

Aus dem Forschungsbereich Genetik und Biometrie der Forschungsinstitut für die Biologie
landwirtschaftlicher Nutztiere

FINE MAPPING OF QUANTITATIVE TRAIT LOCI AFFECTING SOMATIC CELL SCORE ON
BTA02, BTA18 AND BTA27 IN THE GERMAN HOLSTEIN

DISSERTATION

zur Erlangung des Doktorgrades
der Agrar- und Ernährungswissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

CHRISTINE FRANCOISE BAES

M.Sc. (Agr)
aus Stratford,
Ontario Kanada

Kiel, 2008

Dekan: Prof. Dr. Uwe Latacz-Lohmann

1. Berichterstatter: Prof. Dr. Norbert Reinsch

2. Berichterstatter: Prof. Dr. Georg Thaller

Tag der mündlichen Prüfung: 20. November, 2008

I have never let my schooling interfere with my education.
Mark Twain

TABLE OF CONTENTS

	GENERAL INTRODUCTION	PAGE 1
CHAPTER ONE	Computing the condensed conditional gametic QTL relationship matrix and its inverse	PAGE 4
CHAPTER TWO	TIGER: A software system for fine-mapping quantitative trait loci	PAGE 23
CHAPTER THREE	Linkage Disequilibrium at a QTL for Somatic Cell Score on BTA27 in the German Holstein	PAGE 38
CHAPTER FOUR	Refined Mapping of Quantitative Trait Loci for Somatic Cell Score on BTA02 in the German Holstein	PAGE 56
CHAPTER FIVE	Fine mapping of a QTL on BTA18 in the German Holstein	PAGE 75
	GENERAL DISCUSSION	PAGE 95
	GENERAL SUMMARY (ENGLISH)	PAGE 107
	GENERAL SUMMARY (GERMAN)	PAGE 109
	ACKNOWLEDGEMENTS	PAGE 111
	EIDESSTATTLICHE ERKLÄRUNG	PAGE 113
	CURRICULUM VITAE	PAGE 114

GENERAL INTRODUCTION

The vast majority of economically important traits in plants and animals are quantitative in nature and can be influenced by many individual chromosomal regions (quantitative trait loci, QTL). The localisation of QTL which contribute significantly to phenotypic variation is an important goal of both the scientific and the agricultural community, although mapping these QTL is only the first step in the search for the underlying genes. Several QTL for economically important traits in domestic species have been roughly positioned using linkage analysis (LA) (See KHATKAR ET AL., 2004 for a review). However, the confidence intervals achieved in such studies contain far too many genes to allow the identification of causal mutations. The focus has now shifted to fine mapping these QTL as the next step in identifying genes underlying complex traits. A whole genome scan for QTL affecting functional traits including udder health was conducted in 2000 in cooperation with German AI and breeding organizations, scientific institutes for animal breeding and animal computing centers (THOMSEN ET AL., 2000). The project, instigated by the German cattle breeders' federation (ADR), enabled the identification of several QTL affecting functional traits in the German Holstein population, including those on *Bos taurus* autosomes 2, 18 and 27 affecting somatic cell score in milk (KÜHN ET AL., 2003, BENNEWITZ ET AL., 2003). Although extensive information was gained from these experiments, no candidate genes could be identified due to the relatively large confidence intervals typical of such QTL mapping experiments. In 2005, a successor project (FUGATO M.A.S.-Net) was initiated with the goal of further narrowing the localisation of positional and functional candidate genes associated with SCS in the regions previously identified.

This thesis was written in the context of the FUGATO M.A.S.-Net project, and its objectives were two-fold. Firstly, software had to be developed to properly analyse the data produced in the project. This is discussed within the first two chapters, which offer a description and discussion of software for fine mapping experiments using variance component methodology. Secondly, the analyses had to be carried out for marker data collected from the M.A.S.-Net partner laboratories in Giessen, Kiel and Dummerstorf. Phenotypes were contributed by the animal computing center (VIT) in Verden. This result-oriented component of the dissertation presents findings from analyses conducted on BTA02 (chapter 3), BTA18 (chapter 4) and BTA27 (chapter 5).

The joint utilization of marker and phenotype information in current genetic prediction models has evolved rapidly since FERNANDO AND GROSSMAN (1989) first incorporated marked

quantitative trait loci (MQTL) into the existing breeding value estimation model. In splitting the ‘genetic’ portion of the model into the additive effects of the unique QTL gametes and the remaining polygenic effects, calculation of the conditional covariance matrix of QTL allele effects \mathbf{G} and its inverse \mathbf{G}^{-1} is necessary. Calculation of the “condensed” gametic relationship matrix \mathbf{G}^* - a version of \mathbf{G} where linear dependencies have been removed - and its inverse \mathbf{G}^{*-1} is described and discussed in **Chapter 1**.

Chapter 2 presents and describes the software system TIGER, which is a UNIX script linking several individual Fortran programmes. TIGER was specifically developed for comprehensive variance component analysis of fine mapping data. Starting with raw genotype data, pedigree and marker map information and ending with a residual maximum likelihood-based test for each putative quantitative trait locus position, the software provides the user with an ‘all in one’ package capable of linkage analysis, linkage disequilibrium analysis and combined linkage / linkage disequilibrium analysis. The developed programs also proved useful for association analyses where polygenic effects are included in the model and linkage disequilibrium is depicted by a fixed regression on marker alleles.

The chromosomal area affecting SCS on BTA02 was discovered in a joint linkage analysis of French and German Holstein families (BENNEWITZ ET AL., 2003) and showed a 95% confidence interval as large as approx. 100 cM. Further examination of this large region is presented in **Chapter 3** and is based on marker-data from six half-sib families provided from the collaborating colleagues in Gießen. Linkage disequilibrium methods using both variance components and regression on marker alleles were employed, resulting in the identification of multiple QTL regions with chromosome-wide significance, whose joint effects were presumably responsible for the originally detected large QTL-region.

Chapter 4 deals with further investigation of the chromosomal region on BTA18 previously identified by KÜHN ET AL. (2003) and XU ET AL. (2006). Marker data and the underlying map were compiled by the project partners in Dummerstorf. Both single and 2-QTL models were used to investigate the linkage disequilibrium and to narrow down the confidence interval. One significant QTL region was pinpointed using single-QTL analysis. The use of a 2-QTL model confirmed the position of the QTL without much evidence for the existence or position of a second QTL. Ultimately, the QTL position was narrowed down considerably compared to previous results; further analysis can concentrate on an interval of approximately 10.5 cM.

In **Chapter 5**, the chromosomal area affecting SCS on BTA27 (Kühn et al., 2003) around the *defensin*-cluster (ROOSEN ET AL., 2004) was investigated using a much denser marker map and employing linkage disequilibrium methods with both single-QTL and 2-QTL models. Marker data and the underlying map were compiled by the project partners in Kiel. Compared to the point of departure, the region of interest for future fine mapping experiments could be decisively narrowed and further marker intervals of interest were also found. Specific haplotypes associated with both increases and decreases in SCS could be identified and may be useful in further work as well as a denoted number of presumably heterozygous and homozygous animals.

LITERATURE

BENNEWITZ, J.; REINSCH, N.; GROHS, C.; LEVÉZIEL, H.; MALAFOSSE, A.; THOMSEN, H.; XU, N.; LOOFT, C.; KÜHN, C.; BROCKMANN, G.; SCHWERIN, M.; WEIMANN, C.; HIENDLEDER, S.; ERHARDT, G.; MEDJUGORAC, I.; RUSS, I.; FÖRSTER, M.; BREINIG, B.; REINHARDT, F.; REENTS, R.; AVERDUNK, G.; BLÜMEL, J.; BOICHARD, D. AND KALM, E.: 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 (2003) 319-338.

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

KHATKAR, M.S.; THOMSON, P.C.; TAMMEN, I. AND RAADSMA, H.W.: Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet. Sel. Evol.* 36 (2004) 63–190.

KÜHN, C.; BENNEWITZ, J.; REINSCH, N.; XU, N.; THOMSEN, H.; LOOFT, C.; BROCKMANN, G. A.; SCHWERIN, M.; WEIMANN, C.; HIENDLEDER, S.; ERHARDT, G.; MEDJUGORAC, I.; FÖRSTER, M.; BREINIG, B.; REINHARDT, F.; REENTS, R.; RUSS, I.; AVERDUNK, G.; BLÜMEL, J. AND KALM, E.: Quantitative trait loci mapping functional traits in the German Holstein Cattle Population. *J. Dairy Sci.* 86 (2003) 360-368.

ROOSEN, S.; EXNER, K.; PAUL, S.; SCHRÖDER, J. M.; KALM, E. AND LOOFT, C.: Bovine beta-defensins: identification and characterization of novel bovine beta-defensin genes and their expression in mammary gland tissue. *Mamm. Genome.* 15 (2004) 834-42.

THOMSEN, H.; REINSCH, N.; XU, N.; LOOFT, C.; GRUPE, S.; KÜHN, CH.; BROCKMANN, G.; SCHWERIN, M.; LEYHE-HORN, B.; HIENDLEDER, S.; ERHARDT, G.; MEDJUGORAC, I.; RUSS, I.; FÖRSTER, M.; BREINIG, B.; REINHARDT, F.; REENTS, R.; BLÜMEL, J.; AVERDUNK, G. AND KALM, E.: A male bovine linkage map for the ADR granddaughter design. *J. of Anim. Breed. Genet.* 117 (2000) 289-306.

XU, N.; PAUL, S.; BENNEWITZ, J.; REINSCH, N.; THALLER, G.; REINHARDT, F.; KÜHN, C.; SCHWERIN, M.; ERHARDT, G.; WEIMANN, C.; THOMSEN, H.; MISHRA, S. AND KALM, E.: Confirmation of quantitative trait loci for somatic cell score on bovine chromosome 18 in the German Holstein. *Arch. Tierz.* 49 (2006) 111-119.

CHAPTER ONE:
COMPUTING THE CONDENSED CONDITIONAL GAMETIC QTL RELATIONSHIP
MATRIX AND ITS INVERSE

Christine Baes¹ and Norbert Reinsch¹

¹*Forschungsinstitut für die Biologie Landwirtschaftlicher Nutztiere, Forschungsbereich
Genetik und Biometrie*

Published in Archives of Animal Breeding (2007) 50:294-308

ABSTRACT

The inverse of the conditional gametic relationship matrix (\mathbf{G}^{-1}) for a marked quantitative trait locus (MQTL) is required for estimation of gametic effects in best linear unbiased prediction (BLUP) of breeding values if marker data are available. Calculation of the “condensed” gametic relationship matrix \mathbf{G}^* - a version of \mathbf{G} where linear dependencies have been removed - and its inverse \mathbf{G}^{*-1} is described using a series of simplified equations following a known algorithm. The software program COBRA (**c**ovariance **b**etween **r**elatives for a marked QTL) is introduced, and techniques for storing and computing the condensed gametic relationship matrix \mathbf{G}^* and the non-zero elements of its inverse are discussed. The program operates with both simple pedigrees and those augmented by transmission probabilities derived from marker data. Using sparse matrix storage techniques, \mathbf{G}^* and its inverse can be efficiently stored in computer memory. COBRA is written in FORTRAN 90/95 and runs on a variety of computers. Pedigree data and information for a single MQTL in the German Holstein population are used to test the efficiency of the program.

KEYWORDS

marker assisted selection / best linear unbiased prediction / gametic relationship matrix

1.1 INTRODUCTION

The joint utilization of marker and phenotype information in current genetic prediction models is evolving rapidly. FERNANDO AND GROSSMAN (1989) incorporated marked quantitative trait loci (MQTL) information into the existing best linear unbiased prediction (BLUP) breeding value estimation model by splitting the ‘genetic’ portion of the model into the additive effects of the unique QTL gametes and the polygenic effects. In order to include MQTL information in the estimation model, the inverse of the conditional covariance matrix of QTL allele effects \mathbf{G} is required (\mathbf{G}^{-1}). The efficient computation of matrices like \mathbf{G} and \mathbf{G}^{-1} for large pedigrees is crucial for further successful incorporation of marker data in genetic evaluation models.

A numerically efficient algorithm for the calculation of \mathbf{G} and its inverse for an MQTL was developed by ABDEL-AZIM AND FREEMAN (2001) based on the work of FERNANDO AND GROSSMAN (1989), VAN ARENDONK ET AL. (1994) and WANG ET AL. (1995). TUCHSCHERER ET AL. (2004) showed that the calculation of \mathbf{G} depends on the mode of gamete identification (gametes identified by markers *vs.* gametes identified by parental origin), although the final MA BLUP breeding value of each animal is identical irrespective of the gamete identification method employed. They suggested that gamete identification by parental origin may have practical advantages compared to that by markers; for example, fewer values are required to denote marker related transmission probabilities. More importantly, the generalisation developed by TUCHSCHERER ET AL. (2004) showed that under certain circumstances identical rows and columns in \mathbf{G} may occur if parents pass identical copies of their gametes to their offspring (*i.e.* \mathbf{G} may be rank deficient); in such cases, the inverse matrix \mathbf{G}^{-1} is not defined. By excluding duplicate gamete information, the number of gametic effects in \mathbf{G} can be reduced to a smaller set of unique effects in a ‘condensed’ gametic relationship matrix \mathbf{G}^* ; \mathbf{G}^* is always of full rank and \mathbf{G}^{*-1} is defined. Additionally, the smaller \mathbf{G}^* matrix requires less memory and may be calculated faster than a larger one.

In the first section of this article, the calculation of \mathbf{G}^* and its inverse is described to illustrate the practical application of the generalised algorithm presented by TUCHSCHERER ET AL (2004). Secondly, the software program COBRA (**c**ovariance **b**etween **r**elatives at a marked QTL) is introduced, and techniques for determining, storing and computing the condensed gametic relationship matrix \mathbf{G}^* and the non-zero elements of its inverse are discussed. Finally, pedigree data and information for a single marker in the German Holstein population are used to test the efficiency of the program.

1.2 CALCULATING \mathbf{G}^* AND ITS INVERSE

The matrix \mathbf{G} was developed by SMITH (1984) and SMITH AND ALLAIRE (1985) to calculate the probability of any two gametes being identical by descent in an inbred population (cited by SCHAEFFER ET AL. 1989). It is symmetric and contains one row and one column for each gamete (i), which are calculated from the rows and columns belonging to the gametes' predecessor gametes (called the gametes' parents here forth). The diagonal elements of \mathbf{G} are equal to one ($\mathbf{G}_{(i,i)}^* = 1.0$ for gamete i), as the probability of a gamete being identical to itself is always equal to one. The probability that the parental gamete received from the sire of the animal, Pg_a and the parental gamete from the dam Mg_a are identical by descent is the inbreeding coefficient f_a of individual a , $\mathbf{G}_{(Pg_a, Mg_a)}^* = f_a$).

In a pedigree with n animals, the matrix \mathbf{G} has the dimension $2n \times 2n$ because every animal is assumed to have two unique gametes. However, if an exact copy of a parental gamete is passed to its offspring, the effects of that gamete are included twice in \mathbf{G} ; the computation of \mathbf{G}^{-1} fails due to the linear dependencies in \mathbf{G} (see section 4. in TUCHSCHERER ET AL. 2004). If copied alleles are excluded and only unique gametes are assigned rows and columns in \mathbf{G}^* , the linear dependencies caused by identical rows and columns no longer exist. This means that animals with two unique gametes contribute two rows and two columns to \mathbf{G}^* , animals with one unique gamete and one copy of a gamete contribute only one row and column, and for animals with two copied gametes no rows or columns are added. The determination of unique and copied gametes depends on the pedigree and transmission probability information and is outlined below.

1.3 METHODOLOGY

1.3.1 ASSUMPTIONS

Consider one or several marker loci closely linked to a quantitative trait locus (MQTL), with linkage equilibrium between the markers and the MQTL. A recombination rate of zero may occur between an MQTL and one (or more) marker(s). Markers may be single, flanking or multiple for the MQTL, and the number of markers and the distance between them is not limited. The first allele is assumed to be the paternal gamete (Pg) and the second is considered to be the maternal gamete (Mg). The paternal transmission probability ($T_{(Pg)}$) is the probability that the sire of an animal passed his paternal (first) gamete to his offspring. Likewise, the maternal transmission probability ($T_{(Mg)}$) is the probability that the dam of an

animal passed her paternal (first) gamete to her offspring. The probability of the maternal (second) gamete of the sire or dam being passed on to the offspring can be calculated by subtracting the paternal or maternal transmission probability (respectively) from one. Thus two values (paternal and maternal) actually represent all four possible probabilities.

1.3.2 METHOD

The calculation of the gametic relationship matrix using recursive algorithms requires an ordered pedigree in which parental animals occur before their progeny, and transmission probabilities for all non-founder individuals conditional on marker information for the paths sire \rightarrow progeny and dam \rightarrow progeny. The calculation of transmission probabilities is described by MAYER ET AL. (2007, submitted). For missing and non-informative markers the transmission probability is 0.5; the resulting contributions to the relationship matrix are identical to those in the classical numerator relationship matrix.

If an animal has a transmission probability of one for a certain gamete, that animal will receive the paternal gamete from its respective parent with 100% certainty; the gamete received by the animal is an exact copy of the parental gamete and is not unique. Conversely, if a transmission probability of zero occurs for a given gamete, the animal will receive the maternal gamete from its respective parent with 100% certainty.

It is possible to set up a gametic pedigree following the calculation of transmission probabilities; the number of unique gametes can be determined and the size of \mathbf{G}^* can be calculated. All unique paternal and maternal gametes are assigned an integer identification number in ascending order. Predecessor gametes of paternal gametes, $Pg_{(Pg)}$ and $Mg_{(Pg)}$, and predecessor gametes of maternal gametes, $Pg_{(Mg)}$ and $Mg_{(Mg)}$ must also be considered.

In contrast to the gametic identification numbering method employed in the calculation of \mathbf{G} , non-unique gametes do not receive unique identification numbers in \mathbf{G}^* . If a gamete is passed from a parent to its offspring with 100% certainty, it is included in the gametic pedigree with the same identification number as the gamete from which it originated.

1.3.3 CALCULATION OF THE MATRIX \mathbf{G}^*

The calculation of \mathbf{G}^* is described by equation (9) in TUCHSCHERER ET AL. (2004).

If the information in the pedigree is sorted with parents precede their progeny, it is possible to build \mathbf{G}^* recursively starting with the top left corner and working towards the right. The

transmission probability assigned to the gamete depends on whether it is the paternal or maternal gamete of the animal in question.

The matrix \mathbf{G}^* is symmetric, therefore only the upper triangular matrix needs to be calculated; the lower triangular matrix is identical to its upper counterpart. Let (i,j) represent the row and column indices for an element in \mathbf{G}^* and assume that $i>j$. The elements of each column are calculated from row one until the last row before the diagonal ($j-1$), using the information from the predecessor gametes and the transmission probability of the gamete being calculated.

For columns of paternal gametes with known predecessors (*i.e.* j is a paternal gamete),

$$\mathbf{G}_{(i,j)}^* = T_{(Pg)_j} \times \mathbf{G}_{(i,Pg_{(Pg)_j})}^* + (1 - T_{(Pg)_j}) \times \mathbf{G}_{(i,Mg_{(Pg)_j})}^* \quad (1)$$

where $T_{(Pg)_j}$ is the transmission probability for the paternal gamete j , i denotes the row to be calculated, $Pg_{(Pg)_j}$ is the paternal predecessor gamete of the paternal gamete j and $Mg_{(Pg)_j}$ is the maternal predecessor gamete of the paternal gamete j . The calculation of columns of maternal gametes with known parents (*i.e.* j is a maternal gamete) is accomplished similarly,

$$\mathbf{G}_{(i,j)}^* = T_{(Mg)_j} \times \mathbf{G}_{(i,Pg_{(Mg)_j})}^* + (1 - T_{(Mg)_j}) \times \mathbf{G}_{(i,Mg_{(Mg)_j})}^* \quad (2)$$

where $T_{(Mg)_j}$ is the transmission probability for the maternal gamete j , i is the row to be calculated, $Pg_{(Mg)_j}$ is the paternal predecessor gamete of the gamete j and $Mg_{(Mg)_j}$ is the maternal predecessor gamete of the gamete j . If the predecessors of i and j are unknown, $\mathbf{G}_{(i,j)}^* = 0$. In this way, \mathbf{G}^* can be calculated for all unique gametes and no linear dependencies occur.

1.3.4 THE INBREEDING COEFFICIENT

The inbreeding coefficient at the MQTL, f_a , is the probability that the gametes at the MQTL in individual a are identical by descent and can be found in the element (Pg,Mg) of \mathbf{G}^* for each individual. In \mathbf{G} , this value is always located directly above the diagonal of the maternal gamete (or below the diagonal of the paternal gamete) for every individual. Although \mathbf{G}^* also contains inbreeding coefficients for each individual, they are not necessarily located in the same position as in \mathbf{G} (*i.e.* directly above the maternal gamete of the individual) because the matrix is ordered differently due to the exclusion of non-unique gametes.

It is important to realize that the f_a of an animal is a function of its respective transmission probabilities and the relationships between its parental gametes. Although the f_a are the only elements of \mathbf{G}^* required to calculate its inverse, it is apparent that the f_a of an individual depends on certain pre-existing elements of \mathbf{G}^* . Technically, only very specific elements of \mathbf{G}^* are needed for the calculation of its inverse. TIER (1990) described how the minimum subset of matrix \mathbf{G} required for the calculation of \mathbf{G}^{-1} can be determined. However, the computation method presented here is sufficient for medium-sized pedigrees with a half sib structure.

1.3.5 CALCULATION OF THE MATRIX \mathbf{G}^{*-1}

The detailed calculation of \mathbf{G}^{*-1} is described by equation (10) in TUCHSCHERER ET AL. (2004). The matrix \mathbf{G}^{*-1} can be calculated using only the pedigree information (from which the gametic pedigree is derived), the inbreeding coefficient of each individual and the paternal and maternal transmission probabilities. The size of \mathbf{G}^{*-1} is the same as that of \mathbf{G}^* , and each column and row are once again assigned to one gamete. Only the elements of the upper triangular matrix are calculated; the lower triangular matrix is obtained by symmetry. Whereas the calculation of \mathbf{G}^* consisted of adding one column (and therefore one diagonal) for every unique gamete to an existing matrix, \mathbf{G}^{*-1} is calculated by adding one column and one diagonal for every gamete as well as adding values to specific existing elements in \mathbf{G}^{*-1} . It is possible to build \mathbf{G}^{*-1} for one individual at a time by passing through the pedigree from the first individual to the last, where one of four possible situations may occur:

- 1 - the individual has no unique gametes (no columns or rows are added):
 - no changes are made to the existing matrix

- 2 - only one gamete is unique (only one column is added, paternal or maternal):
 - one new diagonal exists,
 - up to two new values exist in the new column and
 - up to three values in the existing matrix are altered

- 3 - both gametes are unique (two columns added):
 - two new diagonals exist,

- up to four new values exist in the new columns and
- up to six values in the existing matrix are altered

Let Pg and/or Mg represent the new column(s) in \mathbf{G}^{*-1} , and therefore also the unique paternal or maternal gamete (respectively) of an individual. If situation 1 occurs, no calculations need to be made and no columns are added because the contributions of such individuals have already been accounted for.

