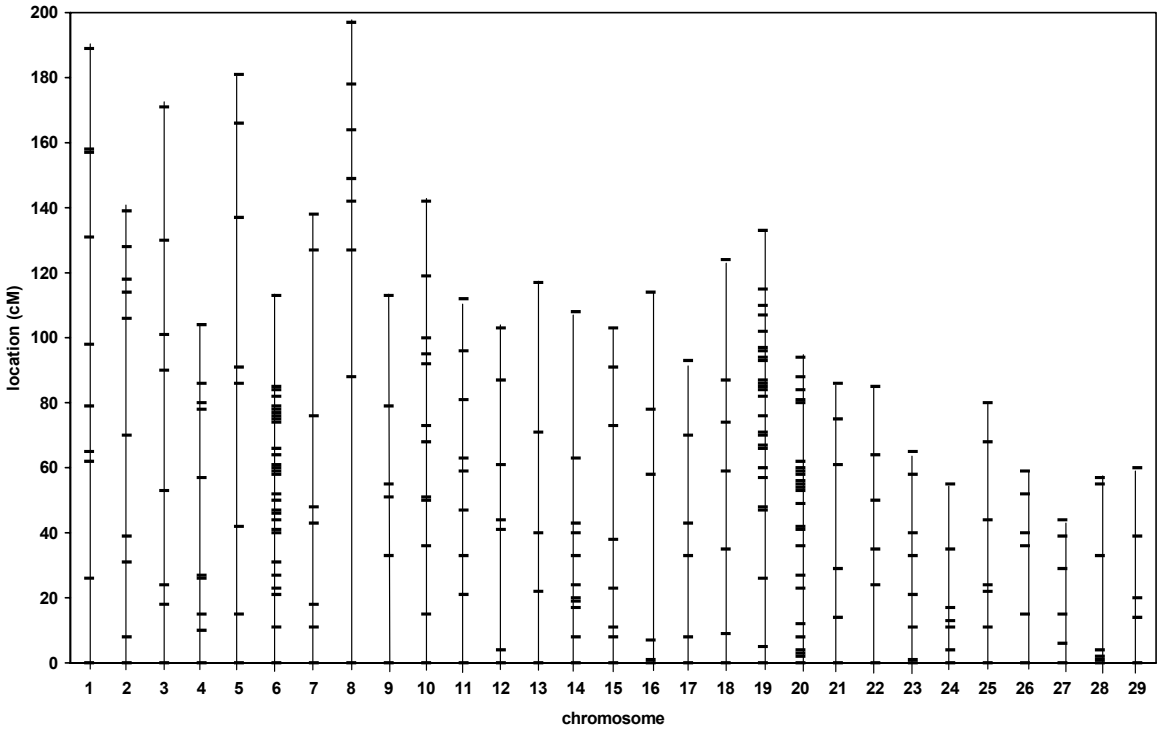

Figure 1. Graphical representation of marker distribution and marker density. Chromosome number is printed on the horizontal axis; location on the chromosomes is printed on the vertical axis and given in centimorgans. Markers on a given chromosome are indicated by small horizontal bars on the vertical line depicting this chromosome.

Data on cell count are collected in the official milk-recording scheme. The interval between two consecutive milk recordings is 3, 4, 5 or 6 wk. The frequency of collection of cell count data varies from once every milk recording to once every 5 milk recordings. Approximately 90% of the cows in the national milk-recording scheme are recorded for cell count (G. de Jong, 1999, personal communication).

A more detailed description of the traits used in the analysis is given in Table 2. Average number of sons with breeding values and number of granddaughters per son for each category of traits is also listed in Table 2.

Table 2. Analyzed traits.[1]

Trait category	Trait	Units[2]	Herita-bility[3]	S/A[4]	# Sons	progeny/son (no.)	
						Range	Median
Birth traits	Gestation length	Days	0.46	S	829	28-397	177
	Birth weight	Kg	0.18	S			
	Calving ease		0.13	S			
	Calving ease daughters		0.07	S			
Fertility	Interval calving to first insemination	Days	0.06	S	831	40 – 42,150	107
	Non-return daughters		0.02	S			
Somatic cell	Somatic cell score (2-log)		0.15	A	838	1 - 56,889	115
Conformation – linear traits[2]	Stature		0.60	A	833	17 – 30,560	57
	Chest width		0.30	A			
	Body capacity		0.35	A			
	Rump angle		0.35	A			
	Rump width		0.30	A			
	Angularity		0.35	A			
	Rear leg set		0.35	A			
	Foot diagonal		0.20	A			
	Fore udder attachment		0.35	A			
	Front teat placement		0.45	A			
	Teat length		0.45	A			
	Udder depth		0.45	A			
	Rear udder height		0.35	A			
	Suspensory ligament		0.25	A			
Conformation – general characteristics[2]	Size		0.60	A	833	17 – 30,560	57
	Dairy character		0.30	A			
	Udder		0.35	A			
	Feet & legs		0.20	A			
Workability[2]	Milking speed		0.30	S	801	1 – 23,319	44
	Temperament		0.15	S	656	4 – 23,319	49

[1]Listed for each trait are units of expression, heritability, model for estimating breeding values, number of sons with data, and number of progeny included in breeding value of a bull

[2]Breeding values for conformation and workability traits are standardized to a scale with mean = 100 and standard deviation = 4.

[3]Heritabilities as used in national evaluation procedure.

[4]Breeding values from sire model (S) or animal model (A).

Statistical Analysis

Multimarker regression (Knott et al., 1994) of trait data was used to detect QTL by applying a weighted across-family analysis to each trait separately with the following model:

$$Y_{ij} = \mu + gs_i + b_{ik}X_{ijk} + e_{ijk} \qquad (1)$$

where Y_{ij} = deregressed estimated breeding value of son j of grandsire i, μ = overall mean, gs_i = fixed effect of grandsire i, b_{ik} = regression coefficient for grandsire i at position k on the chromosome, X_{ijk} = probability that son j receives a chromosomal segment from grandsire i at position k, and e_{ijk} = random residual.

With this model, the probability of a son inheriting a certain chromosomal segment of his sire was calculated. Trait data were regressed on this probability. Contribution of each sire in the regression analysis was weighted based on heritability of the trait and number of daughters (Spelman et al., 1996); the weighting factors were

$$w_{ij} = \frac{n_{ij}}{1 + (n_{ij} - 1)\,¼\,h^2}$$

where w_{ij} = weighting factor for son j of grandsire i, n_{ij} = number of daughters of son j of grandsire i, and h^2 = heritability.

Estimated breeding values were deregressed, to prevent double weighting, before using them in the weighted regression analysis. Deregression factors were based on number of daughters and the heritability of the trait, assuming that daughters were the only information source contributing to the estimated breeding value.

To test for the presence of a QTL, test statistics similar to an F-ratio were calculated for every position on all chromosomes. This test statistic is the ratio of the difference in residual sums of squares under the null hypothesis (no QTL) and residual sums of squares under the QTL model over the residual sums of squares under the QTL model.

Based on results from the across family analysis, a within-family analysis was carried out for areas with significant QTL to identify families segregating for the QTL and to estimate QTL effects.

Significance Thresholds

Chromosomewise significance thresholds were calculated by applying a permutation test as described by Churchill and Doerge (1994). Trait values together with weighting factors were

reassigned randomly to the sons within a family. The new data set was analyzed and a new test statistic was calculated. This procedure was repeated 15,000 times to derive reliable thresholds. For computational reasons, however, permutation was done only for one chromosome. A chromosome was chosen that had average marker spacing and chromosome length, i.e. chromosome 11. Permutation was done for each trait separately, because distribution of the test-statistic was dependent on the trait analyzed. This finding was also observed by Spelman et al. (1996). Genomewise thresholds, accounting for multiple testing across the genome, were calculated from chromosomewise thresholds by using the Bonferroni correction for testing across multiple chromosomes.

Thresholds used in this study were the 10% genomewise threshold and the threshold indicating suggestive linkage. Suggestive linkage was defined by Lander and Kruglyak (1995) as "statistical evidence that would be expected to occur one time at random in a genome scan", i.e., finding one false positive in a genome scan. Considering n = 29 autosomes, the chromosomewise type I error rate α_c to obtain genomewise suggestive linkage for a single trait can be approximated by $n\alpha_c = 1$, which results in $\alpha_c = 0.0345$.

RESULTS

Regression Analysis

Linkage between markers and traits was tested on 29 chromosomes for 27 traits. The 61 trait × chromosome combinations exceeding the threshold for suggestive linkage are presented in Table 3. It should be noted that, according to the definition of suggestive linkage, one false-positive result per trait is expected when there is no QTL. As can be seen from Table 3, a few chromosomes are likely to contain genes for a number of traits. For example, QTL for the traits stature, chest width, body capacity, rump width, udder depth, rear udder height, size, dairy character, birth weight and calving ease all map to chromosome 5.

Figure 2 shows the test statistic profiles of the trait × chromosome combinations that exceeded the 10% genomewise threshold. These results will be discussed in more detail, together with other traits exceeding the threshold for suggestive linkage for that chromosome in the same region. Results from the within-family analysis for trait × chromosome combinations exceeding the 10% genomewise threshold are in Table 4. Tabulated values for the F-distribution were used to determine significance of effects ($P < 0.01$) from the within-family analysis.

Table 3. Results exceeding the threshold for suggestive linkage in the across-family analysis.[1]

Cr[2]	Trait	Marker or marker bracket	Position (cM)	Test statistic	Significance[3] Chromo-somewise P-value	Genomewise P-value[4]
2	Chest width	BM2113	139	2.64	0.0006	0.0145
2	Body capacity	BM2113	139	2.05	0.0202	
2	Rump width	TGLA110-BM6444	117	2.05	0.0193	
2	Calving ease daughters	BM2113	139	2.07	0.0250	
2	Non return daughters 56 d post insemination	BM2113	139	1.91	0.0339	
2	Milking speed	ETH121-BM4440	57	2.98	0.0268	
3	Cell count (log)	BMC5227	171	1.97	0.0317	
3	Milking speed	TGLA127-BMC5227	144	3.10	0.0204	
3	Calving ease daughters	RM19-TGLA263	35	2.01	0.0334	
4	Gestation length	TGLA159-TGLA420	17	2.61	0.0010	0.0279
5	Stature	IGF1-BM315	122	2.40	0.0031	0.0876
5	Chest width	BM315-ETH152	156	2.20	0.0110	
5	Body capacity	BM315-ETH152	154	2.56	0.0008	0.0203
5	Rump width	AGLA22	181	2.09	0.0159	
5	Udder depth	IGF1-BM315	109	2.06	0.0196	
5	Rear udder height	AGLA254-IGF1	88	2.13	0.0116	
5	Size	IGF1-BM315	123	2.54	0.0015	0.0414
5	Dairy character	BP1-AGLA293	27	2.25	0.0075	
5	Birth weight	IGF1-BM315	132	2.28	0.0135	
5	Calving ease	ETH152	166	2.20	0.0210	
6	Stature	MCM53	11	2.15	0.0126	
6	Angularity	BM4311	84	1.94	0.0273	
6	Rear leg set	BP7	85	2.01	0.0220	
6	Size	MCM53	11	2.36	0.0042	
6	Dairy character	ILSTS090	0	2.45	0.0024	0.0684
6	Calving ease	BMS690	44	2.13	0.0296	
6	Interval calving to first insemination	BP7-BM2320	107	2.22	0.0062	
8	Stature	TGLA13	127	2.23	0.0084	
8	Size	TGLA13	127	2.30	0.0066	
9	Non return daughters 56 days post insemination	TGLA73	79	2.12	0.0115	
10	Angularity	CSSM38-BM1237	12	2.10	0.0120	
10	Fore udder attachment	TGLA378	51	1.93	0.0334	
10	Front teat placement	BRRIBOold-BMS861	44	2.03	0.0122	
11	Angularity	BM716	0	2.25	0.0050	

Cr[2]	Trait	Marker or marker bracket	Position (cM)	Test statistic	Significance[3] Chromo-somewise P-value	Genomewise P-value[4]
12	Angularity	TGLA9-AGLA226	31	2.43	0.0015	0.0414
13	Fore udder attachment	TGLA23	0	3.20	< 0.0001	< 0.0001
13	Udder depth	BMC1222-AGLA285	23	2.27	0.0070	
15	Gestation length	BM848	103	2.06	0.0276	
17	Interval calving to first insemination	OARVH98	0	1.97	0.0224	
18	Cell count (log)	BM7109-ILSTS002	70	2.17	0.0103	
19	Fore udder attachment	BMS2503-BMS650	68	2.57	0.0009	0.0241
19	Front teat placement	BMS2503	67	2.25	0.0032	0.0914
19	Udder depth	TGLA94-URB044	34	2.14	0.0119	
19	Rear leg set	BM17132	76	1.98	0.0253	
21	Calving ease	ETH131-TGLA337	33	2.26	0.0149	
23	Fore udder attachment	RM033	11	2.18	0.0070	
23	Milking speed	BM1258-GBCYP21	30	3.20	0.0160	
24	Angularity	CSSM31-AGLA269	16	2.07	0.0135	
24	Rear leg set	AGLA269	17	2.27	0.0061	
24	Foot diagonal	AGLA269	17	2.44	0.0042	
24	Feet & legs	CSSM31-AGLA269	15	2.17	0.0154	
25	Dairy character	BMS1353-AF5	74	2.12	0.0166	
25	Fore udder attachment	BP28-BMS1353	62	1.94	0.0316	
25	Udder	AF5	80	1.99	0.0296	
26	Fore udder attachment	TGLA22-HEL11	3	2.20	0.0067	
26	Udder	TGLA22	0	2.12	0.0152	
27	Udder depth	HUJI13	44	2.04	0.0222	
29	Body capacity	BMC1206	60	2.07	0.0179	
29	Rump angle	BMC3224-BMC1206	53	2.13	0.0082	
29	Rump width	BMC8012	20	2.06	0.0180	
29	Birth weight	BMC8012-BMC3224	29	2.19	0.0222	

[1]Location of highest test statistic for a trait × chromosome combination is given by marker or marker bracket as well as the position expressed in centimorgans

[2]Cr = Chromosome.

[3]determined by permutation test

[4]only presented when below a P-value of 0.10

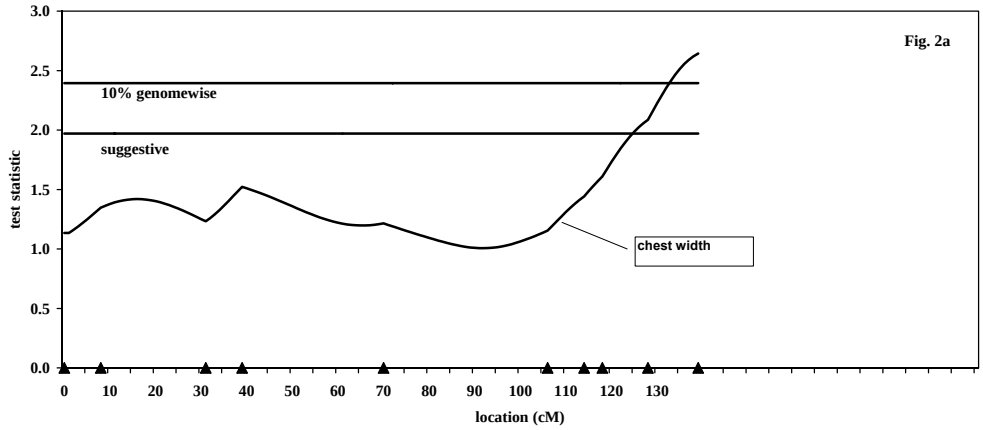
Fig. 2a
10% genomewise
suggestive
chest width
test statistic
location (cM)

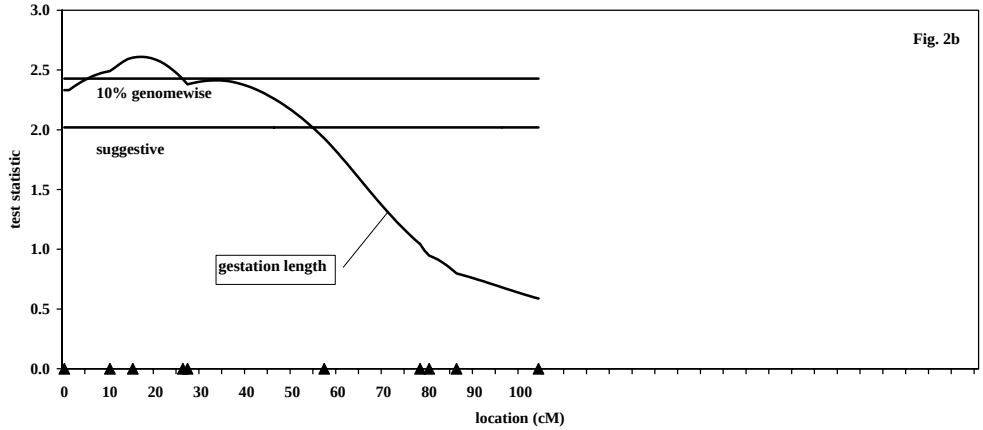
Fig. 2b
10% genomewise
suggestive
gestation length
test statistic
location (cM)

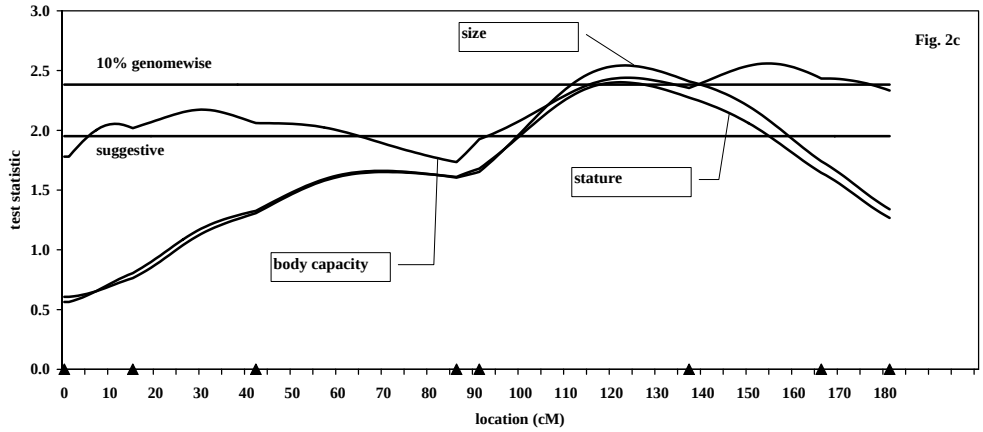
Fig. 2c
size
10% genomewise
suggestive
stature
body capacity
test statistic
location (cM)

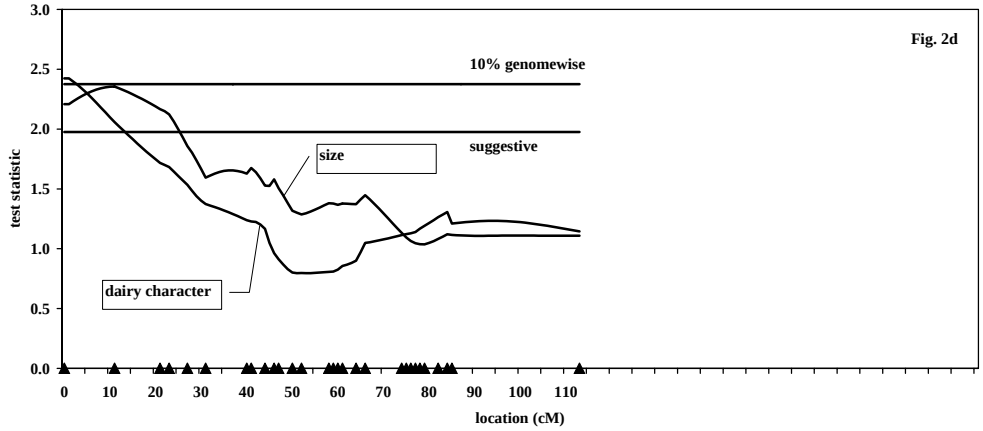
Fig. 2d
10% genomewise
suggestive
size
dairy character
test statistic
location (cM)

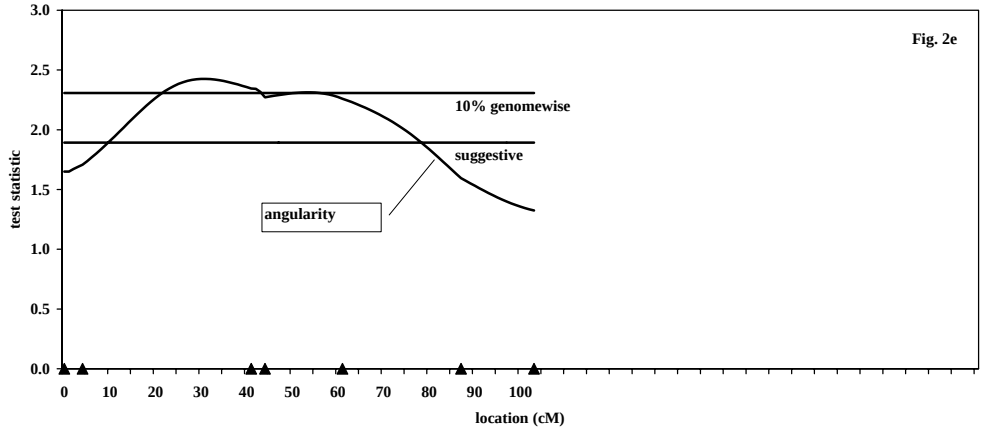
Fig. 2e
10% genomewise
suggestive
angularity
test statistic
location (cM)

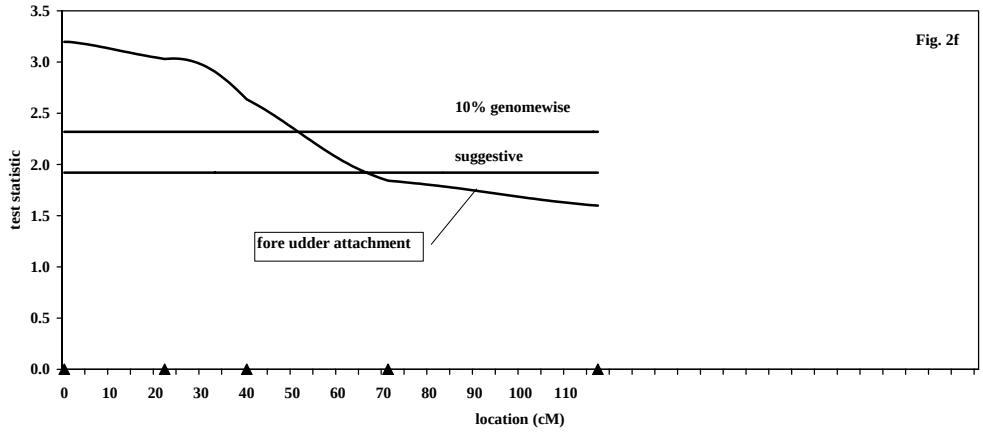
Fig. 2f
10% genomewise
suggestive
fore udder attachment
test statistic
location (cM)

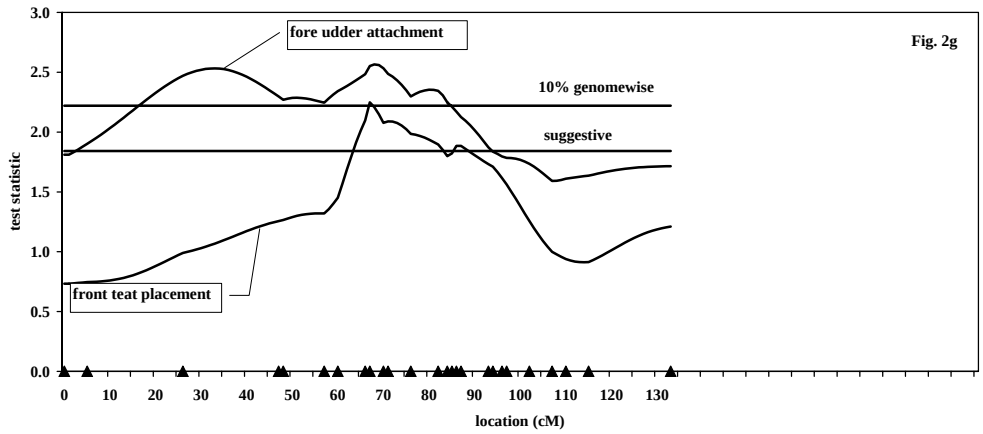

Figure 2. Profile of test statistic for chromosome × trait combinations exceeding the 10% genomewise significance threshold. The value of the test statistic (vertical axis) is given for each location on the chromosome (horizontal axis; location in centimorgans). Position of the markers is indicated by triangles on the horizontal axis. The threshold for suggestive linkage and the 10% genomewise threshold are indicated by horizontal solid lines. Scale of the horizontal axis is equal for all figures. a) Test statistic profile of chest width on chromosome 2. b) Test statistic profile of gestation length on chromosome 4. c) Test statistic profile of stature, body capacity and size on chromosome 5. Indicated threshold levels for stature. Threshold levels for body capacity and size not shown, because they were close to the threshold levels for stature. d) Test statistic profile of size and dairy character on chromosome 6. Indicated threshold levels for dairy character. Threshold levels for size not shown, because they were close to the threshold levels for dairy character. e) Test statistic profile of angularity on chromosome 12. f) Test statistic profile of fore udder attachment on chromosome 13. g) Test statistic profile of fore udder attachment and front teat placement on chromosome 19. Indicated threshold levels are for front teat placement. Threshold levels for fore udder attachment were approximately the same and, therefore, are not shown.

Chromosome 2. The peak test statistic for chest width (Figure 2a) on chromosome 2 was at marker BM2113. The peak test statistic for body capacity was greater than the threshold for suggestive linkage, and this peak was also at marker BM2113. Calving ease of the daughters and nonreturn of daughters 56 d postinsemination also showed suggestive linkage for this marker. The highest test statistic for rump width was between markers TGLA110 and BM6444, which is about 24 cM from marker BM2113.

In the within-family analysis, Families 6 and 9 were segregating for a QTL for chest width (Table 4). The allele substitution effect was 1.94 units for Family 6 and 2.71 units for Family 9, which is 0.43 and 0.61 genetic standard deviation units, respectively.

Chromosome 4. The largest test statistic for gestation length on chromosome 4 was between markers TGLA159 and TGLA420 (Figure 2b). No indications were found for the presence of other QTL on this chromosome. One family was highly significant for the within-family analysis. The allele substitution effect was 0.97 d, which is 0.57 genetic standard deviation units (Table 4).

Chromosome 5. The highest test statistics for stature and size were at location 122, which is between markers IGF1 and BM315 (Figure 2c). The peak test statistics for other traits related to body size, such as chest width, body capacity, and birth weight, were also in the region containing marker BM315. Other indications for QTL in this region were detected for udder depth and rear udder height close to marker IGF1, rump width at the end of the chromosome (marker AGLA22), and calving ease in the same region as chest width and body capacity.

The allele substitution effects estimated in families segregating for a QTL were in the range 0.36 to 0.78 genetic standard deviation units (Table 4).

Chromosome 6. The peak test statistic for dairy character was found at the beginning of the chromosome at marker ILSTS090. Peak test statistics for size and stature were found close to this marker in the interval between ILSTS090 and MCM53 (Figure 2d). The allele substitution effect for dairy character in the three families significant for this trait ranged from 0.39 to 1.02 genetic standard deviation units (Table 4).

Chromosome 12. Chromosome 12 seems to contain a QTL that influences angularity. The highest test statistic was found between markers TGLA9 and AGLA226 (Figure 2e). Four families were significant in the within-family analysis; the effect ranged from 0.29 to 0.50 genetic standard deviation units (Table 4).

Chromosome 13. A QTL influencing fore udder attachment was located at the beginning of this chromosome near marker TGLA23 (Figure 2f). Families 6, 9, and 11 were significant at this location. The most significant family at this location was Family 11. The effect estimated in this family was 7.22 units, which is 1.61 genetic standard deviation units (Table 4).

Table 4. Significant families (nominal significance, P < 0.01) and effect sizes from within-family analysis for most significant quantitative trait loci in across-family analysis.[1]

Chromosome	Trait	Family	Effect size[2,3]	Effect size[4]	Significance
2	Chest width	6	1.94 (0.55)	0.43	**
		9	2.71 (0.77)	0.61	*
4	Gestation length	10	0.97 (0.23)	0.57	***
5	Stature	17	1.86 (0.62)	0.42	*
		20	1.76 (0.61)	0.39	*
5	Body capacity	9	3.48 (0.91)	0.78	**
		10	1.60 (0.49)	0.36	*
5	Size	12	2.09 (0.78)	0.47	*
		20	1.85 (0.56)	0.41	*
6	Dairy character	2	4.54 (1.43)	1.02	*
		17	1.75 (0.54)	0.39	*
		19	4.05 (1.09)	0.91	*
12	Angularity	14	2.25 (0.79)	0.50	*
		15	2.06 (0.63)	0.46	*
		17	1.29 (0.48)	0.29	*
13	Fore udder attachment	6	2.83 (0.94)	0.63	*
		9	4.01 (1.05)	0.90	**
		11	7.22 (1.84)	1.61	**
19	Fore udder attachment	3	3.91 (0.93)	0.87	**
19	Front teat placement	11	2.86 (0.89)	0.64	*
		17	1.11 (0.42)	0.25	*

[1]Effect size given for position of highest test statistic in across-family analysis.

[2]Absolute value of allele substitution effect.

[3]Standard error between brackets.

[4]Expressed in genetic standard deviation units.

*$P < 0.01$. **$P < 0.001$. ***$P < 0.0001$

Chromosome 19. Two udder traits, fore udder attachment and front teat placement, are influenced by QTL located on this chromosome close to marker BMS2503 (Figure 2g). Family 3 had a significant effect of 0.87 genetic standard deviation units on fore udder attachment. Front teat placement was significant in Families 11 and 17, the allele substitution effect being 0.64 and 0.25 genetic standard deviation units, respectively.

DISCUSSION

Regression analysis

The marker map used in the regression analysis had a relatively high marker density. The average marker interval across the genome was 12.8 cM. However, marker density varied between chromosomes because extra genotypes were available for chromosomes on which QTL for production traits were found in previous studies (Spelman et al., 1996; Arranz et al., 1998; Coppieters et al., 1998).

Results were considered significant when a genomewise significance threshold of 10% was exceeded. Some other studies (e.g. Spelman et al., 1996; Zhang et al., 1998) have used experimentwise significance thresholds that account for multiple testing. These studies have considered not only 29 chromosomes but also a number of independent traits. The number of independent traits, however, varies between studies, which makes comparison of results very difficult. To avoid this problem, genomewise thresholds were used, i.e., the thresholds were not adjusted for the number of traits.

The most significant results, with test statistics exceeding the 10% genomewise level, can be grouped into 4 categories. The QTL for traits related to body size were located on chromosomes 2, 5 and 6; QTL for udder traits were located on chromosomes 13 and 19; a QTL for gestation length was located on chromosome 4; and a QTL for angularity was located on chromosome 12.

On chromosome 2, the most significant QTL was for chest width (Table 3). In the same region, QTL for body capacity, rump width, calving ease, and nonreturn of daughters were located. Chest width, body capacity and rump width are related to size of the animal. Body size could have an influence on calving ease. A QTL for chest width, body capacity, and rump width thus might have a pleiotropic effect on calving ease. It should be noted that most data on chest width were derived from other traits, because chest width was not measured before 1996. Traits used to calculate chest width were stature, body capacity, and rump width.

Table 5. Results from studies on detection of quantitative trait loci (QTL) for conformation and functional traits.

Chromosome	Trait	Study[1]	Marker or marker bracket	Position[2]	Position[3]	Significance[4]
1	Somatic cell score	G	MAF46		158	
4	Somatic cell score	F	RM188-TGLA116	43	78-104	***
5	Somatic cell score	C		156		*
6	Somatic cell score	C		84		***
8	Somatic cell score	G	BM3419		48	
9	Foot angle	D	BM4204		59	
9	feet & legs	D	BM4204		59	
11	Somatic cell score	C		46		*
13	Somatic cell score	F	TGLA381-AGLA232	91	71-117	***
13	Somatic cell score	C		104		***
14	Somatic cell score	F	ILSTS11-BM302	21	24-53	***
15	Somatic cell score	C		36		*
16	Somatic cell score	C		30		*
16	Herdlife	E	BM719		111	
18	Somatic cell score	B	BM2078		105	**
19	Somatic cell score	C		64		*
21	Somatic cell score	C		78		
23	Somatic cell score	A	513			*
23	Udder depth	D	513			
23	Somatic cell score	D	513			
23	Herdlife	B	BM1258		21	*
23	Udder depth	D	BM1258		21	
23	Somatic cell score	D	BM1258		21	
23	Somatic cell score	G	RM033		11	
26	Somatic cell score	F	TGLA429-BM804	72	52-59	***
26	Somatic cell score	C		78		*
27	Dairy form	D	BM203		51	

[1]A = Ashwell et al. (1996); single marker regression, across families, 17 markers; B = Ashwell et al. (1997); single marker regression, across families, 16 markers; C = Boichard and Bishop (1997); multimarker regression, across families, 220 markers; D = Ashwell et al. (1998a); single marker regression, within families, 16 markers; E = Ashwell et al. (1998b); single marker regression, within families, 20 markers; F = Zhang et al. (1998); multimarker regression, across families, 206 markers; and G = Reinsch et al. (1998); single marker regression, across families, 45 markers.

[2]Position on the marker map used in the study in which the QTL was detected.