If an individual has one unique paternal gamete, the following calculations are made using the paternal transmission probability $T_{(Pg)}$ of the individual and the inbreeding coefficient of the individuals' sire. The elements of \mathbf{G}^{*-1} where values are to be added are located in the columns of the paternal gametes' predecessor gametes $Pg_{(Pg)}$, $Mg_{(Pg)}$, and the elements where values are to be subtracted are located in the column of the gamete Pg with the exception of the diagonal, which is positive. For the new diagonal element:

$$\mathbf{G}_{(Pg, Pg)}^{*-1} = 1 / d_i \quad (3a)$$

for the new elements in the added column:

$$\mathbf{G}_{(Pg_{(Pg)}, Pg)}^{*-1} = -(T_{(Pg)} / d_i) \quad (3b)$$

$$\mathbf{G}_{(Mg_{(Pg)}, Pg)}^{*-1} = -(1 - T_{(Pg)}) / d_i \quad (3c)$$

and for the elements in the existing matrix:

$$\mathbf{G}_{(Pg_{(Pg)}, Pg_{(Pg)})}^{*-1} = (T_{(Pg)}^2 / d_i) \quad (3d)$$

$$\mathbf{G}_{(Mg_{(Pg)}, Mg_{(Pg)})}^{*-1} = (1 - T_{(Pg)})^2 / d_i \quad (3e)$$

$$\mathbf{G}_{(Pg_{(Pg)}, Mg_{(Pg)})}^{*-1} = (T_{(Pg)})(1 - T_{(Pg)}) / d_i \quad (3f)$$

where $d_i = 2T_{(Pg)}(1 - T_{(Pg)})(1 - f_{sire})$ if the sire is known, and $d_i = 1$ if the sire is unknown (see equation 10 in TUCHSCHERER ET AL. (2004) and appendix A for a proof of d_i).

If an individual has one unique maternal gamete, the same calculations are made, however, the maternal transmission probability $T_{(Mg)}$ of the individual and the inbreeding coefficient of the individuals' dam are used. The elements where values are to be added are located in the columns of the maternal gametes' predecessor gametes $Pg_{(Mg)}$, $Mg_{(Mg)}$, and the elements where values are to be subtracted are located in the column of the gamete Mg with the exception of the diagonal, which is positive. For diagonal elements:

$$\mathbf{G}_{(Mg, Mg)}^{*-1} = 1/d_i \quad (4a)$$

for the elements in the added columns:

$$\mathbf{G}_{(Pg_{(Mg)}, Mg)}^{*-1} = -(T_{(Mg)} / d_i) \quad (4b)$$

$$\mathbf{G}_{(Mg_{(Mg)}, Mg)}^{*-1} = -(1 - T_{(Mg)}) / d_i \quad (4c)$$

and for the elements in the existing matrix:

$$\mathbf{G}_{(Pg_{(Mg)}, Pg_{(Mg)})}^{*-1} = (T_{(Mg)}^2 / d_i) \quad (4d)$$

$$\mathbf{G}_{(Mg_{(Mg)}, Mg_{(Mg)})}^{*-1} = (1 - T_{(Mg)})^2 / d_i \quad (4e)$$

$$\mathbf{G}_{(Pg_{(Mg)}, Mg_{(Mg)})}^{*-1} = (T_{(Mg)})(1 - T_{(Mg)}) / d_i \quad (4f)$$

where $d_i = 2T_{(Mg)}(1 - T_{(Mg)})(1 - f_{dam})$ if the dam is known, and $d_i = 1$ if the dam is unknown.

If both Pg and Mg are unique, all 12 of the above calculations (*i.e.* Equations 3 a-f and 4 a-f) are used. For base animals, two columns are added; the parental gametes are unknown, and only one is added in the diagonal.

The matrix \mathbf{G}^{*-1} is calculated for one individual at a time and therefore ‘layer by layer’ using the equations given above. Once all individual layers have been calculated, they are added to form the final matrix \mathbf{G}^{*-1} .

1.4 COMPUTING \mathbf{G}^* AND ITS INVERSE

The computation of matrices like \mathbf{G} and \mathbf{G}^{-1} require very large amounts of computer memory. ABDEL-AZIM AND FREEMAN (2001) described computational techniques for calculating and storing the minimum possible elements of \mathbf{G} and its inverse using computational methods described by TIER (1990) and linked list storage techniques. Gametes were identified by markers. To the authors’ knowledge, computational techniques for efficiently calculating \mathbf{G}^* and \mathbf{G}^{*-1} have not yet been published.

The following section introduces and describes the software program COBRA, which is a FORTRAN (90/95) program designed to determine, store and compute the non-zero elements of \mathbf{G}^* and \mathbf{G}^{*-1} using the TUCHSCHERER ET AL. (2004) algorithm with gametes identified by parental origin. It is shown that \mathbf{G}^* and its inverse can be efficiently stored in computer memory or saved to file using sparse matrix storage techniques.

COBRA operates in three main steps. In the first step, pedigree information is read in, checked for errors and written to three index matrices. The gametic index matrix is calculated

and the user is informed of the number of animals and ‘unique’ gametes in the pedigree. In step two, \mathbf{G}^* is computed from the index information and the inbreeding coefficients calculated are saved. The user is informed of the number of non-zero elements in \mathbf{G}^* and the fill density of \mathbf{G}^* (in percent). Finally, the index information from step one and the inbreeding coefficients from step two are used to calculate the \mathbf{G}^{*-1} matrix. The user is informed of the number of non-zero elements in \mathbf{G}^{*-1} and the fill density of \mathbf{G}^{*-1} (in percent). Additionally, the number of individuals with no unique gametes, one unique paternal gamete, one unique maternal gamete and two unique gametes is given and the program is complete. COBRA includes the option of saving information on gamete occurrence, inbreeding coefficients with a reference list ordering gametes to animals and the non-zero elements of \mathbf{G}^{*-1} to file.

1.4.1 PREPARATION OF INDEX MATRICES

A text file containing the identification numbers of sire and dam, followed by the transmission probabilities of paternal and maternal gametes is the input pedigree. Each line in the input pedigree file signifies one individual and the pedigree must be ordered such that parents precede their progeny. The column j of all index matrices corresponds to the animal identification number (1 to n). The input pedigree file is first read to determine the number of animals (n) and to test for simple errors in the data (duplicate entries, wrong order, transmission probabilities > 1). The file is rewound, memory is allocated for the first two index matrices (index matrix N1 has dimension $2 \times n$, index matrix N2 has dimension $3 \times n$) and the pedigree is read in.

Index matrix N1 contains the parental identification numbers of animals 1 to n , with the identification number of the sire in the first row and that of the dam in the second. Unknown parents are assigned zeros. At this point, only the first two rows of index matrix N2 are read in; they contain the transmission probabilities of the paternal and maternal gametes of individuals 1 to n . If the sire or dam is unknown, the transmission probability for the paternal or maternal gamete is considered to be 0.5. Linkage equilibrium is assumed. The third row of N2 remains empty at this point, but is filled with the inbreeding coefficients of animals 1 to n once step two is complete.

The third index matrix N3 (dimension $6 \times n$) is now calculated from the information in N1 and N2. This matrix contains the paternal and maternal gamete identification numbers of individual n in rows 1 and 4, respectively. The identification numbers for the paternal pair of predecessor gametes are saved in rows 2 and 3, and those of the maternal pair of predecessor

gametes in rows 5 and 6. All paternal and maternal gametes are assigned an integer identification number in ascending order if the transmission probability of the gamete is not equal to 1.0 or 0.0, starting with the first animal in the pedigree. If the origin of a gamete is unknown, zeros are assigned as parental paternal and maternal gamete identification numbers.

1.4.2 WRITING \mathbf{G}^*

The \mathbf{G}^* matrix is calculated using the generalization of the tabular method of Wang et al. (1995) presented by TUCHSCHERER ET AL. (2004). This formulation adds the covariance information for each gamete column-wise to the upper triangle of \mathbf{G}^* , starting with column one. Columns are calculated from row one to the last row before the diagonal using the equations shown in section two, with the diagonal elements always equal to one.

Only the non-zero elements of the upper half of \mathbf{G}^* are saved in sparse IA, JA, A format (KNUTH, 1997), which reduces the memory requirements of the program. The position of non-zero elements in \mathbf{G}^* is initially unknown and is calculated from information in the index matrices N2 and N3 for each particular gamete using the methods described in section two. As \mathbf{G}^* is calculated, hashing is used to store and retrieve the required non-zero elements of \mathbf{G}^* . The organisation of the hash table is arranged on a “first come first serve” basis; when an element is to be inserted, the locations of its probe sequence are examined sequentially until an empty spot in the vector A is found. The new element is saved to that location and its coordinates are saved in the parallel IA JA vectors.

Diagonals of gametes without offspring are included, however their columns are not calculated. Should certain elements of such columns be required for other calculations, only those specific elements are computed. This ‘economised’ method of calculating \mathbf{G}^* saves a large amount of memory and increases the speed of the program considerably, especially if the input pedigree contains many individuals without offspring.

Once \mathbf{G}^* has been calculated, the inbreeding coefficient of each animal is retrieved from the IA, JA, A vectors and written into row three of the index matrix N2. The IA, JA and A vectors are cleared. As an option a list of animals, their respective paternal and maternal gametes and their inbreeding coefficients can be generated. The IA, JA, A vectors with the \mathbf{G}^* matrix coordinates and values may also be sorted and saved to file before clearing if required. This may be useful when the data are to be processed by other software.

1.4.3 WRITING \mathbf{G}^{*-1}

The inverse \mathbf{G}^{*-1} of the condensed gametic relationship matrix can now be calculated using the TUCHSCHERER ET AL. (2004) generalization of the WANG ET AL. (1995) algorithm. Using index matrices N1, N2 and N3, values are either added to or subtracted from the required elements of \mathbf{G}^{*-1} as described in section two. A simplified flowchart of the COBRA procedure for calculating \mathbf{G}^{*-1} is included in Appendix B. As in the calculation of \mathbf{G}^* , only the non-zero elements of the upper half of the \mathbf{G}^{*-1} matrix are calculated and saved in sparse IA, JA, A format. Hashing methods are used to store and retrieve the required elements. A list including animal identification number, row and column indices and values of non-zero elements is generated and sorted for the further calculation of BLUP breeding values.

1.5 PRACTICAL VALIDATION

Pedigree data from 12008 German Holstein bulls and bull dams (7174 males and 4834 females) was used to test the efficiency of the program. Pedigree and marker information originated from the second phase of the genome analysis project of the German Cattle Breeders Federation (Arbeitsgemeinschaft Deutscher Rinderzüchter, ADR) and is currently used for the MA BLUP evaluation of the trait somatic cell score. The pedigree was obtained from the United Information Systems Animal Production (Vereinigte Informationssysteme Tierhaltung w.V., VIT) in Verden, Germany. Information on a single highly polymorph marker with 15 alleles on chromosome 18 was included for 6520 typed animals (6050 males, 470 females). Allele frequencies ranged from 0.01% to 32.83%.

Transmission probabilities were calculated by extracting standard partial pedigrees (individual, sire, dam, sire's sire, sire's dam, dam's sire, dam's dam) from original pedigree data and applying SimWalk2 software as described by MAYER ET AL. (2007, submitted) and TUCHSCHERER ET AL. (2007, in preparation). The partial pedigrees were grouped by available genotype information. The individual and its male ancestors (sire and dam's sire) were genotyped in 33.6% of all partial pedigrees, followed closely by partial pedigrees in which the individual and its sire were genotyped (32.8%). Only 1.2% of the partial pedigrees contained full marker information for all individuals.

At the marker, 4584 of the typed males were not identical by descent, while the remaining 1466 typed males had an average inbreeding coefficient of 0.0332. Typed females included 373 animals which were not identical by descent at the marker, while the remaining 97

animals had an average inbreeding coefficient of 0.0342. A summary of results, along with maximum and minimum inbreeding coefficients, is shown in Tab. 1.

TABLE 1
Average, minimum and maximum inbreeding coefficients at the marker for typed males and females given observed marker genotype (1,466 male and 97 female typed animals and 12,008 animals including non-typed animals)

		Typed animals ¹		All animals (12008)
		Male (1466)	Female (97)	
Including Marker Information	Average IBC ²	0.0332	0.0342	0.00433
	Minimum	0.0000834	0.000635	0.000
	Maximum	0.455	0.377	0.455

¹ Only typed animals with an inbreeding coefficient > 0 are included.

² IBC = Inbreeding coefficient.

The pedigree was analysed twice; once including marker information and once with transmission probabilities set to 0.5 (*i.e.* not including marker information). Tab. 1 shows selected computational results from the COBRA program using the same pedigree with and without marker information.

It is apparent that the percentage fill in \mathbf{G}^* is not as high as that in \mathbf{G} (see ‘% Fill’, Tab. 2). Although \mathbf{G}^{*-1} seems to be slightly more full than \mathbf{G}^{-1} (0.0278% compared to 0.0232%), it should be noted that only 51,553 non-zero elements are included in \mathbf{G}^{*-1} compared to 66,820 non-zero elements in \mathbf{G}^{-1} , equating to a difference of more than 15,000 elements (see ‘Remaining non-zero elements’, Tab. 2). Furthermore, the condensed matrix has approximately 4,700 fewer columns and rows than the uncondensed counterpart; 19,238 unique gametic effects were included in \mathbf{G}^* and \mathbf{G}^{*-1} compared to 24,016 effects in \mathbf{G} and \mathbf{G}^{-1} (see ‘Diagonals’, Tab. 2).

TABLE 2
Fill density (in percent), storage requirements, number of non-zero elements and execution time requirements for building \mathbf{G}^* and \mathbf{G}^{*-1} using data from a 12008-animal pedigree with (first row) and without (second row) marker information

		% Fill	Elements saved for gametes without offspring	Remaining Non-zero Elements	Execution time ¹ (seconds)	Diagonals
Including marker information	\mathbf{G}^*	2.47	131,739,817	4,567,164	23.869	19,238
	\mathbf{G}^{*-1}	0.0278		51,553	0.0359	
Not including marker information	\mathbf{G}	3.29	212,134,432	9,499,265	46.900	24,016
	\mathbf{G}^{-1}	0.0232		66,820	0.0452	

¹A Dell Precision 630 Workstation with double processor (2x3.6 GHz Xeon, 2x72 GB SCSI hard drive, 8 GB RAM) running Suse-Linux 9.3 (64 bit Version) was used in the evaluation. The program was compiled with Intel-Fortran Version 8.1

The pedigree contained approximately 9,000 animals without offspring, which improved the efficiency of the program greatly since columns of \mathbf{G} and \mathbf{G}^* relating to these gametes were not calculated (see ‘Elements saved for gametes without offspring’, Tab. 2), probably because the non-unique gametes often belong to younger animals which do not yet have offspring.

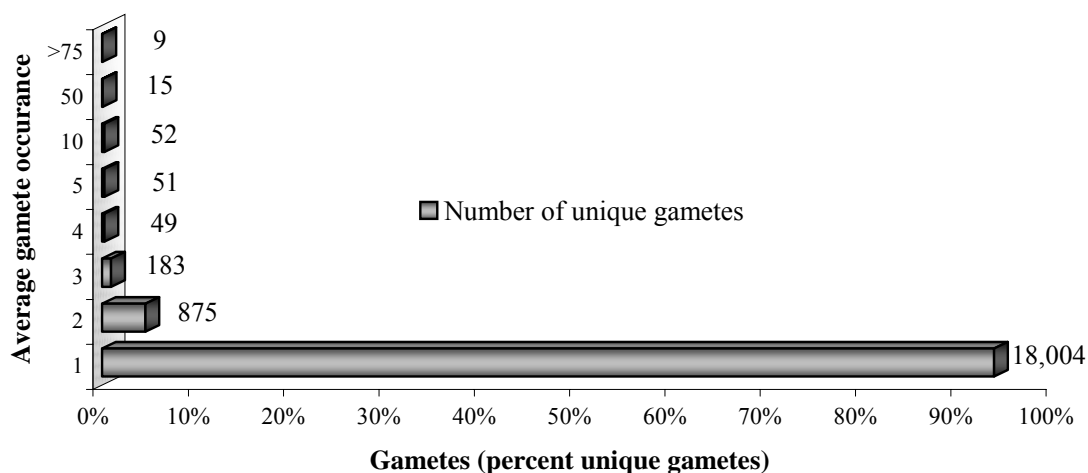


FIGURE 1
Average gamete occurrence (animals per gamete) in a 12,008 animal pedigree containing 19,238 unique gametes and 6,520 genotyped animals

Non-unique gametes have a potentially large effect on the population, especially if their copies occur frequently. Gametes in the current pedigree, which includes both typed and non-typed individuals, occurred an average of 1.25 times, with a maximum occurrence of 396. Fig. 1 shows that 18,004 gametes (93.59%) occur only once, while only 9 gametes (0.05%) have over 75 copies in bulls or bull dams.

1.6 DISCUSSION

This article describes how the conditional gametic relationship matrix and its inverse can be condensed using the TUCHSCHERER ET AL. (2004) algorithm. It describes how ordered pedigree information and transmission probabilities can be used to set up a gametic pedigree for the calculation of \mathbf{G}^* and \mathbf{G}^{*-1} . The computer program COBRA was introduced and its structure was explained. Finally, pedigree data from 12,008 German Holstein animals was analysed and used to test the efficiency of the program.

The concept of reducing the size of \mathbf{G} was proposed by MEUWISSEN AND GODDARD (1996) for multiple markers: parents and offspring sharing the same marker haplotype were both assigned the same gametic QTL effect by assuming that the probability of double recombination was equal to zero for informative animals (transmission probabilities were

rounded to 0, 0.5 or 1.0). In reality, marker information (and therefore the transmission probability for each marker) is variable and may have any value between zero and one; the values approach zero or one for more informative markers (LIU AND MATHUR 2005). The condensing algorithm presented by TUCHSCHERER ET AL. (2004) leads to the same result as the reducing algorithm when transmission probabilities of one and zero occur and no recombination takes place. However, the condensing algorithm uses the original transmission probabilities instead of rounded ones. If transmission probabilities very close to one or zero occur, a predefined threshold (e.g. $\epsilon=0.03$) can be entered in COBRA, causing all transmission probabilities below 0.03 to be treated as zero and all those above 0.97 to be treated as one. This further condenses \mathbf{G}^* and \mathbf{G}^{*-1} and provides similar results to those of MEUWISSEN AND GODDARD (1996).

Economising on the calculation of elements in rows and columns related to gametes without offspring can reduce the number of non-zero elements in \mathbf{G}^* significantly (*i.e.* minimal computation of non-parental gametic effects). The fill density and the number of non-zero elements with (2.47%, 4,567,164 elements) and without (21.31%, 39,430,374) economising on non-parental gametic effects (\mathbf{G}_a^*) for the pedigree described in section 4 underlines the significance of this method with regard to matrix size.

The distribution pattern of non-zero elements in \mathbf{G}_a^* is presented in Fig. 2, where the upper triangular matrix shows the distribution pattern of non-zero elements in \mathbf{G}_a^* . Each non-zero element in the matrix is represented by a single dot; the matrix may appear darker than in reality due to low print resolution.

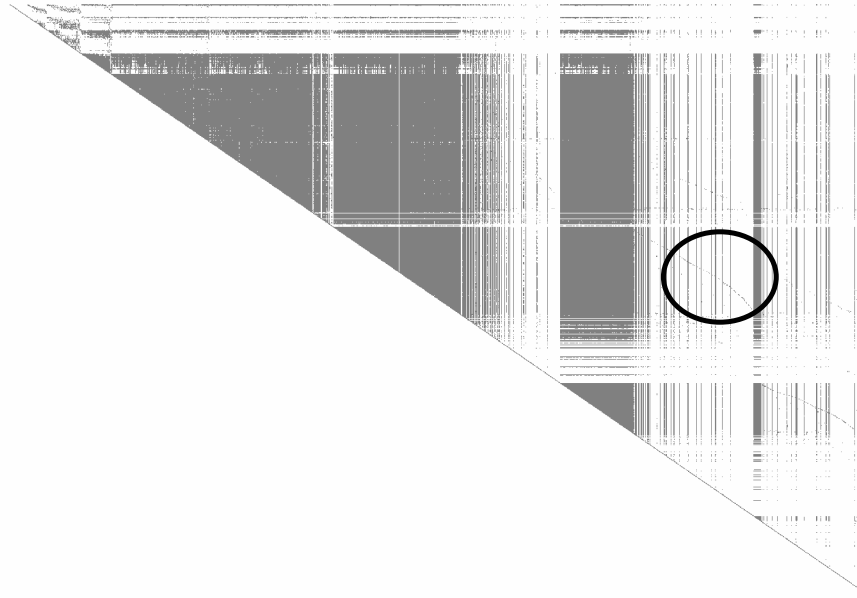


FIGURE 2

Distribution pattern of non-zero elements in \mathbf{G}_a^* (upper triangular matrix); non-zero elements are only calculated for animals with progeny. Empty columns contain inbreeding coefficients of animals without progeny, visible as a diagonal trace of non-zero elements in the circle.

The circle in Fig. 2 draws attention to a faint arrangement of ‘irregular’ dots in the otherwise regular checked-pattern visible in \mathbf{G}_a^* . These irregular elements are computed when certain constituents of economised columns are required for other calculations, such as those required for inbreeding coefficients of animals without unique gametes.

In order to successfully incorporate MQTL information in the BLUP breeding value estimation model, the marker genotype information of as many animals as possible is required. Fig. 1 shows that the vast majority of gametes occurred only once in the pedigree (93.59%), however transmission probabilities for marker genotypes were only included for 6,520 of 12,008 animals. The average gamete occurrence would likely increase if the missing genotype information, especially for bull dams, was included.

The efficient calculation of \mathbf{G} and \mathbf{G}^{-1} is imperative if marker data are to be practically used in genetic evaluation models. The computational techniques proposed by ABDEL-AZIM AND FREEMAN (2001) for constructing the inverse are efficient, however the condensing algorithm proposed by TUCHSCHERER ET AL. (2004) is not included. The relatively uncomplicated programming methods for the calculation of \mathbf{G}^* and \mathbf{G}^{*-1} described in this article prove adequate for medium sized pedigrees with simple structure. However, computation time required may be too high for the calculation of larger pedigrees (>100,000 animals) or

pedigrees with a more complex constitution. Computation time may therefore be improved in later versions of the COBRA program.

1.7 ACKNOWLEDGEMENTS

Financial support from the FUGATO MAS.Net project is gratefully acknowledged. Special thanks to Dr. Manfred Mayer for data preparation and helpful insight on transmission probabilities.