[3]Position on the marker map used in the current study. Indicated only when possible to derive.

*P <= 0.05. **P <= 0.01. ***Above threshold for suggestive linkage (experimentwise, correcting for number of chromosomes and number of traits analyzed).

On chromosome 5, QTL for stature, chest width, body capacity, rump width, and size were found between 122 and 181 cM (Table 3). All traits are related to size of the animal. The QTL for calving ease (at 166 cM) and birth weight (at 132 cM) were located in the same region (Table 3), so a QTL for size- related traits might also affect calving ease and birth weight.

The QTL for stature, size, and dairy character were located in the same region (0 to 11 cM) on chromosome 6. Although the trait dairy character is not a size trait itself, it is strongly correlated with size and stature. Based on estimated breeding values of young bulls sampled from 1987 to 1993 and the repeatabilities of these estimated breeding values, the genetic correlation between stature and dairy character was around 0.80. As on chromosomes 2 and 5, a QTL for calving ease at 44 cM was relatively close to QTL for traits related to size (Table 3).

The most significant QTL on chromosome 13 at 0 cM affected fore udder attachment (Table 3). A QTL for udder depth was located in the same region (23 cM). On chromosome 19, QTL for udder depth, fore udder attachment, and front teat placement were located between markers at 34 and 68 cM. Udder depth and fore udder attachment have a relatively high genetic correlation. Based on estimated breeding values of young bulls sampled from 1987 to 1993 and the repeatabilities of these estimated breeding values, the genetic correlation between fore udder attachment and udder depth was around 0.70.

Literature

Only a few studies on QTL affecting conformation and functional traits in dairy cattle have been published. Results related to the present study are presented in Table 5 and will be discussed here. Exact location of QTL in the present study and in the studies summarized in Table 5 is not known, which makes comparison difficult. It is possible, however, to indicate corresponding results. It should be noted that the studies presented in Table 5 are not completely independent. In the studies by Ashwell et al. (1996, 1997, 1998a, and 1998b), a granddaughter design consisting of seven large US Holstein families was used to detect QTL for milk production and composition, health, and conformation traits. The studies were different with respect to the microsatellite markers or the traits. A single-marker approach was used to detect QTL. In the first two studies (Ashwell et al., 1996; Ashwell et al., 1997), analysis was across families, whereas in the latter two studies (Ashwell et al., 1998a; Ashwell et al., 1998b), only an analysis within families was carried out. In two other studies (Boichard and Bishop, 1997; Zhang et al., 1998), results are presented from a granddaughter design consisting of 14 grandsires with 1794 sons. The main differences between

these studies were an additional trait, productive herd life, in the analysis by Zhang et al. (1998) and the statistical method used to analyze the data. Zhang et al. (1998) used variance components and least-squares methods, whereas Boichard and Bishop (1997) used only least-squares methods.

Somatic cell score and udder traits. Many QTL for somatic cell score were detected (Table 5), partly because somatic cell score was analyzed in six of the seven studies presented in Table 5, whereas conformation traits were analyzed in only two studies. In our study, suggestive QTL for somatic cell count were located on chromosome 3, near marker BMC5227, and at location 70 on chromosome 18, between markers BM7109 and ILSTS002 (Table 3). This location is about 35 cM away from marker BM2078, where a QTL for somatic cell score was detected by Ashwell et al. (1997).

Other reported QTL for somatic cell score (Table 5) were not confirmed in our study. There were, however, indications for QTL near marker BM1258 on chromosome 23, affecting milking speed and fore udder attachment (Table 3). Both traits are genetically correlated with somatic cell score and mastitis resistance (de Jong and Lansbergen, 1996). In two studies (Ashwell et al., 1996; Ashwell et al., 1998a), a QTL for somatic cell score was located on chromosome 23, near markers 513 and BM1258. Reinsch et al. (1998) also reported a QTL for somatic cell count on chromosome 23, near marker RM033. This marker is about 10 cM from marker BM1258.

On chromosome 19, a QTL for somatic cell score was detected at 64 cM by Boichard and Bishop (1997). In our study, a QTL for udder depth was located at 34 cM, near markers TGLA94 and URB044. The QTL for front teat placement and fore udder attachment were also located on chromosome 19, near markers BMS2503 and BMS650, at 67 to 68 cM (Table 3). These udder traits are genetically correlated with mastitis resistance and somatic cell score (de Jong and Lansbergen, 1996), and the position of QTL for these traits might be similar to the position for a QTL for somatic cell score reported by Boichard and Bishop (1997).

Other conformation traits. In two studies (Ashwell et al., 1998a; Ashwell et al., 1998b), a within-family analysis was carried out to detect QTL for conformation traits. Most significant QTL were detected for foot angle and for feet and legs on chromosome 9, udder depth on chromosome 23, and dairy form on chromosome 27 (Table 5). Spelman et al. (1999) identified a QTL for stature on chromosome 14. These results were not confirmed in our study.

Experimental Power

The size of the effects found in this study was generally about 0.5 to 1 genetic standard deviation units. The QTL effect can be overestimated when the test statistic exceeds a certain

significance threshold, especially when power to detect QTL is low (Wang, 1995). Power of an experiment to detect QTL increases with increasing heritability of the trait. Heritability of analyzed traits ranged from 0.02 (nonreturn of daughters) to 0.60 (stature and size). Power was calculated for a similar design as used in this study and is given in Table 6 for various sizes of QTL effect, heritability, and frequency of the favorable QTL allele. Power is close to 1 for a QTL of size 1.0 σ_G (where σ_G = genetic standard deviation) and an allele frequency of 0.5, when heritability is 0.35 or 0.60. Although this situation is not very realistic, it gives an indication of the power of the design in extreme cases. It is hard to detect QTL in this experiment with an effect less than 0.5 σ_G, unless the heritability of the trait and the frequency of the QTL allele are moderate to high (Table 6).

Table 6. Power of a granddaughter design consisting of 20 grandsires with, on average, 42 sons, each having 100 daughters with information.[1]

Heritability	Frequency QTL allele	Size of QTL-effect (σ_g)	Power
0.02	0.20	0.1	0.04
		0.4	0.10
		1.0	0.68
	0.50	0.1	0.04
		0.4	0.16
		1.0	0.90
0.35	0.20	0.1	0.04
		0.4	0.35
		1.0	0.99
	0.50	0.1	0.05
		0.4	0.58
		1.0	1.00
0.60	0.20	0.1	0.05
		0.4	0.38
		1.0	0.99
	0.50	0.1	0.05
		0.4	0.61
		1.0	1.00

[1]The type I error rate was assumed to be 0.0345, recombination rate between a marker and the quantitative trait locus (QTL) was 0.05.

Marker-Assisted Selection

This study suggested the existence of many QTL for conformation and functional traits. The markers associated with the most significant QTL can be used in MAS for these traits. Whether or

not results need to be confirmed before starting MAS, depends on how MAS is applied. Overestimation of the QTL effect would reduce long-term response, but this reduction is less when the QTL effect is reestimated after four generations of MAS (Spelman and van Arendonk, 1998), using BLUP-methods, and confirmation would not be necessary.

When MAS is applied to conformation and functional traits, problems may arise if there are negative effects on other economically important traits, such as production. Some of the QTL reported in this study were located on chromosomes that were reported to contain QTL for production (Georges et al., 1995; Spelman et al., 1996). Before starting MAS for the reported QTL, the effects on other traits should be investigated.

CONCLUSIONS

Chromosomes 2, 5, and 6 may contain QTL for traits related to body size. These QTL may also have an effect on calving ease. Chromosomes 13 and 19 possibly contain QTL for udder traits that may also affect somatic cell score and mastitis resistance. Other indications for QTL were for gestation length on chromosome 4 and angularity on chromosome 12. If there are no negative effects on other economically important traits, MAS with markers associated with these QTL can be applied.

ACKNOWLEDGMENTS

We thank Richard Spelman for providing the regression program and helpful discussions. Genotypes were produced by Paulette Berzi, Nadine Cambisano, Latifa Karim, Fabienne Marcq, Myriam Mni, Laurence Moreau, Carine Nezer, Andreina Shoeberlein, Patricia Simon, Pascal Vanmanshoven, and Danny Wagenaar, Department of Genetics, University of Liège (Liège, Belgium). Their work is greatly appreciated. Holland Genetics and Livestock Improvement are acknowledged for financial support.

REFERENCES

Arranz, J.-J., W. Coppieters, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L. Moreau, C. Nezer, J. Riquet, P. Simon, P. Vanmanshoven, D. Wagenaar, and M. Georges, 1998. A QTL affecting milk yield and composition maps to bovine chromosome 20: a confirmation. Anim. Genet. 29:107-115.

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

Ashwell, M. S., Y. Da, C. P. van Tassell, P. M. vanRaden, R. H. Miller, and C. E. Rexroad, Jr., 1998. Detection of putative loci affecting milk production and composition, health and type traits in a United States Holstein Population. J. Dairy Sci. 81:3309-3314.

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

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

Boichard, D., and M. D. Bishop, 1997. Detection of QTLs influencing milk production and mastitis resistance with a granddaughter design in Holstein cattle. Paper presented at 48th Annu. Mtg. of the Eur. Assoc. of Anim. Prod., Vienna, Austria.

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

CR Delta, 1999. Jaarstatistieken 1998. CR Delta, Arnhem, The Netherlands.

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

De Jong, G., and L. Lansbergen, 1996. Mastitis resistance index: selection for udder health. Proc. of the Open Session of the INTERBULL Annu. Mtg., Veldhoven, The Netherlands. INTERBULL Bull. 14:81-90.

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

Henderson, C. R.,1984. Linear models in animal breeding. University of Guelph Press, Guelph, Canada.

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

Knott, S. A., J. M. Elsen, and C. S. Haley, 1994. Multiple marker mapping of quantitative trait loci in half sib populations. Proc. 5th World Congr. Genet. Appl. Livest. Prod., Guelph, Canada 21:33-36.

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

Lander, E. S., and P. Green, 1987. Construction of multilocus genetic linkage maps in humans. Proc. Natl.

Acad. Sci. USA. 84:2363-2367.

Lander, E. S., and L. Kruglyak, 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat. Genet. 11:241-247.

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

Reinsch, N., N. Xu, H. Thomsen, C. Looft, E. Kalm, S. Grupe, C. Kühn, M. Schwerin, B. Leyhe, S. Hiendleder, G. Erhard, I. Medjugorac, I. Russ, M. Forster, B. Brenig, R. Reents, and G. Averdunk, 1998. First results on somatic cell count loci from the ADR bovine mapping project. Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, Australia 26:426-428.

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

Spelman, R. J., 1998. Detection and utilisation of quantitative trait loci in dairy cattle. Ph.D. Diss., Wageningen Agricultural University, Wageningen, The Netherlands.

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

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

Spelman, R. J., and J. A. M. Van Arendonk, 1998. Effect of inaccurate parameter estimates on genetic response to marker assisted selection in an outbred population. J. Dairy Sci. 80: 3399-3410.

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

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

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

CHAPTER 4

Detection of Pleiotropic Effects of Quantitative Trait Loci in Outbred Populations Using Regression Analysis

C. Schrooten[1,2] and H. Bovenhuis[1]

[1]Animal Breeding and Genetics Group, Wageningen University, PO Box 338, 6700 AH Wageningen, The Netherlands
[2]Holland Genetics, CR Delta, PO Box 5073, 6802 EB Arnhem, The Netherlands

Journal of Dairy Science 85: 3503 – 3513 (2002)

ABSTRACT

In this paper a method is presented to determine pleiotropic quantitative trait loci (QTL) or closely linked QTL in an outbred population. The method is based on results from single-trait analyses for different traits and is derived for a granddaughter design. The covariance between estimated contrasts of grandsires obtained in single-trait regression analysis is computed. When there is no pleiotropic QTL, the covariance between contrasts depends on the heritabilities of the traits involved, the polygenic and environmental correlation between the traits, the phenotypic standard deviations, the number of sires per grandsire, and the number of daughters per sire. A pleiotropic QTL results in a covariance that deviates from this expected covariance. The deviation depends on the size of the effects on both traits and on the fraction of grandsires heterozygous for the QTL. When analyzing experimental data, the expected covariance and the confidence interval for the expected covariance can be determined by permutation of the data. A covariance outside the confidence interval suggests the presence of a pleiotropic QTL or a closely linked QTL. The method is verified by simulation and illustrated by analyzing an experimental data set on chromosome six in dairy cattle.

(**Key words**: pleiotropy, outbred population, regression analysis, quantitative trait loci)

Abbreviation key: **MAS** = marker assisted selection, **QTL** = quantitative trait locus

INTRODUCTION

Genetic improvement of economically important traits in plant and animal species is accomplished by selection based on phenotypic information. Even though the nature of genes influencing these traits is in general not known, selection in this way has been very successful (e.g., Philipsson et al., 1994; Rishell, 1997; Webb, 1998). In the past decade, molecular techniques have become available that allow large-scale genotyping of genetic markers (Gyapay et al., 1996). These markers can be used to map genes affecting quantitative traits (quantitative trait loci, QTL) and many QTL detection studies have been carried out in plant and animal species (e.g., Jansen and Stam, 1994; Georges et al., 1995). This is a first step toward identification and characterization of genes and toward the development of novel selection methods utilizing both phenotypic and genomic information.

Selection for a certain trait can lead to correlated responses for other traits. The main cause for the existence of a genetic correlation between traits in outbred populations is pleiotropy (Falconer

and Mackay, 1996), i.e., genes affect more than one trait. Marker-assisted selection (MAS) for a chromosomal region containing genes with pleiotropic effects will thus have consequences for other traits. To avoid negative side effects of selection for certain chromosomal regions or to exploit positive effects of MAS on other traits of interest, the presence of pleiotropic effects should be studied.

The ultimate goal of QTL detection studies is to locate the mutation responsible for the observed effect. This would not only give insight into the biological background of traits but would also allow direct selection for the favorable alleles. However, resolution of QTL detection studies is such that confidence intervals extend over large chromosomal regions. It is expected that regions of this size contain hundreds of genes, as in the human genome (Lander et al., 2001). Further experiments may reduce the length of the confidence intervals, but it is likely that these regions contain many potential candidate genes. Characterization of the QTL based upon their pleiotropic effects will provide additional clues in the identification of candidate genes. Based on the proteins or other metabolites involved in the expression of each of those traits, selection of the most likely candidates from a list of candidate genes in that region is possible. Thus, knowledge about pleiotropic effects could speed up the identification of the responsible gene.

Most QTL-detection methods are based on single-trait analysis, either using regression or variance component methods (Fernando and Grossman, 1989; Knott et al., 1994; Grignola et al., 1997). However, methods for multiple-trait analysis have been developed (e.g., Jiang and Zeng, 1995; Weller et al., 1996; Knott and Haley, 2000; Hackett et al., 2001; Korol et al., 2001). Jiang and Zeng (1995) mentioned three advantages of multiple-trait methods over single-trait methods: multiple-trait methods have higher power, higher precision of parameter estimation, and give potentially more insight into the nature of genetic correlations between different traits. They presented a multiple-trait method based on a maximum-likelihood approach. Weller et al. (1996) used canonical transformation and subsequently single-trait analysis to look at multiple traits. Knott and Haley (2000) developed a multitrait least-squares method for a three-generation pedigree with fixation of the QTL in the grand-parental generation. This method can be extended to other population structures. Korol et al. (2001) use a transformation of the trait space followed by single-trait analysis and subsequently back transformation. The specific feature of their method is that the multivariate transformations are interval specific. Hackett et al. (2001) extended the regression approach for crosses between lines by Haley and Knott (1992) to a multitrait analysis via multivariate regression and applied this to a doubled haploid population of barley. The application of these methods, however, is not yet common practice. They can present computational

difficulties, especially when they would be applied to an outbred population or if the number of traits is large.

In the present study, a simple and fast method for detecting pleiotropic QTL or closely linked QTL in an outbred population is presented. The method is based on the covariance between marker contrasts from single-trait regression analysis for different traits. First, the theoretical expectation of this covariance is derived. Second, simulation is used to verify the theoretically expected covariances and to determine the power of detecting pleiotropic effects. Third, the method is applied to data on chromosome 6 in dairy cattle.

MATERIALS AND METHODS

Covariance Between Marker Contrasts

Consider a granddaughter design (Weller et al., 1990) and the following statistical model to analyze the data:

$$y_{ijkl} = gs_i + m_{ij} + s_{ijk} + e_{ijkl} \qquad (1)$$

where y_{ijkl} is the phenotype of animal l, daughter of sire k that inherited marker allele j of grandsire i; gs_i is the effect of grandsire i; m_{ij} is the effect of marker allele j nested within grandsire i; s_{ijk} is the effect of sire k that received marker allele j from grandsire i, and e_{ijkl} is the random residual. For each grandsire, two groups of animals can be considered: group 1 consisting of sires that inherited marker allele 1 of the grandsire and group 2 consisting of sires that inherited marker allele 2. The number of sires per grandsire is n_s, and it is assumed that these sires are equally distributed over marker groups 1 and 2. Each sire has n_g daughters. The contrast between marker allele 1 and marker allele 2 within grandsire i can be written as $\overline{y_{i1..}}$ - $\overline{y_{i2..}}$. The variance of the contrast can be written as

$$\mathrm{var}(\overline{y_{i1..}} - \overline{y_{i2..}}) = \left\{ \frac{4 - h^2}{n_s * n_g} + \frac{\frac{3}{4} h^2}{n_s} \right\} * \sigma_p^2 \qquad (2)$$

where h^2 is the heritability and σ_p^2 is the phenotypic variance of the trait (Weller et al., 1990; van der Beek et al., 1995). Analogously, if we consider two traits, say trait v and trait w, the covariance between the marker contrasts for trait v and trait w, for a polygenic situation, is equal to

$$\mathrm{cov}(\overline{v_{i1..}} - \overline{v_{i2..}}, \overline{w_{i1..}} - \overline{w_{i2..}}) = \left\{ \frac{3 r_A h_v h_w + 4 r_E \sqrt{1-h_v^2}\sqrt{1-h_w^2}}{n_s * n_g} + \frac{\frac{3}{4} r_A h_v h_w}{n_s} \right\} * \sigma_{p_v} * \sigma_{p_w} \qquad (3)$$

where r_A is the genetic correlation between trait v and trait w, r_E is the environmental correlation between trait v and trait w, h_v^2 and h_w^2 are heritabilities of trait v and trait w, respectively, and σ_{p_v} and σ_{p_w} are the phenotypic standard deviations of trait v and trait w, respectively. When trait v and trait w are the same traits, i.e., substituting $r_A=r_E=1$, $h_w=h_v$ and $\sigma_{p_v} = \sigma_{p_w}$, equation 3 is equal to equation 2. The derivation of equation 3 can be found in the appendix.

As mentioned, the derived covariance of contrasts applies to a polygenic situation. Now consider a bi-allelic QTL, with a fraction p_{het} of the grandsires heterozygous for the QTL. A fraction $1-p_{het}$ is homozygous for the QTL, and in that case the expected contrast between the first and the second grandsire-allele is zero. For the heterozygous grandsires, the contrast between the first and the second grandsire-allele is either positive or negative. In both cases, the absolute size of the contrast is ½a, where "a" is half the difference between the two homozygous genotypes (Falconer and Mackay, 1996). The variance of contrasts within each of the three groups is given by equation 2. In the absence of dominance effects, the overall variance of the contrast at the QTL can be written as

$$\mathrm{var}(\overline{y_{i1..}} - \overline{y_{i2..}}) = \frac{\sum_{k=1}^{3} \sigma_{p_k}^2}{3} + \sum_{k=1}^{3} p_k (\mu_k - \mu)^2 = \left\{ \frac{4 - h_y^2}{n_s * n_g} + \frac{\frac{3}{4} h_y^2}{n_s} \right\} * \sigma_p^2 + \frac{1}{4} p_{het} a^2, \qquad (4)$$

where k refers to the group (either homozygous, in coupling phase or in repulsion phase), p_k is the fraction of animals belonging to group k, $\sigma_{p_k}^2$ is the variance of contrasts within group k, μ_k is the mean of group k, and μ is the overall mean. It should be noted that in equation 4, the heritability and the phenotypic variance are the heritability and the phenotypic variance exclusive of the QTL effect. The variance caused by the QTL effect is covered by the term $\frac{1}{4} p_{het} a^2$.

Analogously, the covariance between contrasts for two traits, trait v and w, can be written as

$$\mathrm{cov}(\overline{v_{i1..}} - \overline{v_{i2..}}, \overline{w_{i1..}} - \overline{w_{i2..}}) = \left\{ \frac{3 r_A h_v h_w + 4 r_E \sqrt{1-h_v^2}\sqrt{1-h_w^2}}{n_s n_g} + \frac{\frac{3}{4} r_A h_v h_w}{n_s} \right\} \sigma_{p_v} \sigma_{p_w} + \frac{1}{4} p_{het} a_v a_w \quad (5)$$

where a_v and a_w are equal to half the difference between the two homozygous genotypes, for trait v and trait w, respectively. Heritabilities, variances and correlations in equation 5 do not include the effect of the QTL. Equation 5 is similar to equation 3, with the addition of the term $\frac{1}{4}p_{het}a_v a_w$. This term represents the contribution of the QTL to the covariance. If either a_v or a_w is equal to zero, the QTL will not contribute to the covariance of the marker contrasts. Only if both a_v and a_w differ from zero, the covariance is affected. The change in covariance due to the QTL depends on the size and the sign of the QTL effect on each trait and on the fraction of grandsires heterozygous for the QTL. The covariance of marker contrasts can be calculated at each location along the chromosome and will deviate for chromosomal regions containing a pleiotropic QTL or closely linked QTL.

Hypothesis testing. The null hypothesis (H_0) is that the QTL affects only one or none of the two traits under consideration. The alternative hypothesis of a pleiotropic QTL is accepted if the covariance of the contrasts deviates significantly from the covariance under H_0. The distribution of the covariance under H_0 is determined by permutation of the data, as first suggested by Churchill and Doerge (1994) for analysis of a single trait. In the current study, pairs of observations on different traits are permuted within families; each pair of trait values is randomly assigned to another member of the same family. In this way, covariances due to unlinked QTL are not affected. After each permutation, the minimum and maximum covariance between marker contrasts on the chromosome are retained. All minimum and maximum values per chromosome are ranked, and the chromosomewise thresholds are determined. In this way, confidence intervals are obtained for the covariance between the contrasts when there is no association between phenotypes and genotypes (i.e., no QTL, equivalent to equation 3).

Simulated data. Stochastic simulation was used: 1) to verify the theoretically derived covariances for various situations; 2) to calculate the power of QTL detection by single-trait analysis; and 3) to calculate the power to detect pleiotropic QTL by evaluating the covariance between marker contrasts for two different traits.

Table 1. Probability (Prob.) to detect QTL and pleiotropic QTL, covariances before and after permutation and confidence intervals in each alternative (Altern.).

Altern.[1]	r_A, r_P	a_v	a_w	Prob.[2] QTL, trait v	Prob.[3] QTL, trait w	Prob.[3] pleio-tropic QTL[4]	Expec-ted covar. At QTL[5]	Covar. at QTL[6]	[7]Expec-ted covar. after permu-tation	Covar.[8] after permu-tation	Lower[8] threshold	Upper[8] threshold
1	-0.6	0	0	0.06	0.06	0.05	-0.45	-0.46	-0.45	-0.48	-1.25	0.01
2	-0.6	0.5	0	0.87	0.06	0.20	-0.45	-0.47	-0.45	-0.48	-1.28	0.04
3	-0.6	0.5	0.3	0.86	0.32	0.91	0.40	0.44	-0.40	-0.42	-1.20	0.10
4	-0.6	0.5	-0.3	0.87	0.32	0.61	-1.31	-1.35	-0.51	-0.54	-1.38	-0.01
5	0	0	0	0.04	0.05	0.05	0.00	-0.00	0.00	0.00	-0.56	0.56
6	0	0.5	0	0.88	0.05	0.23	0.00	0.00	0.00	0.00	-0.59	0.59
7	0	0.5	0.3	0.86	0.31	0.81	0.86	0.89	0.05	0.06	-0.53	0.68
8	0	0.5	-0.3	0.88	0.34	0.80	-0.86	-0.91	-0.05	-0.06	-0.68	0.53
9	0.6	0	0	0.04	0.05	0.05	0.45	0.46	0.45	0.48	-0.01	1.25
10	0.6	0.5	0	0.84	0.05	0.21	0.45	0.46	0.45	0.48	-0.04	1.28
11	0.6	0.5	0.3	0.87	0.35	0.61	1.31	1.34	0.51	0.54	0.01	1.38
12	0.6	0.5	-0.3	0.89	0.29	0.90	-0.40	-0.43	0.40	0.43	-0.10	1.21

[1]Alternatives are characterized by size of the simulated effects for trait v and trait w (a_v and a_w, respectively), expressed in genetic standard deviation units, and by the phenotypic (r_P) and genetic correlation (r_A).

[2]Type 1 error in alternatives 1, 5 and 9, power in other alternatives.

[3]Type 1 error in alternatives 1, 2, 5, 6, 9 and 10, power in other alternatives.

[4]Determined as the sum of the fraction above the upper threshold and the fraction below the lower threshold.

[5]Calculated using Equation 5.

[6]Average of 1,000 simulated datasets.

[7]Calculated using Equation 3, with parameters adjusted to account for additional variance and covariance caused by the quantitative trait locus (QTL). Adjusted parameters are in Table 2.

[8]Average of 1,000 simulated datasets and 2,000 permutations for each dataset. Presented thresholds are for 95% confidence interval.

Twelve different alternatives were simulated. Each alternative consisted of a granddaughter design with 20 grandsires, 50 sires per grandsire, and 100 daughters per sire. For each grandsire, a chromosome of length 100 cM with 21 equally spaced markers was constructed. Markers were fully informative. Polygenic components for grandsires were generated from normal distributions for trait v as well as for trait w. Heritability for trait v was 0.6 and for trait w the heritability was 0.35. The phenotypic standard deviation for trait v as well as trait w was equal to 10. These heritabilities and

variances do not include a QTL effect. Polygenic components for sires and their daughters were generated based on half the polygenic component of their sire and a randomly generated term accounting for the contribution of the dam and the Mendelian sampling term. Polygenic components for the second trait were generated conditional on the polygenic components for the first trait, i.e., accounting for the genetic correlation between both traits. A QTL was positioned at 30 cM. The probability that a grandsire was heterozygous for the QTL was 0.5. Marker genotypes for the sires were derived, based on genotypes of their sire and distance between the markers, accounting for recombination. Recombinations were generated using Haldane's mapping function. Phenotypic records of the daughters for trait v and trait w were generated based on the polygenic component, a random environmental component, and the QTL contribution. Environmental components were generated accounting for the environmental correlation between the two traits. The QTL alleles of the daughters were simulated based on the transmitted QTL alleles from the grandsires and the assumption that both alleles of the biallelic QTL were present in the population at equal frequencies. Genetic correlations were either –0.6 (alternatives 1 to 4, Table 1), zero (alternatives 5 to 8), or +0.6 (alternatives 9 to 12). Phenotypic correlations were assumed equal to the genetic correlations. In the alternatives where a QTL was present, the effect on trait v was 0.5 σ_a, where σ_a is the genetic standard deviation for the polygenic effects. This QTL explained 7% of the overall phenotypic variance for trait v. The effect on trait w was either 0 or 0.3 σ_a, explaining 1.6% of the phenotypic variance. When both traits were affected by the QTL, the effect on both traits was either antagonistic (alternatives 3, 7, and 11) or synergistic (alternatives 4, 8, and 12). In alternatives 1, 5, and 9, no QTL was simulated.

For each alternative, 1000 independent data sets were generated. To allow for analysis of phenotypes of sires based on phenotypes of their daughters and analysis at every cM along the chromosome, model (1) was replaced by the multi-marker regression method described by Knott et al. (1994),

$$Y_{ij} = \mu + gs_i + b_{ik}X_{ijk} + e_{ijk}, \qquad (6)$$

where Y_{ij} = daughter yield deviations (VanRaden and Wiggans, 1991) of sire j, son of grandsire i, based on 100 daughters, μ = overall mean, gs_i = fixed effect of grandsire i, b_{ik} = regression coefficient for grandsire i at position k on the chromosome, X_{ijk} = probability that sire j receives a chromosomal segment from grandsire i at position k, and e_{ijk} = random residual.

In each family, estimates of the contrast between allelic effects, the b-values in equation 6, were obtained at every centimorgan along the chromosome, for both traits. Covariances based on the contrasts of the 20 grandsire families in the analysis were computed at every centimorgan. To determine whether results from simulation were in agreement with the derived equations, the average covariance at the simulated location of the QTL, based on 1000 data sets, was compared to the covariance according to equation 5. Besides, the average covariance after permutation was compared to the covariance according to equation 3, with parameters adjusted to account for additional variance caused by the QTL. Adjusted parameters are in Table 2.

Table 2. Adjusted parameters to be used in Equation 3 when computing the covariance between contrasts after permutation when a QTL is present; parameters need to be adjusted for additional variance and covariance caused by the QTL.

Alternative[1]	r_A, r_P[2]	a_v	a_w	r_A[3]	r_P[3]	h^2_v[3]	h^2_w[3]
1	-0.6	0	0	-0.60	-0.60	0.60	0.35
2	-0.6	0.5	0	-0.57	-0.58	0.63	0.35
3	-0.6	0.5	0.3	-0.48	-0.54	0.63	0.36
4	-0.6	0.5	-0.3	-0.62	-0.61	0.63	0.36
5	0	0	0	0	0	0.60	0.35
6	0	0.5	0	0	0	0.63	0.35
7	0	0.5	0.3	0.07	0.03	0.63	0.36
8	0	0.5	-0.3	-0.07	-0.03	0.63	0.36
9	0.6	0	0	0.60	0.60	0.60	0.35
10	0.6	0.5	0	0.57	0.58	0.63	0.35
11	0.6	0.5	0.3	0.62	0.61	0.63	0.36
12	0.6	0.5	-0.3	0.48	0.54	0.63	0.36

[1]Alternatives are characterized by size of the simulated effects for trait v and trait w (a_v and a_w, respectively), expressed in genetic standard deviation units, and by the phenotypic (r_P) and genetic correlation (r_A).

[2]Genetic and phenotypic correlation based on polygenic components, assuming no quantitative trait locus (QTL).

[3]Genetic and phenotypic correlation and heritabilities for trait v and trait w corrected for additional variance and covariance caused by the QTL.

To determine the sensitivity of the method for the number of heterozygous grandsires, alternatives 2 and 3 were also studied with 10 and 5 families instead of 20. For the same reason, alternatives 2 and 3 were also studied with 20 families and 25% heterozygous grandsires instead of 50%.

Power. To test the hypothesis that there is no QTL affecting a certain trait, 2000 permutations were applied to each data set. A QTL was considered significant when the test statistic exceeded the 5% chromosomewise threshold. The probability of detecting a QTL for individual traits was obtained by determining the fraction of simulated datasets where the null hypothesis was rejected. When a QTL is simulated, this probability is the power of QTL detection. When no QTL is simulated, this probability is the type I error. To test the hypothesis that the QTL had no pleiotropic effect, the 95% confidence interval for the covariance under the null hypothesis was determined, based on 2000 permutations of each data set. The probability of detecting a QTL with an effect on both traits was computed as the fraction of datasets resulting in a covariance outside the 95% confidence interval under the null hypothesis. When a QTL with effect on both traits is simulated, this probability is the power to detect a pleiotropic QTL. In case the simulated QTL affects only one of the traits, or if there is no QTL, this probability is the type I error.

Experimental data. The method was illustrated using data from a granddaughter design in dairy cattle, consisting of 20 grandsires and 833 sires. The number of sires per grandsire varied from 11 to 147. Genotypes for 29 microsatellite markers on chromosome 6 were available. Pleiotropic QTL effects were studied considering two traits: milk yield and protein percentage. In this data, a QTL affecting protein percentage was identified previously (Spelman et al., 1996). Further, there was a suggestive QTL for milk yield located in the same chromosomal region as the QTL affecting protein percentage, and a second suggestive QTL for milk yield more than 40 cM away from the first QTL.

Estimated breeding values for protein percentage and milk quantity were converted to daughter yield deviations by deregression, and subsequently analyzed using equation 6 (Spelman et al., 1996). The 95 % confidence interval for the covariance under the null hypothesis was based on 15,000 permutations. Variation in number of daughters per sire was accounted for by applying a weighted regression when using equation 6, as described by Spelman et al. (1996). Variation in the number of sires per family was accounted for by weighing the contrasts using the inverse of the standard error, i.e., contrasts with large standard error had less weight in computation of the covariance.

RESULTS

Simulation. In the simulated alternatives, the covariance between contrasts at the QTL was in agreement with the expected covariance. The difference between expected and realized covariances was never larger than 0.05 (Table 1). In Figure 1, the pattern of the covariance in alternative 3,

averaging over 1000 independent replicates, is presented. The average covariance after permutation and the 95% confidence interval, derived from 2000 permutations of each of 1000 datasets, are indicated as well. The largest difference between realized and expected covariance was found at the location of the QTL. The difference between the covariance and the expected covariance decreased with increasing distance to the QTL.