1.8 REFERENCES

ABDEL-AZIM, G. AND FREEMAN, A.: A rapid method for computing the inverse of the gametic covariance matrix between relatives for a marked Quantitative Trait Locus, *Genet. Sel. Evol.* **33** (2001) 153-173.

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

KNUTH, D.: *The Art of Computer Programming, Volume 3: Sorting and Searching.* Addison-Wesley (1997) 513–558.

LIU, Y. AND MATHUR, P.K.: Simplification of marker-assisted genetic evaluation and accounting for non additive interaction effects. *Arch. Tierz.* **48** (2005) 460-474.

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

SCHAEFFER, L.R.; KENNEDY, B.W. AND GIBSON, J.P.: The Inverse of the Gametic Relationship Matrix. *J. Dairy Sci.* **72** (1989) 1266-1272.

SMITH, S.P.: Dominance relationship matrix and inverse for an inbred population. Mimeo, Dep. Dairy Sci., Ohio State Univ. (1984).

SMITH, S.P. AND ALLAIRE, F.L.: Efficient Selection rules to increase non-linear merit: application in mate selection. *Genet. Sel. Evol.* **17** (1985) 387-395.

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

TUCHSCHERER, A.; MAYER, M. AND REINSCH, N.: Identification of gametes and treatment of linear dependencies in the gametic QTL-relationship matrix and its inverse. *Genet. Sel. Evol.* **36** (2004) 621-642.

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

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

1.9 APPENDIX A

From page 634 in TUCHSCHERER ET AL. (2004):

$$d_i^* = (1 - \mathbf{Q}_i^k \hat{\mathbf{G}}_i \mathbf{Q}_i^{k'})$$

For paternal gamete $i = Pg$,

$$\begin{aligned} &= 1 - \begin{bmatrix} T_{(Pg)} & (1 - T_{(Pg)}) & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{G}_{s(i)} & \mathbf{G}_{s(i)d(i)} \\ \mathbf{G}_{d(i)s(i)} & \mathbf{G}_{d(i)} \end{bmatrix} \begin{bmatrix} T_{(Pg)} & 0 \\ (1 - T_{(Pg)}) & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \\ &= 1 - \left[T_{(Pg)} + (1 - T_{(Pg)})f_{sire} \quad T_{(Pg)}f_{sire} + (1 - T_{(Pg)}) \right] \\ &= 1 - \left[T_{(Pg)} + (1 - T_{(Pg)})f_{sire}T_{(Pg)} + (T_{(Pg)}f_{sire}) + (1 - T_{(Pg)})(1 - T_{(Pg)}) \right] \\ &= 1 - \left[T_{(Pg)}^2 + (1 - T_{(Pg)})f_{sire}T_{(Pg)} + T_{(Pg)}f_{sire}(1 - T_{(Pg)}) + (1 - T_{(Pg)}) \right] \\ &= 1 - T_{(Pg)}^2 - (1 - T_{(Pg)})^2 + 2T_{(Pg)}(1 - T_{(Pg)})f_{sire} \\ &= 1 - T_{(Pg)}^2 - (1 - T_{(Pg)})^2 - 2T_{(Pg)}(1 - T_{(Pg)})f_{sire} + 2T_{(Pg)}(1 - T_{(Pg)})(1 - f_{sire}) - 2T_{(Pg)}(1 - T_{(Pg)})(1 - f_{sire}) \\ &= 2T_{(Pg)}(1 - T_{(Pg)})(1 - f_{sire}) \\ &\text{because } T_{(Pg)}^2 + (1 - T_{(Pg)})^2 + 2T_{(Pg)}(1 - T_{(Pg)})f_{sire} + 2T_{(Pg)}(1 - T_{(Pg)})(1 - f_{sire}) = 1. \end{aligned}$$

Therefore for animal a ,

$$\mathbf{D}_a = \begin{bmatrix} 2T_{(Pg)}(1 - T_{(Pg)})(1 - f_{sire}) & 0 \\ 0 & 2T_{(Mg)}(1 - T_{(Mg)})(1 - f_{dam}) \end{bmatrix}$$

and

$$\mathbf{D}_a^{-1} = \begin{bmatrix} \frac{1}{2T_{(Pg)}(1 - T_{(Pg)})(1 - f_{sire})} & 0 \\ 0 & \frac{1}{2T_{(Mg)}(1 - T_{(Mg)})(1 - f_{dam})} \end{bmatrix}$$

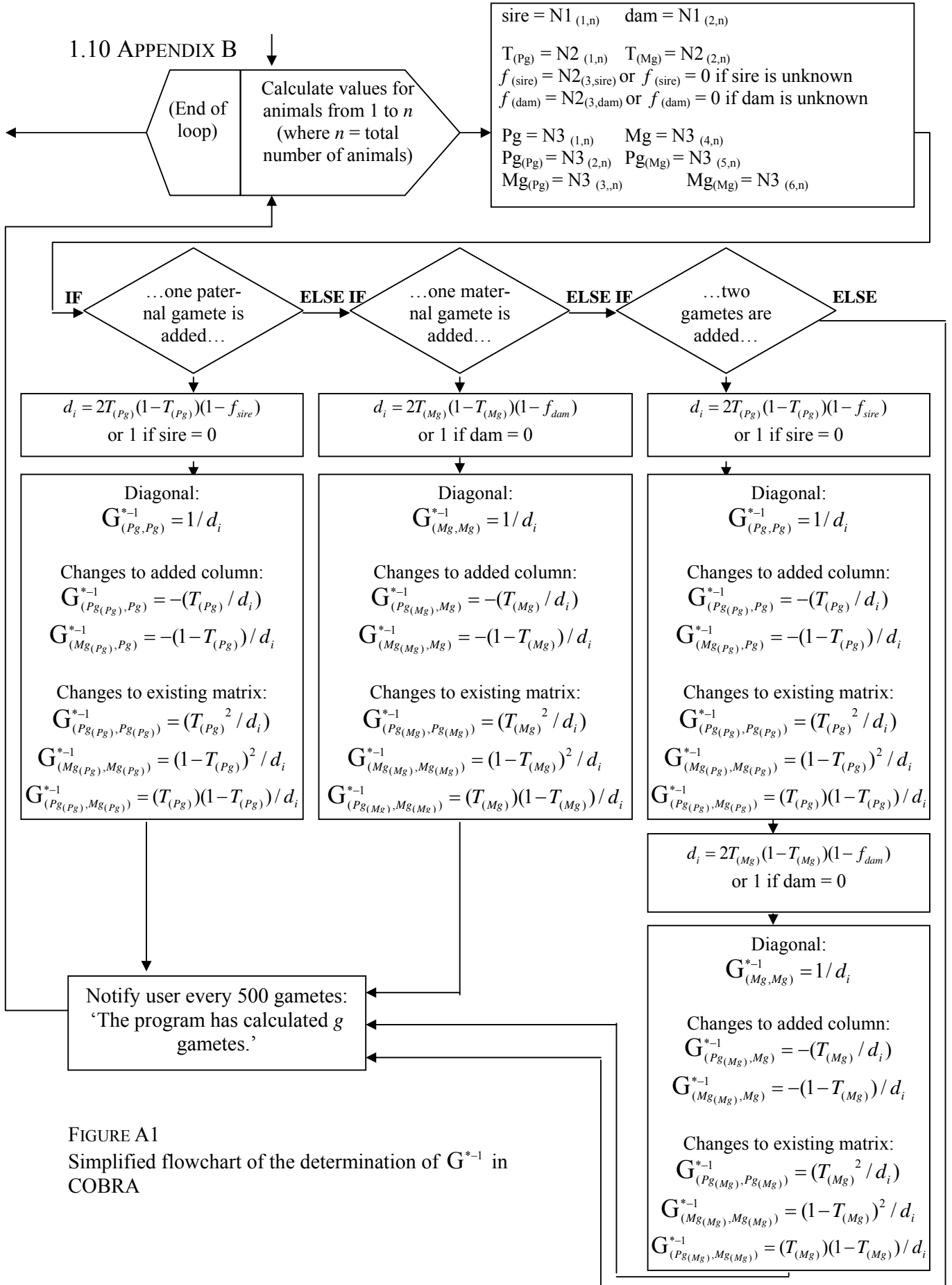


FIGURE A1
Simplified flowchart of the determination of G^{*-1} in
COBRA

CHAPTER TWO:
TIGER: A SOFTWARE SYSTEM FOR FINE-MAPPING QUANTITATIVE TRAIT LOCI

Christine Baes¹ and Norbert Reinsch¹

¹*Forschungsinstitut für die Biologie Landwirtschaftlicher Nutztiere, Forschungsbereich
Genetik und Biometrie*

Published in Archives of Animal Breeding (2008) 51:402-412

ABSTRACT

The localisation of quantitative trait loci which contribute significantly to phenotype variation of economically important traits in domestic species has become an important goal in animal genomics. Several such loci have been roughly identified using linkage analyses; however the focus has now shifted towards fine mapping and pinpointing causal mutations. In the context of a cooperative national research project, the software system TIGER was developed. TIGER is a UNIX script linking several individual Fortran programmes and is used for comprehensive variance component analysis of fine mapping data. Starting with raw genotype data, pedigree and marker map information and ending with a residual maximum likelihood-based test for each putative quantitative trait locus position, the software provides the user with an 'all in one' package capable of linkage analysis, linkage disequilibrium analysis and combined linkage / linkage disequilibrium analysis. The software system has been employed in 4 fine mapping projects on 4 distinct cattle chromosomes.

KEYWORDS

fine mapping / quantitative trait loci / variance component

2.1 INTRODUCTION

The vast majority of economically important traits in plants and animals are quantitative in nature and can be influenced by many individual chromosomal regions (quantitative trait loci, QTL). The localisation of QTL which contribute significantly to phenotypic variation is an important goal of both the scientific and the agricultural community, although mapping these QTL is only the first step in the search for the underlying genes (COPPIETERS ET AL., 1999). Several QTL for economically important traits in domestic species have been roughly positioned using linkage analysis (LA). However, the confidence intervals achieved in such studies contain far too many genes to allow the identification of causal mutations and the results are family-specific. The focus has now shifted to fine mapping these QTL as the next step toward identifying genes underlying complex traits. Methods that utilize historical recombination information on the basis of linkage disequilibrium (LD) allow implementation of fine mapping analyses on existing outbred populations. These methods have become popular for QTL fine mapping in livestock, where the creation of mapping populations such as advanced intercross lines (DARVASI AND SOLLER, 1995) or others (DARVASI, 1998) is not practicable.

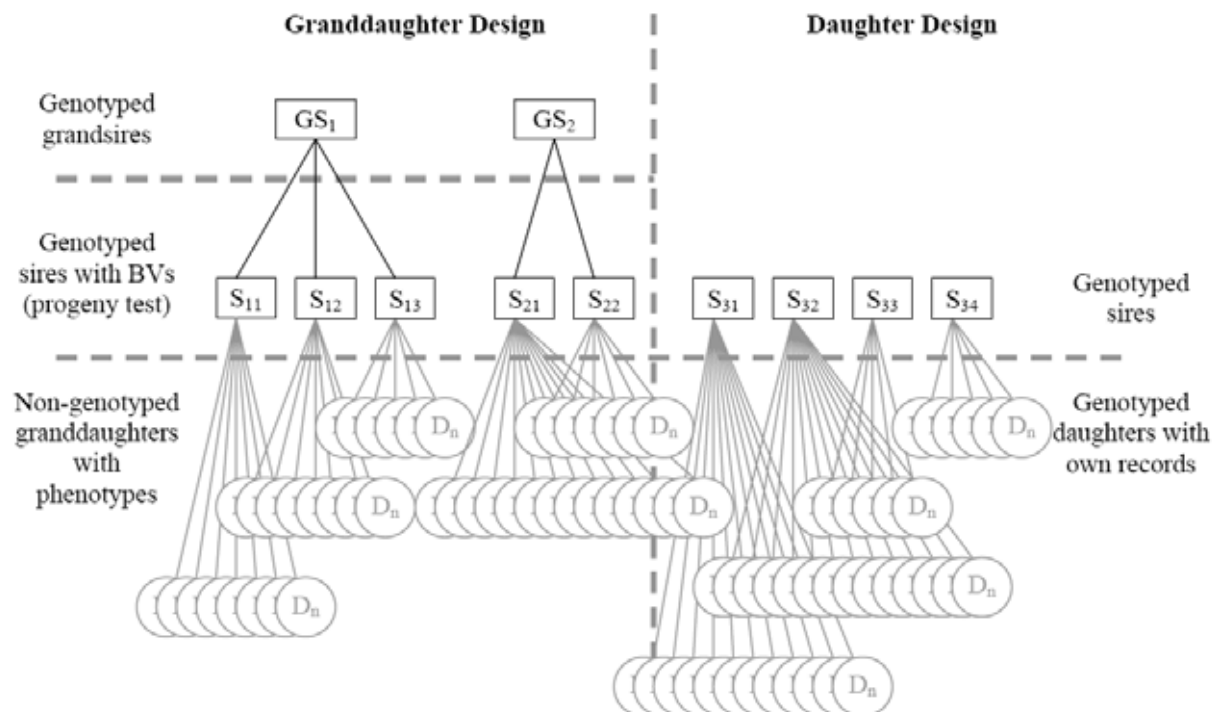


FIGURE 1

The granddaughter design (left) is based on a series of paternal half brother pedigrees. Genotype information is available for grandsires and sires but not for granddaughters. Sire breeding values are estimated through progeny testing of at least 50 granddaughters. The daughter design (right) consists of large sets of paternal half sisters.

The pedigree structure of typical dairy cattle populations is ideally suited for QTL mapping. Through the wide-scale implementation of artificial insemination, the paternal ancestry of many animals in the population can be traced back to a handful of bulls. The effective population size is small, although, as for example in the Holstein Friesian breed, many millions of animals exist. Furthermore, the phenotypic data of female animals for most relevant traits has been systematically recorded for use in routine breeding value estimation. The performance of the sires' progeny can thereby be included in the calculation of the sires' breeding value. This 'granddaughter design' (Fig. 1) lends itself to use in mapping QTL and requires less genotyped animals than a 'daughter design', in which only large sets of paternal half sisters are used (GEORGES ET AL., 1995, WELLER ET AL., 1990).

Various statistical models can be utilized to refine QTL positioning in mapping experiments (FREYER ET AL., 2003), whereby both regression-based models (RM) and variance component models (VC) are usually preferred. Single-marker or multi-marker association studies have been documented (XU, 2003, GRAPES ET AL., 2004, ZHAO ET AL., 2007). Although the RM method is robust, disequilibrium between specific haplotypes and the putative QTL may be larger than between single markers and QTL in some cases.

In VC analyses, a random additive allelic effect is estimated for every unique haplotype and a random polygenic effect for each animal is included to correct for correlations between polygenic and allelic coancestry. The covariance between the effects of any two haplotypes in a pedigree is assumed equivalent to the probability that the haplotypes have identical origin (identical by descent, IBD) and are included in the gametic relationship matrix \mathbf{G} (FERNANDO AND GROSSMAN, 1989). Founder haplotypes in LA (i.e. maternal haplotypes of sires and maternal and paternal haplotypes of grandsires in a granddaughter design) are considered unrelated and their pair-wise IBD probabilities (covariances) are set to zero. For fine mapping experiments, MEUWISSEN AND GODDARD (2000, 2001) suggested incorporating historical recombination as a means of exploiting the non-random allelic association between QTL and closely linked markers (i.e. population-wide LD). The authors showed how to compute IBD sub-matrices containing IBD covariances between founder haplotypes conditional on marker data (see Fig. 2). This method has consequently been extended to include linkage information (MEUWISSEN ET AL., 2002, BLOTT ET AL., 2003, OLSEN ET AL., 2004).

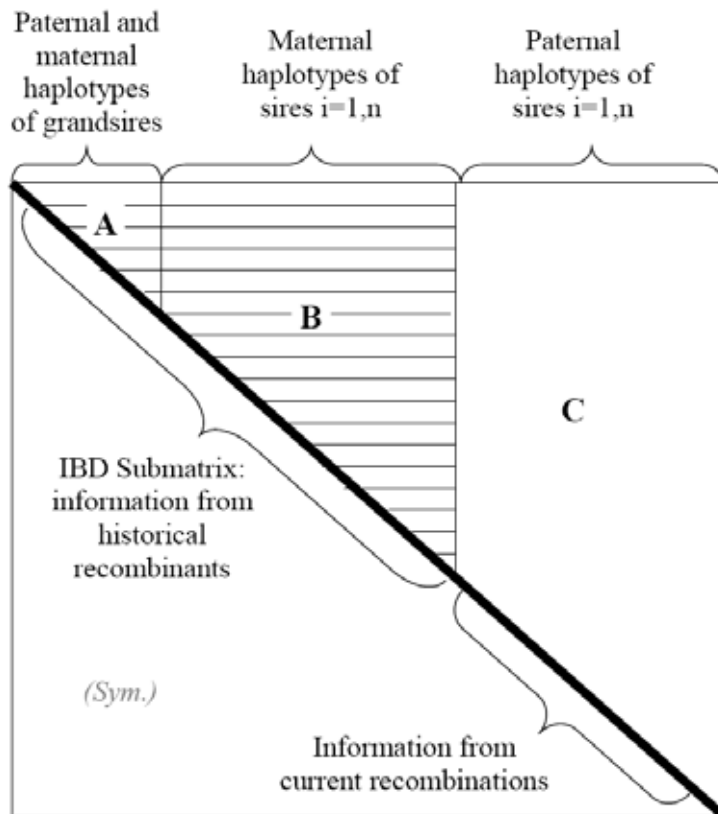


FIGURE 2

Schematic representation of the gametic relationship matrix. Haplotypes are sorted according to generations with founder haplotypes to the left (blocks A and B) and haplotypes with known ancestors to the right (block C). The blocks marked with horizontal lines (A and B, IBD submatrix) are empty in a linkage analysis (i.e. elements = 0) and are dense for linkage disequilibrium and combined linkage / linkage disequilibrium experiments (i.e. elements $\neq 0$). Block C is always dense.

Variance components and individual haplotype effects can be estimated by incorporating the IBD sub-matrix into the gametic relationship matrix \mathbf{G} and applying restricted maximum likelihood methods (REML). The disadvantage of this method is that, as opposed to RM methods, certain assumptions regarding population structure must be made (i.e. number of generations since the mutation and effective population size of founder population, selection, etc.). More importantly, an accurate marker map is mandatory, especially with regard to the correct order of markers. Although the first few drafts of bovine linkage maps appeared as early as 1994 (i.e. BARDENSE ET AL., 1994; BISHOP ET AL., 1994; GEORGES ET AL., 1995, etc.), with later versions using improved techniques (i.e. KAPPES ET AL., 1997; IHARA ET AL., 2004; EVERTS-VAN DER WIND ET AL., 2005), gaps and areas of uncertainty in publicly available maps still persist.

A whole genome scan for QTL affecting functional traits including udder health was conducted in 2001 in cooperation with German AI and breeding organizations, scientific institutes for animal breeding and animal computing centers (THOMSEN ET AL., 2000). This project, instigated by the German cattle breeders' federation (ADR), enabled the identification of several QTL affecting functional traits in the German Holstein population, including those on *Bos taurus* autosomes 2, 18 and 27 affecting somatic cell score in milk (BENNEWITZ ET AL.,

2003; BRINK, 2003; KÜHN ET AL., 2003; XU ET AL., 2006; KÜHN ET AL., 2008). Although extensive information was gained from these experiments, no candidate genes could be identified due to the relatively large confidence intervals typical of QTL mapping experiments. In 2005, a successor project (FUGATO M.A.S.-Net) was initiated with the goal of identifying the positional and functional candidate genes associated with SCS in the regions previously identified. In the context of this second project, a comprehensive software package was developed to process the genotypic data resulting from these efforts using VC methods. The aim of this paper is to present and describe the software system TIGER, which was designed for conducting LA, LD and combined LA/LD analyses for fine mapping experiments. The package is composed of both new and modified pre-existing Fortran programmes within a UNIX shell script. The system has been used for fine mapping on 4 distinct *bos Taurus* chromosomes.

2.2 PROGRAM SEQUENCE

The system operates in 6 main steps:

- 1- Allele frequencies and transmitting probabilities for each putative QTL position are determined.
- 2- The IBD sub-matrices for each putative QTL position are calculated and are stored.
- 3- The IBD sub-matrices are tested for positive definiteness; matrices with negative determinants are bent until all eigenvalues are positive. All IBD sub-matrices are then inverted and stored.

Finally, 3 loops are conducted over all putative QTL positions. The first loop conducts LA, the second LD analysis and the third and final loop provides the used with combined LA / LD analysis information, however the same programmes are used for each loop:

- 4- In the first loop, LA is conducted by setting up and inverting the condensed gametic relationship matrix (BAES ET AL., 2007; TUCHSCHERER ET AL., 2004) using transmitting probabilities calculated in step 1, merging the gametic index information with the phenotypes and running ASReml.
- 5- In the second loop, the IBD sub-matrices calculated in step 2 are incorporated into the gametic relationship matrix. Transmitting probabilities of non-founder gametes (i.e. values calculated in step 1) are set to 0.5 (LD). The gametic index information is then merged with the phenotypes and ASReml is run.

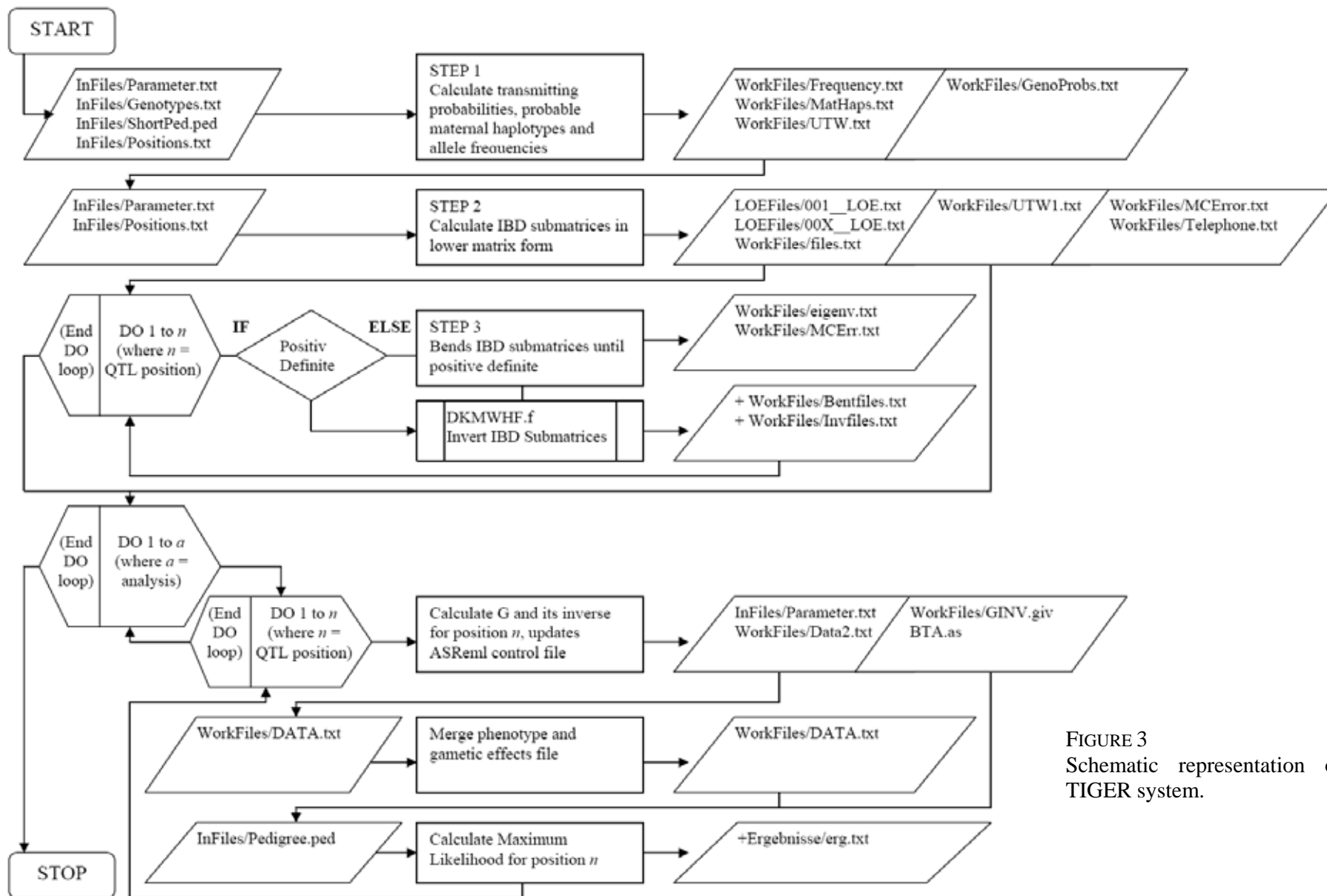


FIGURE 3
Schematic representation of the TIGER system.

- 6- In the third and final loop, the IBD sub-matrices calculated in step 2 are again incorporated into the condensed gametic relationship matrices and the transmitting probabilities of non-founder gametes are calculated as in step 1. As in steps 4 and 5, the gametic index information is then merged with the phenotypes and ASReml is run.

A schematic flowchart of the package is displayed in Fig. 3. The program sequence also describes the acronym ‘TIGER’, which stands for **T**ransmitting probabilities, **I**BD sub-matrices, **G*** matrix and **E**stimation of variance components using **R**EML. The following sections offer a more detailed description of each step.

2.2.1 TRANSMITTING PROBABILITIES AND ALLELE FREQUENCIES

Although QTL genotypes are unobservable, markers linked to the putative QTL can be genotyped and used to infer the unknown QTL IBD status. The problem is that marker information is often incomplete. Unknown linkage phases, non-informative markers and missing marker genotypes complicate the calculation of transmitting probabilities, which refers to the probability that the sire passed his first (second) allele at a given locus to his offspring. These probabilities are required for calculation of **G**, which is in turn required for solving the mixed model equations in VC analysis for LA, LD and LD/LA studies.

In typical fine mapping experiments, markers are dense. Furthermore, due to the size of the half sib families, reconstruction of haplotypes is fairly precise for both parents and progeny. Computation of probabilities for multi-marker haplotypes is implemented as described in REINSCH (2002); the program implemented was developed especially for use in designs in which 2 generations are genotyped (i.e. half sib families, backcross, etc.). Input files include genotype information, a short pedigree (only sires and grandsires) and a file containing marker and putative QTL positions. Once the input files are read in, genotypes of offspring are checked for compatibility with those of the parents using Mendelian laws of inheritance. The pedigree is tested for parent-offspring conflicts and the origin of progeny alleles is determined for each marker locus. The recombination frequencies for all intervals are calculated using Kosambi’s map function. The most probable paternal and maternal haplotypes for parents are then reconstructed. Probabilities for linkage phase 1 are computed from genotypes and pedigree information. After all possible haplotypes of the sires are recorded, the most probable haplotype (linkage phase) produced by the candidates parents is calculated assuming haplotypes of the parents are known and correct. Apart from their use in the recursive calculation of **G**, the transmitting probabilities also provide a measure of marker informativity. The program was further modified to estimate population-wide allele frequencies using all possible maternal haplotypes of offspring:

$$\text{frequency}(k) = \frac{1}{n} \sum_{ij} \delta_{ij}^k \cdot h_{ij}$$

where k is a given marker allele, n is the number of final progeny, δ is an indicator variable equal to 1 if allele k is present in a given haplotype and zero otherwise and h is the probability of haplotype j occurring for animal i . The most probable maternal haplotypes are also saved to file for use in calculation of the IBD sub-matrices.

2.2.2 CALCULATION OF IBD SUB-MATRICES

MEUWISSEN AND GODDARD (2000) described a method to calculate the IBD sub-matrix based on deterministic predictions which takes the number of markers flanking the putative QTL position, the extent of LD in the population based on the expectation under finite population size, and the time point of mutation occurrence into account. A Fortran (90/95) program based on the gene dropping procedure described by MEUWISSEN AND GODDARD (2000, 2001) was developed and incorporated into the TIGER system. The method we employed involves setting up IBD sub-matrices for the maternal haplotypes of the sires and both maternal and paternal haplotypes of the grandsires in the real data set being examined. Parameter options allow the user to define the effective population size, the number of generations of gene-dropping, the number of repetitions and the mapping function to be used. The number of markers left and right of the putative QTL can also be user-defined. GRAPES AT AL. (2006) found that fitting a 4-6 marker haplotype as a sliding window across the region resulted in the greatest accuracy compared to that of the other haplotype sizes tested. However for practical applications it may be advantageous to define the size of the ‘window’ surrounding the putative QTL in cM instead of the number of markers. This allows for more flexibility when the marker distribution over the chromosome is not constant. Haplotypes were sorted into groups according to the number of identical markers to the left and to the right of the putative QTL.

The method and basic principles of gene dropping are relatively simple. Since each parent possesses two alleles at a given locus and since only one of these alleles will be passed on to an offspring, each allele has a one in two chance of being transmitted to each descendant, this probability being independent of the heritage of the other parent. In other words, this is quite simply a simulation of the Mendelian transmission of a gene at a given locus. The method can be extended to the case of several loci, thus making it possible to measure the change in linkage disequilibrium. Application of this method, however, rests on certain assumptions. Selection at the simulated locus is not taken into account, since we are seeking only to replicate the effect of change due to the Mendelian transmission of genes. In reality, besides random variations, allele frequencies are subject to evolutionary forces such as migration, mutation, and natural selection, which produce specific and

non-random variation in allele frequencies. The effect of selection can be considered negligible; however, it is important to realize that it is not accounted for here.

One IBD sub-matrix is calculated for every putative QTL locus given by the user. These matrices are stored individually and are dense. The average inbreeding coefficient within simulated generations and the Monte Carlo errors of each haplotype constellation are saved to file for control purposes.

2.2.3 POSITIVE DEFINITENESS OF IBD COEFFICIENT MATRICES

The numeric instability of the IBD sub-matrices was mentioned but not discussed in depth by MEUWISSEN AND GODDARD (2000). The IBD coefficient sub-matrices in LD and LA/LD analyses are very dense because all founder alleles (paternal and maternal alleles of grandsires and maternal alleles of sires) are assigned covariances. Furthermore, these matrices are not necessarily positive definite, which impedes the inversion step required for their use as a part of the gametic relationship matrix. However, IBD coefficient matrices with determinants less than one can be ‘bent’ until the lowest eigenvalue exceeds a preset limit (i.e. zero to achieve positive definiteness) (HAYES AND HILL, 1981; ESSL, 1991).

We adopted the BENDPDF program of ESSL (1991) and added the matrix inversion subroutine DKMXHF of MEYER (2008, personal communication) to ensure positive definiteness of the IBD sub-matrices, proper calculation of matrix determinants and correct inversion. The numbers of negative eigenvalues in each sub-matrix as well as the matrix determinants are saved to file for control purposes. In our experience, approximately 20% of all IBD sub-matrices within must be bent; the number of negative eigenvalues in our submatrices normally ranged from from 0 to 4.5% per matrix.

2.2.4 THE G MATRIX

Because allelic (gametic) effects are random, their covariance must be included for estimation in a variance component model. This information is included in the gametic relationship matrix \mathbf{G} . TUCHSCHERER ET AL. (2004) showed how to eliminate possible linear dependencies in this matrix by considering haplotypes with extremely high or extremely low transmitting probabilities identical to those of the respective parent. If transmission probabilities very close to one or zero occur, a predefined threshold (e.g. $\varepsilon = 0.03$) can be given and all transmission probabilities below 0.03 are ‘condensed’ to 0 and all those above 0.97 to 1. Haplotypes corresponding to such transmission probabilities are considered ‘non-unique’. The resulting condensed gametic relationship matrix, \mathbf{G}^* , can have between n and $2n$ rows and columns, depending on how many unique gametes occur,

whereby n is the number of genotyped animals in the pedigree. This method prevents numerical difficulties due to quasi-linear dependencies which may occur due to the inclusion of almost identical rows and columns in the gametic relationship matrix.

The condensed covariance matrix therefore contains one row and one column for every unique gamete (haplotype) of genotyped animals and is sorted according to gametic generation (i.e. founder gametes followed by gametes with known ancestors). If only grandsires and sires are genotyped, all maternal gametes of sires are considered founders and therefore unrelated; the pairwise IBD probabilities of these alleles are zero. COBRA (BAES AND REINSCH, 2007) was designed to calculate the condensed matrix \mathbf{G}^* based on the algorithm of TUCHSCHERER ET AL. (2004). The COBRA program was slightly modified to include IBD sub-matrices described above for LD and combined LA/LD analyses (see Fig. 2 for the arrangement of haplotype effects in \mathbf{G}^*). An ASReml parameter file corresponding to the data on the given marker interval (i.e. number of QTL effects in the \mathbf{G}^* matrix, etc.) is created for each position.

2.2.5 LIKELIHOOD CALCULATION

Variance component analyses for continuous traits are normally based on the mixed model, and maximum likelihood or related methods of inference are employed. In particular, the so-called residual or restricted maximum likelihood (REML) is widely used for analyses of continuous traits. The popularity of mixed model analyses by REML has been furthered by the availability of appropriate software. Although implemented in general statistical packages, specific programs for VC estimation in an animal breeding context have been developed (GILMOUR ET AL., 2006; KOVACS AND GROENEVELD, 2003; MADSEN AND JENSEN, 2006). These programs use a mixture of Fisher's scoring and Newton–Raphson to maximize the restricted likelihood, called the average information restricted maximum-likelihood (AI-REML) algorithm.

Due to the somewhat delicate procedure required for setting up and inverting \mathbf{G}^* , a program allowing user-defined covariance structures and their inverses was required. This prerequisite was filled by ASReml (GILMOUR ET AL. 2006). The following mixed model equations were applied:

$$\begin{bmatrix} \mathbf{XD}^{-1}\mathbf{X} & \mathbf{XD}^{-1}\mathbf{Z} & \mathbf{XD}^{-1}\mathbf{W} \\ \mathbf{ZD}^{-1}\mathbf{X} & \mathbf{ZD}^{-1}\mathbf{Z} + \mathbf{A}_u^{-1} & \mathbf{ZD}^{-1}\mathbf{W} \\ \mathbf{WD}^{-1}\mathbf{X} & \mathbf{WD}^{-1}\mathbf{Z} & \mathbf{WD}^{-1}\mathbf{W} + \mathbf{G}_v^{*-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\mu}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{XD}^{-1}\mathbf{y} \\ \mathbf{ZD}^{-1}\mathbf{y} \\ \mathbf{WD}^{-1}\mathbf{y} \end{bmatrix}$$

Where $\mathbf{y}_{(m \times 1)}$ is a vector of m DYDs for n animals, $\boldsymbol{\mu}$ is a fixed effect common to all observations, $\mathbf{u}_{(n \times 1)}$ is a vector of random polygenic effects and $\mathbf{v}_{(\text{gam} \times 1)}$ is a vector of random gametic effects of a

marked QTL that is linked to a single polymorphic marker locus with gam equal to the number of unique gametes. Subscripts in parenthesis of the vectors and matrices denote their dimensions. $\mathbf{X}_{(m \times 1)}$, $\mathbf{Z}_{(m \times n)}$ and $\mathbf{W}_{(m \times gam)}$ are known incidence matrices, \mathbf{A}_u^{-1} is the inverse of the relationship matrix and $\mathbf{G}_v^{*^{-1}}$ is the inverse of the condensed gametic QTL relationship matrix. \mathbf{D}^{-1} is the inverse of a diagonal matrix containing weights for each observation. Expectations of \mathbf{u} , \mathbf{v} and \mathbf{e} and covariances between them are assumed to be zero. The model depicted above contains only the fixed effect $\boldsymbol{\mu}$, however LD can easily be included by regression on one or more markers instead of through the IBD sub-matrix. The number of regression coefficients is then the number of different marker alleles of the given marker(s).

A comprehensive software package was developed to process and analyse the genotypic data within the context of the FUGATO M.A.S.-Net project using VC methods. This paper presented the software system TIGER, which was designed for conducting linkage, linkage disequilibrium and combined linkage / linkage disequilibrium analyses for fine mapping experiments. The TIGER system employs both new and modified pre-existing Fortran programmes within a UNIX shell script and is an 'all in one' solution for analysing genotypic data.

2.3 ACKNOWLEDGEMENTS

Financial support from the FUGATO M.A.S.-Net project is gratefully acknowledged. We would like to thank Dörte Wittenburg for invaluable assistance regarding ASReml.

2.4 REFERENCES

ABECASIS, G.R.; CHERNY, S.S.; COOKSON, W.O. AND CARDON, L.R.: Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* 30 (2002), 97-101.

ANDERSON, S.; DE BRUIJN, M. H.; COULSON, A. R.; EPERON, I.; SANGER, F. AND YOUNG, I. G.: Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. *J. Mol. Biol.* 156 (1982) 4:683-717.

BAES, C. AND REINSCH, N.: Computing the condensed conditional gametic QTL relationship matrix and its inverse. *Arch. Tierz.* 50 (2007) 3, 294-308.

BARENDSE, W.; ARMITAGE, S.M.; KOSSAREK, L.M.; SHALOM, A.; KIRKPATRICK, B.W.; RYAN, A.M.; CLAYTON, D.; LI, L.; NEIBERGS, H.L.; ZHANG, N.; GROSSE, W.M.; WEISS, J.; CREIGHTON, P.; MCCARTHY, F.; RON, M.; TEALE, A.J.; FRIES, R.; MCGRAW, R.A.; MOORSE, S.S.; GEORGES, M.; SOLLER, M.; WOMACK, J.E. AND HETZEL D.J.S.: A genetic linkage map of the bovine genome. *Nature Genet.* 6 (1994) 227-235.

BISHOP, M.D.; KAPPES, S.M.; KEELE, J.W.; STONE, R.T.; SUNDEN, S.L.F.; HAWKINS, G.A.; TOLDO, S.S.; FRIES, R.; GROSZ, M.D.; YOO, J.; AND BEATTIE, C.W.: A genetic linkage map for cattle. *Genetics* 136 (1994) 619-639.

BLOTT, S.; KIM, J.J.; MOISIO, S.; SCHMIDT-KÜNTZEL, A.; CORNET, A., BERZI, P.; CAMBIASO, N.; FORD, C.; GRISART, B.; JOHNSON, D.; KARIM, L.; SIMON, P.; SNELL, R.; SPELMAN, R.; WONG, J.; VILKKI, J.; GEORGES, M.; FARNIR, F. AND COPPIETERS, W.: Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* 163 (2003) 253-266.

COPPIETERS, W.; BLOTT, S.; FARNIR, F.; GRISART, B.; RIQUET, J. AND GEORGES, M.: From phenotype to genotype: towards positional cloning of QTL in livestock? *Arch. Tierz.* 42 (1999) Special Issue, 86-92.

DARVASI, A. AND SOLLER, M.: Advanced Intercross Lines, an Experimental Population for Fine Genetic Mapping. *Genetics* 141 (1995) 3, 1199-1207.

DARVASI, A.: Experimental strategies for the genetic dissection of complex traits in animal models. *Nat. Genet.* 18 (1998) 1, 19-24.

EßL, A.: Choice of an appropriate bending factor using prior knowledge of the parameters. *J. Anim. Breed. Gen.* 108 (1991) 89-101.

EVERTS-VAN DER WIND, A.; LARKIN, D.M.; GREEN, C.A.; ELLIOTT, J.S.; OLMSTEAD, C.A.; CHIU, R.; SCHEIN, J.E.; MARRA, M.A.; WOMACK, J.E. AND LEWIN, H.A.: A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution. *Proc. Natl. Acad. Sci. U S A*, 102 (2005) 51:18526-31.

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

FREYER, G.; KÜHN, C. AND WEIKHARD, R.: Comparison of different statistical-genetic approaches of QTL detection by evaluating results from a real dairy cattle data set. *Arch. Tierz.* 46 (2003) 413-423.

GEORGES, M.; NIELSEN, D.; MACKINNON, M.; MISHRA, A.; OKIMOTO, R.; PASQUINO, A.T.; SARGEANT, L.S.; SORENSEN, A.; STEELE, M.R.; ZHAO, X.; WOMACK, J.E. AND HOESCHELE, I.: Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 139 (1995) 907-920.

GEORGES, M.: Mapping, Fine Mapping, and Molecular Dissection of Quantitative Trait Loci in Domestic Animals. *Annu. Rev. Genomics Hum. Genet.* 8 (2007) 131-162.

GILMOUR, A.; GOGEL, B.; CULLIS, M. AND THOMPSON, R.: ASReml User Guide Release 2.0. (2006) VSN International Ltd., Hemel Hempstead, HP1 1ES, UK.

GRAPES, L.; DEKKERS, J.C.M.; ROTHSCHILD, M.F. AND FERNANDO, R.L.: Comparing linkage disequilibrium-based methods for fine mapping quantitative trait loci. *Genetics* 166 (2004) 1561-1570.

GRAPES, L.; FIRAT, M.Z.; DEKKERS, J.C.M.; ROTHSCHILD, M.F. AND FERNANDO, R.L.: Optimal haplotype structure for linkage disequilibrium-based fine mapping of quantitative trait loci using identity by descent. *Genetics* 172 (2006) 1955-1965.

HAYES, J.F. AND HILL, W.G.: Modification of Estimates of Parameters in the construction of Genetic Selection Indices ('Bending'). *Biometrics* 37 (1981) 483-493.

IHARA, N.; TAKASUGA, A.; MIZOSHITA, K.; TAKEDA, H.; SUGIMOTO, M.; MIZOGUCHI, Y.; HIRANO, T.; ITOH, T.; WATANABE, T.; REED, K.M.; SNELLING, W.M.; KAPPES, S.M.; BEATTIE, C.W.; BENNETT, G.L. AND SUGIMOTO, Y.: A comprehensive genetic map of the cattle genome based on 3802 microsatellites. *Genome Res.*, 14 (2004) 10A:1987-98.

KAPPES, S.M.; KEELE, J.W.; STONE, R.T.; MCGRAW, R.A.; SONSTEGARD, T.S.; SMITH, T.P.L.; LOPEZ-CORRALES, N.L. AND BEATTIE C.W.: A second-generation linkage map of the bovine genome. *Genome Res.*, 7 (1997) 3:235-49.

KOVACS, M.A. AND GROENEVELD, E.: VCE-5 User's Guide and Reference Manual, Version 5.1. (2003) Mariensee, Germany: Institute of Animal Husbandry and Animal Behaviour, Federal Research Center of Agriculture.

KÜHN, C.; BENNEWITZ, J.; REINSCH, N.; XU, N.; THOMSEN, H.; LOOFT, C.; BROCMANN, G.A.; SCHWERIN, M.; WEIMANN, C.; HIENDLEDER, S.; ERHARDT, G.; MEDJUGORAC, I.; FORSTER, M.; BREINIG, B.; REINHARDT, F.; REENTS, R.; RUSS, I.; AVERDUNK, G.; BLUMEL, J.; KALM, E.: Quantitative trait loci mapping of functional traits in the German Holstein cattle population. *J. Dairy Sci.* 86 (2003) 360-368.

KÜHN, C.; REINHARDT, F.; SCHWERIN, M.: Marker assisted selection of heifers improved milk somatic cell count compared to selection on conventional pedigree breeding values. *Arch. Tierz.* 51 (2008) 23-32.

MEUWISSEN, T. AND GODDARD, M.: Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics* 155 (2000) 421-430.

MEUWISSEN, T. AND GODDARD, M.: Prediction of identity by descent probabilities from marker haplotypes. *Genet. Sel. Evol.* 33 (2001) 605-634.

MEUWISSEN, T.; KARLSEN, A.; LIEN, S.; OLSAKER, I. AND GODDARD, M.: Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* 161 (2002) 373-379.

OLSEN, H.G.; LIEN, S.; SVENDSEN, M.; NILSEN, H.; ROSETH, A.; AASLAND OPSAL, M. AND MEUWISSEN, T.H.E.: Fine Mapping of Milk Production QTL on BTA6 by Combined Linkage and Linkage Disequilibrium Analysis. *J. Dairy Sci.* 87 (2004) 690-698.

REINSCH, N.: A multiple-species, multiple-project database for genotypes at codominant loci. *J. Anim. Breed. Genet.* 116 (1999) 425-435.

REINSCH, N.: A general likelihood approach to trait-based multipoint linkage analysis in large groups of half sibs and super sisters. *Genetics* 162 (2002) 413-424.

SOBEL, E. AND LANGE, K.: Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker sharing statistics. *Am. J. Hum. Genet.* 58 (1996) 1323-1337.

THOMSEN, H.; REINSCH, N.; XU, N.; BENNEWITZ, J.; LOOFT, C.; GRUPPE, S.; KUHN, C.; BROCKMANN, G. A.; SCHWERIN, M.; LEYHE-HORN, B.; HIENDLEDER, S.; ERHARDT, G.; MEDJUGORAC, I.; RUSS, I.; FORSTER, M.; BREINIG, B.; REINHARDT, F.; REENTS, R.; BLUMEL, J.; AVERDUNK, G.; KALM, E.: A

whole genome scan for differences in recombination rates among three *Bos taurus* breeds. *Mamm Genome*. 12 (2001) 724-728.

TUCHSCHERER, A., MAYER, M., REINSCH, N.: Identification of gametes and treatment of linear dependencies in the gametic QTL-relationship matrix and its inverse. *Genet. Sel. Evol.* 36 (2004) 621-642.