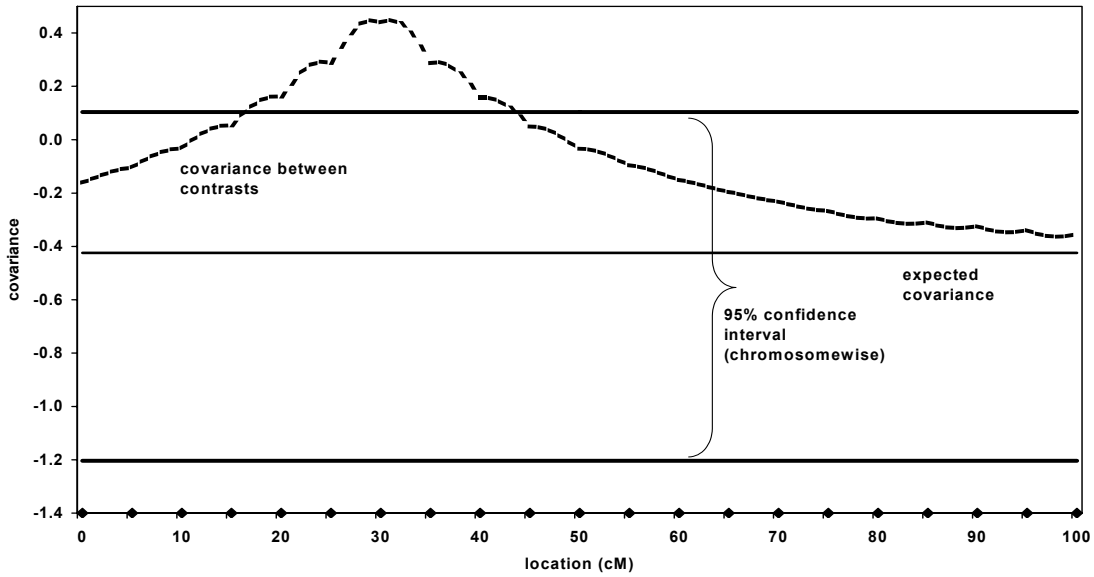

Figure 1. Average covariance and significance thresholds for the 95% confidence interval under the H_0-hypothesis, based on 1,000 replicates of alternative 3, with genetic and phenotypic correlation equal to –0.6. A quantitative trait locus (QTL) with effect 0.5 σ_a on trait v and 0.3 σ_a on trait w was located at 30 cM. Confidence interval (95%, chromosomewise) and expected covariance are based on 2,000 permutations of each dataset and averaged. Position of the markers indicated on the horizontal axis (♦).

Power of QTL detection. Table 1 shows the power of QTL detection, for each simulated alternative and each trait. Power to detect the QTL for trait v, with an effect of 0.5 σ_a, ranged from 0.84 to 0.89, with an average over relevant alternatives of 0.87. In 90% of the cases with a significant QTL affecting trait v, the location of the highest test-statistic was in the interval between 20 and 40 cM, containing the simulated QTL at 30 cM. The power to detect the QTL for trait w, with an effect of 0.3 σ_a, ranged from 0.29 to 0.35, with an average of 0.32. The QTL was positioned in the interval of the simulated QTL, i.e., between 20 and 40 cM, in 66% of the cases in which there was significant evidence for the presence of a QTL. Probability of detecting a QTL when no QTL was simulated (i.e., the type I error), ranged from 0.04 to 0.06, with an average of 0.05 for trait v as well as for trait w, across all relevant alternatives.

Power to detect pleiotropic QTL. In the alternatives where a pleiotropic QTL was simulated, the power to detect a QTL with a pleiotropic effect ranged from 0.61 to 0.91 (Table 1). The power was higher if the covariance due to the QTL and the polygenic and environmental covariance had opposite sign, as in alternatives 3 and 12. If no QTL was simulated, there was a probability of 0.05 to detect a QTL with a pleiotropic effect, which is equal to the expected probability. If a QTL with only an effect on trait v was simulated, the probability of detecting a pleiotropic QTL was 0.20 (alternative 2) to 0.23 (alternative 6), which is higher than the expected probability of 0.05 based on permutation. Threshold levels to determine the power were based on permutations. The average covariance after permutation was in agreement with the expected covariance according to equation 3, with parameters adjusted for additional variance caused by the QTL (Table 2). Size of the confidence intervals ranged from 1.12 in alternative 5 to 1.37 in alternatives 4 and 11. Confidence intervals were larger in alternatives where a QTL-effect was simulated. Results in alternatives 1, 2, 3, and 4 were similar to the results in the corresponding alternatives 9, 10, 12 and 11, respectively.

Number of heterozygous grandsires. Alternatives 2 and 3 were simulated with 10 and 5 families instead of 20, and with 20 families and 25% heterozygous grandsires instead of 50%. These simulations showed, that the method can also be applied if the number of grandsires is less than 20, although the power is reduced (results not shown).

Experimental data. Single-trait analyses revealed a QTL affecting protein percentage and possibly kilograms of milk in the interval between zero and 50 cM. In addition, a suggestive QTL affecting kilograms of milk was found between 75 and 113 cM. The pattern of the covariance between contrasts for kilograms of milk and protein percentage along chromosome 6 is given in Figure 2. At the beginning and the end of the chromosome, the covariance significantly deviates from the expected covariance under the H_0-hypothesis. Figure 2 shows two confidence intervals: one interval is determined by retaining the maximum and minimum value of the covariance on the chromosome in each permutation (chromosomewise thresholds); the other interval is determined by retaining the maximum and minimum value of the covariance at each location separately (nominal thresholds). In the latter case, confidence intervals are not constant along the chromosome.

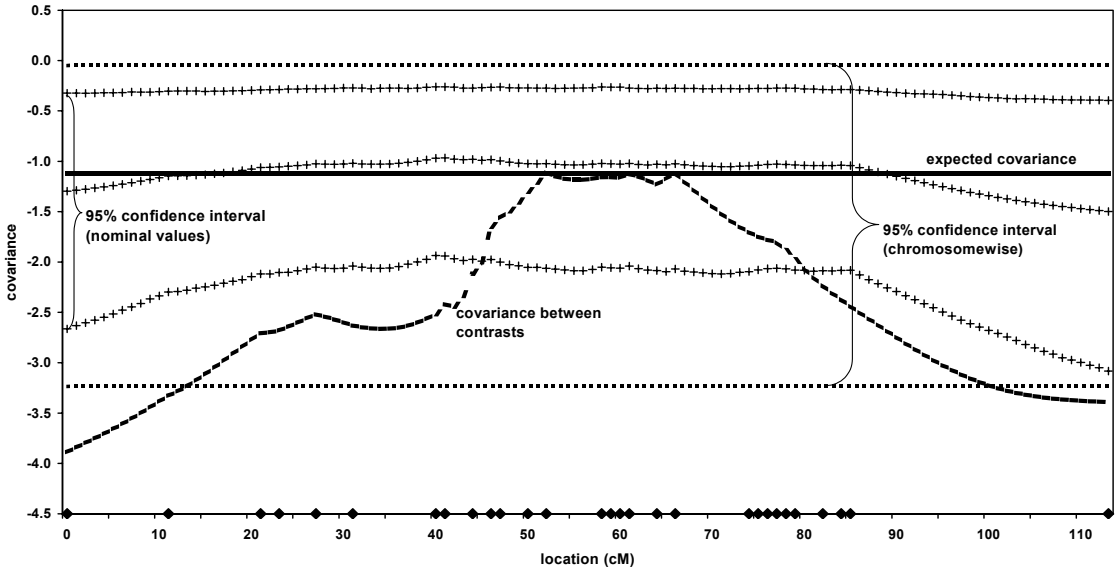

Figure 2. Covariance between contrasts for milk yield and protein percentage on chromosome 6. Confidence interval (95%) and expected covariance under the H_0 –hypothesis (no pleiotropic quantitative trait locus [QTL]), are based on 15,000 permutations of the dataset. Confidence intervals based on chromosomewise and nominal values are indicated in the figure. Expected covariance based on chromosomewise values is indicated by a solid line, expected covariance based on nominal values (based on permutation at each location), is indicated by (+++). Position of the markers indicated on the horizontal axis (♦).

DISCUSSION

The presented method offers an easy way to determine pleiotropic effects of a QTL, as the method makes use of results from single-trait regression analysis. In this paper, the method is derived and is illustrated for a granddaughter design, but the method can also be derived for a two-generation design like a daughter design. Estimated contrasts for grandsires from the regression analysis are used to compute the covariance between these contrasts for different traits at each chromosomal location. Permutation is carried out to determine confidence intervals for the covariance, in order to detect significant deviations from the expected covariance under the null hypothesis of no QTL or a QTL affecting only one trait. The expected covariance in this situation is not necessarily 0, but a function of the genetic and environmental correlation, the heritabilities, the phenotypic variances, and the number of sires and granddaughters involved (equation 3).

Power and Type I Error. In single-trait analysis, power to detect a QTL with an effect of 0.5 σ_a ranged from 0.84 to 0.89. When applying power calculations as presented by Weller et al. (1990), power was 0.93. In alternatives with a QTL effect on trait w of 0.3 σ_a, power ranged from 0.29 to

0.35. Calculation of the power using Weller et al. (1990) gave a power of 0.43. Considering the assumptions underlying the power calculations by Weller et al. (1990), it can be concluded that power based on simulations was in agreement with expectations.

The power to detect a pleiotropic QTL effect was high for the simulated alternatives. If the QTL effect on trait v was 0.5 σ_a and the effect on trait w was 0.3 σ_a, the power to detect the pleiotropic effect ranged from 0.61 (alternatives 4 and 11) to 0.91 (alternative 3). For the alternatives where the QTL had opposite effects on both traits, power ranged from 0.61 (alternative 4) to 0.90 (alternative 12). The highest power was obtained in the alternatives where the covariance due to the QTL and the covariance based on residual genetic and environmental effects had different signs. This indicates that QTL that "break" the overall genetic correlation can be detected more easily. This was also noted by Korol et al. (1995).

It should be noted, however, that the type I error was higher than expected: Type I errors ranged from 0.20 to 0.23. In alternative 2, for example, a type I error of 20% was obtained, where 5% was expected, based on permutation. When no QTL effect was simulated, i.e., in alternatives 1, 5 and 9, type I errors derived from permutation agreed with expected values. The discrepancy between the expected and observed type I error in other alternatives is caused by the fact that permutation does not result in the correct distribution under the null hypothesis in case there is a QTL affecting only one of the traits. This is illustrated in Figure 3 for alternative 2. Figure 3 shows the actual distribution of the covariance (average of 1000 independent datasets) and the distribution of the covariance based on permutation (average of 2000 permutations of these 1000 datasets). Covariances are shown at the QTL. Although expectations of both distributions are equal, the variance of the distribution based on permutations is lower. Consequently, the confidence interval for the covariance based on permutations will be underestimated, resulting in more-than-expected false positives.

Obviously, when analyzing experimental data, this will also result in more false positives than expected based on type I errors set by permutation. Simulation showed that the number of false-positive pleiotropic QTL increases with increasing effect of the QTL. When the QTL-effect was 0.2 σ_a, the observed type I error was 0.10. For QTL effects of size 0.35 σ_a and 0.5 σ_a, the observed type I error was 0.12 and 0.20, respectively. To obtain 5% false positives for a QTL effect of 0.5 σ_a, the type I error based on permutation should be set to 0.2%. Thus, using lower type I errors in constructing confidence intervals can reduce the number of false positives to an acceptable level. This will, however, also reduce the power. To indentify false-positive results, the results from the underlying single-trait analyses could also be considered.

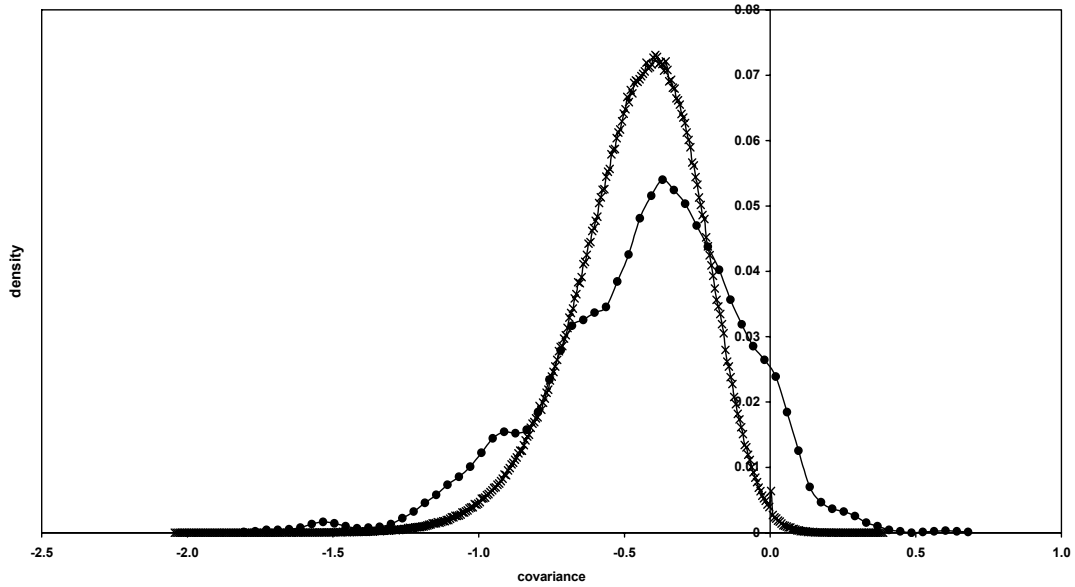

Figure 3. Distribution of covariance at the location of quantitative trait loci (QTL) in simulated (•) and permuted (×) datasets, in alternative 2, with genetic and phenotypic correlation equal to –0.6 and QTL with effect 0.5 σ_a on trait v.

Experimental Data. Based on the pattern of the covariance and the confidence interval in Figure 2, it can be concluded that chromosome 6 contains two pleiotropic QTL, affecting both kilograms of milk and protein percentage. Selection for alleles that increase milk quantity results in a decrease of protein content. This information should be taken into account when deciding whether or not to apply MAS for these QTL.

When confidence intervals for the experimental data were based on nominal values, it appeared that they were not constant along the chromosome. Deviations were observed especially at the beginning and the end of the chromosome. Simulation showed that this effect can be reproduced by simulating data with a low marker density at the beginning and the end of the chromosome. Simulated data containing markers with low information content at the beginning and the end of the chromosome gave similar results. When chromosomewise thresholds are determined for these situations, thresholds will predominantly be influenced by values at the beginning and the end of the chromosome. Therefore, thresholds can be too strict for other regions of the chromosome, resulting in lower power to detect pleiotropic QTL in those regions. This problem can be overcome by adjusting thresholds based on nominal values for repeated testing along the chromosome, i.e., using stringent nominal values.

Biallelic vs. multi-allelic QTL. Both equation 4 and the simulations were based on the assumption that there is an additive biallelic QTL. The contribution of grandsires heterozygous for the QTL to the covariance at the QTL is ¼ * a_v * a_w. Each of the two possible heterozygous grandsires, in coupling or in repulsion phase, contributes equally to the covariance. In reality, QTL can be multi-allelic, or actually consist of two closely linked biallelic QTL, which together act as a multi-allelic QTL. For this situation the contribution of a heterozygous grandsire to the covariance can have a range of values. Depending on the alleles present and their effects, the contribution can be positive, negative, or zero. Summing over all grandsires, the overall effect on the covariance could be close to the expected covariance under the null hypothesis. As a consequence, the pleiotropic QTL might not be detected in all multi-allelic situations.

Pleiotropic vs. linked QTL. Instead of considering a QTL with a pleiotropic effect, one could also consider two linked QTL that are involved in the expression of two traits. This can also cause a significant covariance between the estimated contrast. The probability of detecting a significant covariance between contrasts decreases with increasing distance between the linked QTL. The effect of distance between linked QTL on the covariance was not studied here. Notably, however, even fairly well-spaced QTL could cause a significant covariance, provided the QTL effect on each trait is large enough. The method presented here does not distinguish between pleiotropy and linkage, but if the goal is to look at the possible consequences of MAS for other traits, it is less relevant if the QTL is a pleiotropic QTL or if there are two closely linked QTL.

Covariance vs. correlation. In this study, the covariance between contrasts was used to study QTL with an effect on two traits. It is also possible to look at the correlation between contrasts. Combining equations 3 and 4, it follows that if there is a QTL affecting one trait, the covariance is equal to the covariance when there is no QTL, but the variance of the contrasts increases. As a result, the absolute value of the correlation between contrasts will be lower than in a situation without a QTL (Figure 4). Thus, permutation will not give insight if the deviation of the correlation is due to a QTL affecting two traits, or due to a QTL affecting only one trait. When looking at the covariance, a significantly deviating covariance can be detected by permutation, provided that the right type I errors are used.

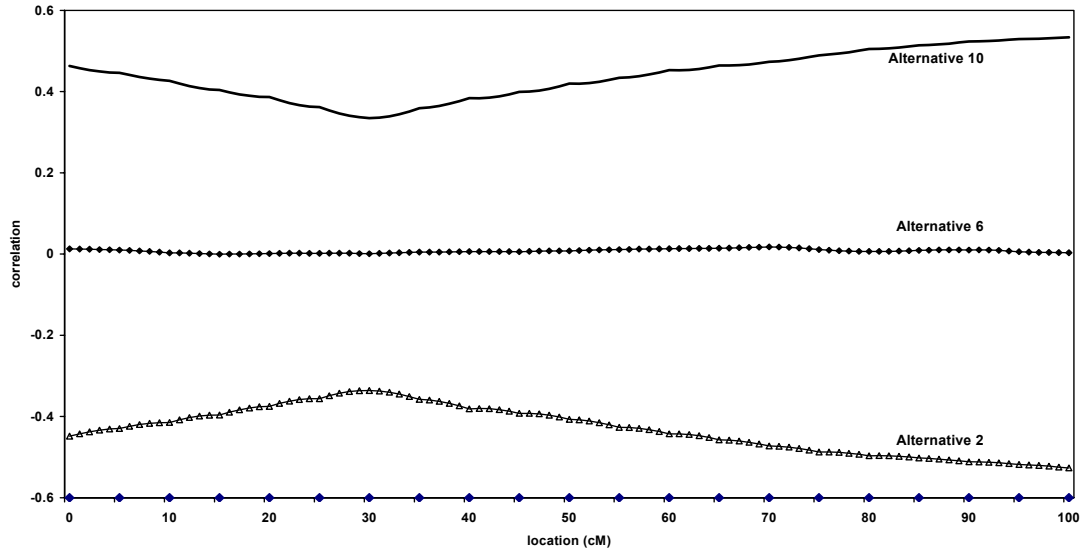

Figure 4. Correlation between contrasts in alternative 2 (r_a and r_p equal to –0.6), alternative 6 (r_a and r_p equal to 0.0), and alternative 10 (r_a and r_p equal to +0.6), based on 1,000 replicates of each alternative. The effect of the quantitative trait locus (QTL) on trait v was 0.5 σ_a in all three alternatives. Position of the markers indicated on the horizontal axis (♦).

CONCLUSIONS

This study has shown that the covariance between contrasts from separate single-trait regression analyses can be used to identify pleiotropic or closely linked QTL, i.e., QTL or linked QTL that act on more than one trait. The presented method gives insight in potential benefits or drawbacks from MAS due to the effect of MAS on nontarget traits. Results from simulations showed that stringent Type I errors should be applied to reduce the number of false-positive results. Application of the method to chromosome 6 indicated two QTL affecting protein percentage as well as milk yield. When MAS is used to increase the level of one of these traits, the level of the other trait will decrease.

ACKNOWLEDGMENTS

We thank Johan van Arendonk, Marco Bink, and Sijne van der Beek for their comments in preparing the manuscript. We appreciate the financial support by Holland Genetics.

REFERENCES

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

Falconer, D. S., and T. F. C. Mackay, 1996. Introduction to quantitative genetics. Longman, Essex, United Kingdom.

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

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

Grignola, F. E., Q. Zhang, and I. Hoeschele, 1997. Mapping linked quantitative trait loci via residual maximum likelihood. Genet. Sel. Evol. 29:529-544.

Gyapay, G., F. Ginot, S. Nguyen, A. Vignal, and J. Weissenbach, 1996. Genotyping procedures in linkage mapping. Methods 9:91-97.

Hacket, C. A., R. C. Meyer, and W. T. B. Thomas, 2001. Multi-trait QTL mapping in barley using multivariate regression. Genet. Res. 77:95-106.

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

Jansen, R. C., and P. Stam, 1994. High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447-1455.

Jiang, C., and Z.-B. Zeng, 1995. Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140:1111-1127.

Knott, S. A., J. M. Elsen and C. S. Haley, 1994. Multiple marker mapping of quantitative trait loci in half sib populations. Proc. 5th World Congr. Genet. Appl. Livest. Prod., Guelph, Canada XXI: 33-36.

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

Korol, A. B., Y. I. Ronin and V. M. Kirzhner, 1995. Interval mapping of quantitative trait loci employing correlated trait complexes. Genetics 140: 1137-1147.

Korol, A. B., Y. I. Ronin, E. Nevo, and P. Hayes, 2001. Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits. Genetics 157:1789-1803.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody , et al., 2001. Initial sequencing and analysis of the human genome. Nature 409: 860-921.

Philipsson, J., G. Banos, and T. Arnason, 1994. Present and future uses of selection index methodology in dairy cattle. J. Dairy Sci. 10:3252-3261.

Rishell, W. A., 1997. Breeding and genetics - historical perspective. Poult. Sci. 76:1057-1061.

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

Van der Beek, S., J. A. M. Van Arendonk, and A. F. Groen, 1995. Power of two- and three-generation QTL mapping experiments in an outbred population containing full-sib or half-sib families. Theor. Appl. Genet. 91:1115-1124.

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

Webb, A. J., 1998. Objectives and strategies in pig improvement: an applied perspective. J. Dairy Sci. 81:36-46.

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

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

APPENDIX

Covariance between contrasts for two traits from regression analysis of a granddaughter design, in the absence of a QTL.

Consider a granddaughter design with the following analysis model

$$y_{ijkl} = gs_i + m_{ij} + s_{ijk} + e_{ijkl} \qquad (1)$$

where y_{ijkl} is the phenotype of animal l, daughter of sire k that inherited marker allele j of grandsire i; gs_i is the effect of grandsire i; m_{ij} is the effect of marker allele j nested within grandsire i; s_{ijk} is the effect of sire k that received marker allele j from grandsire i, and e_{ijkl} is the random residual.

Suppose that the number of sires per grandsire is n_s and that each grandsire has n_g granddaughters per sire. With equal distribution of sires over marker alleles, the number of sires per grandsire per marker allele is ½ n_s

When considering two traits, v and w, the covariance of the contrast of marker group means can be written as:

$$\mathrm{cov}(\overline{v_{i1..}} - \overline{v_{i2..}}, \overline{w_{i1..}} - \overline{w_{i2..}}) =$$

$$\mathrm{cov}(\overline{v_{i1..}}, \overline{w_{i1..}}) - \mathrm{cov}(\overline{v_{i1..}}, \overline{w_{i2..}}) - \mathrm{cov}(\overline{v_{i2..}}, \overline{w_{i1..}}) + \mathrm{cov}(\overline{v_{i2..}}, \overline{w_{i2..}}) \qquad (2)$$

Group i1 (same grandsire and same marker) consists of three different categories when considering variances and covariances: animals themselves, half sibs (same sire, different dam) and animals only related through their paternal grandsire.

$$\mathrm{cov}(\overline{v_{i1..}}, \overline{w_{i1..}}) = \frac{1}{(½ n_s n_g)^2} * \left\{ \tfrac{1}{2} n_s n_g * \mathrm{cov}(v_{i111}, w_{i111}) + \right.$$

$$\left. \tfrac{1}{2} n_s n_g (n_g - 1) * \mathrm{cov}(v_{i111}, w_{i112}) + \tfrac{1}{4} n_s n_s n_g n_g - \tfrac{1}{2} n_s n_g n_g) * \mathrm{cov}(v_{i111}, w_{i121}) \right\} \qquad (3)$$

$$\mathrm{cov}(v_{i111}, w_{i111}) = \text{covariance between records on animal itself} =$$

$$r_p \sigma_{p_v} \sigma_{p_w} = \left(r_A h_v h_w + r_E \sqrt{1 - h_v^2} \sqrt{1 - h_w^2} \right) * \sigma_{p_v} \sigma_{p_w} \qquad (4)$$

$$\mathrm{cov}(v_{i111}, w_{i112}) = \text{covariance between records on half sibs} = ¼ r_A h_v h_w \sigma_{p_v} \sigma_{p_w} \qquad (5)$$

$$\mathrm{cov}(v_{i111}, w_{i121}) = \text{covariance between records on animals with same grandsire, different sire} = 1/16\, r_A h_v h_w \sigma_{p_v} \sigma_{p_w} \qquad (6)$$

$$\mathrm{cov}(\overline{v_{i1..}}, \overline{w_{i2..}}) = 1/16\, r_A h_v h_w \sigma_{p_v} \sigma_{p_w} \qquad (7)$$

$$\mathrm{cov}(\overline{v_{i2..}}, \overline{w_{i1..}}) = 1/16\, r_A h_v h_w \sigma_{p_v} \sigma_{p_w} \qquad (8)$$

$$\mathrm{cov}(\overline{v_{i2..}}, \overline{w_{i2..}}) = \mathrm{cov}(\overline{v_{i1..}}, \overline{w_{i1..}}) \qquad (9)$$

Substituting (3) to (9) into (2) gives:

$$\mathrm{cov}(\overline{v_{i1..}} - \overline{v_{i2..}}, \overline{w_{i1..}} - \overline{w_{i2..}}) = \frac{\sigma_{p_v}\sigma_{p_w}}{\left(\frac{1}{2}n_s n_g\right)^2}\left\{\frac{1}{2}n_s n_g\left(r_A h_v h_w + r_E\sqrt{1-h_v^2}\sqrt{1-h_w^2}\right)\right.$$

$$\left. + \frac{1}{2}n_s n_g(n_g - 1) * \frac{1}{4}r_A h_v h_w + \left(\frac{1}{4}n_s n_g n_s n_g - \frac{1}{2}n_s n_g n_g\right) * \frac{1}{16}r_A h_v h_w\right\} * 2$$

$$- 2 * \frac{1}{16}r_A h_v h_w \sigma_{p_v}\sigma_{p_w}$$

$$= \sigma_{p_v}\sigma_{p_w}\left\{\frac{4r_A h_v h_w + 4r_E\sqrt{1-h_v^2}\sqrt{1-h_w^2}}{n_s n_g} + \frac{r_A h_v h_w}{n_s} - \frac{r_A h_v h_w}{n_s n_g}\right.$$

$$\left. + \frac{1}{8}r_A h_v h_w - \frac{\frac{1}{4}r_A h_v h_w}{n_s} - \frac{1}{8}r_A h_v h_w\right\}$$

$$= \left\{\frac{3r_A h_v h_w + 4r_E\sqrt{1-h_v^2}\sqrt{1-h_w^2}}{n_s * n_g} + \frac{\frac{3}{4}r_A h_v h_w}{n_s}\right\} * \sigma_{p_v} * \sigma_{p_w} \qquad \text{(Equation 3)}$$

CHAPTER 5

Whole Genome Scan to Detect Chromosomal Regions Affecting Multiple Traits in Dairy Cattle

C. Schrooten[1,2], M. C. A. M. Bink[3], and H. Bovenhuis[1]

[1]Animal Breeding and Genetics Group, Wageningen University, PO Box 338, 6700 AH Wageningen, The Netherlands
[2]Holland Genetics, CR Delta, PO Box 5073, 6802 EB Arnhem, The Netherlands
[3]Biometris, PO Box 100, 6700 AC Wageningen, The Netherlands

Journal of Dairy Science (accepted)

ABSTRACT

Chromosomal regions affecting multiple traits (multiple trait quantitative trait regions, MQR) in dairy cattle were detected, using a method based on results from single trait analyses to detect quantitative trait loci (QTL). The covariance between contrasts for different traits in single trait regression analysis was computed. A chromosomal region was considered an MQR when the observed covariance between contrasts deviated from the expected covariance under the null hypothesis of no pleiotropy or close linkage. The expected covariance and the confidence interval for the expected covariance were determined by permutation of the data. Four categories of traits were analyzed: production (five traits), udder conformation (six), udder health (two) and fertility (two). The analysis of a granddaughter design involving 833 sons of 20 grandsires resulted in 59 MQR ($\alpha = 0.01$, chromosomewise). Fifteen MQR were found on BTA 14. Four or more MQR were found on BTA 6, 13, 19, 22, 23 and 25. Eight MQR involving udder conformation and udder health and four MQR involving production traits and udder health were found. Five MQR were identified for combinations of fertility and udder conformation traits, and another five MQR were identified for combinations of fertility and production traits. For 22 MQR, the difference between the correlation due to the MQR and the overall genetic correlation was larger than 0.60. The results from this study can be useful in the decision whether or not to apply marker-assisted selection (MAS) for specific QTL.

(**Key words**: genome scan, multiple trait, quantitative trait locus, dairy cattle)

Abbreviation key: **MAS** = marker-assisted selection, **QTL** = quantitative trait locus, **MQR** = multiple trait quantitative trait region, **BTA** = *Bos taurus* Autosome

INTRODUCTION

Breeding programs for livestock aim at improving the genetic level of several breeding goal traits by selection. These programs rely on phenotypic information on individuals and relationships between individuals. In the past decade, however, molecular techniques have become available that enable the genotypic analysis of animals. Analysis of phenotypic information on a trait and the genotype for genetic markers can lead to the identification of loci involved in the expression of the trait. In dairy cattle, many of these loci (quantitative trait loci, QTL) have been reported, for production traits (e.g. Georges et al., 1995; Spelman et al., 1996), as well as for conformation and functional traits (e.g. Ashwell et al., 1998; Schrooten et al., 2000). Efforts are now undertaken to

identify the mutations responsible for the effects of QTL. Recently, a functional mutation in the DGAT1 gene on BTA 14, responsible for large effects on production traits, has been identified (Grisart et al., 2002). A mutation in the growth hormone receptor gene on BTA 20 is associated with a strong effect on milk yield and composition (Blott et al., 2003). Genotypic information of the functional mutation or information on the genotype of closely linked markers can be incorporated in selection (marker-assisted selection, MAS), which is expected to increase genetic progress (e.g. Kashi et al., 1990; Meuwissen and van Arendonk, 1992).

Selection for a certain trait can lead to genetic changes in other traits, due to the genetic correlations between the traits. One of the possible causes for genetic correlation is pleiotropy. Pleiotropic genes are genes influencing two or more traits. A single QTL may increase, decrease, or not affect the overall genetic correlation. Also, closely linked QTL can contribute to genetic correlation, due to linkage disequilibrium. Pleiotropic effects of QTL, or closely linked QTL, each affecting a different trait, can affect the value of individual QTL for MAS.

The contribution of identified QTL to the overall genetic correlation can be determined by multiple trait QTL analysis. One method to address the effect of QTL on multiple traits is by canonical trait analysis, as first presented by Weller et al. (1996). Later, multiple trait methods have been developed (e.q. Knott and Haley, 2000; Korol et al., 2001), but application to outbred structures is not common practice. These methods can present computational difficulties, especially when applied to outbred populations or to a large number of traits. Schrooten and Bovenhuis (2002) showed that analysis of the covariance between contrasts from single trait analyses for different traits can reveal chromosomal regions affecting multiple traits (multiple trait quantitative trait regions, MQR). They developed a method that utilizes results from single trait QTL analyses to identify MQR in a granddaughter design. The method, however, cannot distinguish between pleiotropic QTL or closely linked QTL.

In this project, the method described by Schrooten and Bovenhuis (2002) is applied to data from a granddaughter design. The goal is to identify chromosomal regions affecting multiple traits (MQR) in dairy cattle.

MATERIALS AND METHODS

Data from a granddaughter design (Weller et al., 1990), consisting of 20 Holstein-Friesian grandsires with 833 sons (Schrooten et al., 2000), were used in the analysis. The number of sons per grandsire ranged from 11 to 147. For each animal, the genotype of 277 microsatellite markers on the 29 autosomes was determined. Per chromosome, the average interval between consecutive

markers ranged from 4 to 29 cM (Schrooten et al., 2000). Estimated breeding values based on progeny information were available for 37 routinely evaluated traits in the Netherlands, and deregressed before analysis. A subset of 15 important breeding goal traits was chosen for the analysis, and they were divided into four categories: production (kg. milk, fat %, protein %, kg. fat and kg. protein), udder conformation (fore udder attachment, front teat placement, teat length, udder depth, rear udder height and central ligament), udder health (somatic cell score and milking speed) and fertility (interval calving – 1st insemination and non return at 56 days after insemination). The number of sons with breeding values was 833, except for milking speed (801), interval calving – 1st insemination (831) and non return at 56 days after insemination (831). For each trait, contrasts between allelic effects were estimated by multi-marker regression (Knott et al., 1994) with the model

$$y_{ij} = \mu + gs_i + b_{ik}X_{ijk} + e_{ijk} \qquad (1)$$

where y_{ij} = deregressed breeding value of son j of grandsire i, μ = overall mean, gs_i = fixed effect of grandsire i, b_{ik} = regression coefficient for grandsire i at location k on the chromosome, X_{ijk} = probability that son j receives a chromosomal segment from grandsire i at position k, and e_{ijk} =random residual.

Significant QTL in the single trait analysis were determined by comparing the test statistic at each location with the 10% genomewise threshold, as described by Churchill and Doerge (1994). In this procedure, phenotypes were shuffled within families and reassigned to the sons, thus breaking the association between phenotypes and genotypes. Significance thresholds were obtained from analysis of 10000 permuted data sets. A QTL was considered suggestive when the threshold would yield one false positive result in a genome scan (Churchill and Doerge, 1994). This threshold corresponds to a chromosomewise type I error (α) of 0.0345.