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

XU, S.: Estimating polygenic effects using markers of the entire genome. *Genetics* 163 (2003) 789-801.

XU, N.; PAUL, S.; BENNEWITZ, J.; REINSCH, N.; THALLER, G.; REINHARDT, F.; KÜHN, CH.; SCHWERIN, M.; ERHARDT, G.; WEIMANN, C.; THOMSEN, H.; MISCHRA, S. AND KALM, E.: Confirmation of quantitative trait loci for somatic cell score on bovine chromosome18 in the German Holstein. *Arch.Tierz.* 49 (2006) 111-119.

ZHAO, H.H.; FERNANDO, R.L. AND DEKKERS, J.C.M.: Power and Precision of Alternate Methods for Linkage Disequilibrium Mapping of Quantitative Trait Loci. *Genetics* 175 (2007) 1975-1986.

CHAPTER THREE:

REFINED MAPPING OF QUANTITATIVE TRAIT LOCI FOR SOMATIC CELL SCORE ON
BTA02 IN THE GERMAN HOLSTEIN

Christine Baes¹, Ines Görtz², Manfred Mayer¹, Christina Weimann², Georg Erhardt², Zengting Liu³, Friedrich Reinhardt³, Norbert Reinsch¹

¹*Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere, Forschungsbereich
Genetik und Biometrie*

²*Institut für Tierzucht und Haustiergenetik, Justus-Liebig-Universität Giessen*

³*Vereinigte Informationssysteme Tierhaltung w.V.*

Submitted

ABSTRACT

The aim of this study was to more precisely map a previously reported quantitative trait locus (QTL) affecting somatic cell score on *Bos taurus* autosome 2 by increasing the number of markers 4-fold, analysing more families and exploiting within-population linkage disequilibrium (LD). A granddaughter design of 10 German Holstein grandsire families with 1121 progeny tested sons was used. Twenty-six markers with an average marker spacing of 3.14 cM were genotyped along 81.6 cM. Linkage analysis (LA) was performed using variance component methodology. The incorporation of LD was first done using variance component methods followed by regression on marker alleles. LA revealed genome-wide significance ($\text{LOD} > 3$) at 15 contiguous marker-intervals, with the maximum test-statistic between DIK2862 and BMS778 and a 1-lod drop-off interval of 38 cM. While the variance-component methods could not detect any LD, 2 individual markers with a significant effect (ILSTS098, $p < 0.05$; BMS778, $p < 0.01$) were found by regression analysis. Compared to previous results QTL-localisation was substantially narrowed; further fine-mapping should focus on the close vicinity of BMS778.

KEYWORDS

fine mapping / linkage disequilibrium / variance components / regression / somatic cell score

3.1 INTRODUCTION

While management is the most effective method for increasing milk quality and curbing mastitis problems in the short term, genetic enhancement must accompany any effective long-term strategy for improving milk quality and udder health. Genetic selection for udder health in Germany, as in many other countries, is achieved through body conformation characteristics and the indicator trait somatic cell score (SCS), which is highly correlated with clinical mastitis and has a moderate heritability of 16-17% (Liu et al., 2004).

There have been many genome scans based on linkage analysis (LA) conducted in the search for QTL affecting production and functional traits. Over 50 QTL affecting SCS have been reported on almost all bovine chromosomes (see Khatkar et al., (2004) for a review), however confidence intervals (CI) are typically of the order of 20-40 cM (Georges et al., 2007) and true association between specific markers and the SCS phenotype cannot be established with such low density maps. Detailed investigation of each individual QTL identified in such studies is required to validate results, to further narrow down chromosomal areas of interest and to identify true candidate genes which may be used in marker assisted selection programmes (MAS).

A promising chromosomal area affecting SCS was found on BTA02 by Bennewitz et al. (2003) after combining designs with data from 5 families in 2 granddaughter designs previously described by Thomsen et al. (2000) and Boichard et al. (2003). These designs were first analysed separately, then jointly for various traits across families to increase statistical power. A chromosome-wide significant QTL for SCS on BTA02 was found in an experiment of Thomsen et al. (2000) conducted using data on the German Holstein population. Although this QTL could not be confirmed in the French population (Boichard et al., 2003), joint analysis of 5 families with offspring in both France and Germany revealed an increased test statistic with genome-wide significance (Bennewitz et al., 2003). The 95% CI identified in the population examined by Thomsen et al (2000) was 128 cM long, while that of the 2 populations examined jointly was 151 cM (Bennewitz et al., 2003).

The purpose of the present study was to further investigate the chromosomal region previously identified by Thomsen et al. (2000) and Bennewitz et al. (2003) affecting udder health on BTA02 in the first lactation by increasing marker density, including more families from the German Holstein population and searching for linkage disequilibrium (LD) using different statistical methods.

3.2 MATERIALS AND METHODS

3.2.1 ANIMALS

Ten paternal half-sib families containing a total of 1121 genotyped bulls (family size ranged from 47 to 350 bulls) were included in our study. The analysed material is a subset of bulls from the German Holstein population currently used in the German MAS breeding programme. Families were mainly chosen because of their size. Three of the families analysed were included in the analysis of Bennewitz et al. (2003), while the remaining 7 families were genotyped in a later phase of the breeding programme. The full pedigree contained 8073 animals and included non-genotyped ancestors of genotyped bulls. Pedigree information was provided by the German genetic evaluation center (VIT) in Verden, Germany.

3.2.2 GENOTYPIC INFORMATION

The analysis was performed with 26 markers, including 24 microsatellites and 2 single nucleotide polymorphisms (SNP). Six of the markers (ILSTS098, ILSTS82, BMS778, MM8D3, TGLA110, BMS1987) included in this study were also investigated in the study by Bennewitz et al (2003).

The microsatellites were chosen from the United States Department of Agriculture's Meat Animal Research Center (USDA MARC) bovine linkage map (<http://www.marc.usda.gov/genome/genome.html>). For genotyping these markers, 5 Multiplex-polymerase chain reactions (PCR) were established with the ABsolute QPCR Mix (Thermo Scientific ABgene, Hamburg, Germany) in a thermocycler (iCycler, BioRad, Munich, Germany) followed by a fragment length analysis using an ABI Prism 377 DNA sequencer (Applied Biosystems, Darmstadt, Germany). For each Multiplex-PCR a 10 x Primermix was compounded with a concentration of each primer between 0.5 - 2.5 pmol/ μ l. The PCR reaction with a final volume of 12 μ l was performed using 6 μ l ABsolute QPCR Mix (Thermo Scientific ABgene, Hamburg, Germany), 3.6 μ l H₂O, 20 ng DNA of genomic bovine DNA and 1.2 μ l 10 x Primermix. The reverse primers were labelled with NED, VIC and 6-FAM. The thermocycler profile was 95°C for 15 min; 30 cycles of 95°C for 30 sec; marker-specific annealing for 90 seconds and 72°C for 60 seconds; and 60°C for 30 min.

The PCR for the SNP SNP_777 was established as described by Goertz et al. (submitted). The PCR was performed with the primers CXCR1_F = 5'-GAGGCCTATCAACCACCGTA-3' and CXCR1_R = 5'-GCGATCAGGACCAGGTTGTA-3' to amplify a 265 bp product containing the SNP. A further PCR was established to verify the SNP SNP_3UTR at position

213 reported by Cobanoglu et al. (2006) with the primers STAT1_fwd: 5'-TAGCCTCAAGTTTGCCAGTG-3' and STAT1_rev: 5'-GGCTCCCTTGATAGAAGTGT-3'. Although various linkage maps are available, a genetic map specific to the population studied was built. To obtain the order of markers with identical linkage map positions in the MARC map, a Basic Local Alignment and Search Tool (BLAST) search was performed against the bovine genome sequence build 3.1. The map distances were then calculated based on this marker order using the /BUILD/ option of the CRIMAP software (Green et al., 1990).

3.2.3 PHENOTYPIC INFORMATION

Somatic cell scores were calculated as follows: $SCS = \log \text{ base } 2 (SCC / 100,000) + 3$, where SCC is the number of somatic cells per millilitre. Lower values represent a lower SCC and therefore favourable effects on udder health. Reliabilities associated with DYD were also obtained from the April 2008 evaluation and were expressed by the number of effective daughter contributions (EDC) following Liu et al. (2004).

Phenotype information was obtained from the VIT and was included for all 1121 genotyped bulls. The phenotypic unit of measurement for SCS in genotyped offspring was the daughter yield deviation (DYD), which is the average phenotype of daughters corrected for fixed effects such as herd, season and calving interval, and corrected for genetic contributions of the daughters' dams. The DYD were obtained from the official release of the January 2008 genetic evaluation and were based on a random regression test-day model (Liu et al., 2001). The DYD for SCS in lactation 1 were selected for analysis, as the number of offspring was highest in this parity.

3.2.4 STATISTICAL METHODS

Linkage phases of all sires and sons were estimated using BIGMAP (Reinsch, 1999) based on marker genotype information. The midpoint of each marker bracket was considered a putative QTL position, thus a total of 25 putative QTL positions were investigated.

3.2.4.1 LINKAGE ANALYSIS (LA)

The model applied to the data for joint LA contained a random polygenic effect as well as a random additive QTL effect (Xu and Atchley, 1995, Hoeschele et al., 1997):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{v} + \mathbf{e} \quad [1]$$

where \mathbf{y} is a $(m \times 1)$ vector of phenotypic observations (DYD) for sires, \mathbf{X} is a $(m \times k)$ design matrix, \mathbf{b} is a fixed mean effect common to all sires, \mathbf{Z} is an $(m \times n)$ incidence matrix relating genotyped animals to phenotypes, \mathbf{u} is a $(n \times 1)$ vector of random polygenic effects, \mathbf{W} is an $(m \times gam)$ incidence matrix, \mathbf{v} is a $(gam \times 1)$ vector of random QTL effects and \mathbf{e} is a $(n \times 1)$ residual vector.

Expectations of the random effects \mathbf{u} , \mathbf{v} and \mathbf{e} and their covariances are assumed to be zero. Their distribution is assumed with the multivariate normal densities $\mathbf{u} \sim N(0, \mathbf{A}\sigma_u^2)$, $\mathbf{v} \sim N(0, \mathbf{G}_q\sigma_q^2)$ and $\mathbf{e} \sim N(0, \mathbf{D}\sigma_e^2)$, where scalar variances σ_u^2 , σ_q^2 and σ_e^2 describe the polygenic variance, the additive variance of the QTL and the residual variance, respectively; \mathbf{A} is the additive genetic relationship matrix; \mathbf{G}_q is the (co)variance matrix for the additive effects of the QTL conditional on marker haplotypes at a given position and the diagonal matrix \mathbf{D} containing 1 over the EDC (Liu et al., 2004). The following values are used as dimensions; m is the number of animals in the sub-pedigree of genotyped animals ($m = 1121$), k equals 1 in our study since only a common mean effect was fitted in the analysis, n is the number of animals in the full pedigree ($n = 8073$) and gam is the number of allelic effects with 1 effect per sire plus 2 times the number of grandsires.

The restricted log likelihood ratio test statistic ($RLRT^{Pos} = -2[\ln(L_0) - \ln(L_1^{Pos})]$) was calculated for each putative QTL position ($Pos=1-25$) using ASReml (Green et al., 1990), whereby L_1^{Pos} is the RLRT of model 1 for position Pos and L_0 is that of the same model without random gametic effects for a given QTL.

3.2.4.2 INCORPORATING LD USING VARIANCE COMPONENTS (LDLA-V)

Combined linkage and linkage disequilibrium analysis using variance component methodology was conducted using the approach described by Meuwissen et al. (2002), in which similar marker haplotypes have an increased probability of having identical QTL alleles. This method models (co)variances between individuals by deriving identical by descent (IBD) probabilities for QTL alleles carried by alternate marker haplotypes (Meuwissen and Goddard, 2000) and is dependant on the marker map. If a common ancestor occurs within the known pedigree, the IBD probabilities can be calculated from the marker information (i.e. LA). However, if no common ancestor exists, the probability of being IBD at the QTL is calculated from the similarity between the marker haplotypes (i.e. LD). The \mathbf{G}_q matrix in model 1 was supplemented with an IBD sub-matrix containing (co)variances

between founder marker haplotypes with no known ancestors. The dimension gam was then equal to $2(m) + 2$ times the number of grandsires. In this way, the LA model described above becomes a combined LDLA-V analysis when the IBD sub-matrices are included.

3.2.4.3 INCORPORATING LD USING REGRESSION ON MARKER ALLELES (LDLA-R)

Following Grapes et al. (2004) and Zhao et al. (2007), all 26 markers were tested individually for significance using the following model:

$$\mathbf{y} = \mathbf{XB} + \mathbf{Qm} + \mathbf{Zu} + \mathbf{e} \quad [2]$$

where \mathbf{m}_i is a $(a \times 1)$ vector of fixed allelic effects for a given marker, i , in which each marker allele is assigned an individual regression coefficient and \mathbf{Q} is a $(m \times a)$ known incidence matrix relating marker alleles to animals; a is the number of different alleles for i . All other effects are as described for model 1.

Markers showing a nominal significance level of 5% were identified and further analysed over the 26 putative QTL positions with model 3, which contains the significant marker (QTL) as a fixed effect (Fulker et al., 1999):

$$\mathbf{y} = \mathbf{XB} + \mathbf{Qm} + \mathbf{Zu} + \mathbf{Wv} + \mathbf{e} \quad [3]$$

with terms defined as in models 1 and 2. Once again, RLRT were calculated at each putative QTL position. The RLRT ($RLRT^{Pos} = -2[\ln(L_0^*) - \ln(L_2^{Pos})]$), was calculated, whereby L_2 is the RLRT of model 3 and L_0^* is that of the same model without random allelic effects at position Pos . The difference between L_0 and L_0^* lies in the inclusion of the fixed marker effect in L_0^* .

For all models discussed above, the (co)variance matrices for the additive effects of the QTL conditional on marker haplotypes were calculated by treating highly correlated haplotypes in the same way: if very close correlations between haplotypes occur, a predefined threshold (e.g. $\varepsilon = 0.03$) was given and all haplotypes with correlations above 0.97 were clustered together (Tuchscherer et al., 2004; Calus et al. 2008). Haplotype groups with such correlations are considered ‘non-unique’. The resulting condensed gametic relationship matrix can have between n and $2n$ rows and columns, whereby n is the number of genotyped animals in the pedigree plus 2 times the number of grandsires. This method prevents numerical difficulties due to quasi-linear dependencies which may occur due to the inclusion of almost identical rows and columns in the QTL (co)variance matrices. The matrices were computed

using a modified version of the Fortran 90 software COBRA (Baes and Reinsch, 2007) as part of the TIGER software (Baes and Reinsch, 2008)

For simplicity, significance thresholds were calculated in logarithm of odds (LOD) units, where a LOD score > 3 indicates genome-wide significance (Lander and Botstein, 1989). For linkage, the significance level required is traditionally set at a LOD score of 3 (Morton, 1955), which is equivalent to a RLRT value of 13.8 and equates to a nominal significance level of 0.0001 (Ott, 1991). The LOD drop-off method of Lander and Botstein (1989) was likewise employed to estimate an approximate CI, which can be calculated by identifying the highest point in the RLRT profile and subtracting 1 unit (i.e. an RLRT of 4.6). All positions within 1 LOD unit from the maximum on either side of this point are considered within the CI, which is asymptotically approximately equivalent to 96.8% (Mangin et al., 1994).

3.3. RESULTS

3.3.1 MAP AND MARKER INFORMATION

Microsatellite markers had an average of 6.38 alleles; a maximum of 11 alleles was observed for the marker MNB48 and a minimum of 3 alleles was observed for the marker CSSM53. The average marker interval including SNP was 3.14 cM with a maximum interval of 9.6 cM and a minimum interval of 0.01 cM between the markers. Two markers in our map (DIK2862 and BMS778) were calculated at interchanged positions in comparison to the MARC map, however the distance between these markers was only 0.486 in the MARC map and the number of informative meioses used in our calculations was more than twice that of the MARC map for both of these markers. The average information content for linkage was measured as the mean of $|1 - 2p_{ijk}|$ over all sires, where p_{ijk} is the transmitting probability of the putative QTL from grandsire i to sire j at marker interval k . The mean average information content over all marker intervals was 90.6%.

To achieve high data quality, all genotypes were recorded independently twice and checked for Mendelian inheritance and unlikely double recombinants. Suspicious genotypes were re-typed. After a second control, sires with conflicting genotypes were excluded from the analysis for the given marker. After data cleansing, a total of 28,275 genotypes remained for analysis, equating to 95.6% of all possible genotypes. Marker positions, the number of alleles, the number of informative meioses and the sire heterozygosity across markers are shown in Figure 1.

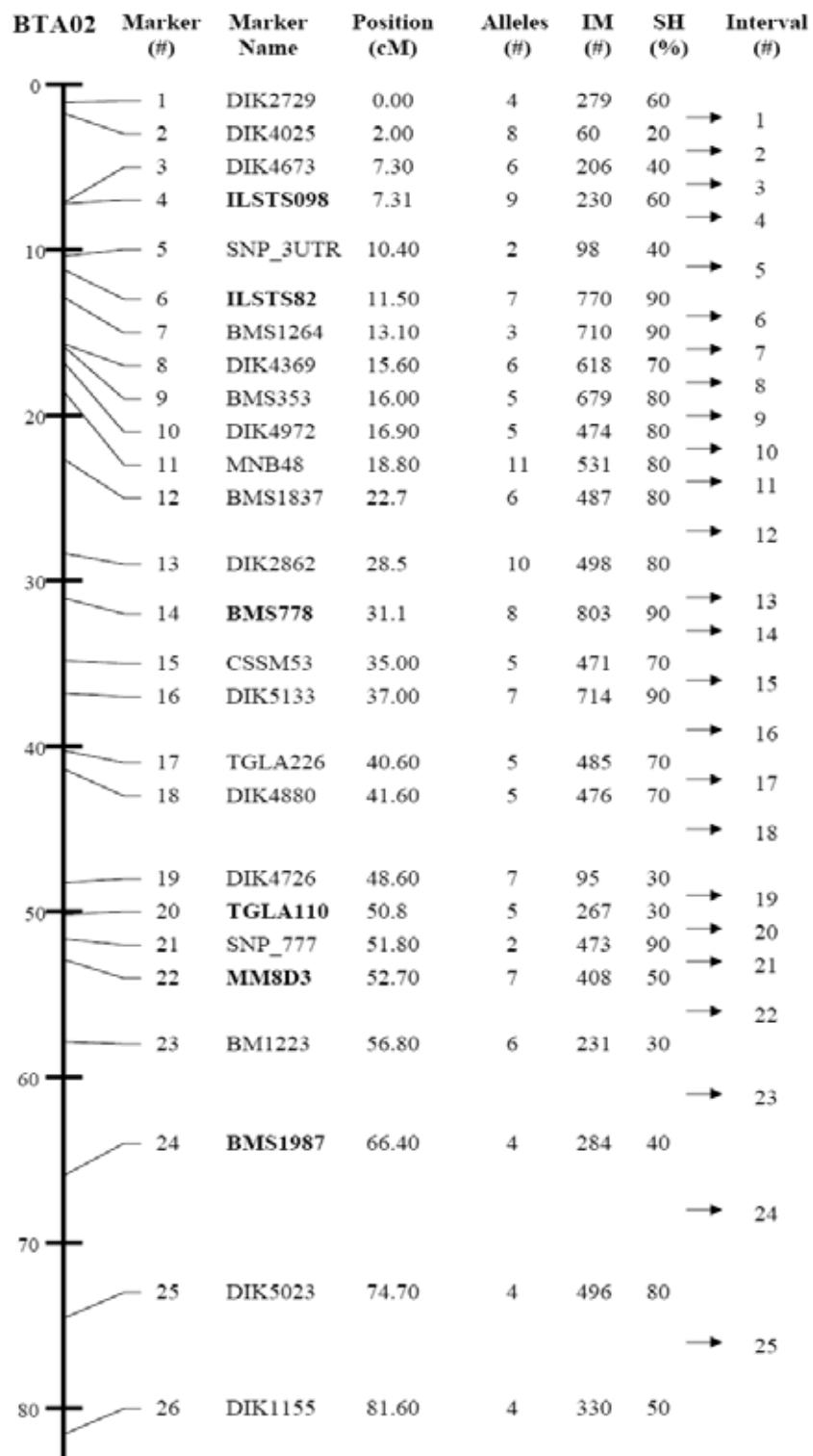


FIGURE 1.

Characteristics of microsatellite markers and SNPs used in this study. Markers are numbered from 1 to 26, with marker intervals from 1 to 25. IM = the number of informative meioses per marker, SH = sire heterozygosity per marker. Markers in bold-face print were included in the study of Bennewitz et al. (2003).

3.3.2 LA USING VARIANCE COMPONENTS

The RLRT profile of a model containing a QTL for SCS and random polygenic effects vs. a model containing only random polygenic effects is shown in Figure 2a. The RLRT profile is plotted for the putative QTL located at the midpoint between every pair of markers. The likelihood curve reached a global maximum RLRT of 20.29 at marker interval 13 between DIK2862 and BMS778. Twenty-two positions were significant at the chromosome-wide level, 15 positions also reached significance at a genome-wide level, and 12 positions were within a 1-LOD drop CI of the global maximum.

3.3.3 COMBINED LDLA-V

The analysis based on LD provided the likelihood profile shown in Figure 2b. The profile is completely flat on a very low level, indicating a complete lack of linkage disequilibrium over the entire area under consideration. The combined LDLA-V resulted in two main peaks, one of which surpassed both chromosome-wide and genome-wide significance thresholds (between the markers MNB48 and TGLA226 and between DIK4880 and DIK4726). Two further putative QTL positions also reached the chromosome-wide significance level (between the markers ILSTS098 and SNP_3UTR and between BM1223 and BMS1987). The main peaks were within the significant region identified using LA, however the RLRT in the combined LDLA-V analysis was lower than the RLRT-profile from LA alone at all positions.

3.3.4 COMBINED LDLA-R

Two markers (ILSTS098, $p \leq 0.05$ and BMS778, $p \leq 0.01$) showed a nominal significance level of at least 5% for model 2 and were further analysed with model 3. The resulting RLRT profiles are shown in Figure 3. The RLRT of model 3 and that of the same model without random allelic effects at position *Pos* was examined in order to see how much QTL variance could be explained by the inclusion of the individual markers. In contrast to a LA RLRT profile, the amount of QTL variance the single marker explains can be observed by the difference in RLRT profiles (this is more thoroughly explained later). The marker ILSTS098 explained only a very small portion of the QTL variance (Figure 3a), while BMS778 had a much greater effect (Figure 3b). A model including both markers as fixed effects plus an interaction between these markers was also applied, however the interaction between the two markers was not significant (data not shown).