To detect chromosomal regions affecting multiple traits, the covariance between estimated contrasts for pairs of traits, i.e. the regression coefficient in equation (1), was determined at each location separately. As shown by Schrooten and Bovenhuis (2002), the expected covariance can be written as

$$\mathrm{cov}(\overline{v_{i1..}} - \overline{v_{i2..}}, \overline{w_{i1..}} - \overline{w_{i2..}}) = \left\{ \frac{3r_A h_v h_w + 4r_E \sqrt{1-h_v^2}\sqrt{1-h_w^2}}{n_s n_g} + \frac{\frac{3}{4} r_A h_v h_w}{n_s} \right\} \sigma_{p_v} \sigma_{p_w} + \frac{1}{4} p_{het} a_v a_w \qquad (2)$$

where $\overline{v}_{i1..} - \overline{v}_{i2..}$ and $\overline{w}_{i1..} - \overline{w}_{i2..}$ are the contrasts for trait v and w for grandsire i, r_A is the genetic correlation between trait v and trait w, r_E is the environmental correlation between trait v and trait w, h_v^2 and h_w^2 are heritabilities of trait v and trait w, respectively, n_s is the number of sires per grandsire, n_g is the number of daughters per sire, σ_{p_v} and σ_{p_w} are the phenotypic standard deviations of trait v and trait w, respectively, p_{het} is the fraction that is heterozygous for the QTL, and a_v and a_w are equal to half the difference between the two homozygous genotypes (Falconer and Mackay, 1996), i.e. the allele substitution effect when there is no dominance. Note that parameter b_{ik} in equation (1) is an estimate of "$a_.$" in equation (2). Variances, heritabilities and correlations do not include the effect of the QTL. A more detailed description of the method and discussion of its properties can be found in Schrooten and Bovenhuis (2002).

When there is no QTL, or the QTL affects only one trait, equation 2 shows that the covariance does not depend on the values of a_v and a_w, but solely on the polygenic and environmental parameters. Deviations from this expected covariance indicate the presence of an MQR, i.e. a pleiotropic QTL, or closely linked QTL, each affecting one of the traits. It is not possible to distinguish between QTL with a pleiotropic effect and closely linked QTL in a statistical analysis. In this paper we will therefore consider chromosomal regions affecting multiple traits (MQR), because pleiotropic QTL or closely linked QTL may be involved.

In the analysis, the expected covariance under the null hypothesis, i.e. a situation without a pleiotropic QTL or closely linked QTL, can be determined by calculating the average covariance in 10000 permuted data sets. Permuted data sets were obtained as described for the single trait analysis. Here, permutation was also used to construct a 99%-confidence interval for the average or expected covariance. Covariances outside the confidence interval indicate that this chromosomal region is affecting both traits. Effects can have the same direction of change, i.e. the value of both traits is either increased or decreased, or have opposite direction of change. This can in some cases result in a large difference between the correlation due to the MQR and the overall genetic correlation. To detect chromosomal regions where this is the case, trait combinations with large differences between the MQR correlation and the overall genetic correlation were identified. Overall genetic correlations were not based on the data set used in the current project, but on data used for national evaluations (Harbers, 2003, personal communication).

RESULTS

Example

The method to detect MQR is based on the covariance between contrasts from single trait analyses. To illustrate the method, Figure 1 shows the results for fat % and protein % on BTA 6. The covariance between contrasts is shown at 1 cM intervals. The expected covariance, determined from analyzing 10000 permuted data sets, was 0.037, with a 99% confidence interval ranging from 0.006 to 0.117. The covariance in the region between 0 and 9 cM exceeded the confidence interval. The largest deviation from the expected covariance was found at 0 cM. An increase in fat % due to the QTL was accompanied by an increase in protein %, resulting in a positive covariance. Results from single trait analysis revealed a significant QTL for protein %. Test statistics were high between 0 and 43 cM, with the highest test statistic at 13 cM. A suggestive QTL for fat % was found in the same region, with the highest test statistic at 14 cM.

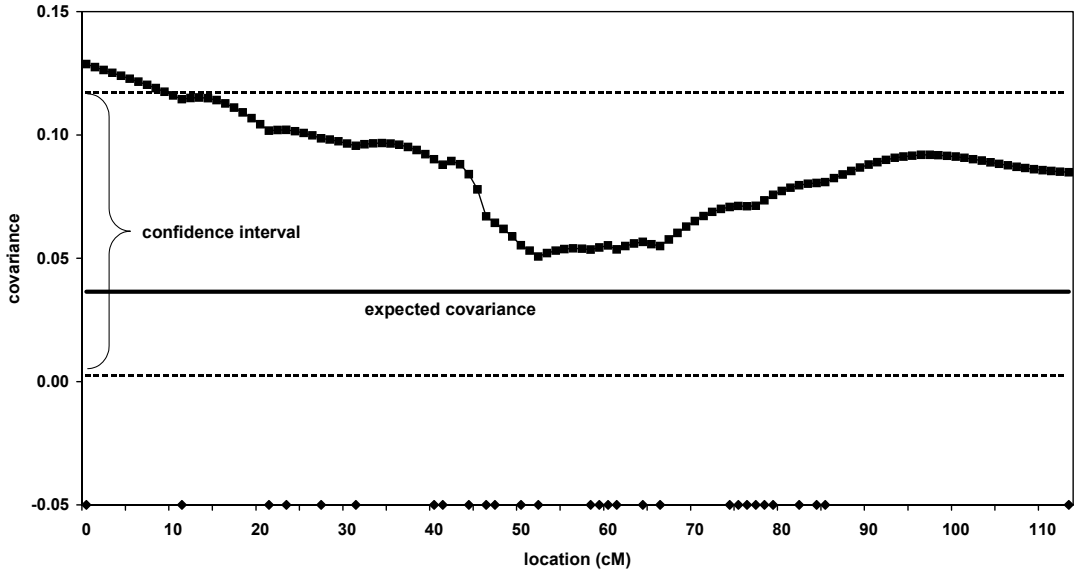

Figure 1. Covariance between contrasts for fat % and protein % on BTA 6. Location of markers indicated by (♦). Expected covariance (solid line) and 99% confidence interval (broken lines) based on chromosomewise values.

Identified MQR

In the same way as described for fat % and protein % on BTA 6, 105 combinations of traits on 29 chromosomes were tested for the presence of multiple trait quantitative trait regions (MQR).

Based on chromosomewise thresholds, 59 covariances between contrasts were outside the 99% confidence interval under the null hypothesis. As an overview of the results, the number of MQR per combination of trait categories is presented in Table 1. The number of MQR and QTL in Table 1 depends on trait characteristics and the number of traits in a trait category. Therefore, to put results in perspective, the number of traits in each trait category and the number of QTL from single trait analysis are presented in Table 1 as well. The identified MQR and associated results of relevant single trait analyses are in Table 2, characterized by location and correlation due to the MQR, and by the type I error (α) that was used to construct the confidence interval for the covariance.

Table 1. Number of traits per trait category, number of QTL[1] from single trait analysis for each trait category and number of multiple trait quantitative trait regions (MQR)[2] for combinations of trait categories. In parentheses, the number of traits in each combination of trait categories is listed. Diagonals show the number of MQR involving two traits that belong to the same trait category.

Trait category	# traits in category	# single trait QTL[1]	# MQR for combinations of trait categories[3]			
			Production	Udder conformation	Udder health	Fertility
Production	5	23	9 (10)	19 (30)	4 (10)	5 (10)
Udder conformation	6	17		8 (15)	8 (12)	5 (12)
Udder health	2	8			0 (1)	0 (4)
Fertility	2	4				1 (1)

[1]Genomewise significant or suggestive QTL

[2]MQR were reported when the covariance between contrasts was outside the 99% confidence interval

[3]Number of evaluated combinations of traits is in brackets

For combinations between production and udder conformation traits, 19 MQR were found (Table 1). Most of these MQR were located on BTA 13 (four, Table 2) and BTA 14 (five). Nine MQR involving two production traits were found, mainly on BTA 14 (six MQR) and BTA 6 (two MQR). For combinations among udder conformation traits, eight MQR were found. BTA 19 (three MQR) and BTA 23 (two) mainly contributed to this number. BTA 14, BTA 19 and BTA 26 each contained two MQR affecting an udder conformation trait and an udder health trait.

Table 2. Chromosomes and trait combinations where MQR were found, with location[1] of extreme covariance between contrasts. Correlation due to the MQR (r_{MQR}) and the overall genetic correlation (r_g) are indicated. QTL from single trait analysis are indicated by their location[1].

							Single trait analysis[2]	
BTA	Location (cM)	Trait 1	trait 2	r_{MQR}[3]	r_g[3]	α[4]	trait 1	Trait 2
1	54	Teat length	Non return daughters 56 days	-0.67[a]	0.12[a]	0.008	NS	NS
4	67	Milking speed	Kg. protein	0.66[b]	-0.07[b]	0.002	NS	NS
5	57	Rear udder height	Protein %	-0.69	-0.20	0.004	88	123
5	39	Kg. fat	Non return daughters 56 days	-0.77	-0.19	0.002	166	NS
6	0	Fat %	Protein %	0.89	0.78	0.004	14	13
6	34	For udder attachment	Front teat placement	-0.45[a]	0.44[a]	0.002	NS	NS
6	101	Rear udder height	Non return daughters 56 days	0.76[a]	-0.18[a]	0.002	NS	NS
6	113	Protein %	Kg. protein	-0.76[b]	0.04[b]	0.004	87	NS
7	83	Milking speed	Kg. milk	0.50	-0.02	0.010	NS	NS
9	113	Central ligament	Non return daughters 56 days	-0.77	-0.25	0.002	NS	79
10	30	Fore udder attachment	Kg. fat	-0.71[b]	-0.09[b]	0.004	NS	36
10	26	Teat length	Kg. protein	-0.55[a]	0.11[a]	0.006	NS	NS
13	0	Fore udder attachment	Kg. milk	-0.63	-0.05	0.010	0	NS
13	0	Udder depth	Protein %	0.75[b]	0.03[b]	0.010	22	NS
13	0	Fore udder attachment	Protein %	0.57[b]	-0.03[b]	0.004	0	NS
13	0	Udder depth	Fat %	0.71[b]	-0.01[b]	0.010	22	NS
14	7	Kg. milk	Kg. fat	-0.24[a]	0.37[a]	0.008	0	0
14	7	Fore udder attachment	Fat %	0.55	-0.04	0.002	NS	0
14	10	Fore udder attachment	Kg. milk	-0.63	-0.05	0.006	NS	0
14	11	Protein %	Kg. protein	-0.62[b]	0.04[b]	0.002	10	0
14	11	Kg. milk	Protein %	-0.86	-0.59	0.002	0	10

BTA	Location (cM)	Trait 1	trait 2	r_{MQR}[3]	r_g[3]	α[4]	Single trait analysis[2] trait 1	Trait 2
14	11	Kg. milk	Fat %	-0.91	-0.65	0.002	0	0
14	11	Front teat placement	Kg. milk	-0.61[a]	0.13[a]	0.006	NS	0
14	11	Fat %	Protein %	0.93	0.78	0.002	0	10
14	11	Udder depth	Fat %	0.61[b]	-0.01[b]	0.002	NS	0
14	11	Front teat placement	Fat %	0.46[a]	-0.19[a]	0.008	NS	0
14	12	Fat %	Kg. protein	-0.77	-0.22	0.002	0	0
14	13	Milking speed	Kg. milk	-0.46	-0.02	0.006	17	5
14	13	Milking speed	Fat %	0.44	-0.09	0.004	17	0
14	13	Udder depth	Milking speed	0.56	0.00	0.006	NS	17
14	13	Fore udder attachment	Milking speed	0.50	0.00	0.004	NS	17
18	68	Fat %	Interval calving – 1st insem.	0.49	-0.05	0.006	NS	NS
18	67	Protein %	Interval calving – 1st insem.	0.58[a]	-0.14[a]	0.004	105	NS
19	0	Fore udder attachment	Somatic cell score	-0.62	-0.30	0.004	66	NS
19	0	Rear udder height	Somatic cell score	-0.61	-0.14	0.004	NS	NS
19	0	Udder depth	Fore udder attachment	0.91	0.74	0.008	35	66
19	0	Rear udder height	Fore udder attachment	0.66	0.21	0.002	NS	66
19	0	Rear udder height	Front teat placement	0.82[b]	0.10[b]	0.006	NS	66
19	39	Interval calving – 1st insemination	Non return daughters 56 days	-0.51[a]	0.36[a]	0.004	NS	NS
20	60	Udder depth	Fore udder attachment	-0.19[a]	0.74[a]	0.004	NS	NS
22	0	Kg protein	Non return daughters 56 days	0.64[a]	-0.28[a]	0.008	NS	NS
22	0	Kg fat	Non return daughters 56 days	0.58[a]	-0.19[a]	0.006	NS	NS
22	18	Front teat placement	Protein %	-0.59	-0.13	0.010	3	NS
22	85	Fore udder attachment	Non return daughters 56 days	-0.69	-0.20	0.008	NS	NS

BTA	Location (cM)	Trait 1	trait 2	r_{MQR}[3]	r_g[3]	α[4]	Single trait analysis[2] trait 1	Trait 2
23	8	Fore udder attachment	Kg. protein	-0.58	-0.08	0.004	14	18
23	8	Fore udder attachment	Kg. milk	-0.55	-0.05	0.008	14	NS
23	12	Udder depth	Front teat placement	0.67	0.35	0.010	NS	NS
23	14	Fore udder attachment	Front teat placement	0.77	0.44	0.008	14	NS
24	29	Udder depth	Interval calving – 1st insem.	-0.74	-0.28	0.008	NS	NS
25	80	Fat %	Kg. protein	-0.76	-0.22	0.008	NS	NS
25	80	Central ligament	Kg. milk	0.63	0.12	0.006	24	NS
25	80	Central ligament	Somatic cell score	0.50	-0.04	0.006	24	NS
25	80	Rear udder height	Central ligament	0.80	0.36	0.004	11	24
25	80	Central ligament	Kg. protein	0.70[b]	0.05[b]	0.004	24	NS
26	0	Fore udder attachment	Milking speed	0.43	0.00	0.010	3	NS
26	0	Udder depth	Milking speed	0.57	0.00	0.004	NS	NS
27	44	Udder depth	Somatic cell score	-0.74	-0.35	0.006	44	NS
27	29	Udder depth	Protein %	-0.40	0.03	0.010	44	NS
28	24	Fore udder attachment	Protein %	-0.56	-0.03	0.010	NS	NS

[1]Only locations with most extreme covariance (MQR analysis) or highest test statistic (for single trait analysis) are shown. Table A1 of the Appendix contains markers or flanking markers at these locations.

[2]NS = no significant or suggestive QTL detected.

[3]Differences between r_{MQR} and r_g larger than 0.60 are indicated by superscript "b". When, besides, $|r_g| > 0.10$ and r_g and r_{MQR} have opposite sign, these are indicated by superscript "a".

[4]For each test: smallest type I error (α) that yielded a significant result (smallest tested was 0.002).

Chromosomes with four or more MQR are presented graphically in Figure 2 and listed in Table 2. Results for these chromosomes will be discussed in more detail. Results for other chromosomes are only presented in Table 2.

BTA 6. On BTA 6, four MQR were detected. These were located at 0 cM (fat % and protein %), 34 cM (fore udder attachment and front teat placement), 101 cM (rear udder height and non return at 56 days of daughters) and at 113 cM (protein % and kg protein). In the single trait analysis, significant QTL for fat % and protein % were found around 13 cM. The correlation between the contrasts of fore udder attachment and front teat placement (-0.45) and the overall genetic correlation between these traits (0.44) differ considerably and have opposite sign. The same applies to the MQR affecting rear udder height and non return of daughters at 56 days (r_{MQR} = 0.76, r_g = -0.18). The correlation due to the MQR affecting protein % and kg. protein was far more negative (-0.76) than the overall genetic correlation, which was close to zero.

BTA 13. On BTA 13, four MQR involving production (kg milk, protein % and fat%) and udder conformation traits (fore udder attachment and udder depth) were located at 0 cM. In the single trait analysis, significant QTL for fore udder attachment and udder depth were found in this region. Overall genetic correlations were close to zero for all four trait combinations. MQR correlations were positive for combinations of udder conformation traits and percentage traits (0.57 to 0.75). The correlation due to the MQR affecting fore udder attachment and kg. milk was negative (-0.63).

BTA 14. BTA 14 contained 15 MQR. Covariances exceeded the confidence interval in the chromosomal region between 0 and 17 cM, with most extreme covariances between 7 and 13 cM (Table 2, Figure 2). In Figure 2, only this region of BTA 14 is shown.

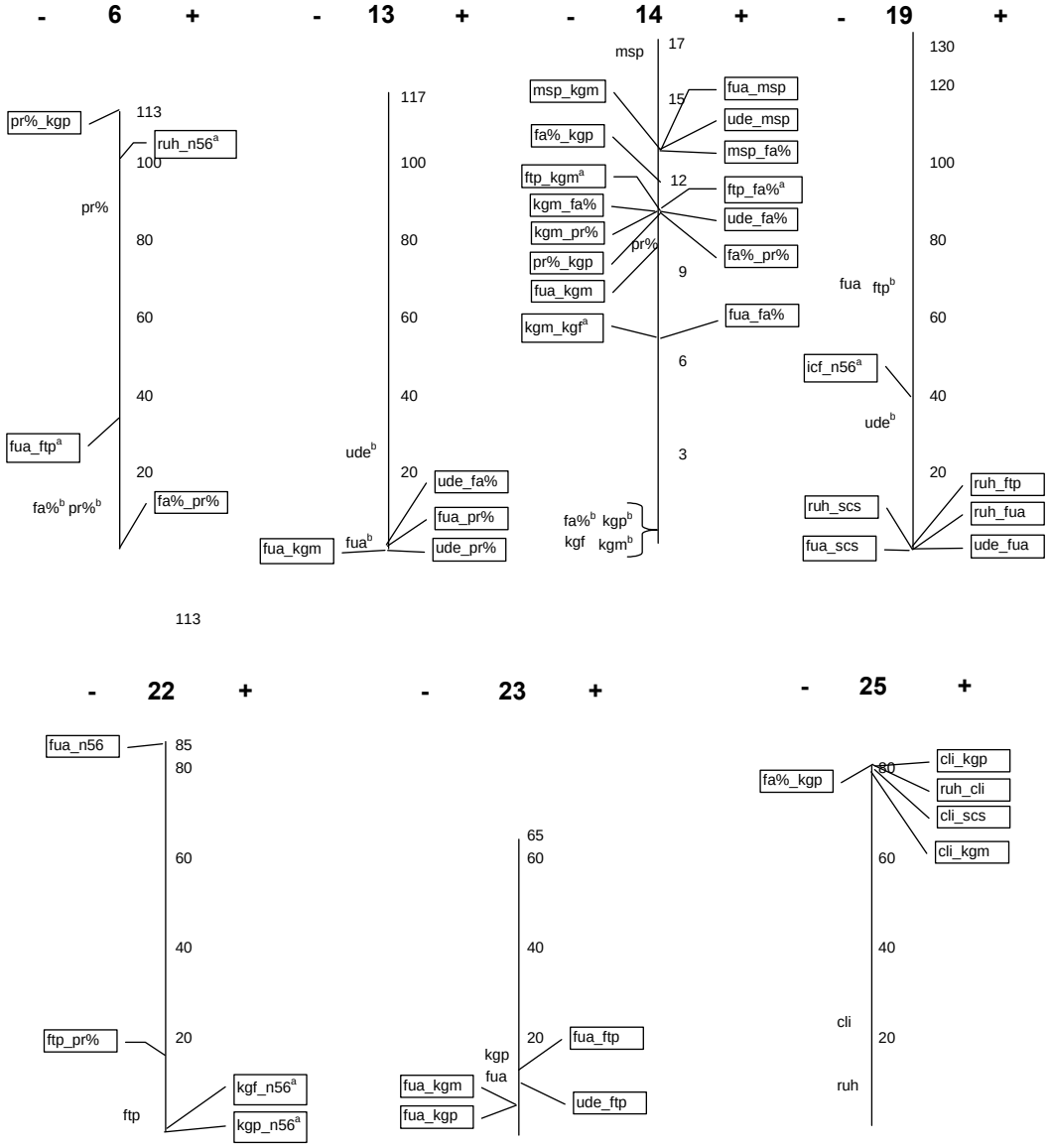

Figure 2. BTA 6, 13, 14, 19, 22, 23 and 25, with regions affecting multiple traits (MQR) as well as significant and suggestive QTL from single trait analysis for traits affected by MQR. Traits are indicated by abbreviations: kgm = kg. milk; fa% = fat%; pr% = protein %; kgf = kg. fat; kgp = kg. protein; fua = fore udder attachment; ftp = front teat placement; tle = teat length; ude = udder depth; ruh = rear udder height; cli = central ligament; scs = somatic cell score; msp = milking speed; icf = interval calving – 1st insemination; N56 = non return of daughters at 56 days after insemination

The vertical bar represents the chromosome, with locations indicated in cM. Trait combinations with identified MQR are in boxes. MQR contributing positively to the overall genetic covariance are positioned to the right of the chromosome bar, whereas MQR contributing negatively to the overall genetic covariance are positioned to the left. Trait combinations with superscript "a" indicate MQR where the sign of the covariance due to the MQR was opposite to the sign of the overall genetic covariance. The location of QTL from the single trait analysis is given by the abbreviation of the trait to the left of the chromosome bar. Trait abbreviations with superscript "b" indicate significant QTL, trait abbreviations without subscript indicate suggestive QTL.

Note the different scale for BTA 14, only the beginning of this chromosome is shown.

Thirteen of the 15 MQR involved one or two production traits, and six of these involved two production traits. Traits other than production that were affected by these MQR were fore udder attachment, front teat placement, udder depth and milking speed. Most overall genetic correlations between the traits involved were close to zero, except for the correlation between kg. milk and kg. fat (0.37), among percentage traits (0.78) and between milk yield and percentage traits (-0.59 and –0.65). Correlations due to the MQR were positive between udder conformation and percentage traits (0.46 to 0.55), and negative between udder conformation traits and kg. milk (-0.61 to –0.63). MQR correlations among yield traits were comparable to overall genetic correlations, with the exception of the MQR correlation between kg. milk and kg. fat (-0.24), and kg. protein with protein % and fat % (-0.62 to –0.77). MQR correlations with milking speed were positive for udder conformation traits and fat % (0.44 to 0.56) and negative for milk yield (-0.46). In the single trait analysis, significant QTL were found for fat %, kg milk and kg protein, whereas suggestive QTL were identified for protein %, kg fat and milking speed, in the same region as the MQR.

BTA 19. Five of the six MQR on this chromosome are located at 0 cM and affect one of the udder conformation traits and somatic cell score, or two udder conformation traits. The sixth MQR was at 39 cM. The correlation due to this MQR affecting the two fertility traits, interval calving to first insemination and non return of daughters at 56 days after insemination, was moderately negative (-0.51), whereas the overall genetic correlation was moderately positive (0.36). The MQR correlation between rear udder height and front teat placement was strongly positive (0.82), whereas the overall genetic correlation between these traits was close to zero. For udder depth, fore udder attachment and front teat placement the most probable location for QTL, from the single trait analysis, was at 35, 66 and 66 cM, respectively.

BTA 22. On BTA 22, MQR were found at locations 0 cM (non return of daughters at 56 days after insemination, with both kg fat and kg protein), 18 cM (front teat placement and protein %) and 85 cM (fore udder attachment and non return at 56 days after insemination). The single trait analysis only revealed a suggestive QTL at 0 cM for front teat placement. For the two MQR at 0 cM, there was a large difference between the correlation due to the MQR (0.58 to 0.64), and the overall genetic correlation (-0.19 to –0.28).

BTA 23. MQR on BTA 23 affected production and udder conformation traits. These MQR were found in the region between 8 and 14 cM. In the single trait analysis, suggestive QTL for fore udder attachment and kg protein were found in the same region.

BTA 25. On BTA 25, five MQR were located at 80 cM. Four of these MQR involved central ligament. Positive effects on central ligament, due to the MQR, were associated with positive effects for yield traits, somatic cell score and rear udder height. The overall genetic correlations for the relevant combinations ranged from around zero to moderately positive. There was no significant or suggestive QTL for central ligament at this location, nor for the other trait involved in these four MQR. The highest test statistic for central ligament was found at 24 cM. Also for the fifth MQR on this chromosome, involving fat % and kg protein, there was no significant or suggestive QTL in the single trait analysis.

MQR with Correlations Deviating from the Overall Genetic Correlation

For a number of MQR, the correlation due to the MQR differed largely from the overall genetic correlation. In Table 2, correlations are indicated where the difference between MQR correlation and the overall genetic correlation was 0.60 or higher. Also indicated are MQR where, in addition, the MQR correlation and the overall genetic correlation had opposite sign, and the overall genetic correlation had an absolute value of 0.10 or higher. This was the case for 12 MQR (Table 2).

Most of the MQR with a correlation deviating from the overall genetic correlation have been indicated in the previous sections already. In addition, positive MQR correlations were observed between milking speed and kg. protein (BTA 4), and protein % and interval calving – 1st insemination (BTA 18). Negative MQR correlations were observed between teat length and non return of daughters 56 days after insemination (BTA 1), fore udder attachment and kg. fat (BTA 10), teat length and kg. protein (BTA 10) and udder depth and fore udder attachment (BTA 20). The correlation due to the MQR on BTA 20 affecting udder depth and fore udder attachment was –0.19. On BTA 19, another MQR affecting udder depth and fore udder attachment was observed. The correlation due to this MQR was +0.91, which differs largely from the correlation due to the MQR on BTA 20.

DISCUSSION

Method

Pleiotropic QTL and Closely Linked QTL. The method that was used in this paper can identify MQR. The method, however, cannot distinguish between pleiotropic and closely linked QTL. Therefore, the regions where the covariance between contrasts is outside the confidence interval for the expected covariance under the null hypothesis, might contain pleiotropic QTL or it might

involve regions where two different QTL are located, each affecting a different trait. For example, consider the results found on BTA 14, where the single trait analysis for kg milk gave a peak test statistic at 0 cM, and the single trait analysis for milking speed gave a peak test statistic at 17 cM (Figure 2). The covariance between contrasts for kg milk and milking speed deviated most from the expected covariance at 13 cM, i.e. in between the most likely locations from the single trait analyses. Given that the confidence interval for the location of a QTL in this type of analysis extends over a wide region, it can not be excluded that actually this QTL is not a pleiotropic QTL, but two linked QTL in a 20 cM region. Methods to distinguish between pleiotropic and closely linked QTL have been developed, for example by Lebreton et al. (1998). They concluded, however, that in common QTL mapping designs it is unlikely to be able to distinguish between pleiotropic QTL and linked QTL at locations 10 to 30 cM apart. Recently, Lund et al. (2003) presented a method, based on combined linkage and linkage disequilibrium analysis, that can distinguish two closely linked QTL that are 5 cM apart, from a pleiotropic QTL. This method, however, requires a dense linkage map, which is not available in the genome scans applied in dairy cattle so far. With even denser marker maps (marker intervals of e.g. 0.25 cM), they expect that more closely linked QTL (e.g. 1 cM apart) can be distinguished from pleiotropic QTL. When the goal is to use the QTL for MAS, the issue of pleiotropy or close linkage is less important than in the case where efforts are undertaken to find the gene or genes. Due to recombination, correlated effects resulting from linkage will disappear over time. In the short term, however, the implications for MAS will not be very different, whether there is pleiotropy or close linkage. It might take many generations before correlated effects due to linkage will have disappeared, although that does depend on the distance between the QTL, the magnitude of the QTL effects and the selection pressure on the QTL.

Permutation test. In the method used in this paper, a permutation test is used to identify MQR. Covariances between contrasts are determined in the original data set, as well as in a number of permuted data sets. The applied method consists of computation of the covariance between contrasts and comparison of this covariance with the covariance in permuted data sets. Permutation takes into account the characteristics of the data set, e.g. pedigree structure, marker density, trait distribution etc. It is, however, not possible to obtain permuted data sets with a QTL affecting only one trait, which would probably be the ideal situation. As indicated by Schrooten and Bovenhuis (2002), the standard deviation of the covariance may be higher in a data set containing a QTL affecting only one trait, as compared to the permuted data sets having no QTL. Depending on the characteristics of the QTL, this phenomenon can lead to higher type I errors than expected. Elevated covariances between contrasts normally indicate the presence of MQR, but in some cases may also

be due to one QTL affecting one trait. Figure 3 shows the relationship between QTL effect and type I error for a trait with heritability of 0.60, and for QTL effects up to 0.5 σ_a. These results were obtained by simulating a granddaughter design of 20 grandsires with 50 sons per grandsire. Each son had 100 daughters with records on two traits. Heritability of the traits was 0.6 and 0.35, respectively, and the effect of the QTL on trait 1 ranged from 0.1 to 0.5 σ_a. The second trait was not affected by the QTL. Half of the grandsires were heterozygous for the bi-allelic QTL. Grandsires were informative for markers at 5 cM interval on a chromosome of 100 cM. The QTL was located at 30 cM. 1000 data sets were generated and each data set was permuted 1000 times, and analyzed to obtain 95% and 99% confidence intervals for the covariance between contrasts. The number of data sets where the covariance between contrasts was outside the confidence interval was counted to obtain the real type I error. The figure shows that type I error increases with increasing effects of the QTL. The impact on type I error rates is small for QTL effects up to 0.2 σ_a.

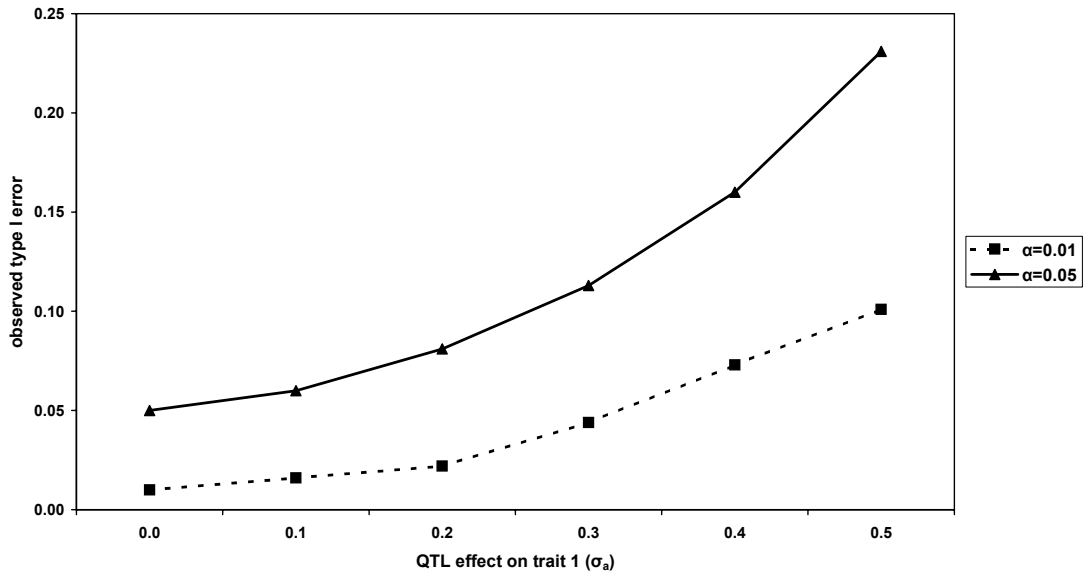

Figure 3. Probability (= observed type I error) to detect an MQR in a typical granddaughter design (20 grandsires, 50 sons per grandsire, 100 daughters per son). A QTL-effect was simulated only for trait 1, the effect ranged from 0 to 0.5 σ_a.Type I error is set to 0.01 or 0.05.

Literature

Lipkin et al. (2002) examined pleiotropic effects of QTL on milk yield, protein percentage and protein yield using selective DNA-pooling, in regions of the genome previously found to be affecting protein percentage. Because of the algebraic dependence of the analyzed traits, it is likely

that a relatively high number of QTL would be declared to be pleiotropic. Combining results, Lipkin et al. (2002) concluded that there might be about 40 suggestive QTL on the genome affecting all three traits. Chromosomes and markers involved were not specified and therefore, it is not possible to compare the results found by Lipkin et al. (2002) on pleiotropic QTL with results from the current study.

For a large number of MQR presented in Table 2, QTL for one of the traits involved have been reported in literature. Most studies only reported on a limited number of traits, therefore only a limited number of MQR found in the current research could be confirmed from literature, especially those on BTA 6 and BTA 14 affecting production traits. These will be mentioned here, and the QTL from literature affecting one of the traits involved in the MQR from Table 2 will not be discussed.

BTA 6. Numerous authors have studied QTL on BTA 6, indicating the presence of multiple QTL on this chromosome (reviewed by Mosig et al., 2001). Spelman et al. (1996) found a QTL affecting both protein % and milk yield in two grandsire families. Using interval mapping and multiple regression analysis in a daughter design, Ron et al. (2001) found results supporting the hypothesis that one QTL is segregating in two different families and that this QTL affects 5 production traits: kg. milk, fat %, protein %, kg. fat and kg. protein. This QTL was located close to 55 cM, which corresponds to 40 cM on the map used in our study. Freyer et al. (2002) reported a QTL on BTA 6 affecting both protein yield and fat yield. This QTL was located at 70 cM, which corresponds to 52 cM on the map used in the current study. Freyer et al. (2002) also found QTL at the casein cluster affecting several production traits. The casein cluster corresponds to a location of 95 cM on the map used in the current study. Compared to the MQR involving production traits in our study, QTL presented by Freyer et al. (2002) were located more towards the center of the chromosome.

BTA 14. Recently, a mutation in the DGAT1 gene on BTA 14 was reported, with an effect on production traits (Grisart et al., 2002). This gene has an effect on several milk production traits and is located in the region between the first two markers used in this study, close to 0 cM. The substitution effects for milk production traits were: kg milk –158; fat% 0.17; protein % 0.04; kg fat +5.23; kg protein –2.82 (Grisart et al., 2002). Effects were expressed in terms of daughter yield deviations. Based on these effects, an increase of kg milk, for example, is expected to result in a decrease of fat % and a decrease of kg fat. In the current study, which consists of partly the same data as used by Grisart et al. (2002), a region on BTA 14 affecting multiple milk production traits was reported (Figure 2). For all combinations of production traits presented in Figure 2, the effects on the traits involved were in agreement with the effects of the gene as presented by Grisart et al.

(2002), which supports the results presented in this study. The MQR for production traits on BTA 14 in the current study showed most extreme covariances at 7 cM, 11 cM and 12 cM which agrees with the location of the DGAT1 gene close to 0 cM.

MQR with Correlations Deviating from the Overall Genetic Correlation

In Table 2, a number of MQR are shown, for which the covariance due to the MQR and the overall genetic covariance differ largely. The overall genetic covariance is the result of the summation of all covariances due to individual QTL, either pleiotropic or closely linked. The contribution of each MQR to the overall genetic covariance depends on the size and the sign of the effects on the traits involved. If the genetic correlation is strongly positive, it is expected that the covariance at most of the individual MQR is positive. There might, however, be a few MQR with a negative contribution to the overall genetic covariance, i.e. MQR with correlations that strongly deviate from the overall genetic correlation. As was shown by Schrooten and Bovenhuis (2002), the statistical power to detect these MQR is increased compared to MQR that do not deviate from overall genetic correlations. For pairs of traits with high genetic correlations, it is expected that there are relatively few MQR that strongly deviate from the overall genetic correlation, but the power to find these MQR is higher (Korol et al, 1995; Schrooten and Bovenhuis, 2002). For cases in which the overall genetic correlation between traits is unfavorable, MAS for MQR that strongly deviate from the overall genetic correlation offers excellent opportunities to achieve progress for both traits. This was demonstrated by de Koning and Weller (1994).

Comparison with Results from Single Trait Analysis

The analysis described in this paper resulted in regions affecting multiple traits, in some cases affecting traits for which no QTL were detected in the single trait analysis. As was discussed by Korol et al. (1995), methods that use information from multiple traits have higher power, which can account for a number of QTL not previously identified. However, in the present study we did not account for multiple testing when identifying MQR. In total 3045 tests were performed, and MQR were reported when the covariance was outside the 99% chromosomewise confidence interval. The type I error of 0.01 was divided over both sides of the interval when constructing the confidence interval. In the single trait analysis, 435 tests were performed, and QTL were considered suggestive when the chromosomewise type I error was smaller than 0.0345. Combining the number of detected QTL and MQR (Table 1) with the respective number of tests, yields a false discovery rate (Benjamini and Hochberg, 1995) of (0.0345*435)/52 = 0.29 for the single QTL analysis and

(0.01*3045)/59 = 0.52 for the MQR analysis. The latter is relatively high. Therefore, MQR reported here should be considered suggestive, rather than significant. However, it was considered important to present these results, in order to assess potential consequences of using the given MQR in MAS.

To compare results from single trait analysis and the analysis described in this paper, consider for example the results for BTA 25. Four MQR involving central ligament were detected at 80 cM, whereas in the single trait analysis for this trait, a suggestive QTL was located at 24 cM. At first glance, these results do not seem to be in agreement. A within-family analysis on this chromosome for central ligament, however, indicated two families with suggestive QTL at 80 cM, besides other families with suggestive QTL for central ligament at or around 24 cM (results not shown). Within family contrasts for other traits, like rear udder height, were significant in these families. As a result, the covariance between contrasts for these traits was outside the confidence interval for the expected covariance under the null hypothesis. Although these QTL were not detected in the across family analysis, they can be detected in a within-family analysis and also by the method used in this paper.

The MQR identified in the current study include a number of potentially interesting regions from the point of view of selection. For example, the MQR on BTA 25 should be explored further: favorable effects on kg milk and kg protein are associated with favorable effects on udder traits, especially central ligament. Udder traits like rear udder height, front teat placement, fore udder attachment and udder depth are affected by MQR on BTA 19, which could be exploited in MAS. Favorable effects of MQR on BTA 13, on udder traits like udder depth and fore udder attachment, are associated with favorable effects on fat % and protein %.

CONCLUSIONS

In this research, a number of chromosomal regions affecting multiple traits in dairy cattle have been identified. These regions might contain a QTL with a pleiotropic effect on both traits or they might contain two different QTL, each affecting one of the traits. Chromosomal regions influencing multiple traits were found on almost all chromosomes, but especially on BTA 6, 13, 14, 19, 22, 23 and 25. Eight MQR involving udder conformation and udder health and four MQR involving production traits and udder health were found. Five MQR were identified for combinations of fertility and udder conformation traits, and another five MQR were identified for combinations of fertility and production traits. These results are of great value for determining the importance of chromosomal regions for MAS.

ACKNOWLEDGMENTS

We thank Johan van Arendonk and Sijne van der Beek for their comments on the manuscript and Michel Georges' group, department of Genetics, University of Liège (Belgium), for providing the genotypes. We appreciate the financial support by Holland Genetics.

REFERENCES

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

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

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

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

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

Falconer, D. S., and T. F. C. Mackay, 1996. Introduction to quantitative genetics. Longman, Essex, United Kingdom.

Freyer, G., C. Kühn, R. Weikard, Q. Zhang, M. Mayer, and I. Hoeschele, 2002. Multiple QTL on chromosome six in dairy cattle affecting yield and content traits. J. Anim. Breed. Genet. 119:69-82.

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

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

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

Knott, S. A., J. M. Elsen and C. S. Haley, 1994. Multiple marker mapping of quantitative trait loci in half sib populations. Proc. 5th World Congr. Genet. Appl. Livest. Prod., Guelph, Canada XXI: 33-36.

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

Korol, A. B., Y. I. Ronin and V. M. Kirzhner, 1995. Interval mapping of quantitative trait loci employing correlated trait complexes. Genetics 140: 1137-1147.

Korol, A. B., Y. I. Ronin, E. Nevo, and P. Hayes, 2001. Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits. Genetics 157:1789-1803.

Lebreton, C. M., P. M. Visscher, C. S. Haley, A. Semikhodskii and S. A. Quarrie, 1998. A nonparametric bootstrap method for testing close linkage vs. pleiotropy of coincident quantitative trait loci. Genetics 150:931-943.

Lipkin, E., G. Grosman, E. Tchourzyna, A. Friedmann and M. Soller, 2002. QTL affecting milk protein percent also affect milk yield and protein yield. Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France 31:55-58.

Lund, M. S., P. Sørensen, B. Guldbrandtsen and D. A. Sorensen, 2003. Multitrait fine mapping of quantitative trait loci using combined linkage disequilibria and linkage analysis. Genetics 163: 405-410.

Meuwissen, T. H., 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.

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

Ron, M., D. Kliger, E. Feldmesser, E. Seroussi, E. Ezra, and J. I. Weller, 2001. Multiple quantitative trait locus analysis of bovine chromosome 6 in the Israeli Holstein Population by a Daughter Design. Genetics 159: 727-735.

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

Schrooten, C., and H. Bovenhuis, 2002. Detection of pleiotropic effects of quantitative trait loci in outbred populations using regression analysis. J. Dairy Sci. 85: 3503-3513.

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

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

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

APPENDIX

Table A1. Markers on a number of chromosomes, at or close to locations relevant for the current study. These locations are listed in Table 2.

BTA	location (cM)	marker or flanking markers	BTA	location (cM)	marker or flanking markers
1	54	BMS711, TGLA57	14	12	BULGE034, BULGE017
4	67	BM6458, TGLA116	14	13	BULGE017
5	39	BP1, AGLA293	14	17	CSSM66
5	57	AGLA293, AGLA254	18	67, 68	BM7109, ILSTS002
5	88	AGLA254, IGF1	18	105	IDVGA55, TGLA227
5	123	IGF1, BM315	19	0	BM9202
5	166	ETH152	19	35, 39	TGLA94, URB044
6	0	ILSTS090	19	66	BMS2503
6	13, 14	MCM53, URB016	20	60	BMS703, BM5004
6	34	BMS2508, BM143	22	0	CSSM26
6	101	BP7, BM2320	22	3, 18	CSSM26, CSSM6
6	113	BM2320	22	85	HMH1R
7	83	TGLA164, AGLA260	23	8	BM47, RM033
9	79	TGLA73	23	12,14,18	RM033, BM1258
9	113	CSSM56	24	29	AGLA269, BMS66
10	26, 30	BM1237, BRRIBOold	25	11	RM074
10	36	BRRIBOold	25	24	TGLA40
13	0	TGLA23	25	80	AF5
13	22	BMC1222	26	0	TGLA22
14	0	BULGE011	26	3	TGLA22, HEL11
14	5	BULGE030	27	29	CSSM43
14	7	BULGE030, BULGE036	27	44	HUJI13
14	10	BULGE036, BULGE034	28	24	BMS362
14	11	BULGE034			

CHAPTER 6

Genetic Progress in Multi-Stage Dairy Cattle Breeding Schemes Using Genetic Markers

C. Schrooten[1,2], H. Bovenhuis[1], J. A. M. van Arendonk[1], and P. Bijma[1]

[1]Animal Breeding and Genetics Group, Wageningen University, PO Box 338, 6700 AH Wageningen, The Netherlands
[2]Holland Genetics, CR Delta, PO Box 5073, 6802 EB Arnhem, The Netherlands

Submitted (2004)

ABSTRACT

In this paper, closed nucleus breeding schemes in dairy cattle that use information on quantitative trait loci (QTL), have been evaluated by deterministic simulation. In the base scheme, the selection index for dams consisted of pedigree information and own performance. The selection index for sires consisted of pedigree information and performance of 100 daughters. In alternative breeding schemes, information on a QTL was accounted for by simulating an additional index trait. The fraction of the variance explained by the QTL determined the correlation between the additional index trait and the breeding goal trait. Information on the QTL became available either at birth or at the embryo level. Response in progeny test schemes relative to a base breeding scheme without QTL information ranged from +4.5% (QTL explaining 5% of the additive genetic variance) to +21.2% (QTL explaining 50% of the additive genetic variance). A QTL explaining 5% of the additive genetic variance would allow to progeny test 130 young bulls and maintain genetic response at the level of the base scheme, consisting of 200 progeny tested young bulls. For schemes with increased embryo production and selection on QTL information at embryo level, genetic progress was up to 31.3% higher. This provides opportunities to change the design of the breeding program.

INTRODUCTION

In the past 10 years, many efforts have been undertaken to detect genes underlying economically important traits in dairy cattle. Quantitative trait loci (QTL) have been identified, i.e. chromosomal regions acting on a certain trait have been localized without knowing the exact position or the mutation responsible for the effect. First results were obtained for milk production traits (e.g. Georges et al., 1995), followed by results for conformation and functional traits (e.g. Ashwell et al., 1996). In some cases the QTL has been fine-mapped, up to the point where the mutation responsible for the effect has been identified. This has been accomplished for monogenic traits like BLAD (Shuster et al., 1992), and, more recently, also for quantitative traits like milk production (Grisart et al., 2002; Blott et al., 2003).

QTL have been mapped with the aid of genetic markers. Markers for which linkage to specific QTL has been established, can be used for marker assisted selection (MAS). They give information on the inheritance of the QTL alleles and thus give information about the genetic potential of an animal. This can be used to increase genetic progress, either by increasing the reliability of estimated breeding values, by reduction of the generation interval or by increasing the selection

intensity. Soller and Beckman (1982) suggested screening of sires to be progeny tested, based on genetic polymorphisms. After that, various studies have evaluated a range of breeding schemes with MAS. Some of these studies looked at two-stage selection schemes where candidate young bulls were selected based on pedigree and marker information and subsequently the most promising animals entered the progeny test (e.g. Kashi et al., 1990; MacKinnon and Georges, 1998; Spelman and Garrick, 1998). Another group of studies looked at selection in a single stage, combining all available information, including marker information (e.g. Meuwissen and van Arendonk, 1992; Meuwissen and Goddard, 1996), and allowing selection of parents across all age classes. Estimates for additional gain ranged from –6% (Spelman and Garrick, 1997) to +105% (Spelman et al., 1999). Profitability of MAS breeding schemes depends on several factors, e.g. fraction of the genetic variation explained by the QTL, frequencies of QTL alleles, design of the breeding scheme and time horizon of evaluation.

Current dairy cattle breeding schemes rely on progeny testing to assess the genetic value of bulls. Progeny testing increases the generation interval, but this is compensated by an increase in accuracy. Progeny testing would not be necessary if markers would explain a substantial fraction of the genetic variance. This will not be realized in the near future, however, because breeding goals in dairy cattle consist of many traits. MAS of sires will therefore most likely be applied as an initial selection step in a multi-stage setting. Marker information will increase the accuracy of selection, especially in the first stage. This can lead to an increased rate of gain when the breeding scheme is changed to enable preselection based on marker information. Alternatively, it offers the possibility to progeny test less young bulls while maintaining the rate of gain.

Genotyping techniques can be applied to newborn animals, but they can also be applied to embryos (Bredbacka, 2001). Selection at the embryo stage offers additional opportunities, because the number of implanted embryos can be reduced, compared to a scheme based on selection of calves. To date, these opportunities have not been explored in detail.

The current study will first look at general aspects of multi-stage selection and then address

- the possibilities to reduce the number of progeny tested young bulls in a MAS-scheme, while maintaining the rate of gain
- the additional genetic progress from MAS in breeding schemes with increased embryo production and genotyping of embryos

MATERIAL AND METHODS

General Characteristics of Multi-Stage Selection

Marker Assisted Selection can be part of multi-stage selection: markers are typed early in life and used in the first stage of selection, for example in combination with pedigree information. In the second stage, the remaining selection candidates are selected on a combination of first stage information, and phenotypic information that has come available in the meantime.

Several selection schemes are evaluated to illustrate the general characteristics of multi-stage selection compared to selection in a single step, i.e. after all information is available. In these schemes, selection is for one trait and in one sex only. The number of selection candidates and the total number of animals selected after the final selection step are fixed, i.e. the overall selected fraction (p_m for males, p_f for females) is constant. The response of multi-stage selection relative to the response of single-stage selection depends on the fractions selected in stage 1 (p_1) and in stage 2 (p_2), and on the ratio between the accuracies of the selection index in stage 1 and in stage 2 (Saxton, 1983). Fractions selected in the first and the second stage are varied, from all selection after stage 2 (p_1=1, p_2=p) to all selection after stage 1 (p_1=p, p_2=1). In all schemes, accuracy in the second stage was set to 0.93, representing an index consisting primarily of information on a large group of progeny. In the first set of alternatives, the ratio of accuracies in stage 1 and in stage 2 is set to 0.22, reflecting a situation with relatively low accuracy of the stage 1 index. In a second set of alternatives this ratio is set to 0.69, reflecting a situation with relatively high accuracy of the stage 1 index. Overall selected fraction in both sets of schemes is either 0.01, 0.02, 0.05, 0.10 or 0.20. The first and second set of alternatives reflect schemes with a fixed number of selection candidates at the start of stage 1, and a fixed number of selected animals after stage 2. In a third set of alternatives, selected fractions in the first stage are varied while keeping the number of selection candidates and the selected fraction in stage 2 at a fixed level. Accuracies are equal to accuracies in the second set of alternatives. The third set of alternatives represents situations where the number of candidates available before selection in stage 1 are increased with the degree of first-stage selection. Selected fraction in stage 2 is either 0.01, 0.02, 0.05, 0.10 or 0.20.

Responses are computed using an exact method developed by Ducrocq and Colleau (1986). This method is based on principles described by Tallis (1961) and Dutt (1973).

Breeding Schemes

The response in a number of schemes applying MAS-preselection is compared to the response in a breeding scheme with preselection on pedigree information only. The comparison is carried out by deterministic simulation of multi-stage selection schemes, using the program SelAction (Rutten et al., 2002). This program predicts the rate of genetic gain based on selection index theory. The program accounts for reduction in variance due to selection (Bulmer, 1971), and corrects selection intensities for finite population size and for the correlation between index values of family members (Meuwissen, 1991). Full pedigree information as with BLUP selection is accounted for (Villanueva et al., 1993), and the program is therefore an accurate approximation of stochastic simulation with BLUP. Deviations from normality after the first stage of selection are dealt with, and selection response is predicted for the equilibrium situation (Rutten et al., 2002).

Base breeding program. The breeding program is a closed nucleus scheme, with discrete generations. The breeding goal consists of one trait, with heritability 0.35 and phenotypic variance of 100. Selection of males and females in stage 1 is for an index based on pedigree information only. In stage 2, selection in females is for an index based on pedigree information and phenotypic information of the candidate and its half sibs and full sibs. In males, selection in stage 2 is for an index based on pedigree information, progeny information and sib information. This situation is comparable to selection for a milk production trait in dairy cattle. Each generation, 20 sires and 200 dams are selected to produce 4000 embryos. Half of these embryos result in live calves, with a sex ratio of 0.50. In stage 1, 200 males are selected out of 1000 to enter the progeny test. In stage 2, 20 males are selected out of 200 to produce the next generation. Selection in females reduces numbers from 1000 to 700 in stage 1 and from 700 to 200 in stage 2. Involuntary culling and death are ignored. The base breeding scheme is summarized in Figure 1.

Breeding schemes with MAS. In the breeding schemes with MAS, additional information, consisting of information on Quantitative Trait Loci (QTL), is available in the first stage. The QTL explains either 5, 10, 20, or 50% of the genetic variance, referred to as Q05, Q10, Q20 and Q50, respectively. The remaining genetic variation results from polygenes (i.e. not marked). The QTL information is modeled as a trait that is correlated with the breeding goal trait and has a heritability of 1. The correlation between the QTL and the breeding goal trait depends on the amount of variation that is explained by the QTL. The correlation between the QTL and the polygenic component is 0 in the base generation (i.e. prior to selection). Genetic correlations of the breeding goal trait with the QTL and polygenic component are $\sqrt{q}$ and $\sqrt{1-q}$, respectively, where q is the

fraction of the genetic variance explained by the QTL. Correlations of the breeding goal trait with the QTL and polygenic component are $\sqrt{q * h_o^2}$ and $\sqrt{1-q * h_o^2}$, respectively, where h_o^2 is the heritability of the overall trait. The heritability of the polygenic component is $\frac{1-q}{\frac{1}{h_o^2}-q}$.

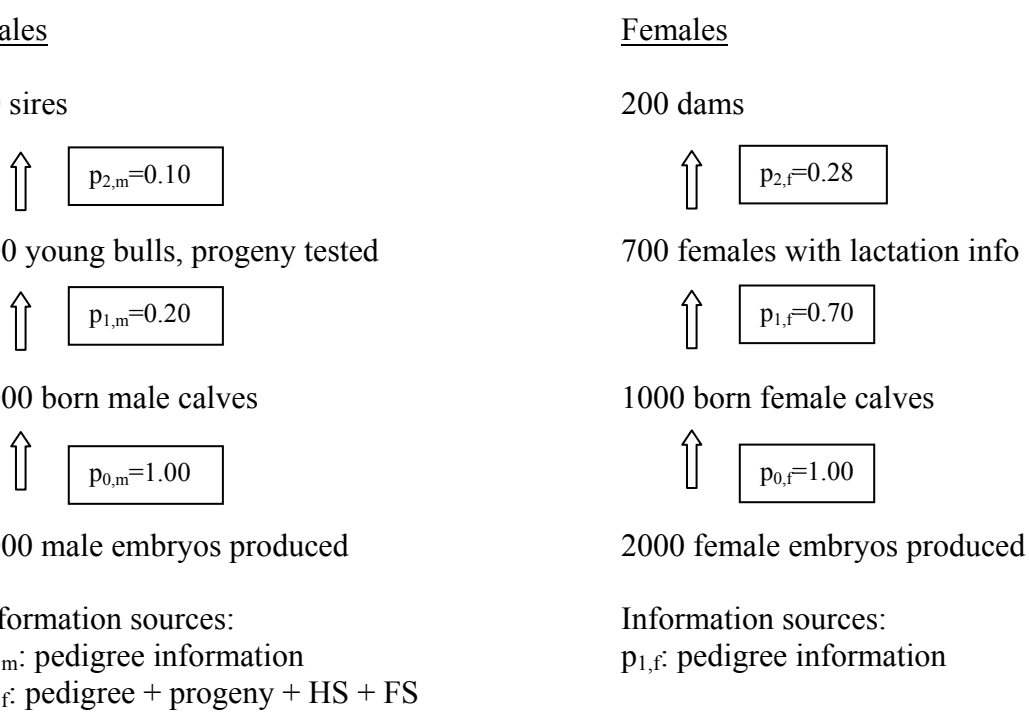

Figure 1. Selection steps, selected fractions and information sources for males and females in the base breeding scheme. Selected fractions are given as $p_{a,b}$, with a = 0, 1 or 2, representing selection stage, and b equals m or f, indicating males and females. There is no selection in embryos, indicated by $p_{0,m} = p_{0,f} = 1.00$.

Fixed number of candidates. The number of candidates before selection in stage 1 and the number of selected animals are fixed. Selection in stage 1 is on an index with information on parents (BLUP breeding values) and information on the QTL. In stage 2, information on sibs and own performance is added to the index for females, and information on sibs and progeny is added to the index for males.

Preselection in males. The level of preselection in bulls is varied from $p_{1,m} = 0.20$ (preselection equal to preselection in the base scheme) down to $p_{1,m} = 0.02$ (all selection in stage 1), with steps of 0.01. In this way, the level of preselection where genetic progress is equal to genetic progress in the base scheme, can be determined. The selected fractions for the females in stage 1 and stage 2 are equal to selected fractions in the base scheme. For the scheme with all selection in stage 1 ($p_{1,m} =$

0.02), generation intervals are reduced, because there is no need to wait for progeny test results. In comparing results for this scheme with other schemes, genetic progress is corrected for the difference in generation intervals. Schemes will be referred to as M_{Q05}, M_{Q10}, M_{Q20} and M_{Q50}, where M refers to varying selection in males, and Qxx refers to the amount of variation explained by the QTL.

Preselection in females. The advantage of additional information is expected to be different for dams and sires. To quantify this difference, schemes are simulated where the level of preselection in females is varied from $p_{1,f} = 0.70$ (equal to preselection in the base scheme) down to $p_{1,f} = 0.20$ (all selection in the first stage), with steps of 0.05. The selected fractions for the males in stage 1 and stage 2 are equal to the selected fractions in the base scheme. These schemes will be referred to as F_{Q05}, F_{Q10}, F_{Q20} and F_{Q50}, where F refers to varying selection in females, and Qxx refers to the amount of variation explained by the QTL.

Varying numbers of candidates. So far the number of embryos (i.e. number of candidates before selection) has been kept constant. However, application of genotyping techniques on embryos offers additional opportunities for preselection without increasing the number of transferred embryos. This is modeled by simulating breeding schemes with increased reproductive rates, i.e. 8000 in stead of 4000 embryos are produced by 20 sires and 200 dams. The number of progeny tested bulls is kept constant. Therefore, selection in the first stage can be more intense, i.e. $p_{1,m} = 0.10$ for males, and $p_{1,f} = 0.35$ for females. These schemes will be referred to as E_{Q05}, E_{Q10}, E_{Q20} and E_{Q50}, where E refers to extra embryos produced.

Overview. The general characteristics of the schemes are summarized in Table 1.

Table 1. Characteristics of simulated breeding schemes.

Scheme[1]	# breeding goal traits	Selected fractions males		Selected fractions females	
		$p_{1,m}$	p_m	$p_{1,f}$	p_f
M_{Qxx}	1	0.20 – 0.02	0.02	0.70	0.20
F_{Qxx}	1	0.20	0.02	0.70 – 0.20	0.20
E_{Qxx}	1	0.10 – 0.01	0.01	0.35	0.10

1) $_{Qxx}$ refers to the amount of genetic variance explained by the QTL. This is 0%, 5%, 10%, 20% or 50%

Results of these schemes are compared to a corresponding base scheme with the same characteristics, but without a QTL. Schemes are compared for genetic progress and contribution of males and females to genetic progress.

RESULTS

General Characteristics of Multi-Stage Selection

Figure 2 shows the response in multi-stage selection schemes with a high ratio between accuracies in stage 1 and stage 2 (i.e. a relatively high accuracy in stage 1). The response is expressed as a fraction of the response obtained when all selection is in stage 2 (R_0), which is the maximum response. Note that the horizontal axis contains values for $1\text{-}p_1$, i.e. larger values indicate more selection in stage 1.

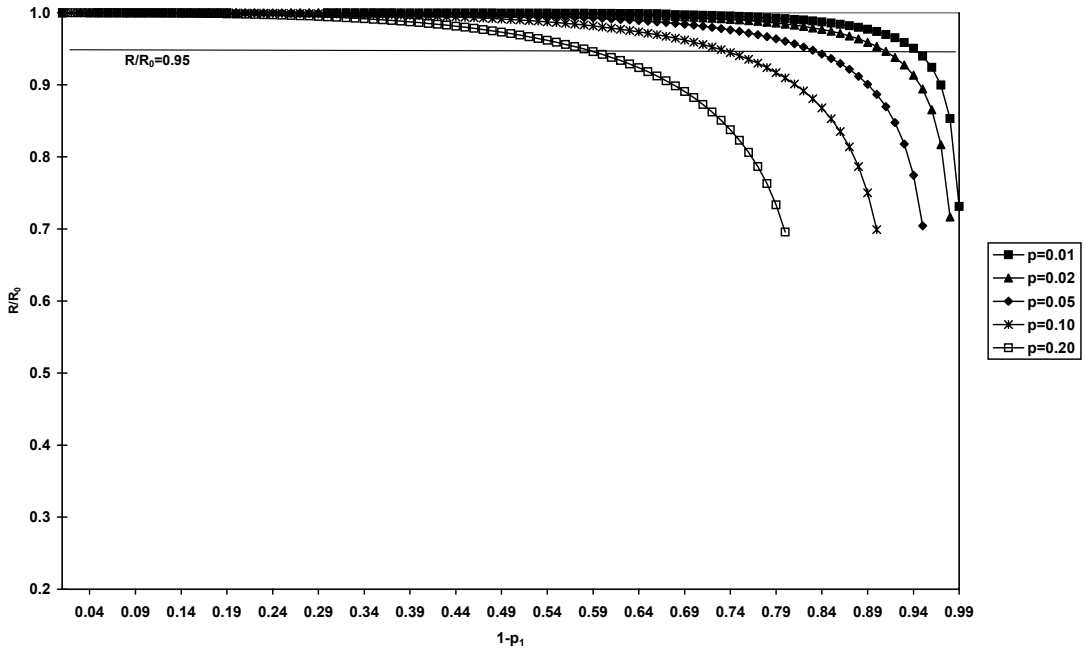

Figure 2. Response of schemes with different selected fractions in stage 1 (p_1), expressed as fraction of response in a scheme with all selection in stage 2. Responses are given for 5 different levels of overall selection (p), for high ratio (0.69) of accuracies of the index in stage 1 and in stage 2. Note that the horizontal axis contains values for $1\text{-}p_1$, i.e. larger values indicate stronger selection in stage 1. The horizontal line ($R/R_0 = 0.95$) indicates responses equal to 95% of the response in the scheme with all selection in stage 2.

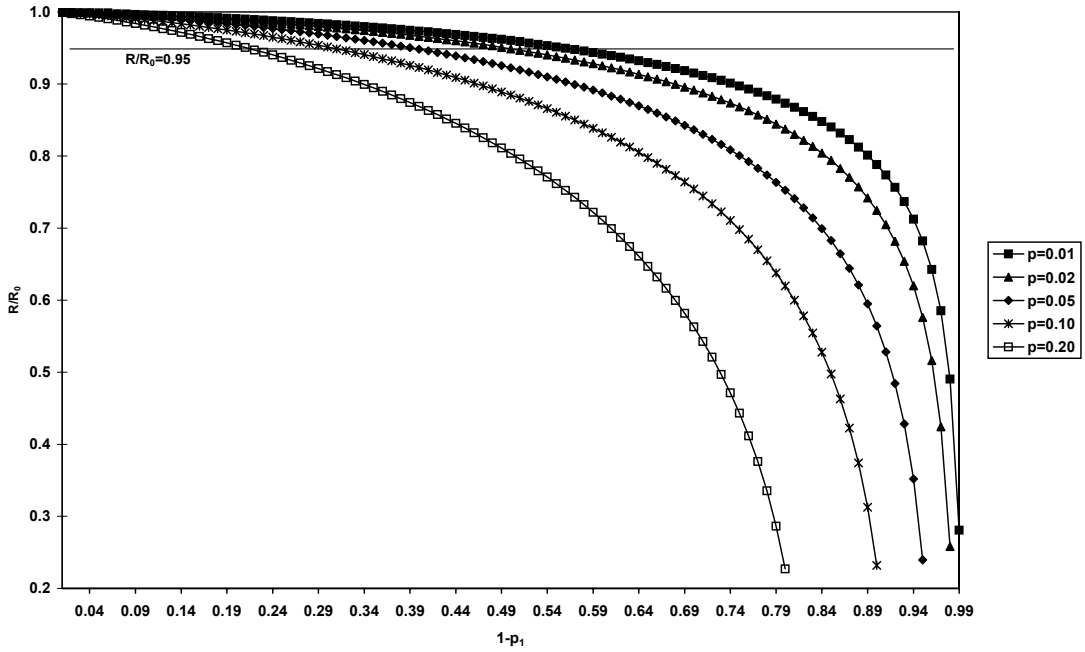

Figure 3. Response of schemes with different selected fractions in stage 1 (p_1), expressed as fraction of response in a scheme with all selection in stage 2. Responses are given for 5 different levels of overall selection (p), for low ratio (0.22) of accuracies of the index in stage 1 and in stage 2. Note that the horizontal axis contains values for $1-p_1$, i.e. larger values indicate more selection in stage 1. The horizontal line (R/R_0 = 0.95) indicates responses equal to 95% of the response in the scheme with all selection in stage 2.

Response is highest when there is no preselection. There is only a slight reduction in response when limited preselection is applied. The reduction in response is due to culling of valuable animals that would have been selected when all information would have been available. The reduction in response is small and approximately linear with increasing selection in stage 1 until a certain level of preselection, where response drops at increasing rate. For higher levels of overall selection, more preselection can be applied before overall response drops below 95% of the maximum response. For example, in Figure 2, when the overall selected fraction is 0.01, response drops below 95% when the selected fraction in stage 1 equals 0.05 ($1-p_1 = 0.95$). This means that 95% of the candidates does not have to be retained until the second selection step, with only 5% loss in response. When the overall selected fraction is 0.20, response drops below 95% when the selected fraction in stage 1 equals 0.42 ($1-p_1 = 0.58$, Figure 2).

The reduction in response depends on the ratio between accuracies of the selection index in stage 1 and stage 2. This can be seen by comparing results in Figure 2 with results in Figure 3. The ratio of accuracies of the selection index in stage 1 and stage 2 is 0.69 in Figure 2 and 0.22 in Figure

3, respectively. In Figure 3, a reduction of 5% of the response for overall selected fractions of 0.01 and 0.20 is obtained for selected fractions in stage 1 of 0.44 ($1-p_1=0.56$) and 0.78 ($1-p_1=0.22$), respectively. The difference between Figures 2 and 3 depends on the ratio of accuracies, and not on the absolute level of accuracies (Tallis,1961).

In Figures 2 and 3, the selected fractions in stage 1 and stage 2 were varied, keeping the overall selected fraction constant. In Figure 4, selection in stage 2 is kept at the original level, and the number of selection candidates at the beginning of the selection process is varied. This results in increased overall selection intensity. When the selected fraction is currently 0.01 (selection in one stage), the increase in response is 8.5% when an additional preselection step is introduced with a selected fraction of 0.50 (Figure 4). For example, initially, 10 out of 1000 animals are selected in one stage. With preselection, 1000 out of 2000 animals are selected in stage 1, and 10 out of 1000 in stage 2. The increase in response is 24.1% when preselection with a selected fraction of 0.50 is introduced in a scheme with a selected fraction of 0.20 (Figure 4). So, when the selected fraction is already small, increase in response is relatively small when a preselection step is introduced. In the latter case, the increase in overall selection intensity is relatively higher than in the first case.