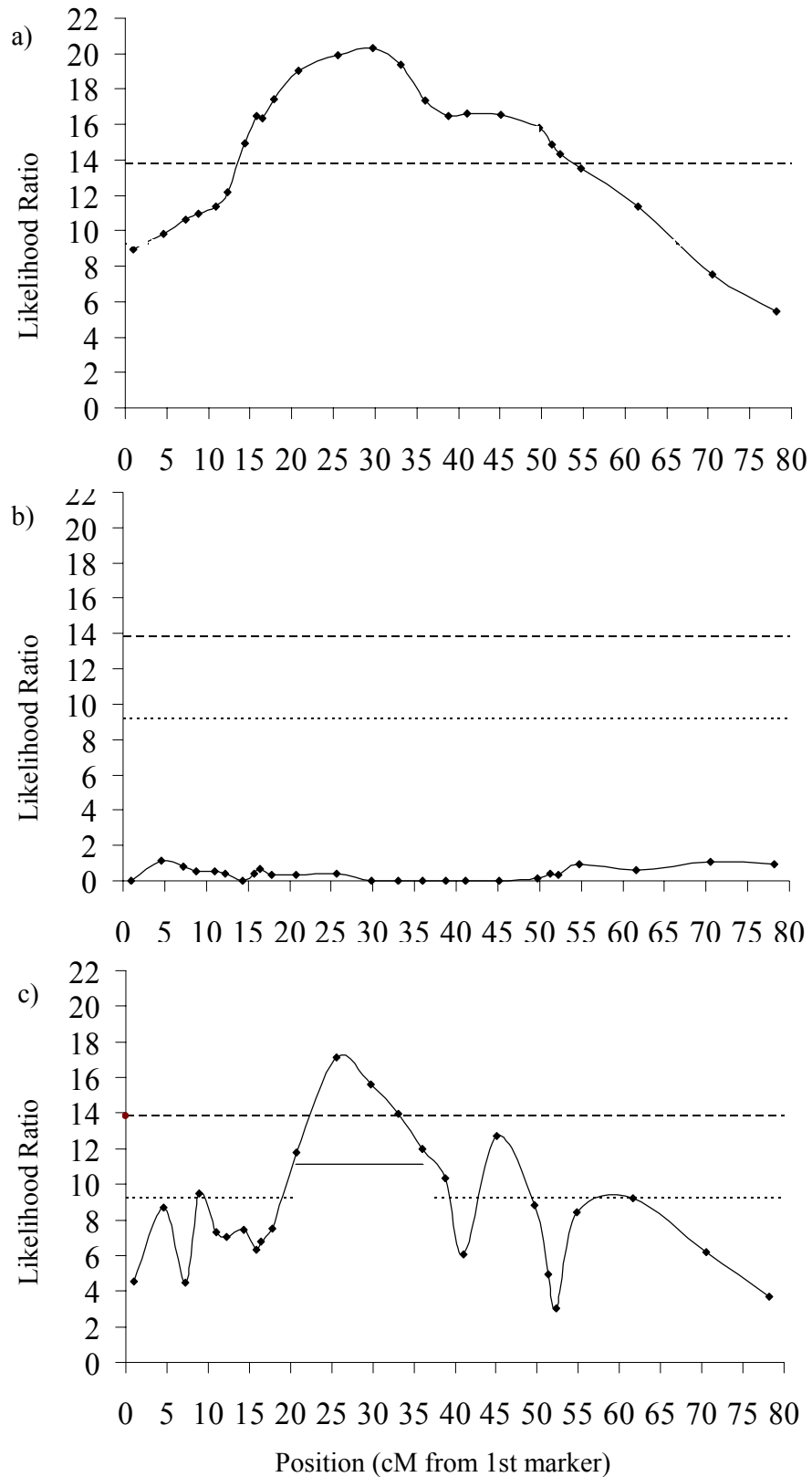


FIGURE 2. Restricted log-likelihood ratio of linkage analysis (a), linkage disequilibrium analysis (b) and combined linkage and linkage disequilibrium analysis (c). The dashed horizontal line denotes a genome-wide significance threshold; putative QTL positions are indicated with (•)

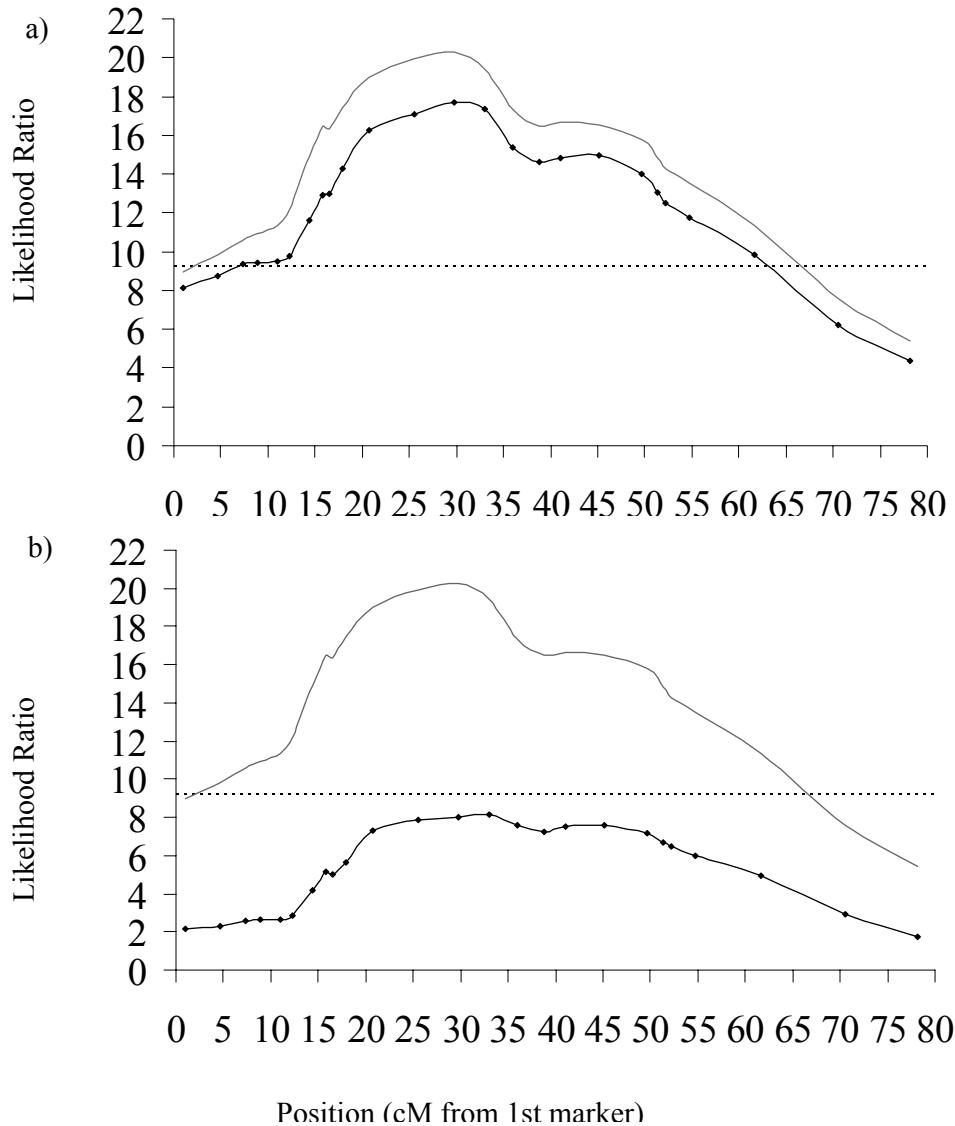


FIGURE 3. Restricted log-likelihood profiles of linkage analysis with regression on additive marker effects across families: ILSTS098 (a) and BMS778 (b). Grey profiles provide a reference to linkage analysis results; putative marker locations are indicated with (•). The dashed horizontal line denotes a chromosome-wide significance threshold.

3.4 DISCUSSION

As mentioned in the introduction, many QTL influencing SCS have been mapped in dairy cattle. However, confidence intervals for QTL locations are typically of the order of 20-40 cM, corresponding in mammals to ~20-40 million base pairs (Georges, 2007). On BTA02, the CI found by Bennewitz et al. (2003) covered approximately 80% (128 cM) of the entire

chromosome in the design of Thomsen et al (2000), and approximately 96% (151 cM) of the chromosome in the joint design (total map length was 158 cM). The number of genes within a CI of such length is expected to exceed 1000, and is, of course, far too high for effective use in the search for candidate genes. With an improved LA we first succeeded in reducing the confidence interval from 128 cM (Thomsen et al., 2000) to 38 cM. We were additionally able to identify two individual markers with a significant effect on SCS in the German Holstein.

3.4.1 LD VIA VARIANCE COMPONENT ESTIMATION

None of the putative QTL positions found in LA were significant in the LD analysis using variance components. Obviously the haplotypes we used in the variance component approach were not in LD with the QTL. The results of a simulation study let Grapes et al. (2004) postulate that similar IBD probability matrices for adjoining QTL positions in IBD-based LALD methods could cause insensitivity regarding QTL localization. Zhao et al. (2007) also came to the conclusion that the use of haplotype information does not necessarily provide the best precision for detecting QTL when using maps of limited density and when compared to single-marker regression. This is in concordance with our results, as significant LD could be found in more proximal regions, namely at ILSTS098 and BMS778, using single-locus marker regression.

3.4.2 COMBINED LDLA-V

In accordance with the completely flat shape and low level of the LD-profile the LDLA-V profile was completely enveloped by the LA-profile – all test statistics calculated using a combined variance component analysis were lower than those calculated using simple LA. The inserted correlation structure of haplotype effects was less adequate at all QTL-positions when the presence of LD was assumed additionally to LA. As often observed in similar studies, the profile calculated with combined LALD variance component analysis showed many more fluctuations along the chromosome than the LA profile. Upon closer inspection, a striking similarity between these fluctuations and the variability of the degree of certainty with which the allele-combination on the maternal haplotypes of sons could be determined within the window around each particular QTL-position was observed – though this certainty was on a high level. Therefore the local peaks of the LDLA-V profile do not provide any additional information on the precise QTL-position in our case. It is, however, interesting to note that the regression approach has the advantage of being independent of the correct resolution of genotypes into haplotypes.

A further important factor may be the definition of haplotype length. In contrast to the simulation studies of Meuwissen et al. (2002), Grapes et al. (2004, 2006) and Zhao (2007), we used chromosomal distance as a limiting window around the putative QTL position as opposed to a fixed number of markers. This was done to account for the uneven spacing of markers in our study. By using a window including all markers 5 cM to the left and 5 cM to the right of a given putative QTL locus, the number of markers in our haplotypes varied between 2 and 7 markers, depending on marker spacing in the chromosomal area. Analyses conducted using windows of various length did not alter the peak position (data not shown), which leads us to believe that the analysis is relatively robust with regards to haplotype length. The use of an equivalent number of SNP as opposed to microsatellites would have enabled a more even marker spacing, which is likely beneficial for the characterisation of haplotypes in the variance component-based LALD method. Varilo et al. (2002) compared the extent of inter-marker LD detected by microsatellite markers with that detected by SNP using data on various human subpopulations in Finland. Single, informative microsatellites were shown to provide more power to detect inter-marker LD than SNP, even when information from 3 to 5 SNP was combined. Therefore, one microsatellite may roughly be equivalent to 5 SNP and 100 SNP can be regarded equivalent to our microsatellite map.

3.4.3 COMBINED LDLA-R

Due to the inability of the variance component approach to find LD between marker haplotypes and putative QTL, a polygenic effect was added to the single-locus regression model of Grapes et al. (2004) to identify significant markers (model 2). This was done to correct for residual polygenic co-ancestry that might be correlated with the co-ancestry at the tested locus (Zhao et al., 2007). Markers found to be significant using model 2 were further analysed by incorporating LD using regression on marker alleles (model 3), which combines QTL LA and association analysis. A similar model was employed by Fulker et al. (1999), who used full sib pairs for simultaneous analysis of QTL-marker association and linkage for quantitative traits in humans. The major advantage of this model is that linkage and association parameters are modelled simultaneously; linkage being modelled within the \mathbf{G}_q matrix as the (co)variance structure of the random paternal haplotype effects of the sires and the association parameters modelled with regression coefficients on marker alleles. Determination of whether the marker is in close disequilibrium (or a candidate) or in weak disequilibrium with a trait locus can thereby be tested. If significant linkage is detected while modelling association, the putative locus is likely a locus in weak disequilibrium with the trait

locus but is not very close to the functional gene itself (Fulker et al., 1999). If linkage is observable before the association parameter is incorporated into the model, the extent that this linkage evidence is diminished by the inclusion of association into the model can be used to indicate the strength of the LD between the QTL and the marker. In effect, modeling of association parameters cancels out any variance accounted for by the alleles themselves. Ideally, the linkage evidence vanishes completely, as would be the case when the marker is the QTL itself. We can therefore conclude that the microsatellite ILSTS098 is only in weak disequilibrium with the QTL (Figure 3a) while BMS778 is in pronounced disequilibrium with the QTL because of the notable decrease in the RLRT profile (Figure 3b).

Several studies investigated associations between SNP_777 and SCS, (Youngerman et al., 2004; Leyva-Baca et al., 2007, 2008; Goertz et al., submitted). Results obtained with LA and regression based LD analysis support those of Goertz et al. (submitted), who found that no population-wide LD exists between this SNP and the trait SCS. This is not surprising, as the 3 families included in the study of Goertz et al. (submitted) were also included in our study. Nevertheless, SNP_777 is within the region where the test-statistic is above the genome-wide significance threshold in our LA (see Figure 2a).

This study was undertaken with the goal of further investigating the chromosomal area affecting SCS on BTA02 identified by Bennewitz et al. (2003) with regards to LD. Compared to the point of departure, the region of interest for future fine mapping experiments could be decisively narrowed to the vicinity of BMS778. Further investigation of this area using more evenly, densely-spaced SNP is required to identify a tighter disequilibrium with the QTL.

3.5 ACKNOWLEDGEMENTS

The authors gratefully acknowledge financial support from the Förderverein Biotechnologieforschung (FBF) and the German Ministry of Education and Research (BMBF) within the FUGATO M.A.S.-Net Project.

3.6 REFERENCES

- BAES, C. AND REINSCH, N.: Computing the condensed conditional gametic QTL relationship matrix and its inverse. *Arch. Tierz.* 50 (2007) 294-308.
- BAES, C. AND REINSCH, N.: TIGER: A software system for fine-mapping quantitative trait loci. *Arch. Tierz.* 51 (2008) 402-412.
- BENNEWITZ, J.; REINSCH, N.; GROHS, C.; LEVÉZIEL, H.; MALAFOSSE, A.; THOMSEN, H.; XU, N.; LOOFT, C.; KÜHN, C.; BROCKMANN, G.; SCHWERIN, M.; WEIMANN, C.; HIENDLEDER, S.; ERHARDT, G.; MEDJUGORAC, I.; RUSS, I.; FÖRSTER, M.; BREINIG, B.; REINHARDT, F.; REENTS, R.; AVERDUNK, G.; BLÜMEL, J.; BOICHARD, D. AND KALM, E.: 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 (2003) 319-338.
- BOICHARD, D.; GROHS, C.; BOURGEOIS, F.; CERQUEIRA, F.; FAUGERAS, R.; NEAU, A.; RUPP, R.; AMIGUES, Y.; BOSCHER, M.Y. AND LEVEZIEL, H.: Detection of genes influencing economic traits in three French dairy cattle breeds. *Genet. Sel. Evol.* 35 (2003) 77-101.
- CALUS, M.; MEUWISSEN, T.; DE ROOS, A. AND VEERKAMP, R.: Accuracy of Genomic Selection Using Different Methods to Define Haplotypes. *Genetics* 178 (2008) 553-561.
- COBANOGLU, O.; ZAITOUN, I.; CHANG, Y.; SHOOK, G. AND KHATIB, H.: Effects of the signal transducer and activator of transcription 1 (STAT1) gene on milk production traits in Holstein dairy cattle. *J. Dairy Sci.* 89 (2006) 4433-4437.
- DRUET, T.; FRITZ, S.; BOUSSAHA, M.; BEN-JEMAA, S.; GUILLAUME, F.; DERBALA, D.; ZELENKA, D.; LECHNER, D.; CHARON, C.; BOICHARD, D.; GUT, I.; EGGEN, A. AND GAUTIER, M.: Fine Mapping of Quantitative Trait Loci Affecting Female Fertility in Dairy Cattle on BTA03 Using a Dense Single-Nucleotide Polymorphism Map. *Genetics* 178 (2008) 2227-2235.
- FULKER, D.; CHERNY, S.; SHAM, P. AND HEWITT, J.: Combined Linkage and Association Sib-Pair Analysis for Quantitative Traits. *Am. J. Hum. Genet.* 64 (1999) 259-267.
- GEORGES, M.: Mapping, Fine Mapping, and Molecular Dissection of Quantitative Trait Loci in Domestic Animals. *Annu. Rev. Genomics Hum. Genet.* 8 (2007) 131-162.
- GRAPES, L.; DEKKERS, J.; ROTHSCHILD, M. AND FERNANDO, R.: Comparing linkage disequilibrium-based methods for fine mapping quantitative trait loci. *Genetics* 166 (2004) 1561-1570.
- GRAPES, L.; FIRAT, M.; DEKKERS, J.; ROTHSCHILD, M. AND FERNANDO, R.: Optimal Haplotype Structure for Linkage Disequilibrium-Based Fine Mapping of Quantitative Trait Loci Using Identity by Descent. *Genetics* 172 (2006) 1955-1965.
- GREEN, P.; FALLS, K. AND CROOKS, S.: Documentation for CRIMAP, version 2.4. (1990) Washington University School of Medicine, St. Louis, MO, USA.
- HOESCHELE, I.; UIMARI, P.; GRIGNOLA, F.; ZHANG, Q. AND GAGE, K.: Advances in statistical methods to map quantitative trait loci in outbred populations. *Genetics* 147 (1997) 1445-1457.

- KHATKAR, M.S.; THOMSON, P.C.; TAMMEN, I. AND RAADSMA, H.W.: Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet. Sel. Evol.* 36 (2004) 163–190.
- LANDER, E. AND BOTSTEIN, D.: Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps. *Genetics* 121 (1989) 185-199.
- LEYVA-BACA, I.; SCHENKEL, F.; MARTIN, J. AND KARROW, N.A.: Polymorphisms in the 5' upstream region of the CXCR1 chemokine receptor gene, and their association with somatic cell score in Holstein cattle in Canada. *J. Dairy Sci.* 91 (2008) 407-17.
- LEYVA-BACA, I.; SCHENKEL, F.; SHARMA, B.S.; JANSEN, G.B. AND KARROW, N.A.: Identification of single nucleotide polymorphisms in the bovine CCL2, IL8, CCR2 and IL8RA genes and their association with health and production in Canadian Holsteins. *Anim. Genet.* 38 (2007) 198-202.
- LIU, Z.; REINHARDT, F.; BÜNGER, A. AND REENTS, R.: Application of a random regression model to genetic evaluation test-day yields and somatic cell scores in dairy cattle. *Interbull Bull.* 27 (2001) 159-166.
- LIU, Z.; REINHARDT, F.; BÜNGER, A. AND REENTS, R.: Derivation and Calculation of Approximate Reliabilities and Daughter Yield-Deviations of a Random Regression Test-Day Model for Genetic Evaluation of Dairy Cattle. *J. Dairy Sci.* 87 (2004) 1896-1907.
- MANGIN, B.; GOFFINET, B. AND REBAÏ, A.: Constructing confidence intervals for QTL location. *Genetics* 138 (1994) 1301–1308.
- MEUWISSEN, T. AND GODDARD, M.: Fine Mapping of Quantitative Trait Loci Using Linkage Disequilibria With Closely Linked Marker Loci . *Genetics* 155 (2000) 421-430.
- MEUWISSEN, T.; KARLSEN, A.; LIEN, S.; OLSAKER, I. AND GODDARD, M.: Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* 161 (2002) 373-379.
- MORTON, N.: Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* 7 (1955) 277–318.
- OTT, J.: *Analysis of Human Linkage*. John Hopkins University Press. (1991) Baltimore USA.
- REINSCH, N.: A multiple-species, multiple-project database for genotypes at codominant loci. *J. Anim. Breed. Genet.* 116 (1999) 425-435.
- THOMSEN, H.; REINSCH, N.; XU, N.; LOOFT, C.; GRUPE, S.; KÜHN, C.; BROCKMANN, G.; SCHWERIN, M.; LEYHE-HORN, B.; HIENDLEDER, S.; ERHARDT, G.; MEDJUGORAC, I.; RUSS, I.; FÖRSTER, M.; BRENIG, B.; REINHARDT, F.; REENTS, R.; BLÜMEL, J.; AVERDUNK, G. AND KALM, E.: A male bovine linkage map for the ADR granddaughter design. *J. of Anim. Breed. Genet.* 117 (2000) 289-306.
- TUCHSCHERER, A.; MAYER, M. AND REINSCH, N.: Identification of gametes and treatment of linear dependencies in the gametic QTL-relationship matrix and its inverse. *Genet. Sel. Evol.* 36 (2004) 621-642.

VARILO, T.; PAUNIO, T.; PARKER, A.; PEROLA, M.; MEYER, J.; TERWILLIGER, J. AND PELTONEN, L.: The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. *Hum. Mol. Genet.* 12 (2003) 51-59.

XU, S. AND ATCHLEY, W.: A random model approach to interval mapping of quantitative genes. *Genetics* 141 (1995) 1189-1197.

YOUNGERMAN, S.M.; SAXTON, A.M.; OLIVER, S.P. AND PIGHETTI, G.M.: Association of CXCR2 polymorphisms with subclinical and clinical mastitis in dairy cattle. *J. Dairy Sci.* 87 (2004) 2442-2448.

ZHAO, H.H.; FERNANDO, R.L. AND DEKKERS, J.C.M.: Power and Precision of Alternate Methods for Linkage Disequilibrium Mapping of Quantitative Trait Loci. *Genetics* 175 (2007) 1975-1986.

CHAPTER FOUR:

REFINED POSITIONING OF A QUANTITATIVE TRAIT LOCUS AFFECTING SOMATIC
CELL SCORE ON CHROMOSOME 18 IN THE GERMAN HOLSTEIN USING LINKAGE
DISEQUILIBRIUM

Christine Baes^{1†}, Bodo Brand^{2†}, Manfred Mayer¹, Christa Kühn², Zenting Liu³, Friedrich
Reinhardt³, Norbert Reinsch¹

¹*Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere, Forschungsbereich
Genetik und Biometrie*

²*Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere, Forschungsbereich
Molekularbiologie*

³*Vereinigte Informationssysteme Tierhaltung w.V.*

[†]*Both authors contributed equally to this article*

Submitted

ABSTRACT

Combined linkage and linkage disequilibrium analysis was conducted to more accurately map a previously reported quantitative trait locus (QTL) affecting somatic cell score on bovine chromosome 18. A granddaughter design consisting of 6 German Holstein grandsire families with 1,054 progeny tested genotyped sons was used in this study. Twenty microsatellite markers, 5 single nucleotide polymorphisms and an erythrocyte antigen marker with an average marker spacing of 1.95 cM were analysed along a chromosomal segment of 50.80 cM. Variance components were estimated and restricted maximum likelihood test statistics were calculated at the midpoint of each marker interval. The test statistics calculated in single-QTL linkage analysis exceeded the genome-wide significance threshold at several putative QTL positions. Using combined linkage and linkage disequilibrium analysis (LALD) we were successful in assigning a genome-wide significant QTL to a confidence interval of 10.8 cM between the markers ILSTS002 - BMS833. The QTL in this marker interval was estimated to be responsible for between 5.89 and 13.86 % of the genetic variation in somatic cell score. In contrast to the single-QTL linkage analysis model, LALD analyses with a 2-QTL model confirmed the position of 1 QTL, but gave no conclusive evidence for the existence or position of a second QTL. Ultimately, the QTL position was narrowed down considerably compared to previous results due to the refined confidence interval of less than 11 cM.

KEYWORDS

mastitis / somatic cell score / linkage disequilibrium / BTA18

4.1 INTRODUCTION

Mastitis is one of the most persistent and expensive problems in the dairy production industry (Seegers et al., 2003). Somatic cell score (SCS) is commonly used as an indicator trait for mastitis and is implemented in routine sire evaluation in many countries (Rupp and Boichard, 2003). Although management practices play the most important role in mastitis prevention, the heritability for SCS has been estimated as high as 16-17% (e.g. Liu et al., 2004). Because of the unfavourable correlation between SCS and milk production traits (Rupp and Boichard, 2003), SCS has proven difficult to improve genetically. The opportunity to address shortcomings in conventional selection for SCS, low selection intensity and the low reliability of estimated breeding values for SCS at an early age makes this trait a prime candidate for use in marker assisted selection (MAS) programs.

A number of whole genome scans have been conducted in the search for QTL underlying the genetic variance of SCS. Over 50 independent regions have been found in various populations on almost all bovine chromosomes (see Khatkar et al. (2004) for a review). The telomeric end of the *Bos taurus* autosome (BTA) 18 has been repeatedly identified as a likely region for harbouring significant SCS QTL in various populations, including the United States Holstein (Ashwell et al. 1997; Rodriguez-Zas et al., 2002), Finnish Ayrshire (Schulman et al., 2004; Lund et al., 2007), Swedish Red and White (Holmberg and Andersson-Eklund, 2004; Lund et al., 2007), Swedish Holstein (Holmberg and Andersson-Eklund, 2004), Danish Red (Lund et al., 2007) and German Holstein (Brink, 2003; Kühn et al., 2003; Xu et al., 2006). Despite these efforts to pinpoint the specific chromosomal region affecting phenotypic variation in SCS, the causal mutation has not yet been identified. The existence of more than 1 QTL affecting SCS in the telomeric part of BTA18 has also been suggested (Brink, 2003, Xu et al., 2006).

The efficient implementation of MAS requires identification of marker(s) tightly linked to QTL, or discovery of the specific causative mutation. An initial family-based MAS pilot project proved the usefulness of existing marker information regarding selection on the QTL for SCS on BTA18 in the German Holstein population (Kühn et al., 2008). Nevertheless, there is a need to narrow down the QTL position on BTA18 to increase the efficiency of MAS and to provide a reasonable basis for the search for causal mutations. The purpose of this study was to further investigate the chromosomal regions affecting udder health on BTA18 in the German Holstein population. Marker density was considerably increased in comparison with previous studies (Brink, 2003; Kühn et al., 2003; Xu et al., 2006), and statistical models for single and 2-QTL analysis incorporating linkage disequilibrium methods were used to

improve localization of the QTL within the context of a national research effort initiated by the German Förderverein Biotechnologieforschung (FBF).

4.2 MATERIALS AND METHODS

4.2.1 FAMILIES

Six paternal half-sib families comprising a total of 1,054 genotyped bulls were analysed. These animals are a subset of the German Holstein population and originate from the granddaughter designs used in the studies of Kühn et al. (2003) and Xu et al. (2006) which were conducted as part of the national QTL mapping project of German breeding organisations, several German animal breeding institutes and animal computing centers. The number of bulls in each family ranged from 60 to 353, with an average of 175 bulls per family. A full pedigree including non-genotyped ancestors of genotyped animals was also available for estimating random polygenic effects (7,627 animals). Pedigree information was provided by the German genetic evaluation center (VIT) in Verden, Germany.

4.2.2 PHENOTYPES

The VIT provided phenotype information for genotyped animals, with the daughter yield deviation (DYD) used as the phenotypic unit of measurement for SCS. The DYD for each bull were calculated as the average of the daughters' performance adjusted for the fixed and non-random effects of the daughters and the genetic effects of the bull's mates following Liu et al. (2004). Compared to estimated breeding values, DYD provide a more independent measure of phenotypic performance, because DYD describe the genetic value of the sires more accurately due to the adjustment of the daughters dam (Bennewitz et al., 2004). The DYD were obtained from the official release of the January 2008 genetic evaluation and were calculated based on a random regression test day model (Liu et al., 2004). The DYD for the first lactation were analysed, as the number of offspring was highest in this parity.

Reliabilities associated with the DYD used in this study were obtained from the official release of the January 2008 genetic evaluation and were expressed as the number of effective daughter contributions (EDC) following Liu et al. (2004).

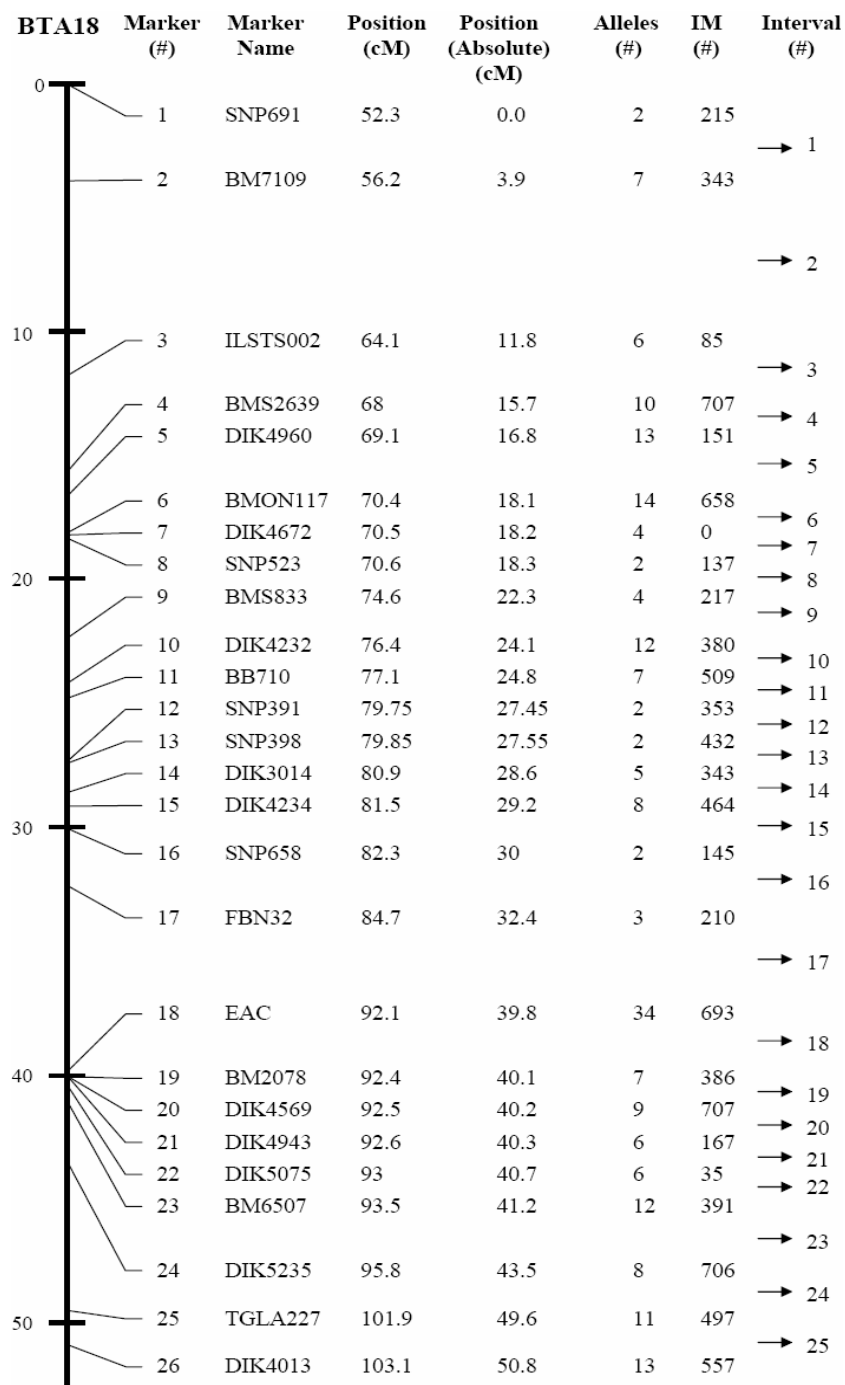


FIGURE 1.
 Marker map for the middle – telomeric region of *Bos taurus* autosome 18, including marker number, marker name, marker position in cM, absolute marker position (cM), number of alleles, number of informative meioses used for calculating the marker map and the interval number, which corresponds to the putative QTL positions.

4.2.3 MARKER SET AND GENOTYPING

The marker set (Figure 1) included a total of 26 markers (15 newly selected microsatellite markers, 5 single nucleotide polymorphisms (SNP) and 6 markers that were genotyped within previous QTL mapping studies, 1 of which is a erythrocyte antigen (Kühn et al., 2003; Xu et al., 2006 and Brink, 2003)). The 20 newly selected microsatellite and SNP markers were chosen based on the information of the putative QTL positions for somatic cell score on BTA18 reported by Brink and Kühn et al. in 2003 and Xu et al. in 2006. For the selection of microsatellite markers the bovine linkage map of the United States Department of Agriculture's Meat Animal Research Center (MARC USDA) was used (<http://www.marc.usda.gov/genome/genome.html>). DNA of all sires was isolated from sperm samples provided by German Artificial Insemination stations. Genotyping of the microsatellite markers was performed by PCR or Multiplex-PCR with fluorescence (FAM, HEX, TET) labelled primers and a subsequent fragment length analysis using the MegaBACE™1000 DNA Analysis System and MegaBACE Fragment Profiler Version 1.2 software (GE Healthcare, Munich, Germany). The SNP were genotyped by standard RFLP techniques (Seyfert and Kühn, 1994). Thus, a PCR specific for the SNP and its flanking region was established and after amplification and restriction digest the fragments were analysed by agarose gel electrophoresis. To achieve high data quality, all genotypes were recorded independently twice and checked for Mendelian inheritance and unlikely double recombinants. Suspicious genotypes were re-typed. After a second control, sires with conflicting genotypes were excluded from the analysis for the given marker.

4.2.4 LINKAGE MAP

To evaluate the most likely marker order on BTA18 information from published linkage-maps (Ihara et al., 2004; Snelling et al., 2005), RH-maps (Itoh et al. 2005), human and bovine sequence assemblies (NCBI Map Viewer; <http://www.ncbi.nlm.nih.gov/mapview/>) as well as own linkage-mapping results were used to define a refined marker order. The marker order in the current bovine sequence assembly Btau4 is in accordance with our marker order, whereas there are several discrepancies to the previous bovine sequence assembly Btau3.1. The genetic linkage map was calculated based on our genotyping data using CRIMAP software (Green et al., 1990). For some markers (BMON117, DIK4672, SNP523; SNP391, SNP398; BM2078, DIK4569, DIK4943) a recombination rate of 0 was calculated. A marker spacing of 0.1 cM was used in these cases to alleviate technical difficulties arising in the calculation of

transmitting probabilities; the correct marker order was ensured through sequence information mentioned above.

4.2.5 TRANSMISSION PROBABILITIES

The most probable linkage phases of genotyped sires were calculated using all marker genotypes in order to determine paternally and maternally inherited haplotypes. Sire transmission probabilities were calculated at the midpoint of each marker interval for later determination of the gametic relationship matrices. Both transmission probabilities and marker allele frequencies were calculated using BIGMAP (Reinsch, 1999). The midpoint of each marker bracket was considered a putative QTL position, resulting in a total of 25 putative QTL positions.

4.2.6 LINKAGE ANALYSIS

A model containing a random polygenic effect and a random additive QTL effect (Xu and Atchley, 1995; Hoeschele et al., 1997) was applied to the data for linkage analysis:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{v} + \mathbf{e} \quad [1]$$

where $\mathbf{y}_{(m \times 1)}$ is a vector of phenotypic observations (DYD) for sires, $\mathbf{X}_{(m \times k)}$ is a design matrix, $\boldsymbol{\beta}_{(1 \times 1)}$ is a fixed mean common to all observations, $\mathbf{Z}_{(m \times n)}$ is an incidence matrix relating animals to phenotypes, $\mathbf{u}_{(n \times 1)}$ is a vector of random polygenic effects, $\mathbf{W}_{(m \times gam)}$ is an incidence matrix, $\mathbf{v}_{(gam \times 1)}$ is a vector of random QTL effects and $\mathbf{e}_{(m \times 1)}$ is a residual vector. Subscripts in parenthesis of the vectors and matrices denote their dimensions; m is the number of animals in the sub-pedigree of genotyped animals, k equals 1 in our study for the common mean, n is the number of animals in the extended pedigree and gam is the number of unique allelic effects. Random effects \mathbf{u} , \mathbf{v} and \mathbf{e} and their covariances have an expected value of zero. Their distribution is assumed with the multivariate normal densities: $\mathbf{u} \sim N(0, \mathbf{A}_p \sigma_p^2)$, $\mathbf{v} \sim N(0, \mathbf{G}_q \sigma_q^2)$ and $\mathbf{e} \sim N(0, \mathbf{D} \sigma_e^2)$, where σ_p^2 , σ_q^2 and σ_e^2 describe the polygenic variance, the additive variance of the QTL, and the residual variance, respectively; \mathbf{A}_p is the additive genetic relationship matrix; \mathbf{G}_q is the covariance matrix for the additive effects of the QTL conditional on marker haplotypes within a 10 cM window around a given position with haplotypes sorted according to their generation (i.e. founder haplotypes followed by

haplotypes with known origin) and the diagonal matrix **D** containing 1 over the EDC (Liu et al., 2004).

The covariance matrices (\mathbf{G}_q) for the additive effects of the QTL conditional on marker haplotypes in all analyses were calculated using the condensing algorithm of Tuchscherer et al. (2004), which has an effect similar to that of haplotype clustering (e.g. Calus et al., 2008). Haplotype groups are clustered if their correlation is high (>0.97); such haplotypes are considered 'non-unique'. The dimension of the resulting condensed gametic relationship matrices are no longer twice the number of genotyped animals, but contain 1 row and column for every unique haplotype effect. This method prevents quasi-linear dependencies which can occur when nearly identical rows and columns in the QTL covariance matrices occur.

The QTL test performed in all analyses was a likelihood ratio test, where the REML of the full model was compared with the REML of the model missing the QTL effect. The restricted log likelihood ratio test statistic ($RLRT^{Pos} = -2[\ln(L_0) - \ln(L_1^{Pos})]$) was calculated for each putative QTL position ($Pos=1$ to 25), whereby L_1^{Pos} is the REML of the full model (model 1) for position Pos and L_0 is that of the same model without random gametic effects for a given QTL. The additive relationship matrix \mathbf{A}_p was the same for both models and included all relevant animals in the pedigree (7,627 animals). All calculations were done with a combination of own programs (Baes and Reinsch, 2007) and the ASReml package (Gilmour et al., 2006) as described in Baes and Reinsch (2008).

4.2.7 INCORPORATING LINKAGE DISEQUILIBRIUM WITH VARIANCE COMPONENTS

In addition to the LA analysis, an LD analysis and a combined LA and LD analysis (LALD) was performed. Both analyses involving LD are based on the assumption that similar marker haplotypes are more likely to have identical QTL alleles (Meuwissen et al., 2002). Covariance between founder haplotypes (i.e. maternal haplotypes of bulls and maternal and paternal haplotypes of their sires) was modelled by deriving the probability that QTL alleles carried by different marker haplotypes are identical by descent (IBD) as described by Meuwissen et al. (2002). If a common ancestor occurs within the known pedigree, IBD probabilities are calculated from the marker information (i.e. as in LA). However, if no common ancestor exists (e.g. founder haplotypes), the probability of being IBD at the QTL is estimated from the similarity between founder marker haplotypes. In this way, IBD sub-matrices containing the

covariances between founder haplotypes were used to supplement the \mathbf{G}_q matrix with LD information.

IBD sub-matrices were estimated for every putative QTL position in our study. The IBD probabilities contained in these matrices was used to fill the rows and columns corresponding to the respective founder haplotypes in the subsequent \mathbf{G}_q matrices. The IBD probability for a given haplotype combination was therefore considered equivalent to the covariance between the specific haplotypes.

In a simulation study with evenly spaced markers, Grapes et al. (2006) found that fitting a 4-6 marker haplotype as a sliding window across the region resulted in the greatest accuracy compared to that of the other haplotype sizes tested. For practical applications, however, it may be advantageous to define the size of the window surrounding the putative QTL in cM instead of the number of markers (i.e. when markers are unevenly spaced). In this study a haplotype was defined as the marker combination within a 10 cM window (5 cM left and right of the putative QTL position), which equated to an average haplotype size of 6 markers and ensured that each window contained at least 1 marker on each side of the putative QTL. We tried other window sizes (5 cM and 20 cM), however the results did not differ markedly in these analyses.

One IBD-submatrix for every putative QTL position was formed, resulting in 25 different IBD submatrices. These matrices were tested for positive definiteness; matrices with negative eigenvalues were appropriately adjusted using the BENDPDF program (Eβl, 1991). Bending was applied to 8 of 25 IBD submatrices. The average bending factor was 0.077. The \mathbf{G}_q matrix in model 1 was supplemented with the IBD sub-matrices containing the covariances between founder marker haplotypes with no known ancestors as described by Meuwissen et al (2002). In this way, the LA model described in the previous section becomes a combined LALD model when the IBD sub-matrices are included and the \mathbf{G}_q matrix becomes 100% dense. The LD analysis was carried out similarly to the combined analysis, however information on marker transmission from parents to offspring was ignored by setting paternal transmitting probabilities to 0.5.

4.2.8 2-QTL MODEL

A 2-QTL model was calculated using a similar model:

$$\mathbf{y} = \mathbf{XB} + \mathbf{Zu} + \mathbf{W}_1\mathbf{v}_1 + \mathbf{W}_2\mathbf{v}_2 + \mathbf{e} \quad [2]$$

where $\mathbf{W}_{i(m \times gam)}$ is an incidence matrix and $\mathbf{v}_{i(gam \times 1)}$ is a vector of random QTL effects for the putative QTL i ($i = 1-2$). Dimensions and other effects are as in model 1. The RLRT ($RLRT^{Pos} = -2[\ln(L_0) - \ln(L_1^{Pos})]$) was calculated for each pair of putative QTL positions.

4.2.9 STATISTICAL INFERENCE

Significance thresholds were calculated as RLRT-equivalents of logarithm of odds (LOD) units, where a LOD score > 2 indicates chromosome-wide significance and a LOD score > 3 indicates genome-wide significance (Lander and Botstein, 1989). A LOD score > 2 is equivalent to a RLRT value of 9.2 and a LOD score > 3 is equivalent to a RLRT value of 13.8 (Ott, 1991, pp. 65). Likewise, the LOD drop-off method of Lander and Botstein (1989) was employed to estimate approximate confidence intervals, which can be calculated by identifying the highest point in the RLRT profile and subtracting 1 LOD unit (equivalent difference in RLRT = 4.6); all positions within 1 LOD unit from the maximum are considered within the confidence interval. Asymptotically, these are approximately equivalent to a 96 % confidence interval (Mangin et al., 1994).

4.3 RESULTS

4.3.1 MARKERS AND MAPS

Marker spacing ranged from 0.10 to 7.9 cM, with an average marker spacing of 1.95 cM. The markers had an average of 8.19 alleles, ranging from 2 for SNP markers to 34 for the erythrocyte antigen marker EAC. Average marker heterozygosity was 0.59 across all markers (range of 0.19 to 0.91). The number of informative meioses for markers and the number of alleles per marker is shown in Figure 1. There was an average of 365 informative meioses per marker with a range from 0 to 707.

4.3.2 LINKAGE ANALYSIS

The RLRT profile of linkage analysis conducted using a model containing a single QTL for SCS and random polygenic effects vs. a model containing only random polygenic effects is shown in Figure 2 a). The profile is plotted for putative QTL located at the midpoint of every marker interval. Putative QTL positions within a 1-LOD drop-off interval are shown in Table 1. The global maximum was located in marker interval 4 at position 68.55 cM (RLRT =

19.53) between BMS2639 and DIK4960, followed closely by interval 5 between DIK4960 and BMON117 at position 69.75 cM (RLRT = 19.34).

TABLE 1.

Marker intervals exceeding the 1-LOD drop-off threshold using linkage analysis (LA) and combined linkage and linkage disequilibrium analysis (LALD) for somatic cell score QTL on BTA 18 in the German Holstein. Restricted likelihood ratio test statistics as well as polygenic variance, QTL variance, and the ratio of QTL / genetic variance values are shown.

Method	Interval	Position (cM)	Marker 1 ³	Marker 2 ⁴	RLRT ⁵	Polygenic Variance	QTL Variance	QTL / Genetic Variance	
LA ¹	3	66.05	ILSTS002	BMS2639	17.734	0.1681	0.03071	15.45%	
	4	68.55	BMS2639	DIK4960	19.53	0.1661	0.03287	16.52%	
	5	69.75	DIK4960	BMON117	19.344	0.1624	0.03729	18.68%	
	6	70.45	BMON117	DIK4672	18.266	0.1669	0.03236	16.24%	
	7	70.55	DIK4672	SNP523	18.328	0.1665	0.03283	16.47%	
	8	72.6	SNP523	BMS833	18.14	0.1654	0.03374	16.94%	
	9	75.5	BMS833	DIK4232	16.044	0.1664	0.03332	16.68%	
	LALD ²	3	66.05	ILSTS002	BMS2639	19.788	0.1807	0.02415	11.79%
		4	68.55	BMS2639	DIK4960	18.888	0.1892	0.01185	5.89%
5		69.75	DIK4960	BMON117	20.184	0.1878	0.01223	6.11%	
6		70.45	BMON117	DIK4672	21.616	0.1851	0.01318	6.65%	
7		70.55	DIK4672	SNP523	21.89	0.1850	0.01423	7.14%	
8		72.6	SNP523	BMS833	21.784	0.1778	0.02860	13.86%	

¹ LA = Linkage Analysis

² LALD = Combined linkage analysis + linkage disequilibrium analysis

³ Marker 1 = first marker of marker interval

⁴ Marker 2 = second marker of marker interval

⁵ RLRT = Restricted Likelihood Test Ratio statistic

4.3.3 LINKAGE DISEQUILIBRIUM AND COMBINED ANALYSIS

The analysis based on linkage disequilibrium provided the likelihood profile shown in Figure 2 b). One distinct peak between markers SNP523 and BMS833 at marker interval 8 can be observed, although the RLRT at this position does not reach the significance threshold (RLRT = 7.86). The likelihood profile of the combined linkage and linkage disequilibrium analysis (Figure 2 c)) is similar to that of the linkage analysis, although the maximum test statistics are slightly higher in the combined analysis. Maxima were located in marker interval 7 at position 70.55 cM (RLRT = 21.89) and interval 8 at position 72.55 cM (RLRT = 21.78). A 1-LOD drop-off interval was calculated, which encompasses marker intervals 3 to 8 (10.50 cM total length). Putative QTL positions within the 1-LOD drop-off interval are shown in Table 1. The

ratio of QTL variance to total genetic variance was estimated between 5.89 % and 13.86 % of the additive genetic variation in somatic cell score in the German Holstein in this region.