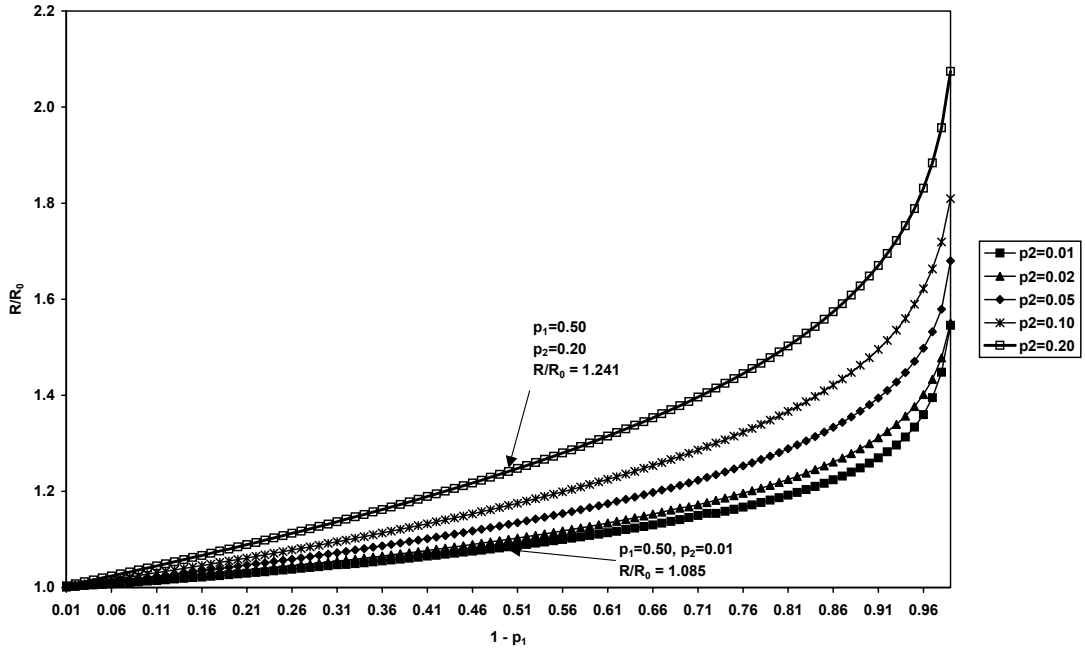

Figure 4. Response of schemes with different selected fractions in stage 1 (p_1), expressed as fraction of response in a scheme with all selection in stage 2. Responses are given for 5 different levels of selection in stage 2 (p_2), for high ratio (0.69) of accuracies of the index in stage 1 and in stage 2. Compared to previous schemes, there are more selection candidates at the beginning. Note that the horizontal axis contains values for $1-p_1$, i.e. larger values indicate more selection in stage 1.

MAS Breeding Schemes with Varying Preselection in Males

Figure 5 shows genetic response per generation relative to response in the base scheme, for varying levels of preselection in bulls and various sizes of the QTL effect. Information on the QTL is available in both males and females, but preselection is only applied in the males. When the QTL explains 5% of the genetic variance (M_{Q05}), genetic progress is 4.5% higher, for selected fractions equal to selected fractions in the base scheme ($p_{1,m}$ = 0.20). For M_{Q10}, M_{Q20} and M_{Q50}, genetic response per generation is 7.7%, 12.5% and 21.2% higher, respectively (Figure 5).

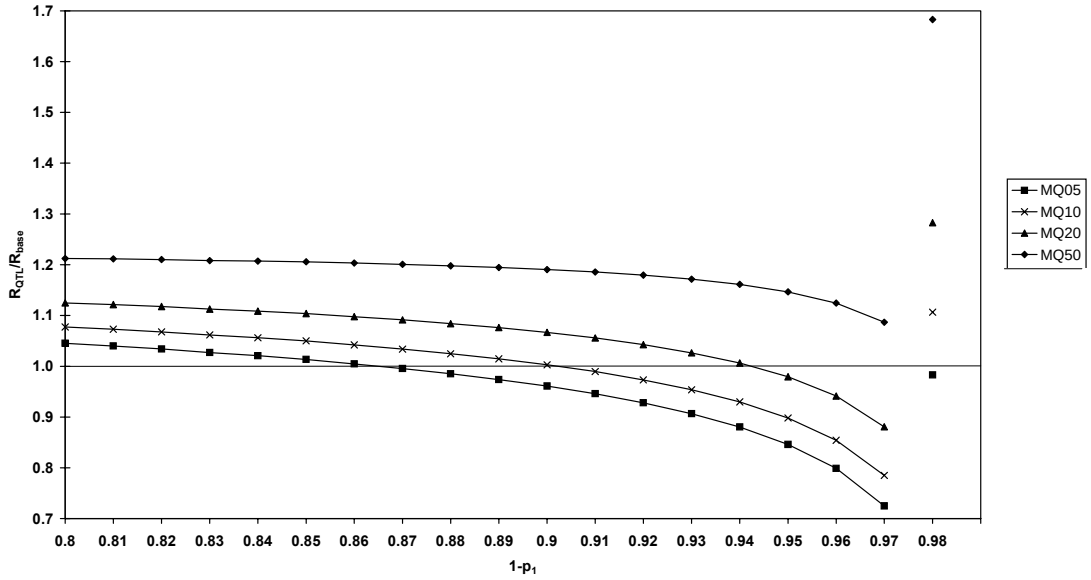

Figure 5. Genetic response per generation relative to response in the base scheme, when information on a QTL explaining 5, 10, 20 or 50% of the genetic variance is used. Selection in males in stage 1 is varied from 0.20 (equal to the base scheme; $1\text{-}p_{1,m}$ = 0.80) down to 0.02. The different generation interval for $p_{1,m}$ = 0.02 (no progeny testing) has been accounted for. Number of selection candidates and overall selected fraction are constant and equal to the base scheme.

As can be seen in Figure 5, increasing the selection intensity in stage 1 while decreasing selection intensity in stage 2 leads to decreased genetic progress. However, the decrease is small when changing the selected fraction in stage 1 from 0.20 to 0.10, especially for large QTL. Genetic progress compared to the base scheme is -3.9%, +0.3%, +6.7% and +19.0%, for M_{Q05}, M_{Q10}, M_{Q20} and M_{Q50}, respectively (Table 2, Figure 5), when the selected fraction in stage 1 is 0.10. Generation interval in males and females was assumed to be 5 for all values of $p_{1,m}$, except for $p_{1,m}$ = 0.02, where the generation interval was assumed to be 3. Due to the different generation interval when

$p_{1,m}$ = 0.02 (all selection in males in stage 1, i.e. no progeny testing of young bulls is needed), these results deviate from the curve. Responses for the schemes without progeny testing ranged from –1.7% for M_{Q05} to +68.3% for M_{Q50}, relative to the base scheme.

The patterns of response for M_{Q50} and M_{Q05} were comparable to the patterns shown in Figure 2 and 3, respectively, for an overall selected fraction of 0.02. For the schemes in Figure 2 and Figure 3, the ratios between accuracies in stage 1 and stage 2 were 0.69 and 0.22, respectively. The ratio between accuracies in stage 1 and stage 2 was 0.72 for M_{Q50} and 0.31 for M_{Q05} (Table 2). Table 2 shows genetic progress at two levels of preselection in males ($p_{1,m}$ = 0.20 and $p_{1,m}$ = 0.10). Also, the level of preselection resulting in responses equal to response in the base scheme is listed, and the contribution of females to genetic progress is shown.

Table 2. Genetic progress in MAS breeding schemes relative to base breeding scheme, for selected fraction in bulls in stage 1 ($p_{1,m}$) of 0.20 and 0.10, selected fraction in stage 1 resulting in equal genetic progress as in the base scheme, and contribution of females to genetic progress.

MAS-scheme[1]	Ratio of accuracies[2]	Genetic progress[3], $p_{1,m}$ = 0.20 (%)	Genetic progress[3], $p_{1,m}$ = 0.10[4] (%)	$p_{1,m}$ with equal genetic progress as in base scheme	Contribution of females to genetic progress[5] (%)
M_{Q00}	0.22	0	-9.7	0.20	28.0
M_{Q05}	0.31	+4.5	-3.9	0.13	28.2
M_{Q10}	0.37	+7.7	+0.3	0.10	28.5
M_{Q20}	0.48	+12.5	+6.7	0.06	29.1
M_{Q50}	0.72	+21.2	+19.0	0.02	31.6

[1]MAS-scheme: M refers to marker assisted preselection in males, $_{Qxx}$ indicates percentage of additive genetic variance explained by the QTL

[2]Ratio of accuracies in stage 1 and stage 2

[3]Relative to base breeding scheme

[4]This corresponds to a scheme where the number of progeny tested bulls is only 50% of the number of progeny tested bulls in the base scheme

[5]When selected fraction in stage 1 is equal to stage 1 selected fraction in base scheme; in the base scheme, selection is on pedigree info only

For M_{Q05}, genetic response is equal to response in the base scheme, when the selected fraction in stage 1 is 0.13 (Table 2). This means that progeny testing 130 bulls in a MAS-scheme with a QTL explaining 5% of the genetic variance would yield equal genetic response as a scheme where 200 bulls are progeny tested and no QTL is used. For M_{Q10} and M_{Q20}, response is equal to response

in the base scheme when selected fractions in stage 1 are approximately 0.10 and 0.06, respectively, corresponding to 100 and 60 progeny tested bulls. Genetic response for M_{Q50} is always higher than genetic response in the base scheme.

In the base scheme, 28.0% of genetic progress is contributed by selection in females. When information on a QTL is used, the contribution of selection in females increases slightly. For M_{Q05}, M_{Q10}, M_{Q20} and M_{Q50}, the contribution of selection in females was 28.2%, 28.5%, 29.1% and 31.6% respectively, when selected fractions were equal to selected fractions in the base scheme. The contribution of QTL information leads to increased accuracy of selection in stage 1 and stage 2. The increase in accuracy in stage 1 is equal for males and females, while the increase in accuracy in stage 2 is relatively higher for females than for males. This results in a relatively higher contribution of females to the genetic progress for schemes with a QTL explaining a higher percentage of the additive genetic variance.

MAS Breeding Schemes with Varying Preselection in Females

Figure 6 shows genetic response relative to the base scheme when preselection in females is varied while keeping the selected fraction in stage 1 of males at 0.20. Information on the QTL is available in both males and females, but preselection is only applied in the females. This represents breeding schemes where preselection leads to less females that need to be performance tested, in order to reduce the size of the nucleus and reduce the costs of recording. Curves are relatively flat, which indicates that genetic response is only slightly decreasing with increased preselection. The ratio between accuracies in stage 1 and stage 2 is relatively high, ranging from 0.50 for alternative F_{Q05} to 0.91 for alternative F_{Q50}. Consequently, a relative flat curve is expected, based on general characteristics of multi-stage selection. With a selected fraction in stage 1 of 0.35 instead of 0.70 (i.e. halving the size of the nucleus), the genetic progress compared to the base scheme is -0.1%, +4.2%, +10.3% and +21.0%, for F_{Q05}, F_{Q10}, F_{Q20} and F_{Q50}, respectively. When the selected fraction in stage 1 of males is 0.10 instead of 0.20, curves look similar, but genetic progress is at a lower level. When $p_{1,f}$ in females is 0.35, the genetic progress compared to the base scheme is -8.5%, -3.3%, +4.5% and +18.8%, for F_{Q05}, F_{Q10}, F_{Q20} and F_{Q50}, respectively.

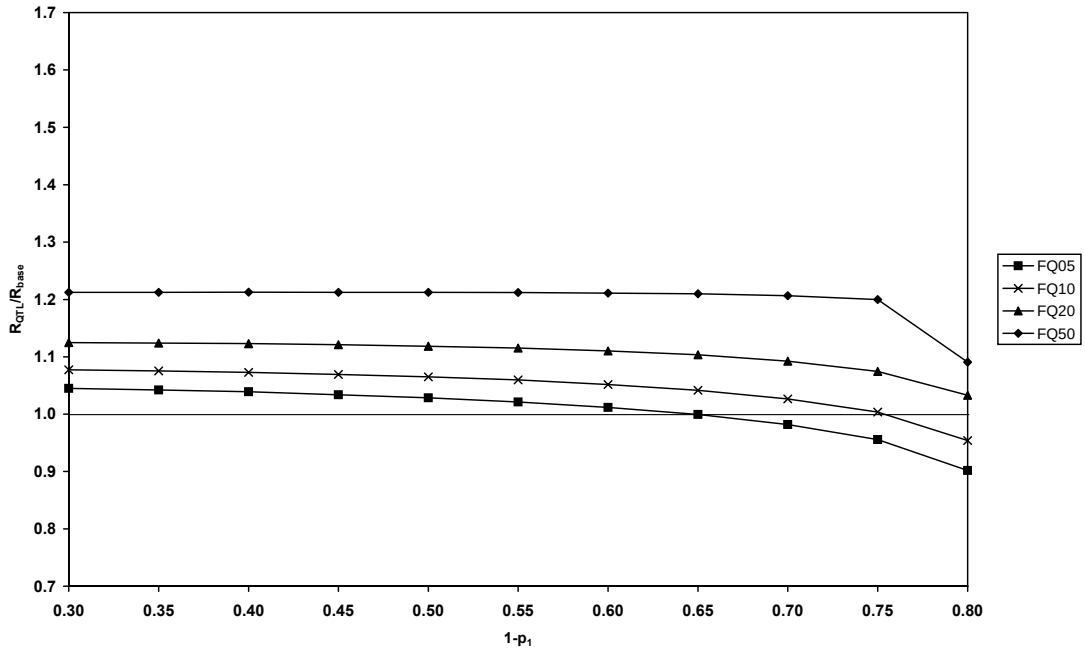

Figure 6. Genetic response per generation relative to response in the base scheme (no QTL), when information on a QTL explaining 5, 10, 20 or 50% of the genetic variance is used. Selection in females in stage 1 is varied from 0.70 (equal to the base scheme; $1\text{-}p_{1,f} = 0.30$) down to 0.20. Overall selection and selection in males is constant and equal to selection in the base scheme.

Increased Embryo Production

We have evaluated the consequences of doubling the number of embryos produced by the same number of parents, from 4000 to 8000. In these schemes, selection in stage 1 is more intense, with $p_{1,m}$ and $p_{1,f}$ equal to 0.10 and 0.35, respectively. Genetic progress is 12.1%, 17.1%, 24.5% and 37.6% higher than in the base scheme with 4000 embryos, for E_{Q05}, E_{Q10}, E_{Q20} and E_{Q50}, respectively (Table 3).

Part of the additional genetic progress in these breeding schemes is caused by increased production of embryos. Therefore, these schemes were also compared to a base scheme with increased embryo production and increased preselection. Doubling the number of embryos increased genetic progress by 4.8%. This means, for example, that the increased genetic gain of 12.1% of E_{Q05} is caused by 4.8% increase due to doubling the number of embryos and 7.0% due to MAS (Table 3). To obtain responses equal to the alternative breeding scheme, the number of progeny tested young bulls could be reduced from 200 in the base scheme to 120, 80 or 40 in schemes E_{Q05}, E_{Q10} and E_{Q20}, respectively.

Table 3. Genetic progress in MAS breeding schemes relative to base breeding scheme and alternative base breeding scheme with increased embryo production, for selected fraction in bulls in stage 1 ($p_{1,m}$) of 0.10 and 0.05, selected fraction in stage 1 resulting in equal genetic progress as in the alternative base scheme, and contribution of females to genetic progress.

MAS-scheme[1]	Genetic progress[2], $p_{1,m} = 0.10$ (%)	Genetic progress[3], $p_{1,m} = 0.10$ (%)	Genetic progress[3], $p_{1,m} = 0.05$[4] (%)	$p_{1,m}$ with equal genetic progress as in alternative base scheme	Contribution of females to genetic progress[5] (%)
E_{Q00}	0	0	-9.7	0.10	29.8
E_{Q05}	+12.1	+7.0	-1.6	0.06	30.6
E_{Q10}	+17.1	+11.8	+4.1	0.04	31.1
E_{Q20}	+24.5	+18.9	+12.6	0.02	32.0
E_{Q50}	+37.6	+31.3	+28.7	-	34.6

[1]MAS-scheme: E refers to marker assisted preselection in embryos, $_{Qxx}$ indicates percentage of additive genetic variance explained by the QTL

[2]Relative to base breeding scheme

[3]Relative to alternative base breeding scheme with increased embryo production

[4]This corresponds to a scheme where the number of progeny tested bulls is only 50% of the number of progeny tested bulls in the base scheme

[5]When selected fraction in stage 1 is equal to stage 1 selected fraction in base scheme; in the base scheme, selection is on pedigree info only

DISCUSSION

Simulation Model

In this study, breeding schemes for dairy cattle were compared by deterministic simulation of multi-stage selection schemes, using the program SelAction (Rutten et al., 2002). Information sources for the index in phase 1 are included in the index for phase 2 as well, resulting in a correlation between index 1 and 2. This correlation is equal to the ratio of the accuracies of index 1 and index 2 (Saxton, 1983). Xie and Xu (1998) compared the efficiency of multistage MAS to efficiency of conventional phenotypic selection. They applied a similar procedure, but to avoid numerical multiple integration, they imposed a constraint on the solutions for multistage MAS by forcing the covariance between the index in phase 1 and phase 2 to be 0. This has an effect similar to that of a restricted index, resulting in less than optimal gain. They did not account for the effect of finite population size on selection intensities and for reduction in variance due to selection. The

program SelAction used in the current study accounts for these effects, and constructs an optimum index using full pedigree info with an animal model (Villanueva et al., 1993).

MAS-schemes tend to have high responses initially, due to strong emphasis on marked genetic variance (e.g., Spelman and Garrick, 1997). The variation due to the QTL decreases in the first generations of MAS but reaches a constant value after a few generations. This reduction in variance leads to a reduction of the response. In our study we calculated equilibrium rates of response. The response shortly after the implementation of MAS is not predicted, but this response will be higher than the equilibrium response.

In this study, the QTL is modeled as a trait correlated to the breeding goal trait. We have assumed that the QTL behaves as a polygenic trait which remains normally distributed over time. For the QTL, this will probably only be valid if the QTL is in fact not one QTL, but several unlinked QTL, where each QTL has multiple alleles. The detection of new QTL over time will make the assumption more realistic.

The simulation model ignores inbreeding (Rutten et al., 2002). Ideally, schemes with equal rates of inbreeding should be compared. The rate of inbreeding is proportional to the square of the long-term genetic contributions of selected animals. This is likely to differ between the schemes evaluated in this study, resulting in different rates of inbreeding. It is a considerable challenge, however, to determine the expected rate of inbreeding for multi-stage selection schemes that are simulated deterministically.

Genetic Response in MAS-schemes

Several authors have studied MAS-schemes for dairy cattle. Responses from MAS obtained in these studies varied largely, depending on size of the QTL, type of breeding scheme, and the way of calculating response, e.g. equilibrium rates of response or response from one round of selection, response in young bulls or response in population females. In the following section, results from a number of MAS-studies that looked at preselection will be discussed.

The use of markers to select young bulls has first been suggested by Soller and Beckman (1982). Extra gain in their scheme was estimated (by Smith and Simpson, 1986) at 46% for direct selection on 20 QTL explaining 50% of the genetic variance. Kashi et al. (1990) found up to 26% increase in genetic merit of bulls entering the progeny test, resulting from one round of marker assisted selection of candidate bulls before entering the progeny test. In this case, selection was based on haplotypes of polyallelic markers associated with 20 segregating QTL explaining all the genetic variance. They did not account for the loss in genetic variation as a result of MAS.

Mackinnon and Georges (1998) re-evaluated the approach by Kashi et al. (1990). They found 8% increase in response in young bulls from one round of selection in a bottom up scheme, when the QTL explained 10% of the genetic variance. For a QTL explaining 50% of the genetic variance, increase in response was 23%. In the current study, equilibrium response was increased with 7.7% and 21.2% for QTL explaining 10% and 50% of the genetic variance, respectively, which is in the same range as response obtained by MacKinnon and Georges (1998). Spelman and Garrick (1998) found 9% increase in rate of gain from one round of MAS-selection in a combination of a top down and a bottom up scheme (MacKinnon and Georges, 1998). The QTL explained 16.7% of the genetic variance, and bull dams produced 40 progeny. Only one bull per full sib group was selected. QTL information was used on both the dam and the sire selection paths, in a population that had been selected for 1 to 2 generations. In the current study, 5 male progeny were produced per bull dam and selection was across families, not distinguishing between full sibs and unrelated animals. When the QTL explained 10% or 20% of the genetic variance, equilibrium response was increased with 7.7% and 12.5%, respectively, which is comparable to response obtained by Spelman and Garrick (1998).

Implications for the Breeding Program

Results in Figure 5 and Figure 6 indicate, that increased accuracy of selection in stage 1 facilitates a reduction of the number of animals entering into progeny testing or performance testing. For example, a QTL explaining 5% of the genetic variance would allow a reduction from 200 to 130 progeny tested young bulls without affecting the genetic response. Doubling the number of embryos produced in combination with preselection for a QTL explaining 5% of the genetic variance would require the progeny testing of only 80 young bulls to obtain the genetic response of the base breeding scheme. This demonstrates that QTL information can be used in various ways, but the cost structures differ.

QTL information can be measured accurately, on newborn animals or even on embryos (Bredbacka, 2001), which offers good opportunities for preselection and reduction of the number of progeny tested or performance tested animals by MAS. Schemes with increased selection intensity for males in stage 1 seem most promising, because reducing the number of progeny tested young bulls can reduce the cost of the breeding program considerably. Comparing Figure 6 with Figure 5, a relatively higher reduction in the number of performance tested females would be possible, but this is likely to have less impact on cost reduction. Whether implementation of MAS in a breeding scheme is economically beneficial, depends on the additional cost of the breeding program and the

benefits from increased genetic progress. Monetary cost and benefits from changes to the breeding program need to be tailored towards the particular breeding program. Economic evaluation of the MAS-schemes has therefore not been carried out in the current study.

Breeding goals in dairy cattle consist of a variety of traits. Although QTL have been mapped for quite a few of these traits, in some cases explaining approximately 50% of the genetic variation of a particular trait (e.g. Grisart et al., 2002), the fraction of the variation of the overall breeding goal explained by QTL is likely to be moderate. For breeding schemes in the current study, the fraction of the genetic variance explained by the QTL has been assumed to be 5% or higher. Even for the QTL explaining 5% of the genetic variance, there is a considerable impact on the genetic response.

REFERENCES

Ashwell, M. S., C. E. Rexroad Jr, R. H. Miller, and P. M. VanRaden, 1996. Mapping economic trait loci for somatic cell score in Holstein cattle using microsatellite markers and selective genotyping. Animal Genetics 27: 235-242.

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

Bredbacka, P., 2001. Progress on methods of gene detection in preimplantation embryos. Theriogenology 55: 23-34.

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

Ducrocq, V. and J. J. Colleau, 1986. Interest in quantitative genetics of Dutt's and Deak's methods for numerical computation of multivariate normal probability integrals. Genet. Sel. Evol. 18: 447-474.

Dutt, J. E., 1973. A representation of multivariate probability integrals by integral transforms. Biometrika 60: 637-645.

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

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon,

R. Spelman, M. Georges, and R. Snell, 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res. 12: 222-231.

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

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

Meuwissen, T. H. E., 1991. Reduction of selection differentials in finite populations with a nested full-half sib family structure. Biometrics 47: 195-203.

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.

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

Rutten, M. J. M., P. Bijma, J. A. Woolliams, and J. A. M. van Arendonk, 2002. SelAction: Software to predict selection response and rate of inbreeding in livestock breeding programs. The Journal of Heredity 93: 456-458.

Saxton, A. M., 1983. A comparison of exact and sequential methods in multistage selection. Theor. Appl. Gen. 66: 23-28.

Shuster, D. E., M. E. Kehrli Jr, M. R. Ackermann, and R. O. Gilbert, 1992. Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. Proc. Natl. Acad. Sci. 89: 9225-9229.

Smith, C., and S. P. Simpson, 1986. The use of genetic polymorphisms in livestock improvement. J. Anim. Breed. Genet. 103: 205-217.

Soller, M., and J. S. Beckmann, 1982. Restriction fragment length polymorphisms and genetic improvement. Proc. 2nd World Congr. Genet. Appl. Livest. Prod., Madrid, Spain VI: 396-404.

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.

Spelman, R. J., D. J. Garrick, and J. A. M. van Arendonk, 1999. Utilisation of genetic variation by marker assisted selection in commercial dairy cattle populations. Livest. Prod. Sci. 59: 51-60.

Tallis, G. M., 1961. The moment generating function of the truncated multi-normal distribution. J. R. Stat. Soc. B 23: 223-229.

Villanueva, B., N. R. Wray, and R. Thompson, 1993. Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors. Anim. Prod. 57: 1-13.

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

Chapter 7

General Discussion

Already in the sixties, it was shown that breeding programs could benefit from knowledge on genetic factors associated with economically important traits (Neimann-Sorenson and Robertson, 1961; Smith, 1967). Renewed interest in knowledge on genetic factors was generated by the discovery of restriction fragment length polymorphisms (RFLPs, Soller and Beckman, 1982), followed by minisatellites and microsatellites. This has triggered numerous studies to detect loci underlying quantitative traits (quantitative trait loci, QTL), in dairy cattle as well as in other domesticated animals. Because these studies were successful for production traits as well as non-production traits, breeding organisations have started to implement marker-assisted selection (MAS) into their breeding program. In the first part of this chapter, these applications of MAS will be discussed.

Although substantial benefits can be realised with MAS, it has not been implemented on a large scale yet, and implementation is not a straightforward process. Issues related to implementation are discussed in this chapter.

The fraction of the variance in the breeding goal explained by markers plays an important role in the anticipated benefits of MAS. Part 2 of this chapter will deal with the variance that needs to be explained by markers to start MAS, the variance that is explained to date and the possibilities of current designs and methods to detect additional QTL variance.

Most QTL have been detected by linkage analyses, resulting in relatively unprecise location estimates of the QTL. Part 3 of this chapter will deal with the required accuracy of the location of the QTL, the need for fine-mapping and/or gene-detection, and the strategy to follow, once potentially interesting QTL have been detected.

New developments like sequencing the bovine genome might change the way of QTL and gene detection. Part 4 of this chapter will discuss this development and the opportunities this will generate.

MAS – current status of implementation

MAS-applications. Most benefit can be expected from MAS on traits with low to moderate heritability (e.g. fertility), traits that are difficult or costly to measure (e.g. disease traits), traits that are not available on all selection candidates before selection (e.g. carcass traits), or traits that can only be recorded in one sex (e.g. milk production, mastitis) (Soller, 1994; Meuwissen and Goddard, 1996). Initial genome scans in dairy cattle to map QTL (e.g. Georges et al., 1995) primarily dealt with milk production traits, and consequently most commercial applications of MAS up until now are for these traits. Because of the increased importance of non-production traits for the breeding

goal, these traits have also been investigated in QTL detection studies, and results for these traits can be implemented in breeding programs.

Production traits still constitute a large part of the breeding goal, but the importance of production traits in the breeding goal has decreased over the past decade. In the Netherlands, for example, the fraction of the variance in the breeding goal that is explained by production traits has decreased from 0.66 to 0.58 (H. Geertsema, 2003, personal communication; A. Harbers, 2003, personal communication). Nevertheless, MAS for milk production traits can be beneficial, because milk production can only be recorded in one sex. Pre-selection of bulls before progeny testing can be augmented by information on markers.

MAS can be applied in different ways. In the top down approach (Kashi et al., 1990), sires segregating for a QTL are identified based on genotypes and phenotypes of their sons. Subsequently, grandsons receiving the favourable QTL variants or not receiving the unfavourable QTL variants from their grandsire, can be selected. In the bottom up approach (MacKinnon and Georges, 1998), daughters of progeny tested sires are genotyped for a limited number of regions, based on previous findings. This information is used to identify segregating sires. Sons that received favourable QTL alleles from their sire can be selected. A third approach is to integrate all available information (pedigree, phenotypes, and genotypes) into one procedure to estimate breeding values (whole pedigree approach). This has first been proposed by Fernando and Grossman (1989). Modifications to this method have been proposed (e.g., Goddard, 1992; Bink, 1998), for example to account for multiple markers and missing genotypes.

At present, the following information is available on the practical implementation of MAS in dairy cattle breeding programs.

New Zealand. Spelman (2002) reported on MAS in the breeding program operated by Livestock Improvement Corporation in New Zealand, in 1998 and 1999. A bottom up approach (MacKinnon and Georges, 1998) was applied. Young bulls were selected before progeny testing, based on 25 markers linked to six QTL affecting milk production traits. Spelman (2002) indicated that implementation has been less successful than expected due to problems associated with the introduction of new reproductive techniques like multiple ovulation and embryo transfer (MOET). These techniques were needed because selection was practised within families, which makes large full sib ships necessary. By now, for two of the utilised QTL the mutation has been identified, and information on a single SNP test can be used in evaluating the selection candidates, instead of having to rely on information on linked markers. Efforts are undertaken to identify the polymorphisms underlying other important QTL, thus facilitating MAS across families.

France. In France, the MAS-program is carried out by a combination of eight AI-companies and 3 research organisations. The goal is to apply MAS to all breeding goal traits (Boichard et al., 2002). MAS in France consists of genotyping selection candidates and their relatives for at least 33 markers associated with 12 different QTL. Selection candidates are males and females between 1 to 12 months of age. The number of markers was a compromise between efficiency of MAS and genotyping cost. The accuracy of the location estimate does not allow the use of population-wide linkage disequilibrium. The MAS-strategy applied in France can be regarded as a mixture of a top-down and a bottom-up approach (Boichard et al., 2002). The utilised QTL had an effect on one or more of the following traits: milk, fat and protein yield, fat and protein content, somatic cell score, and female fertility. Each QTL explained 8-20% of the genetic variance of a particular trait, with the exception of a QTL on chromosome 14 affecting fat %, that explained approximately 50% of the genetic variance. Female and male selection candidates as well as historical animals (i.e., ancestors of selection candidates) are genotyped, with approximately 10,000 animals genotyped each year, consisting of approximately 5,000 candidates and 5,000 historical animals (Boichard et al., 2002).

Germany. In Germany, a top-down MAS-scheme is expected to start in 2003. It will be carried out as a joint effort by most of the 10 different breeding organisations for dairy cattle (Bennewitz et al., 2003b). MAS is applied on candidate bulls for progeny testing and on candidate bull dams. Three QTL-regions with 13 markers in total are involved, affecting milk production traits and somatic cell score. In addition, selection is for the DGAT1 gene, which also has an effect on milk production traits (Grisart et al., 2002; Winter et al., 2002). Approximately 5,000 animals have been genotyped so far. A model is under construction for evaluation of animals (genotyped animals as well as animals that are related to genotyped animals), following a marker-assisted best linear unbiased prediction (MA-BLUP) procedure (Fernando and Grossman, 1989).

The Netherlands. The MAS-program in the Netherlands started in 1999. The program is based on a whole pedigree approach. New-born calves born in the nucleus breeding program of Holland Genetics, are genotyped for 12 microsatellite markers associated with 2 QTL, and for 2 identified functional mutations (single nucleotide polymorphisms, SNP's). This information is used for preselection of candidate bulls for progeny testing. Dams and sires of selection candidates are also genotyped. The number of animals with genotypes ranges from 2,600 (SNP's) to 5,000 (microsatellites). The QTL and genes involved have an effect on production traits. Genetic evaluation is carried out by using a modification of the MA-BLUP procedure described by Bink and

van Arendonk (1999), which uses an MCMC-algorithm. This procedure samples missing genotypes and utilises all relationships in the pedigree.

USA. Several dairy cattle breeding organisations in the USA (ABS, Accelerated Genetics) also apply MAS for preselection of young bulls before progeny testing, for genes known to have an effect on production traits. These include the prolactin gene (Cowan et al., 1990), and beta-lactoglobulin and kappa-casein (Bovenhuis et al., 1992).

Implementation issues

Calculations on the potential benefits from MAS (e.g. Soller and Beckman, 1982) showed that additional genetic progress in MAS breeding schemes can be substantial. These calculations were based on optimal situations, with all animals having known genotypes, high marker density, or high proportion of variance explained by the QTL. However, organisations that have implemented MAS, have experienced a number of problems during implementation, resulting in benefits that are lower than expected. The problems are related to the following factors:

1. the available QTL do not explain enough genetic variation. Additional genetic progress in the breeding goal that can be obtained by MAS is not large enough to offset the additional costs.
2. a number of QTL are available that affect individual traits included in the breeding goal, but the number of QTL with a significant effect on the total breeding goal is smaller. For example, the effect of the DGAT1 polymorphism on BTA 14 is largest for fat percentage, explaining almost 50% of the genetic variance. Other milk production traits are influenced by this gene as well (Grisart et al., 2002). In the Netherlands, however, the effect of DGAT1 on the overall index for milk production traits (INET, including milk yield, fat yield and protein yield) is negligible. This makes utilisation of this polymorphism less attractive for selection in the Netherlands.
3. costs of genotyping have remained relatively high. As a consequence, the number of genotyped markers is less than in the ideal situation, resulting in sub-optimal benefits from MAS.
4. difficulties in infrastructure. For example, different protocols are used by research labs and routine labs. This complicates transfer of protocols from research to every-day routine, and complicates the analysis of markers determined in different laboratories. A sophisticated database is needed for processing, checking, and storing genotypes, and to handle genotypes from different origin.
5. introduction of associated techniques. Implementation in New Zealand, for example, has been faced by problems related to the introduction of ovum pick up and in vitro fertilisation

(Spelman, 2002). This technique was needed to increase full sib family size for MAS within-family.

6. implementation of information on linked markers in genetic evaluation is difficult on a large scale. Software is not available that can be applied to optimally handle this situation and to deal with missing marker information.
7. reluctance to accept new technologies.

Item 1 and 2 deal with the variance explained by currently detected or detectable QTL. Item 6 deals with how MAS is used. These topics will be discussed further on in this chapter.

QTL-DETECTION

Based on the topics identified in the previous section, the following questions arise that are related to the variance explained by QTL:

1. which fraction of the genetic variation can be explained by current strategies or designs?
2. which fraction of the genetic variation explained by QTL should we aim for, to make MAS worthwhile?

Strategies

Currently, there are two main strategies for QTL-detection: genome scans and the candidate gene approach (Andersson, 2001). Genome scans will identify QTL with a major effect, but are less likely to detect QTL with a small effect (Andersson, 2001). In dairy cattle, this approach is attractive, because pedigree structures that facilitate this approach are already available, like the daughter and granddaughter designs (Weller et al., 1990). However, genome scans are also applied in designs specifically set up for QTL detection. For example, crossbred designs common to pig and poultry breeding (Andersson et al., 1994) have been applied to dairy cattle as well (Larroque et al., 2002; Spelman, 2002).

Utilisation of the granddaughter design for genome scans has resulted in a large number of QTL, initially only for production traits (e.g. Georges et al., 1995), followed by QTL for other traits (e.g. Ashwell et al., 1996; Schrooten et al., 2000). To reduce the cost for genotyping, selective genotyping (Lander and Botstein, 1989) and selective DNA-pooling (Lipkin et al., 1998) have been applied. Animals with extreme values for the phenotype (e.g. top and bottom 10%) are genotyped individually (selective genotyping) or DNA-pools are made and genotyped (pooling). A strong

reduction in the number of genotypes leads to a relatively small reduction in power, compared to a situation where all individuals are typed (Lander and Botstein, 1989).

Selective DNA-pooling has successfully been applied in analyzing milk protein % in Israeli Holsteins (Mosig et al., 2001), resulting in numerous QTL. Mosig et al. (2001) claim that these QTL together explain 100% of the genetic variance for this trait. This is unexpectedly high, which might be due to overestimation of allelic effects, or due to assigning average heterozygosity to every QTL when calculating the explained genetic variance. In selective genotyping, different sets of animals have to be selected and analysed for different traits. However, strategies have been developed to analyse correlated traits in a set of animals with extreme values for one of the traits (Bovenhuis and Spelman, 2000). The application of selective genotyping or selective DNA-pooling to the tails of the distribution for the total breeding goal could be considered as an alternative. However, the effect of individual QTL on the total breeding goal may be small, resulting in low power to detect these QTL, unless these QTL have favourable pleiotropic effects on traits in the breeding goal.

In the candidate gene approach, associations between known genes and relevant traits are explored. Candidate genes are selected based on the function of these genes. This is a powerful approach, which can even detect genes with small effects. However, due to linkage disequilibrium with the causative gene, the candidate gene may erroneously be considered the causative gene. If the candidate gene is not the causative gene, efficiency of MAS will reduce over generations. This is, however, conditional on the amount of linkage disequilibrium between the causative gene and the candidate gene. The presence of the effect in the population undergoing MAS should be tested before MAS is applied, and verified in subsequent generations.

There are, however, many candidate genes, and it is time-consuming and costly to test all of these genes for associations with relevant traits. It would result in a genome scan with candidate genes. Appropriate correction for multiple testing is necessary, to restrict the number of false positive results (Andersson, 2001). As time proceeds, prior knowledge about functions of genes increases, and candidate genes can be chosen more accurately. However, there is still a chance to overlook some of the interesting genes, due to lack of knowledge about the function of these genes. This may especially be the case for, e.g., regulatory regions like promotors and enhancers, that also are involved in the expression of traits. Knowledge about regulatory regions is limited (Stein, 2001).

It is hard to predict which genes have the largest influence on a particular trait. Due to limited financial resources, or due to lack of available tests, only a fraction of all candidate genes can be

tested, and we may not be testing the candidate genes that explain the highest fraction of the additive genetic variance in the breeding goal.

Studies on the associations between milk protein polymorphisms and milk production traits (e.g. Bovenhuis et al., 1992) are examples of applying the candidate gene approach to dairy cattle. More recently, Liefers et al. (2002) studied associations between the leptin gene and dairy cattle traits, and found an effect of RFLP-polymorphisms, located at the leptin gene locus, on milk yield.

In a genome scan based on a daughter or granddaughter design, QTL cannot be mapped very precisely, with confidence intervals for the QTL in typical designs as large as 30 cM or more (Darvasi and Soller, 1997). There may be a need to identify the location of the QTL with higher precision, before results can be applied in marker-assisted selection. In that case, further efforts are necessary, for example identity-by-descent mapping (Riquet et al., 1999) and combined linkage and linkage disequilibrium mapping (Farnir et al., 2002; Meuwissen et al., 2002), which has been applied to a QTL affecting milk production traits on chromosome 14. This reduced the confidence interval for the QTL considerably, to approximately 3 cM. The positional candidate gene approach was applied to this region, eventually resulting in detection of functional mutations (Grisart et al., 2002; Blott et al., 2003). This is an example of combining the genome scan and the candidate gene approach, with candidate genes restricted to a relatively small region of the genome.

Large fractions of the genetic variance for particular traits have been attributed to one gene (e.g. Grisart et al., 2002; fat percentage, 64 percent of genetic variance explained), or multiple QTL (Mosig et al., 2001; protein %, all genetic variance explained). However, this may not be the case in all populations or for all individual traits, or for the total breeding goal. This indicates that it might be necessary to increase the fraction of the genetic variance that can be explained by QTL.

Power

The granddaughter design is only one of the possible designs in one of the possible strategies, but probably it will continue to serve as a starting strategy to detect QTL in dairy cattle. Power of only this design will therefore be evaluated in this section. Granddaughter designs have been applied to dairy cattle data, with medium sized granddaughter designs consisting of about 15 grandsires with 70 sons each, and large granddaughter designs consisting of about 20 grandsires and 75 sons per grandsire (Bovenhuis and Schrooten, 2002). These designs cover designs as analyzed, e.g., in France, Germany, and the Netherlands / New Zealand, and will be referred to as M_15_70 and L_20_75, respectively. Because of the potential increase in power, it is attractive to combine granddaughter designs in different countries, but this may be realised only for a very limited

number of designs, due to competition between countries and organisations running these designs. Nevertheless, the French and German granddaughter designs (Boichard et al., 2002; Thomsen et al., 2000) have been combined, although only data of five grandsire families present in both designs were analysed jointly, for markers on nine chromosomes (Bennewitz et al., 2003a). In the combined data set, 13 QTL identified in one or both of the separate designs were confirmed in the combined design, with higher significance at genomewise thresholds. Four new QTL were identified and four QTL identified in the German design could not be confirmed in the joint design (Bennewitz et al., 2003a).

In each dairy cattle breeding program, information for a granddaughter design is continuously building up, with new grandsire families with progeny tested sons coming available all the time. These are, however, not continuously exploited to detect new QTL, but could offer possibilities to increase power substantially (van der Beek, 2003, personal communication), especially when all existing relationships in the design are utilised. To show what size of design and what power would be feasible in the Netherlands, family sizes of Black and White Holstein-Friesian young bulls tested in the Netherlands over the past 10 years were analysed. When families of 30 sons or more were selected, a design of 38 grandsire families could be set up. Family size ranged from 30 to 154, with average family size of 60. This design will be referred to as NL_38_60. Calculations of power were based on Weller et al. (1990), and assuming allele frequencies of 0.5, equal family sizes, type I error of 0.05, and markers every 15 cM. When requiring a power of at least 0.80, QTL explaining 6% or more of the genetic variance could be detected, for a trait with heritability 0.35 (Table 1). In that case, effect size is about 0.2 σ_p. For traits with heritability 0.05 and 0.10, QTL explaining 10% and 8% of the genetic variance can be detected with power ranging from 0.65 to 0.80. Power to detect QTL increased dramatically when utilising design NL_38_60, especially when compared to design M_15_70. Design NL_38_60 would require generating approximately 531,000 genotypes. The cost for this are estimated to be approximately 1 million Euro.

It should be noted that the assumptions underlying the power calculations influence the results presented in Table 1. For example, the assumption that allele frequencies are 0.5 may not hold for part of the QTL, resulting in lower power than presented in Table 1. On the other hand, results in Table 1 were obtained by using a chi-square approximation (Weller et al., 1990) and taking only the paternal relationships into account. New analysis models have been developed, that result in higher power than the “standard” analyses of a granddaughter design by single-trait regression analysis. For example, the use of co-factors in regression analysis (de Koning et al., 2001) and including information on multiple traits (Knott and Haley, 2000) have been described. Bink and van

Arendonk (1999) showed that power could be increased considerably, especially for QTL with small effect, when maternal relationships were taken into account as well. Power was increased from 0.3 to 0.7 in a granddaughter design of 800 sons, originating from 20 grandsires and 400 dams, with a QTL explaining 10% of the genetic variance. This could also considerably increase the power in design NL_38_60 (Table 1).

Table 1. Power to detect QTL, with variance due to the QTL (σ_q^2) expressed as fraction of the genetic variance. This fraction is 0.04, 0.06, or 0.10. Results are for three granddaughter designs of different sizes[1], for different sizes of heritability (h^2).

Design[2]	GS[3] (no.)	sons (no.)	h^2=0.05			h^2=0.10			h^2=0.35		
			σ_q^2			σ_q^2			σ_q^2		
			0.04	0.06	0.10	0.04	0.06	0.10	0.04	0.06	0.10
M_15_70	15	70	0.22	0.33	0.56	0.30	0.46	0.71	0.40	0.60	0.84
L_20_75	20	75	0.27	0.42	0.69	0.38	0.58	0.83	0.51	0.73	0.93
NL_38_60	38	60	0.31	0.49	0.78	0.44	0.67	0.91	0.59	0.82	0.98

[1]Each sire had 100 daughters with phenotypic information. A bi-allelic QTL was assumed, with allele frequency 0.50. Type I error was assumed to be 0.05 and recombination fraction between marker and QTL was 0.075.

[2]Names of designs refer to number of grandsires and number of sons in the design. Design M_15_70 and L_20_75 refer to medium and large designs as mentioned by Bovenhuis and Schrooten (2002), and cover designs as analyzed, e.g., in France, Germany, and the Netherlands / New Zealand. Design NL_38_60 refers to a design that could be set up in the Netherlands, utilizing grandsire families with at least 30 sons that were progeny tested in the last 10 years.

[3]Number of grandsires.

Variance explained by QTL

Most QTL that have been detected to date, have an effect of 0.5 genetic standard deviation or higher. The QTL detected in the French granddaughter design, for example, have an effect of 0.5 to 1.0 genetic standard deviation (Boichard et al., 2003), and most QTL explains 8-20% of the genetic variance for that trait, except the QTL for fat percentage on chromosome 14, which explains up to 50% (Boichard et al., 2002). Although effects in QTL detection experiments may be overestimated (e.g., Georges et al., 1995), these are probably the largest QTL, and the number of QTL with small effect is expected to be much larger (Hayes and Goddard, 2001).

How much variance explained by QTL do we need? Application of QTL-information in the breeding scheme can increase genetic progress, and can also be combined with changes to the design of the breeding program. Possible changes are, for example, reduction of the number of progeny tested bulls, or replacing the progeny test scheme by a scheme without progeny testing. Schrooten et al. (2003b) showed that a QTL or a number of QTL explaining 50% of the genetic variance could increase genetic progress in a typical nucleus breeding scheme with 21% when applied in the existing breeding structure, or 31% when combined with increased embryo production. QTL explaining that much genetic variance offer opportunities to stop progeny testing, or to progeny test a limited number of bulls for marketing reasons only. In this case, genetic progress per year increases with up to 68% (Schrooten et al., 2003b), partly due to selection at younger age, resulting in reduction of the generation interval.

If the goal is to increase genetic progress by 5-10%, one or a few QTL explaining 5-10% of the genetic variance are sufficient. If the goal is to stop progeny testing completely, while maintaining accuracies of breeding values at an acceptable level in order to guarantee acceptance by customers, the total genetic variance explained by the QTL has to be probably 80%. According to Hayes and Goddard (2001), explaining 80% of the genetic variance would mean the detection of QTL with size of at least 0.2 σ_p. For a typical milk production trait with heritability 0.35, this corresponds to QTL of size 0.34 σ_g or larger. Individual QTL explain 6% or more of the genetic variance in this case. Power to detect these QTL is approximately 0.82 or higher, in design NL_38_60 (Table 1). This indicates that a large fraction of the QTL that are necessary to abandon progeny testing, could be detected with this design.

The breeding goal usually consists of a large number of traits. QTL for part of these traits may be hard to detect. Therefore, although QTL may explain large fractions of the genetic variance for individual traits, these QTL may explain a smaller fraction of the genetic variance in the breeding goal. Besides, individual genes or chromosomal regions harbouring a number of genes can affect multiple traits. This has been shown in literature (e.g. Grisart et al., 2002; Schrooten et al., 2003b). If we suppose that the QTL affects two traits in the breeding goal, then various situations can be distinguished, with regard to nature of the QTL and direction of effects. The QTL or chromosomal region can be a pleiotropic QTL, or consist of two linked QTL, with degree of linkage varying from close linkage to almost completely unlinked. The QTL may favour both traits, which is very advantageous, or the effect on the traits can be opposite. If the effects of the QTL are opposite, and for example 10% of the genetic variance in each trait is explained, this QTL will explain a smaller fraction of the genetic variance in the breeding goal. It is therefore less likely that a sufficiently

large fraction of the variance of the breeding goal has been explained already, to quit progeny testing completely.

QTL-detection – concluding remarks

The currently available pedigrees in dairy cattle offer excellent opportunities to detect QTL, also for traits with relatively low heritability, and especially if the available methodologies to increase power are utilized. QTL-detection is limited, however, to routinely collected traits, in pedigree structures already available. With "new" traits or traits that are not routinely collected (e.g. disease traits) becoming more important, additional phenotyping may be required.

Is there a need for completely new models for QTL detection? The ultimate situation would be to have dense marker maps, with markers every cM for example. These markers can be used in a candidate marker approach, covering the whole genome. The feasibility of this approach is conditional on low cost for genotyping, because approximately 3,000 markers need to be genotyped per animal. Another way to utilize dense marker maps is to skip the detection of QTL-effects, and to account for every location on the genome in the evaluation procedure. This idea has been put forward by Meuwissen et al. (2001) and is referred to as genomic selection. Accuracies of breeding values obtained in this way could be as high as 0.85, which is equivalent to the accuracy of breeding values of bulls for production traits, based on approximately 27 daughters.

MAS USING POPULATION-WIDE LINKAGE DISEQUILIBRIUM

The application of MAS may be faced with difficulties to use information on linked markers for genetic evaluation, due to, e.g., missing genotypes. This section deals with the use of markers in disequilibrium with the functional mutation, that allow direct selection on marker genotypes across the population.

A genome scan, for example through a granddaughter design, will yield QTL with relatively low precision of location, i.e. confidence intervals extend over regions of 20-30 cM or more. As reported by Darvasi and Soller (1997), 95% confidence intervals for QTL mapped in a granddaughter design can be computed as

$$CI_{95} = \frac{3000}{k * N * \alpha^2} \qquad (1)$$

where α = allele substitution effect expressed in phenotypic standard deviation units, k = design parameter (k = 1 for granddaughter design), and N = number of animals in segregating families.

For a large granddaughter design (L_20_75, Table 1), with 50% heterozygote grandsires, the confidence interval for a trait with $h^2 = 0.35$ and a QTL explaining 10% of the genetic variance would be approximately 57 cM (Equation 1). Increasing the size of the design can reduce the confidence interval. However, to obtain a confidence interval of 20 cM, a granddaughter design consisting of approximately 4,300 sons would have to be analysed. For traits with heritability 0.05 or 0.10, confidence intervals for QTL explaining up to 10% of the genetic variance encompass the whole chromosome, even in the larger designs explored (Table 1). The initial linkage analysis, with low marker density, could be followed by increasing the marker density in selected chromosomal areas. The new data should then be analysed by a combined linkage analysis – linkage disequilibrium approach, that can substantially reduce the size of the confidence interval (Meuwissen et al., 2002).

Although precision of location for detected QTL in a linkage analysis is limited, these QTL can be utilised in selection. Markers are in population-wide linkage equilibrium with the gene affecting the trait, and linkage phase needs to be established. This results in sub-optimal selection efficiency. Spelman and Bovenhuis (1998) studied MAS-schemes with size of the marker bracket varying from 2 to 15 cM, using information from markers in a region of 22 to 35 cM, respectively. The QTL explained approximately 14.3% of genetic variance (5% of phenotypic variance, and $h^2 = 0.35$). Cumulative genetic superiority after 7 generations of selection with a 15 cM marker bracket was only 49% of response obtained for a 2 cM marker bracket. Loss in response was lower for larger QTL, but for a QTL explaining 10% of the phenotypic variance, response was still approximately 25% lower for the 15 cM marker bracket, when compared with response to selection in a 2 cM marker bracket.

Uncertainty about the exact location of a QTL can also lead to selection for an incorrectly positioned QTL. Spelman and van Arendonk (1997) showed that, for a QTL explaining 10% of the phenotypic variance, response after 7 generations of selection was reduced by 19% for a 5 cM location error, and by 42 % and 44% for location errors of 10 and 15 cM, respectively. Response with MAS for incorrectly positioned QTL (Spelman and van Arendonk, 1997) or with large marker brackets surrounding the QTL (Spelman and Bovenhuis, 1998) is higher than in a scheme without MAS. However, efficiency of MAS-selection can be increased considerably when the location of the QTL is known with higher accuracy. For QTL explaining 4%, 6%, 10% or 20% of the genetic variance of a trait with heritability 0.35, the cost associated with obtaining a given confidence interval are depicted in Figure 1. A large granddaughter design (L_20_75, Table 1) would allow obtaining a confidence interval of around 30 cM for a QTL explaining 20% of the genetic variance.

A larger granddaughter design (NL_38_60, Table 1) would be able to reduce the confidence interval to 20 cM for this particular QTL. QTL explaining 10% of the genetic variance would be mapped to an interval of around 40 cM.

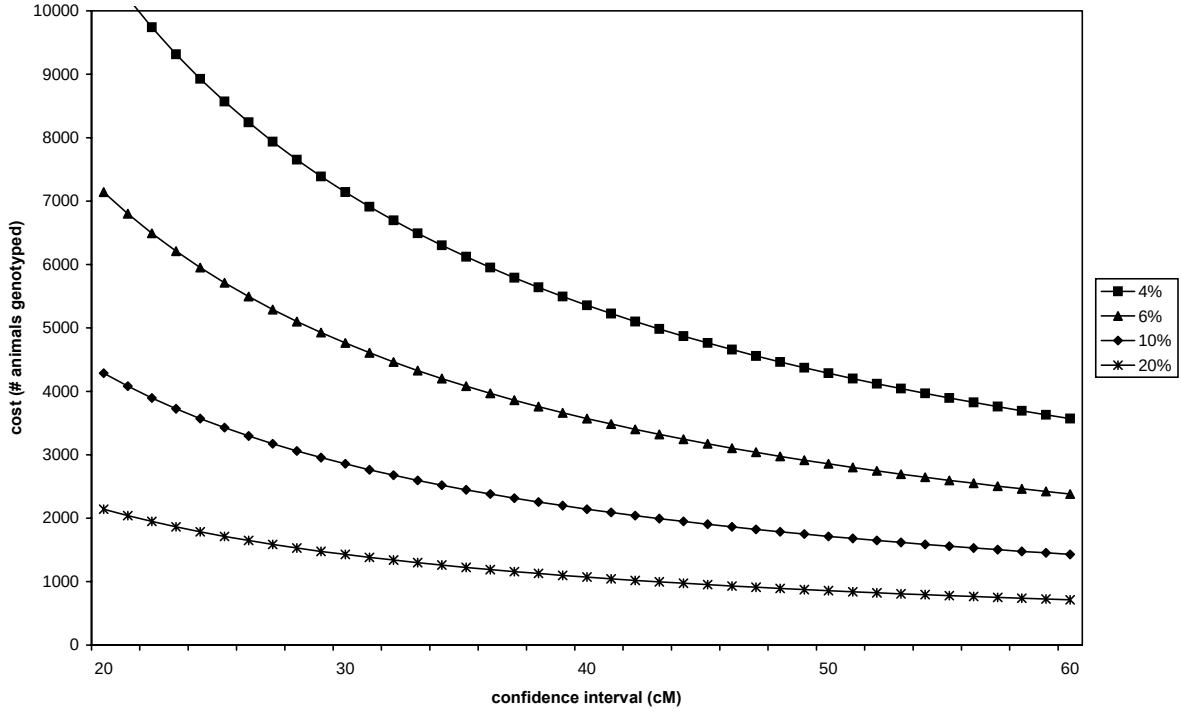

Figure 1. Cost (# animals genotyped) as a function of desired 95% confidence interval for the QTL, for four different sizes of the QTL (i.e. 4, 6, 10 or 20% explained genetic variance).

If markers that are in population-wide linkage disequilibrium with the gene can be detected, selection for marker genotypes or haplotypes across the population is possible. This requires additional efforts to narrow down the QTL-region, e.g. through high-resolution mapping. Even more accurate selection is possible when the functional mutation is known, requiring identification of positional candidate genes, mutation screening and functional analysis (Andersson, 2001). Dekkers (2003) referred to the various categories of markers as LE-markers (LE-MAS, markers in population-wide linkage equilibrium with the functional mutation), LD-markers (LD-MAS, markers in population-wide linkage disequilibrium with the functional mutation) and the functional mutation itself (GAS, gene assisted selection. Dekkers (2003) pointed out that the ease and ability to utilise markers in selection is opposite to their ease of detection. LE-markers are easiest to detect, and functional mutations are easiest to select for, and potentially give highest response to selection.

Which type of MAS can or should be used, depends on a number of factors, e.g. type of trait, variance explained by linked markers and cost of fine-mapping.

Type of trait. With markers in population-wide linkage equilibrium with the functional mutation (LE-markers), phenotypes are essential. If traits are not routinely measured, application of MAS requires a precision of location that allows for at least LD-MAS and ideally GAS.

Variance explained by linked markers. To get an idea of the efficiency of LD-MAS relative to gene-assisted selection, consider the variance explained by linked markers, as a fraction of the variance explained by the gene. Suppose a QTL has alleles Q^1 and Q^2, and the alleles of a linked marker are M^1 and M^2. Frequencies of haplotypes M^1Q^1, M^2Q^2, M^1Q^2 and M^2Q^1 are r, u, s and t, respectively. The amount of linkage disequilibrium in the population is then defined as:

$$D = ru - st \qquad (1)$$

and the variance associated with the linked marker is (Bovenhuis and Meuwissen, 1996):

$$\sigma^2_{Marker} = \left[\frac{2D^2a^2}{(r+s)(t+u)} \right] \qquad (2)$$

If there is complete linkage disequilibrium, only haplotypes M^1Q^1 and M^2Q^2 are present. This results in s = t = 0, and D = ru. The variance explained by the marker is equal to $2*r*u*a^2$, which is equal to the variance explained by the QTL.

The amount of linkage disequilibrium decreases with time and distance between loci. Nsengimana (2003) expressed the level of linkage disequilibrium as a function of the distance as

$$D' = S + (1 - S)e^{\left(\frac{-3x}{R}\right)} \qquad (3)$$

with

D’ = linkage disequilibrium as a fraction of complete linkage disequilibrium

S = residual D’ (for unlinked loci).

R = distance at which D’ reaches value of S.

x = genetic distance

This equation describes an equilibrium situation, resulting from various LD-influencing factors like mutation, recombination, drift, selection and admixture (Ardlie et al., 2002). Farnir et al. (2000) gave values for the Dutch dairy cattle population of 0.16 for S and 50 cM for R. These values can be substituted in Equation (3).

Linkage disequilibrium depends on frequencies of alleles. It can be expressed as a fraction of the maximum linkage disequilibrium for a given set of allelic frequencies:

$$D'_{ij} = \frac{D_{ij}}{D_{max}} \qquad (4)$$

where D_{max} is the maximum linkage disequilibrium, and D_{ij} the linkage disequilibrium for alleles i and j at different loci. These formula make it possible to relate the variance explained by a genetic marker (due to a linked QTL) to the distance between the marker and the QTL. Combining equations (1) to (4), the variance explained by the linked marker as a fraction of the variance explained by the QTL is equal to $(D')^2$. The QTL variance explained by a genetic marker can be related to genetic progress. It should be noted, however, that this equation is only valid when allele frequencies for M^1 and Q^1 (or Q^2) are equal. For example, when the frequency of M^1 is ¾ and the frequency of Q^1 is ½, the equation is $1/3\ (D')^2$.

Table 2 gives the variance explained by a linked marker as a function of the distance in Morgan for a bi-allelic QTL linked to a bi-allelic marker, with similar allele frequencies for the QTL and the marker. The variance explained by the linked marker is expressed as the fraction of the variance explained by the QTL. When the distance between the marker and the QTL is approximately 7 cM, the marker explains 50% of the QTL variance. When taking the recently identified DGAT1 gene as an example, initial fine-mapping efforts localised the QTL in a region of 4 cM (Riquet et al., 1999). This means that a marker in that region explains at least 67% of the QTL variance.

Table 2. Variance explained by linked markers relative to variance explained by the gene, as a function of the distance between linked marker and the gene

Distance	D’	% QTL Var. explained
0.00	1.00	100
0.01	0.95	90
0.02	0.91	82
0.03	0.86	74
0.04	0.82	67
0.05	0.78	61
0.06	0.75	56
0.07	0.71	51
0.08	0.68	46
0.09	0.65	42
0.10	0.62	39

To illustrate the variance explained by linked markers, a data set of 1,500 sires with breeding

values for milk fat% and genotypes for the DGAT1 gene (Grisart et al., 2002) and for linked markers was analysed. The fixed effect in the statistical model was either DGAT1 genotype, or genotype of the linked marker. Sums of squares explained by the different models were compared. Figure 2 shows the variance explained by the linked marker as a fraction of the variance explained by the DGAT1 genotype.

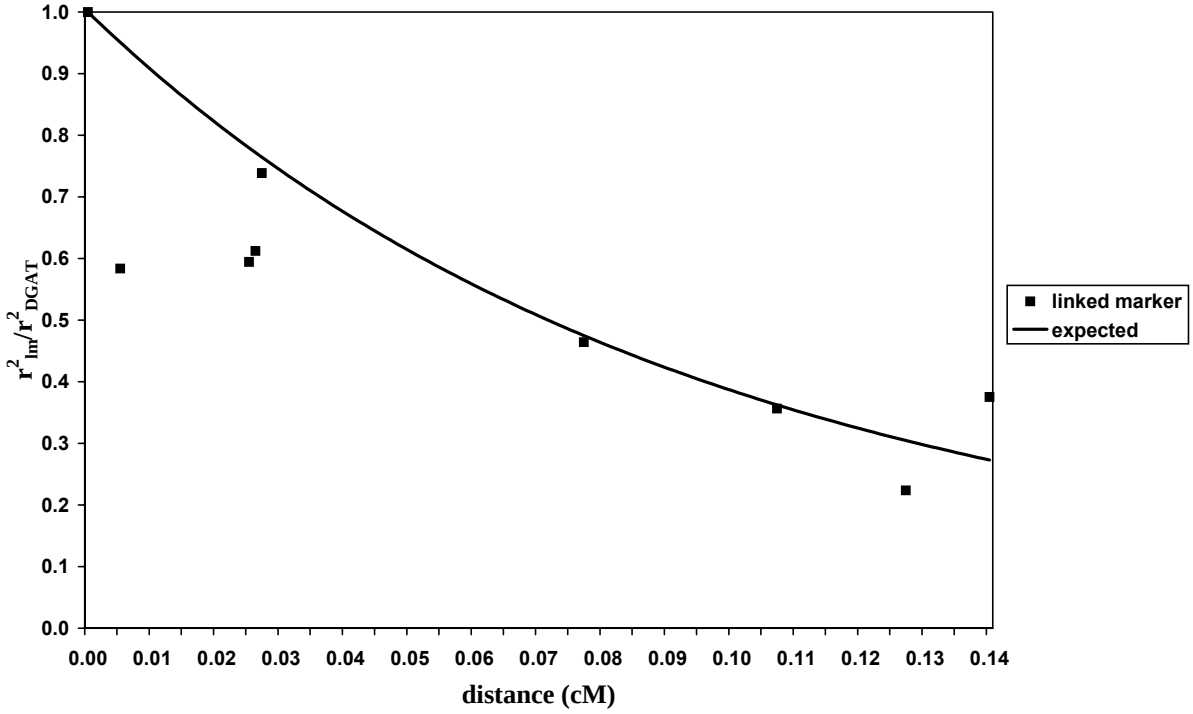

Figure 2. Variance explained by linked marker genotype as a fraction of the variance explained by the DGAT1 genotype. The line indicates the expectation.

The variance explained by linked markers is approximately equal to the explained variance that is theoretically expected. For markers within 3 cM of the DGAT1 gene, however, observed values are slightly lower than the expectation. This could be caused by higher recombination rates in the current data set, compared to the data set used for map construction, or by differences between frequencies of the DGAT1 alleles and frequencies of the alleles at the marker, resulting in less variance explained by the marker.

MAS using population-wide LD – concluding remarks

There are a number of reasons that stimulate the detection of markers in population-wide linkage disequilibrium with the functional mutations, e.g., difficulties to set up an evaluation system with linked markers, or restricted availability of phenotypes for continuous evaluation. However,

LD-markers should be in a region that is restricted to approximately 2 cM, because the variance that is explained by LD-markers quickly decreases with increasing distance between the marker and the functional mutation. Cost to identify the functional mutation may not offset the additional benefits compared to using LD-markers. The availability of many markers in identified candidate regions will facilitate the detection of markers in linkage disequilibrium with the functional mutation. These may become available from the bovine genome sequencing project, which will be discussed in the next section.

BOVINE GENOME SEQUENCING

A project to sequence the bovine genome is about to start (http://hgsc.bcm.tmc.edu/projects/bovine, last accessed 18-11-03). After human, mouse and rat, bovine will be the fourth mammal with the genomic sequence available, and, after chicken, the second domestic animal. It is expected that a working draft sequence of the bovine genome will be available in two years, which is relatively fast. The availability of BAC (bacterial artificial chromosome) fingerprints and BAC end sequences before the start of the project speeds up the sequencing process (http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/BovineSEQ.pdf, last accessed 18-11-03). A working draft sequence means a sequence that is not yet complete, because of gaps and regions with inaccurate sequence. However, the incomplete sequence is already extremely valuable for preliminary analyses (Green, 2001).

The bovine genome sequence is of interest for animal genetics as well as for human genetics. Knowledge about the bovine genome sequence will assist in comparative mapping and can help in determining the function of human genes. This is especially helpful, because there is better conservation between the human genome and the bovine genome, than between the human genome and the mouse genome. Bovine genetics can benefit from additional markers and positional candidate genes through comparative mapping, linking the bovine genome to other genomes, like the human and mouse genomes.

Once the genome has been sequenced, the location of genes will be determined, either by gene prediction techniques (i.e. sequences known to be elements of genes will be located), or by comparing the sequence with the sequence of known genes, partly of other species (Stein, 2001). Also, known elements like genetic markers will be located. This will result in a large number of genes, without exactly knowing the function of these genes. A next step will be to identify the

proteins these genes are coding for, and the processes in which these genes are involved (Stein, 2001).

The QTL and gene detection process will benefit from the results of the bovine sequencing project. Microarrays with tens of thousands expressed DNA-sequences can be prepared (Schena et al., 1995). Labelled DNA from specific tissues (e.g. udder gland of a cow with high protein yield and of a cow with low protein yield) is hybridised to the array. In this way, genes that are differentially expressed in for example cows with high and low protein yield, can be detected. Another benefit will be that abundant polymorphisms will become available, facilitating the fine-mapping of QTL. Once a QTL has been identified in a region of approximately 5 cM, it will be very easy to increase marker density in that area and apply fine-mapping techniques exploiting linkage disequilibrium to further narrow down the region. In the next step, increased knowledge about candidate genes in this small region will facilitate positional candidate cloning, to identify the functional mutation.

In human genetics, new techniques for high-throughput genotyping have been developed and applied (e.g. Oliphant et al., 2002). These techniques can reduce genotyping cost, down to approximately 10 cents per SNP-genotype, provided that at least 1000 SNP genotypes on hundreds of individuals are determined.

The availability of numerous genes, and of numerous SNP's in combination with new developments that allow cheap SNP genotyping, will lead towards new types of genome scans. Besides, it paves the way towards implementation of methodology for genetic evaluation using genomic information without actually identifying QTL or genes (Meuwissen et al., 2001). Provided large-scale genotyping can be conducted at similar, low cost as in the human field, this will change evaluation of animals completely. Assuming a dense marker map with 1 cM marker spacing, genotyping cost per animal would be approximately $ 300, to obtain breeding values with accuracy 0.85. This is considerably lower than the cost to obtain accurate breeding values for bulls, which are roughly estimated to be US$ 30,000 per bull (Spelman, 2002). Emphasis in evaluation will then change (again) to collection of phenotypic data, especially for "new" traits, i.e. traits that are not yet routinely collected.

Bovine genome sequencing – concluding remarks

Information generated by the bovine genome project will facilitate the identification of markers in linkage disequilibrium with functional mutations, or the functional mutations themselves. In conjunction with cost-effective genotyping, this will facilitate large-scale implementation of MAS.

As a result, breeding programs will undergo tremendous changes in the next 5 to 10 years and will rely heavily on genomic information.

ACKNOWLEDGMENTS

Information, ideas and comments were provided by Henk Bovenhuis, Johan van Arendonk, Sijne van der Beek, Marco Bink, Wiepk Voskamp and Martien Groenen. Their input is greatly appreciated.

REFERENCES

Andersson, L., 2001. Genetic dissection of phenotypic diversity in farm animals. Nature Rev. Genet. 2: 130-138.

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

Ardlie, K. G., L. Kruglyak, and M. Seielstad, 2002. Patterns of linkage disequilibrium in the human genome. Nature Rev. Genet. 3: 299-309.

Ashwell, M. S., C. E. Rexroad Jr, R. H. Miller, and P. M. VanRaden, 1996. Mapping economic trait loci for somatic cell score in Holstein cattle using microsatellite markers and selective genotyping. Animal Genetics 27: 235-242.

Bennewitz, J., N. Reinsch, C. Grohs, H. Levéziel, A. Malafosse, H. Thomsen, N. Xu, C. Looft, C. Kühn, G. A. Brockmann, M. Schwerin, C. Weimann, S. Hiendleder, G. Erhardt, I. Medjugorac, I. Russ, M. Förster, B. Brenig, F. Reinhardt, R. Reents, G. Averdunk, J. Blümel, D. Boichard, and E. Kalm, 2003. Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle. Genet. Sel. Evol. 35: 319-338.

Bennewitz, J., N. Reinsch, H. Thomsen, J. Szyda, F. Reinhardt, Z. Liu, C. Kühn, A. Tuchscherer, M. Schwerin, C. Weimann, G. Erhardt, and E. Kalm, 2003. Marker assisted selection in German Holstein dairy cattle breeding: outline of the program and marker assisted breeding value estimation. Paper presented at 54th Annu. Meeting of the European Assoc. of Animal Production, August 31 - Sept. 3 2003, Rome, Italy.

Bink, M. C. A. M., 1998. Complex pedigree analysis to detect quantitative trait loci in dairy cattle. Ph.D. Diss., Wageningen University, Wageningen, the Netherlands.

Bink, M. C. A. M., and J. A. M. van Arendonk, 1999. Detection of quantitative trait loci in outbred

populations with incomplete marker data. Genetics 151: 409-420.

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

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

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

Bovenhuis, H., and C. Schrooten, 2002. Quantitative trait loci for milk production traits in dairy cattle. Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France 31: 27-34.

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

Bovenhuis and Meuwissen, 1996. Detection of quantitative trait loci. Ph.D. course Armidale, Australia.

Bovenhuis, H., J. A. M. van Arendonk and S. Korver, 1992. Associations between milk protein polymorphisms and milk production traits. J. Dairy Sci. 75: 2549-2559.

Cowan, C. M., M. R. Dentine, R. L. Ax, and L. A. Schuler, 1990. Structural variation around prolactin gene linked to quantitative traits in an elite Holstein sire family. Theor. Appl. Genet. 79: 577-582.

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

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

de Koning, D. J., N. F. Schulman, K. Elo, S. Moisio, R. Kinos, J. Vilkki, and A. Maki-Tanila, 2001. Mapping of multiple quantitative trait loci by simple regression in half-sib designs. J. Anim. Sci. 79: 612-622.

Dekkers, J. C. M., 2003. Commercial applications of marker- and gene-assisted selection in livestock: strategies and lessons. Paper presented at 54th Annu. Meeting of the European Assoc. of Animal Production, August 31 - Sept. 3 2003, Rome, Italy.

Farnir, F., W. Coppieters, J. J. Arranz, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L. Moreau,

M. Mni, C. Nezer, P. Simon, P. Vanmanshoven, D. Wagenaar, and M. Georges, 2000. Extensive genome-wide linkage disequilibrium in cattle. Genome Res. 10: 220-227.

Farnir, F., B. Grisart, W. Coppieters, J. Riquet, P. Berzi, N. Cambisano, L. Karim, M. Mni, S. Moisio, P. Simon, D. Wagenaar, J. Vilkki, and M. Georges, 2002. Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. Genetics 161: 275-287.

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

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

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

Green, E. D., 2001. Strategies for the systematic sequencing of complex genomes. Nature Rev. Genet. 2: 573-583.

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

Hayes, B., and M. E. Goddard, 2001. The distribution of the effects of genes affecting quantitative traits in livestock. Genet. Sel. Evol. 33: 209-229.

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

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

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

Larroque, H., Y. Gallard, L. Thaunat, D. Boichard, and J. J. Colleau, 2002. A crossbreeding experiment to detect quantitative trait loci in dairy cattle. Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France 29: 219-222.

Liefers, S. C., M. F. W. te Pas, R. F. Veerkamp, and T. van der Lende, 2002. Associations between leptin

gene polymorphisms and production, live weight, energy balance, feed intake, and fertility in Holstein heifers. J. Dairy Sci. 85: 1633-1638.

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

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.

Meuwissen, T. H. E., A. Karlsen, S. Lien, I. Olsaker, and M. E. Goddard, 2002. Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. Genetics 161: 373-379.

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

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

Neimann-Sorensen, A., and A. Robertson, 1961. The association between blood groups and several production characteristics in three Danish cattle breeds. AAS 11: 163-196.

Nsengimana, J., 2003. Contribution to the analysis of linkage disequilibrium in livestock: effects of selection and inbreeding. Ph.D. Diss., Univ.de Louvain, Louvain-la-Neuve, Belgium.

Oliphant, A., D. L. Barker, J. R. Stuelpnagel, and M. S. Chee, 2002. BeadArray^{TM} technology: enabling an accurate, cost-effective approach to high-throughput genotyping. Biotechniques 32 (Suppl.): 56-61.

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

Schena, M., D. Shalon, R. W. Davis, and P. O. Brown, 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467-470.

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

806.

Schrooten, C., H. Bovenhuis, J. A. M. van Arendonk, and P. Bijma, 2003a. Genetic progress in multi-stage dairy cattle breeding schemes using genetic markers. (submitted)

Schrooten, C., M. C. A. M. Bink, and J. A. M. van Arendonk, 2003b. Whole genome scan to detect chromosomal regions affecting multiple traits in dairy cattle. (submitted)

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

Soller, M., 1994. Marker assisted selection - an overview. Anim. Biot. 5: 193-207.

Soller, M., and J. S. Beckmann, 1982. Restriction fragment length polymorphisms and genetic improvement. Proc. 2nd World Congr. Genet. Appl. Livest. Prod., Madrid, Spain VI: 396-404.

Spelman, R. J., 2002. Utilisation of molecular information in dairy cattle breeding. Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France 33: 11-17.

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

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

Stein, L., 2001. Genome annotation: from sequence to biology. Nature Rev. Genet. 2: 493-503.

Thomsen, H., N. Reinsch, N. Xu, C. Looft, S. Grupe, C. Kühn, G. A. Brockmann, M. Schwerin, B. Leyhe-Horn, S. Hiendleder, G. Erhardt, I. Medjugorac, I. Russ, M. Förster, B. Brenig, F. Reinhardt, R. Reents, J. Blümel, G. Averdunk, and E. 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.

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

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

Summary

The availability of molecular techniques and their potential benefit for genetic progress have led to numerous experiments to detect QTL. In dairy cattle, these studies have mainly focused on milk production traits. **Chapter 2** discusses the design of experiments to detect QTL for milk production traits, and the power of these designs. Most of the QTL mapping studies that are presently being carried out use the granddaughter design. Even the largest designs that are being carried out at this moment are expected to detect a limited fraction of the QTL underlying milk production traits. Therefore, power needs to be increased, either by increasing the size of designs, or by using sophisticated methods, e.g., using all relationships in the pedigree. A survey of the results from QTL mapping studies aiming at milk production traits reveals that especially for protein content significant QTL have been reported. Linkage has been confirmed for QTL located on chromosomes 3, 6, 14 and 20. The gene located on chromosome 14 has a major effect on fat content and has been cloned. Chromosomes 3, 6 and 20 have effects on protein content where chromosome 6 is likely to contain more than one QTL. Chromosomal regions showing effects on milk production traits also seem to have effects on other traits. Multiple trait analyses should be used to determine whether pleiotropic gene effects play a role.

Although milk production traits constitute a substantial part of the breeding goal, traits like fertility and health have become increasingly important. Compared to QTL for production traits, QTL for fertility and health may be relatively more important, for example due to low heritability of these traits, resulting in low accuracy of the current selection criterion, or due to difficulties to obtain phenotypes. Therefore, information on QTL for these traits is essential for efficient selection. **Chapter 3** describes results from a granddaughter design that was used to locate quantitative trait loci determining conformation and functional traits. In this granddaughter design, consisting of 20 Holstein Friesian grandsires and 833 sons, genotypes for 277 microsatellite markers covering the whole genome were determined. Breeding values for 27 traits regarding conformation (18), fertility (2), birth (4), workability (2) and udder health (1), were evaluated in an across-family analysis using multimarker regression. Significance thresholds were determined using a permutation test.

The across-family analysis suggested the presence of 61 quantitative trait loci, where 27 (i.e. one for each trait) were expected by chance. The test statistic exceeded the genomewise significance threshold for the following traits and chromosomes: chest width on chromosome 2, gestation length on chromosome 4, stature, body capacity and size on chromosome 5, dairy character on chromosome 6, angularity on chromosome 12, fore udder attachment on chromosome 13, and fore udder attachment and front teat placement on chromosome 19. The quantitative trait loci for size traits on chromosomes 2, 5 and 6 may also have an effect on calving ease. The

quantitative trait loci for udder traits on chromosomes 13 and 19 may also affect somatic cell score and mastitis resistance.

The QTL that have been detected for milk production traits, conformation traits, fertility and health, can be used in marker assisted selection (MAS). However, selection for a specific QTL may also influence other traits, due to pleiotropy (i.e. the QTL affects multiple traits), or because the QTL is linked to QTL affecting other traits. To avoid negative side effects of selection for certain chromosomal regions or to exploit positive effects of MAS on other traits of interest, insight into the pleiotropic effects of a QTL, or the effect of a chromosomal region on multiple traits, is needed. **Chapter 4** describes a method to determine pleiotropic QTL or closely linked QTL in an outbred population. The method is based on results from single trait analyses for different traits and is derived for a granddaughter design. The covariance between estimated contrasts of grandsires obtained in single trait regression analysis is computed. When there is no pleiotropic QTL, the covariance between contrasts depends on the heritabilities of the traits involved, the polygenic and environmental correlation between the traits, the phenotypic standard deviations, the number of sires per grandsire and the number of daughters per sire. A pleiotropic QTL results in a covariance that deviates from this expected covariance. The deviation depends on the size of the effects on both traits and on the fraction of grandsires heterozygous for the QTL. When analyzing experimental data, the expected covariance and the confidence interval for the expected covariance can be determined by permutation of the data. A covariance outside the confidence interval suggests the presence of a pleiotropic QTL or a closely linked QTL. The method is verified by simulation and illustrated by analyzing an experimental data set on chromosome six in dairy cattle. In **chapter 5**, the method is applied to a genome scan in a granddaughter design involving 833 sons of 20 grandsires. Four categories of traits were analyzed: production (five traits), udder conformation (six), udder health (two) and fertility (two). In total, 59 chromosomewise significant multiple trait quantitative trait regions (MQR) (α = 0.01) were identified. Fifteen MQR were found on chromosome 14. Four or more MQR were found on chromosome 6, 13, 19, 22, 23 and 25. Eight MQR involving udder conformation and udder health and four MQR involving production traits and udder health were found. For fertility, the number of MQR was 5 for both the combinations with udder conformation and production traits. For 22 MQR, the difference between the correlation due to the MQR and the overall genetic correlation was larger than 0.60. The results from this study can be useful in the decision whether or not to apply marker-assisted selection (MAS) for specific QTL.

Most benefit can be expected from MAS on traits with low to moderate heritability (e.g. fertility), traits that are difficult or costly to measure (e.g. disease traits), traits that are not available on all selection candidates before selection (e.g. carcass traits), or traits that can only be recorded in one sex (e.g. milk production, mastitis). MAS of sires will most likely be applied as an initial selection step in a multi-stage setting, before entering a progeny test. Marker information will increase the accuracy of selection, especially in the first stage. In **chapter 6**, closed multi-stage nucleus breeding schemes in dairy cattle that use information on quantitative trait loci (QTL), have been evaluated by deterministic simulation. In the base scheme, the selection index for dams consisted of pedigree information and own performance. The selection index for sires consisted of pedigree information and performance of 100 daughters. In alternative breeding schemes, information on a QTL was accounted for by simulating an additional index trait. The fraction of the variance explained by the QTL determined the correlation between the additional index trait and the breeding goal trait. Information on the QTL became available either at birth or at the embryo level. Response in progeny test schemes relative to a base breeding scheme without QTL information ranged from +4.5% (QTL explaining 5% of the additive genetic variance) to +21.2% (QTL explaining 50% of the additive genetic variance). A QTL explaining 5% of the additive genetic variance would allow to progeny test 130 young bulls and maintain genetic response at the level of the base scheme, consisting of 200 progeny tested young bulls. For schemes with increased embryo production and selection on QTL information at embryo level, genetic progress was up to 31.3% higher. This provides opportunities to change the design of the breeding program.

The general discussion (**Chapter 7**) lists a number of MAS-applications. Implementation is faced with a number of issues, related to, e.g., the amount of variance explained and the precision of the location estimate. The currently available pedigrees in dairy cattle offer excellent opportunities to detect QTL, also for traits with relatively low heritability, and especially if the available methodologies to increase power are utilized. In this way, the amount of variation explained by QTL can be increased, and so will benefits from MAS. Besides, this offers possibilities to reduce the number of progeny tested bulls, to make the breeding program more cost-effective.

A more precise location of the QTL would allow the use of markers in population-wide linkage disequilibrium with the functional mutations However, LD-markers should be in a region that is restricted to approximately 2 cM, because the variance that is explained by LD-markers quickly decreases with increasing distance between the marker and the functional mutation. Cost to identify the functional mutation may not offset the additional benefits compared to using LD-markers. The availability of many markers in identified candidate regions will facilitate the detection of markers

in linkage disequilibrium with the functional mutation. These may become available from the bovine genome sequencing project, which is about to start and which is expected to change identification of genomic variation and, as a result, dairy cattle breeding programs, considerably in the next 5-10 years.

Samenvatting

In de fokkerij gaat het om het verbeteren van de erfelijke aanleg voor belangrijke kenmerken. De erfelijke aanleg wordt bepaald door het DNA, wat verdeeld is over een aantal chromosomen. Alle chromosomen tezamen worden aangeduid als het genoom. De beschikbaarheid van technieken om het DNA te analyseren en de mogelijkheden daarvan voor toepassing in fokprogramma's, hebben geleid tot vele studies om de plaatsen op het DNA te localiseren die verantwoordelijk zijn voor de erfelijke aanleg van belangrijke kenmerken. Deze plaatsen worden ook wel aangeduid als QTL (quantitative trait loci). Indien de locatie exact bekend is, wordt gesproken over genen. Dit proefschrift gaat over het opsporen van QTL en de mogelijkheden om die te gebruiken in de fokkerij, en richt zich uitsluitend op melkvee.

De studies bij melkvee waren in eerste instantie vooral gericht op melkproductie kenmerken. In **hoofdstuk 2** wordt de opzet van deze studies besproken en wordt tevens ingegaan op de kansen (power), die deze studies bieden om QTL te vinden. De meeste QTL detectie studies in melkvee maken gebruik van de typische structuur van melkvee fokprogramma's. Hierin is veelal een aantal stiervaders aanwezig met een redelijk tot groot aantal zonen. De verschillende varianten van makkelijk te bepalen delen van het DNA (merkers) van stiervaders en zonen worden hierbij bepaald in het laboratorium. De zonen hebben fokwaardes (schatting van de erfelijke aanleg) voor een groot aantal kenmerken, gebaseerd op een groot aantal (gemiddeld meer dan 100) dochters. De informatie van merkers en fokwaardes wordt vervolgens gekoppeld, om te bepalen welke merkers (welke delen van het DNA) van invloed zijn op de fokwaardes. Een dergelijke opzet, waarbij informatie van 3 generaties wordt meegenomen, wordt veelal aangeduid met granddaughter design (GDD).

Zelfs de grootste studies die uitgevoerd zijn, zullen slechts een beperkte fractie kunnen detecteren van de QTL die verantwoordelijk zijn voor melkproductie kenmerken. De kans van dergelijke studies op het vinden van QTL dient daarom vergroot te worden, ofwel door grotere studies op te zetten, ofwel door het gebruik van verfijnde methoden, bijv. door de beschikbare gegevens efficiënter te analyseren middels het meenemen van alle relaties tussen dieren in de afstamming. Een overzicht van de resultaten van QTL-studies gericht op melkproductie kenmerken laat zien, dat er vooral voor eiwit percentage QTL zijn gevonden. De QTL op chromosomen 3, 6, 14 en 20 zijn met een vrij grote zekerheid aangetoond. Op chromosoom 14 is de exacte locatie bekend van een gen met een grote invloed op vet percentage. De QTL op chromosomen 3, 6, en 20 beïnvloeden het eiwit %. Op chromosoom 6 liggen waarschijnlijk meerdere QTL. Een aantal van de chromosomale gebieden die van invloed zijn op melkproductie kenmerken, lijken tevens effect te hebben op andere kenmerken. Indien een QTL effect heeft op meerdere kenmerken, wordt

gesproken over pleiotropie. Door meerdere kenmerken in de analyse op te nemen, kan vastgesteld worden of hier sprake is van pleiotrope effecten.

Hoewel het fokdoel voor een (groot) deel gericht is op melkproductie kenmerken, neemt het belang van kenmerken als vruchtbaarheid en gezondheid steeds meer toe. Vergeleken met QTL voor melkproductie kenmerken, zijn QTL voor vruchtbaarheid en gezondheid mogelijk relatief belangrijker. Dit heeft bijv. te maken met de lage erfelijkheidsgraad voor deze kenmerken, waardoor een schatting van de erfelijke aanleg van dieren op basis van waarnemingen voor deze kenmerken relatief onnauwkeurig is. Ook kan het lastig zijn om dergelijke kenmerken te meten. Dit geldt bijvoorbeeld voor de gezondheidskenmerken. Efficiënte selectie op deze kenmerken is dan ook sterk gebaat bij informatie omtrent QTL voor deze kenmerken. In **hoofdstuk 3** worden de resultaten beschreven van een studie gericht op het vinden van QTL voor exterieurkenmerken en functionele kenmerken. In deze studie, die gebaseerd was op 20 Holstein Friesian stiervaders en 833 zonen van deze vaders, is het genotype (= verschijningsvorm) van 277 zgn. microsatelliet merkers bepaald. Deze merkers lagen verspreid over alle chromosomen, m.u.v. de geslachtschromosomen. Tevens was de fokwaarde van de zonen beschikbaar voor 18 exterieur kenmerken, 2 vruchtbaarheidskenmerken, 4 geboortekenmerken, 2 gebruikskenmerken en één uiergezondheidskenmerk. De informatie over de vererving van merkers werd vervolgens vergeleken met deze fokwaardes. Hierbij werd informatie van meerdere merkers en van alle families gecombineerd.

Uit de analyse over families heen kwamen 61 mogelijke QTL naar voren, waarbij er 27 vals-positieven verwacht werden (i.e., één vals-positief resultaat per kenmerk). De sterkste aanwijzingen voor QTL werden gevonden voor: borstomvang op chromosoom 2, draagtijd op chromosoom 4, kruishoogte, inhoud en ontwikkeling op chromosoom 5, type op chromosoom 6, bespiering op chromosoom 12, vooruieraanhechting op chromosoom 13, en vooruieraanhechting en voorspeenplaatsing op chromosoom 19. De QTL voor lichaamsmaten op de chromosomen 2, 5, en 6 hebben mogelijk ook een effect op geboortegemak. De QTL voor uierkenmerken op de chromosomen 13 en 19 hebben mogelijk ook invloed op celgetal en weerstand tegen uierontsteking.

De QTL die zijn gevonden voor melk productie kenmerken, exterieur kenmerken, vruchtbaarheid en (uier)gezondheid, kunnen gebruikt worden bij merker-ondersteunde selectie (ook wel aangeduid met MAS, marker-assisted selection). Informatie over de aanwezigheid en vererving van merkers die gekoppeld zijn aan QTL, wordt hierbij gebruikt om een betere schatting te krijgen van de erfelijke aanleg. Selectie op een bepaald QTL kan echter ook een effect hebben op andere kenmerken, door bijv. pleiotropie (i.e., het QTL beïnvloedt meerdere kenmerken), of doordat het

QTL op het chromosoom nauw gekoppeld is aan QTL met een effect op andere kenmerken. Aangezien deze QTL veelal gezamenlijk overerven, leidt selectie op het ene QTL tevens tot een effect op een ander kenmerk. De negatieve effecten op andere kenmerken van selectie op een bepaald QTL moeten zoveel mogelijk voorkomen worden, en positieve effecten op andere kenmerken moeten zoveel mogelijk benut worden. Hiervoor is het noodzakelijk inzicht te hebben in de pleiotrope effecten van QTL, dan wel het effect van een chromosomaal gebied op meerdere kenmerken. **Hoofdstuk 4** beschrijft een methode om pleiotrope QTL of gebieden met een effect op meerdere kenmerken te bepalen. De methode is gebaseerd op de resultaten van QTL-studies gericht op individuele kenmerken, en is afgeleid voor een granddaughter design. De QTL-studies voor individuele kenmerken resulteren in een schatting van de QTL-effecten voor iedere stiervader en elk kenmerk. Voor elke combinatie van kenmerken wordt vervolgens de covariantie tussen deze schattingen berekend. Indien er geen sprake is van pleiotropie, dan wordt de covariantie tussen de schattingen bepaald door de erfelijkheidsgraden van de betrokken kenmerken, de polygene correlatie (wordt bepaald door alle genen gezamenlijk) en de milieu correlatie tussen de kenmerken, de variatie in de kenmerken, het aantal zonen per stiervader, en het aantal dochters per stier. Indien er wel sprake is van pleiotropie, dan zal de covariantie tussen de schattingen van de QTL-effecten afwijken van de verwachte covariantie. Deze afwijking hangt af van de grootte van de effecten van het pleiotrope QTL, en van de fractie stiervaders die heterozygoot zijn voor dit QTL (= twee verschillende varianten hebben van dat QTL). Bij analyse van werkelijk data, kunnen de verwachte covariantie indien er geen QTL is, en een betrouwbaarheidsinterval voor deze verwachte covariantie, bepaald worden. Wanneer de covariantie in de werkelijke data buiten het betrouwbaarheidsinterval ligt, is er sprake van een pleiotroop QTL of enkele nauw gekoppelde QTL met een effect op verschillende kenmerken. De methode is getoetst met een simulatie studie en geïllustreerd aan de hand van data van chromosoom 6 bij rundvee.

In **hoofdstuk 5** is de ontwikkelde methode toegepast op een dataset waarin informatie van alle chromosomen was opgenomen. Het betrof hier een granddaughter design met 20 stiervaders en 833 zonen. Vier kenmerk-categorieën werden geanalyseerd: melk productie (vijf kenmerken), uierkenmerken (exterieur, zes), uiergezondheid (twee) en vruchtbaarheid (twee). Er werden in totaal 59 chromosomale gebieden geïdentificeerd met een effect op meerdere kenmerken (met een kans van 0.01 op een vals-positief resultaat, niet gecorrigeerd voor het analyseren van 29 chromosomen). Deze gebieden worden vanaf nu aangeduid met MQR (multiple trait quantitative trait region = gebied met een effect op meerdere kenmerken). Vijftien MQR werden gevonden op chromosoom 14. Ook op de chromosomen 6, 13, 19, 22, 23, en 25 werden vier of meer MQR gevonden. Acht

MQR hadden zowel invloed op uier conformatie als uiergezondheid, en vier MQR beïnvloedden zowel melkproductie kenmerken als uier gezondheid. Het aantal MQR met een effect op zowel vruchtbaarheid als uier conformatie was vijf. Ook werden vijf MQR gevonden met een effect op zowel vruchtbaarheid als melk productie kenmerken. Voor 22 MQR week de correlatie veroorzaakt door de MQR sterk (meer dan 0.60) af van de genetische correlatie. De resultaten van dit onderzoek kunnen gebruikt worden bij het besluit om al dan niet MAS gericht op bepaalde QTL toe te passen.

Het meeste voordeel van MAS wordt verwacht voor kenmerken met een lage tot redelijke erfelijkheidsgraad (bijv. een kenmerk als vruchtbaarheid), kenmerken die moeilijk of alleen tegen hoge kosten te meten zijn (bijv. ziektes), kenmerken die niet beschikbaar zijn voor alle kandidaten voor selectie (bijv. karkaskenmerken), of kenmerken die uitsluitend meetbaar zijn bij één van beide geslachten (bijv. melkproductie, uierontsteking). Wanneer MAS bij stieren wordt toegepast, dan zal dit in de meeste gevallen worden toegepast als een eerste selectie stap bij meer-fase selectie, voordat stieren worden ingezet als proefstier. Merker-informatie verhoogt de nauwkeurigheid van selectie, vooral in de eerste stap, als nog weinig andere informatie beschikbaar is. In **hoofdstuk 6** zijn met een simulatie programma een aantal fokprogramma's bestudeerd die QTL-informatie gebruiken. In het basis schema werd in de vrouwelijke dieren geselecteerd op een combinatie van afstammings informatie en metingen aan het dier zelf. In de selectie index voor de mannelijke dieren zaten afstammings informatie en metingen aan 100 dochters. In de eerste fase bestond de index in beide gevallen uitsluitend uit afstammings informatie. In de alternatieve fokprogramma's werd de QTL-informatie meegenomen als een extra kenmerk in de index. De correlatie van dit extra kenmerk met het fokdoel kenmerk werd bepaald door het aandeel van het QTL in de erfelijke variatie van kenmerken. QTL-informatie was beschikbaar bij geboorte van een dier, of op embryo-niveau. De erfelijke vooruitgang in de diverse fokprogramma's, ten opzichte van de erfelijke vooruitgang in een fokprogramma zonder QTL-informatie, varieerde van +4.5% (wanneer het QTL een aandeel van 5% had in de totale erfelijke variatie) tot +21.2% (bij een QTL met een aandeel van 50% in de erfelijke variatie). Een QTL met een aandeel van 5% biedt de mogelijkheid om minder proefstieren te testen (130 in plaats van 200, in het onderzochte programma), en daarbij dezelfde erfelijke vooruitgang te behalen. Indien het aantal geproduceerde embryo's in het fokprogramma vergroot werd, en selectie op basis van QTL-informatie op embryo-niveau werd toegepast, kon de erfelijke vooruitgang met 31.3% verhoogd worden. Dit biedt mogelijkheden om de opzet van het fokprogramma drastisch te wijzigen. Een afweging van de kosten is hiervoor echter ook van groot belang.

In **hoofdstuk 7** wordt een aantal toepassingen van MAS in hedendaagse fokprogramma's voor melkvee gegeven. Bij toepassing van MAS komen een aantal zaken naar voren die te maken hebben met bijv. het aandeel van de gevonden QTL in de totale erfelijke variatie, en de nauwkeurigheid waarmee de plaats van het QTL bekend is. Deze bemoeilijken een efficiënte toepassing, en illustreren de noodzaak om meer QTL te vinden, en de plaats van deze QTL nauwkeuriger te bepalen. De datasets die nu beschikbaar zijn, bieden uitstekende mogelijkheden om QTL op te sporen, ook voor kenmerken met een relatief lage erfelijkheidsgraad, en zeker indien alle mogelijkheden qua analyse methoden worden benut. Hiermee kan de kans dat aanwezige QTL daadwerkelijk worden gevonden, worden vergroot. De gevonden QTL verklaren dan een groter deel van de erfelijke variatie, en MAS kan efficiënter worden toegepast. Bovendien kan het fokprogramma anders worden opgezet, bijv. door minder proefstieren te testen, en kan het fokprogramma tegen lagere kosten worden uitgevoerd.

Wanneer een QTL met grote waarschijnlijkheid in een beperkt gebied ligt, dan kan geselecteerd worden op informatie van één enkele merker of een combinatie van merkers in dat gebied, zonder eerst het belang van die merkers in verschillende families vast te hoeven stellen. De waarde van dergelijke merkers, ten opzichte van de waarde van het QTL zelf, neemt echter snel af met toenemende afstand tussen de merker en het QTL. Deze merkers zullen dus zeer dicht bij het QTL moeten liggen om ze op deze manier te kunnen gebruiken. In het ideale geval is het gen zelf bekend, maar de kosten om het gen te vinden wegen mogelijk niet op tegen de extra voordelen, wanneer deze kosten en voordelen vergeleken worden met het gebruik van nauw gekoppelde merkers.

De beschikbaarheid van grote aantallen merkers in gebieden waar mogelijk genen liggen, vergemakkelijkt het vinden van de meest geschikte merkers voor selectie. Door nieuwe initiatieven, zoals het in kaart brengen van de samenstelling van het genoom bij het rund, zal het aantal beschikbare merkers enorm toenemen, en daarmee het in kaart brengen van de genen veranderen en vergemakkelijken. Deze ontwikkelingen zullen ervoor zorgen dat selectie in rundveefokprogramma's aanzienlijk gaat veranderen in de komende 5-10 jaar.

Nawoord

Na afloop van 5 jaar werken en een half jaar zwoegen, wil ik graag de mensen bedanken die mij daarin hebben gesteund.

Johan van Arendonk en Henk Bovenhuis waren het meest betrokken bij de begeleiding van het onderzoek. Johan, jouw positieve benadering in voorspoed en bij tegenslag in het onderzoek heeft mij enorm gemotiveerd. Henk, jouw kennis en ideeën over het onderwerp hebben in belangrijke mate bijgedragen aan de invulling van mijn onderzoek. Beiden, hartelijk dank daarvoor.

Naast Johan en Henk als directe begeleiders hebben Sijne van der Beek en Marco Bink, als vaste leden van de begeleidingscommissie, hun bijdrage geleverd. Piter Bijma was met name betrokken bij het in hoofdstuk 6 beschreven onderzoek. Sijne, Marco en Piter, bedankt voor jullie kritische blik op mijn onderzoek.

Het onderzoek is uitgevoerd bij de leerstoelgroep Fokkerij en Genetica. De sfeer was altijd prima, en ik vind het jammer dat ik niet altijd aanwezig kon zijn bij de inhoudelijke en sociale activiteiten van de leerstoelgroep. Iedereen bedankt. Liesbeth, gedurende bijna de volledige periode in Wageningen waren we kamergenoten. De gesprekken over werk en andere zaken vormden een welkome afwisseling met reken- en schrijfwerk. En we hebben als kamergenoten toch maar mooi de Intermediair gehaald.

Ongeveer zes jaar geleden kreeg ik van Holland Genetics de kans om dit onderzoek uit te voeren, als onderdeel van de werkzaamheden bij de afdelingen foktechniek en R&D. Ate Lindeboom en Nanke den Daas, jullie hebben mij toen die kans geboden, en ook van Janneke van Wagtendonk, Lucia Kaal en Wiepk Voskamp heb ik daarna de ruimte en tijd gekregen om dit onderzoek uit te voeren. De MKZ-crisis en daaropvolgende reorganisatie hebben de beschikbare tijd voor “Wageningen” wel wat teruggebracht, maar ook daarna kon ik doorgaan met het onderzoek. Erik, Annelies, Karen, Adrie, Erwin en Joke, als collega’s bij de afdeling R&D hebben jullie rekening gehouden met mijn Wageningen-dagen en mij qua werkzaamheden van tijd tot tijd ontzien, vooral in het laatste half jaar. Hierbij alle waardering en dank voor de steun vanuit Arnhem.

Pa en ma, ik ben toch nog een beetje “in de koeien” gebleven, ook al is het niet helemaal zoals ik zelf (en misschien jullie ook) ruim 20 jaar geleden in gedachten had. Maar het is wel duidelijk waar mijn belangstelling voor het vakgebied vandaan komt. Bedankt.

Anita, toen we bijna 6 jaar geleden verhuisden naar Elst en de zolder gingen inrichten, sprak ik de historische woorden: “en hier ga ik dus promoveren”. Ik had toen namelijk juist besloten de kans te grijpen die mij werd geboden. Je hebt me de afgelopen jaren wel eens naar de zolder moeten sturen, omdat je bang was dat het anders wel een erg langdurige geschiedenis zou worden, dat onderzoek van mij. Ik dank je voor deze steun en voor al het andere, het is af!

Overigens is de zolder de zolder niet meer, door de komst van onze beide dochters is het huis inmiddels behoorlijk anders ingericht. Stefanie en Marit, bedankt voor jullie vrolijkheid en afleiding en voor het ontzien van jullie promoverende vader. Huilen in de nacht is gelukkig niet jullie grootste hobby. En papa heeft vanaf nu weer wat meer tijd….

Chris

Curriculum Vitae

Christiaan Schrooten werd op 12 oktober 1963 geboren te Zwollerkerspel, en groeide op op het ouderlijk melkveebedrijf aldaar. In 1982 voltooide hij de VWO-opleiding aan het Meander College te Zwolle, waarna hij begon met de studie Zoötechniek aan de Landbouwhogeschool te Wageningen. Deze studie, met als afstudeervakken veefokkerij en gezondheids- en ziekteleer, rondde hij in 1987 af. Sindsdien heeft hij gewerkt bij diverse coöperatieve rundveeverbeteringsorganisaties in Nederland, in functies op het gebied van automatisering, foktechniek en onderzoek. In 1989 – 1991 verrichtte hij gedurende ruim anderhalf jaar onderzoek naar nucleus-fokprogramma's voor melkvee, bij de vakgroep Veefokkerij van de Landbouwuniversiteit te Wageningen. Sinds 1993 is hij werkzaam bij Holland Genetics te Arnhem, momenteel als senior onderzoeker bij de afdeling R&D. Vanaf september 1998 was hij in deeltijd gedetacheerd bij de leerstoelgroep Fokkerij en Genetica te Wageningen, voor het in dit proefschrift beschreven onderzoek.

The research described in this thesis was financially supported by Holland Genetics BV, Arnhem, and through a grant from Senter (division of the Dutch Ministry of Economic Affairs)

Cover design: Frans Boers – CR Delta VRV Holding BV, Arnhem, The Netherlands

Printed by Grafisch Bedrijf Ponsen & Looijen BV, Wageningen, The Netherlands