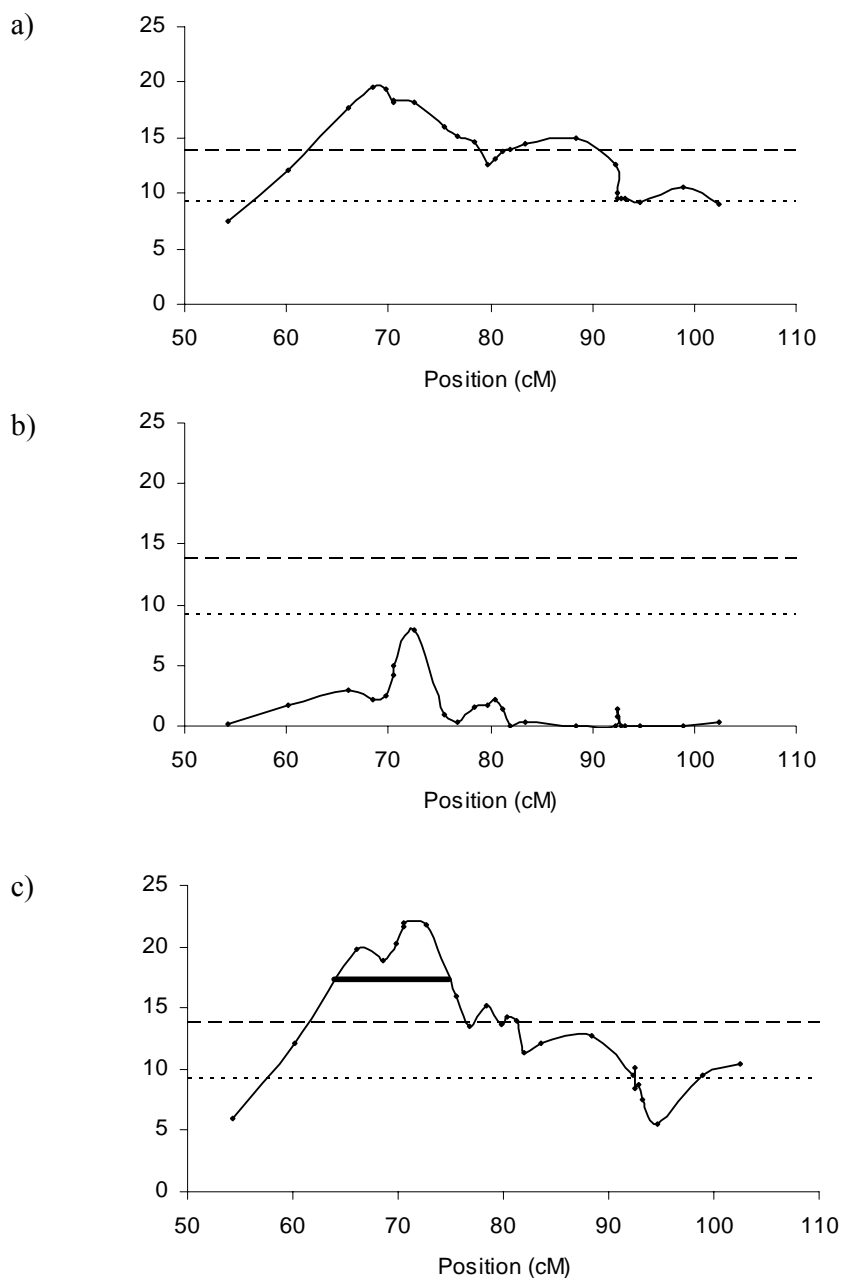


FIGURE 2.

Restricted Log Likelihood Ratio profiles for linkage analysis (a), linkage disequilibrium analysis (b) and combined linkage and linkage disequilibrium analysis (c) for a putative QTL affecting somatic cell score on *Bos taurus* autosome 18. The lower dashed line (---) indicates the chromosome-wide significance threshold, the upper dashed line (—) indicates the genome-wide significance threshold. The thick black line in (c) indicates an approximate 96.8% confidence interval.

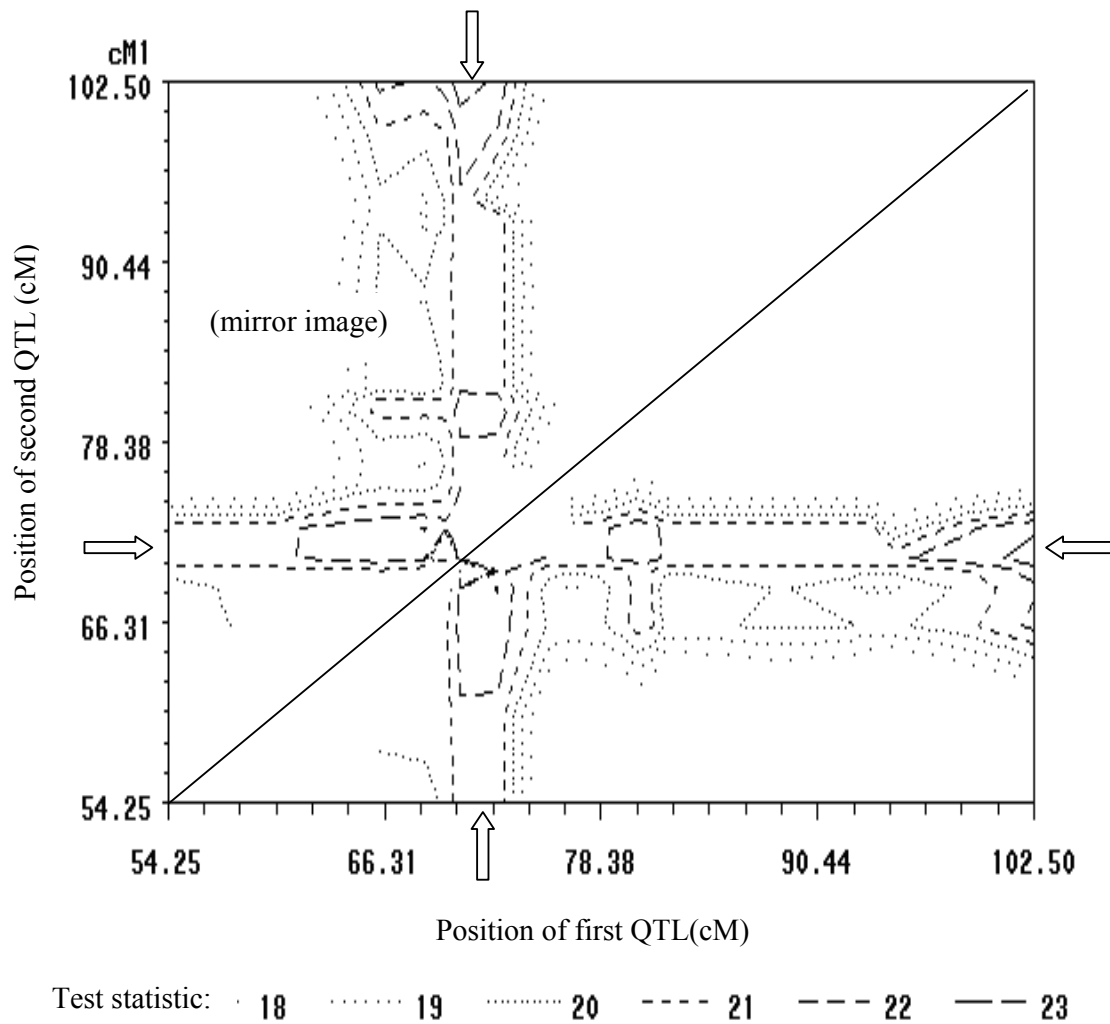


FIGURE 3.

Contour plot of the Restricted Log Likelihood Ratio profile using combined linkage and linkage disequilibrium analysis for a 2-QTL model affecting somatic cell score on *Bos taurus* autosome 18. QTL-positions in cM are given on the x- and y-axis, each point on the plane represents a particular combination of QTL-positions (for better readability each point is mirrored across the diagonal). Combinations with equal height of the test-statistics are connected by contour lines, where dark lines are used to indicate the top levels of the surface. A clear likelihood ridge is seen between positions 70.55 - 72.55 cM (see arrows).

4.3.4 2-QTL COMBINED LINKAGE AND LINKAGE DISEQUILIBRIUM ANALYSIS

A contour plot of the likelihood profile for an analysis using a 2-QTL LALD model is shown in Figure 3. The QTL located between markers DIK4672 and BMS833 is again obvious due to the clear 'likelihood ridge' between positions 70.55 and 72.55 cM. This ridge is visible in Figure 3 as dark horizontal and nearly parallel contour lines. Test statistics for the marker

intervals 7 and 8 exceed the genome-wide significance threshold for all possible putative second QTL positions. The highest likelihood was found for the QTL pair at intervals 7 and 25 (RLRT statistic = 23.51), however the RLRT statistics at interval 6 and 25 and intervals 8 and 3, 4, 6, 7, 13 and 25 were all comparably high (RLRT statistic ranged from 22.67 to 23.45) with no obvious clear position for a second QTL. Moreover the difference between likelihoods in the single-QTL LALD model and the 2-QTL LALD model was too small (difference in RLRT of 1.62) to provide evidence for a 2-QTL hypothesis.

4.4 DISCUSSION

Previous studies used LA analysis to investigate data for SCS on BTA18; this is, to the best of the authors' knowledge, the first study which applies LD and combined LALD methods to information on BTA18. Applying these methods in combination with an increased marker density, we were able to narrow down the QTL region with effect on SCS previously identified by Brink (2003) and Xu et al. (2006) to a 10.5 cM interval. Xu et al. (2006) and Kühn et al. (2003) identified the telomeric end of BTA18 with marker TGLA227 as being linked to a QTL for SCS in an LA. However, in our LA, there was no longer a peak near TGLA227. In agreement with this the 2-QTL LALD analysis showed no indication of a second QTL in the interval DIK5235-TGLA227 or elsewhere. In all previous studies testing for a linked QTL, TGLA227 was considered the last marker on the map; only 1 marker (DIK4013) is located more distally in the MARC USDA bovine linkage map. The inclusion of DIK4013 in our study could have had an effect on the significance of the LD analyses in the telomeric region of BTA18. Markers 5 cM left and right of the putative QTL position were considered a haplotype in the LD analyses, therefore the haplotype for interval 24 (between DIK5235 and TGLA227) included DIK4013. At the end of the chromosome, the window used for haplotype characterisation is limited on 1 side and contains fewer markers than those for markers in the middle of the chromosome. It may be more difficult to detect linkage disequilibrium to a QTL at the borders of the investigated region. The use of SNP markers, which are much more frequent in the genome, will provide a good possibility for increasing marker density around TGLA227 to further investigate the telomeric region of BTA18.

Any comparison between our results and previously published studies is hampered by the lack of information on confidence intervals and limited marker densities in previous analyses. Ashwell et al. (1997) found a QTL (nominal $P \leq 0.01$) on BTA18 at the marker BM2078 using single marker regression across families in the US Holstein population, which is likely

the same QTL reported in 1 family of the same population by Rodriguez-Zas et al. (2002). Although Muncie et al. (2006) also analysed US Holsteins, they did not report this QTL, nor did they report any chromosome-wide significant regions within the region under investigation in the present study.

In a study by Schrooten et al. (2000) on Dutch Holstein, 7 markers spanning 124 cM (average marker interval = 20.7 cM) on BTA18 were included in a whole genome scan for conformation and functional traits. A QTL affecting SCC with a chromosome-wide $P \leq 0.01$ was found between BM7109 and ILSTS002. This chromosomal region was included in our study, however our results do not support these findings. The marker map of Schrooten et al. (2000) is, however, difficult to compare to our map since their calculations were based on only 7 markers; it is likely that marker positions are quite different to ours in terms of interval size and total chromosomal length.

In contrast to other countries, breeding programs for udder health traits in Denmark, Finland, Norway and Sweden include clinical mastitis (CM) in addition to SCC / SCS. Clinical mastitis is a binary trait based on whether or not the cow had at least 1 veterinary treatment for clinical mastitis shortly before or during the lactation (Klungland et al., 2001). In a study by Holmberg and Andersson-Eklund (2004) on Swedish Red and White and Swedish Holstein animals, QTL for both CM and SCC were found on BTA18, however the QTL were not significant in an across-family analysis. Schulman et al. (2004) studied the Finnish Ayrshire population and found similar results: genome-wide significant QTL for both SCS and mastitis were localized near TGLA227. Lund et al. (2007) examined BTA18 in the search for CM and SCS QTL in Finnish Ayrshire, Swedish Red and White and Danish Red animals using within-family (LA) variance component analysis. They did not detect QTL for CM, however the RLRT test statistics for SCS and a 2-trait model with a pleiotropic QTL affecting both CM and SCS reached chromosome-wide significance over the entire telomeric end of the chromosome (35.1 cM). Seven of 8 markers analysed by Lund et al. (2007) (BM7109, ILSTS002, BMS2639, BMON117, BM2078, BM6507 and TGLA227) were also included in our study.

Compared to the previously mentioned studies, we increased marker density and incorporated linkage disequilibrium methods in single and 2-QTL models to improve localization of the QTL affecting SCS repeatedly reported on BTA18. A large chromosomal region (>30 cM) was identified using LA, whereby LD narrowed the test statistic profile down to a single peak at position 72.55 cM between the markers SNP523 and BMS833. Likelihood values were generally higher in a 2-QTL model than in the single QTL model ($RLRT^7_{2-QTL} = 23.51$ vs.

RLRT⁷_{1-QTL} = 21.89), however no conclusion regarding further QTL affecting udder health can be drawn.

4.5 CONCLUSION

This study was undertaken with the goal of further investigating the chromosomal area affecting SCS on BTA18 identified by Kühn et al. (2003) and Xu et al. (2006) using a much denser marker map and employing linkage disequilibrium methods. Compared to the point of departure, a region of interest for future fine mapping experiments was narrowed to the marker interval ILSTS002 and BMS833, with an approximate confidence interval spanning less than 11 cM. Although a 2-QTL model showed genome-wide significance for a putative QTL between DIK5235 and TGLA227 in connection with the putative QTL between DIK4672 and SNP523, no conclusive evidence for the existence or position of a second QTL could be found. Further studies regarding the QTL at the telomeric end of BTA18 are required, whereby the use of densely-spaced SNP in the vicinity of TGLA227 may provide a clearer picture of the linkage disequilibrium in this region.

4.6 ACKNOWLEDGEMENTS

The authors would like to thank the Federation of German Cattle Breeders (Arbeitsgemeinschaft Deutscher Rinderzüchter, ADR) for their supply of data. Financial support from the FUGATO.M.A.S.-Net project and the Förderverein Biotechnologieforschung (FBF) is gratefully acknowledged.

4.7 REFERENCES

- ASHWELL, M.S.; REXROAD JR., C.E.; MILLER, R.H.; VANRADEN, P.M. AND DA, Y.: Detection of loci affecting milk production and health traits in an elite US Holstein population using microsatellite markers. *Anim. Genet.* 28 (1997) 216–222.
- BAES, C. AND REINSCH, N.: Computing the condensed conditional gametic QTL relationship matrix and its inverse. *Arch. Tierz.* 50 (2007) 294-308.
- BAES, C. AND REINSCH, N.: TIGER: A software system for fine-mapping quantitative trait loci. *Arch. Tierz.* 51 (2008) 402-412.
- BENNEWITZ, J.; REINSCH, N.; REINHARDT, F.; LIU, Z. AND KALM, E.: Top down preselection using marker assisted estimated breeding values in dairy cattle. *J. Anim. Breed. Genet.* 121 (2004) 307-318.
- BRINK, M.: Ein Beitrag zur Feinkartierung von QTL-Regionen für Eutergesundheit beim Rind. Schriftenreihe des Instituts für Tierzucht und Tierhaltung der Christian-Albrechts-Universität zu Kiel (2003) 142.
- CALUS, M.; MEUWISSEN, T.; DE ROOS, A. AND VEERKAMP, R.: Accuracy of Genomic Selection Using Different Methods to Define Haplotypes. *Genetics* 178 (2008) 553-561.
- EBL, A.: Choice of an appropriate bending factor using prior knowledge of the parameters. *J. Anim. Breed. Genet.* 108 (1991) 89–101.
- GILMOUR, A.; GOGEL, B.; CULLIS, M. AND THOMPSON, R.: ASReml User Guide Release 2.0. (2006) VSN International Ltd., Hemel Hempstead, HP1 1ES, UK.
- GRAPES, L.; FIRAT, M. Z.; DEKKERS, J. C. M.; ROTHSCHILD, M. F. AND FERNANDO, R. L.: Optimal haplotype structure for linkage disequilibrium-based fine mapping of quantitative trait loci using identity by descent. *Genetics* 172 (2006) 1955-1965.
- GREEN, P.; FALLS, K. AND CROOKS, S.: Documentation for CRIMAP, version 2.4. (1990) Washington University School of Medicine, St. Louis, MO, USA.
- IHARA, N.; TAKASUGA, A.; MIZOSHITA, K.; TAKEDA, H.; SUGIMOTO, M.; MIZOGUCHI, Y.; HIRANO, T.; ITOH, T.; WATANABE, T.; REED, K.M.; SNELLING, W.M.; KAPPES, S.M.; BEATTIE, C.W.; BENNETT, G.L. AND SUGIMOTO, Y.: A Comprehensive Genetic Map of the Cattle Genome Based on 3802 Microsatellites. *Genome Res.* 14 (2004) 1987-1998.
- ITOH, T.; WATANABE, T.; IHARA, N.; MARIANI, P.; BEATTIE, C.W.; SUGIMOTO, Y. AND TAKASUGA, A.: A comprehensive radiation hybrid map of the bovine genome comprising 5593 loci. *Genomics* 85 (2005) 413-424.
- HOESCHELE, I.; UIMARI, P.; GRIGNOLA, F.; ZHANG, Q. AND GAGE, K.: Advances in statistical methods to map quantitative trait loci in outbred populations. *Genetics* 147 (1997) 1445-1457.
- HOLMBERG M. AND ANDERSSON-EKLUND, L.: Quantitative Trait Loci Affecting Health Traits in Swedish Dairy Cattle. *J. Dairy Sci.* 87 (2004) 2653-2659.

- KHATKAR, M.S.; THOMSON, P.C.; TAMMEN, I. AND RAADSMA, H.W.: Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet. Sel. Evol.* 36 (2004) 163–190.
- KLUNGLAND, H.; SABRYIU, A.; HERINGSTAD, B.; GRO OLSEN, H.; FOMEZ-RAYA, L.; INGE VÅGE, D.; OLSAKER, I.; ØDEGÅRD, J.; KLEMETSDAL, G.; SCHULMAN, N.; VILKKI, J.; RUANE, J.; AASLAND, M.; RØNNINGEN, M. AND LIEN, S.: Quantitative trait loci affecting clinical mastitis and somatic cell count in dairy cattle. *Mammalian Genet.* 12 (2001) 837-842.
- KÜHN, CH.; BENNEWITZ, J.; REINSCH, N.; XU, N.; THOMSEN, H.; LOOFT, C.; BROCKMANN, G.A.; SCHWERIN, M.; WEIMANN, C.; HIENDLEDER, S.; ERHARDT, G.; MEDJUGORAC, I.; FÖRSTER, M.; BREINIG, B.; REINHARDT, F.; REENTS, R.; RUSS, I.; AVERDUNK, G.; BLÜMEL, J. AND KALM, E.: Quantitative trait loci mapping functional traits in the German Holstein Cattle Population. *J. Dairy Sci.* 86 (2003) 360-368.
- KÜHN, CH.; REINHARDT, F. AND SCHWERIN, M.: Marker assisted selection of heifers improved milk somatic cell count compared to selection on conventional pedigree breeding values. *Arch. Tierzucht* 51 (2008) 23-32.
- LANDER, E. AND BOTSTEIN, D.: Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps. *Genetics* 121 (1989) 185-199.
- LIU, Z.; REINHARDT, F.; BÜNGER, A. AND REENTS, R.: Derivation and Calculation of Approximate Reliabilities and Daughter Yield-Deviations of a Random Regression Test-Day Model for Genetic Evaluation of Dairy Cattle. *J. Dairy Sci.* 87 (2004) 1896-1907.
- LUND, M.S.; SAHANA, G.; ANDERSSON-EKLUND, L.; HASTINGS, N.; FERNANDEZ, A.; SCHULMAN, N.; THOMSEN, B.; VIITALA, S.; WILLIAMS, J.L.; SABRY, A.; VIINALASS H. AND VILKKI, J.: Joint Analysis of Quantitative Trait Loci for Clinical Mastitis and Somatic Cell Score on Five Chromosomes in Three Nordic Dairy Cattle Breeds. *J. Dairy Sci.* 90 (2007) 5282-5290.
- MANGIN, B.; GOFFINET B. AND REBAÏ, A.: Constructing confidence intervals for QTL location. *Genetics* 138 (1994) 1301–1308.
- MEUWISSEN, T.; KARLSEN, A.; LIEN, S.; OLSAKER I. AND GODDARD, M.: Fine Mapping of a Quantitative Trait Locus for Twinning Rate using Combined Linkage and Linkage Disequilibrium Mapping. *Genetics* 161 (2002) 373-379.
- MUNCIE, S.A.; CASSADY, J.P. AND ASHWELL, M.S.: Refinement of quantitative trait loci on bovine chromosome 18 affecting health and reproduction in US Holsteins. *Animal Genetics* 37 (2006) 273-275.
- OTT, J.: *Analysis of Human Linkage*. John Hopkins University Press. (1991) Baltimore Maryland.
- REINSCH, N.: A multiple-species, multiple-project database for genotypes at codominant loci. *J. Anim. Breed. Genet.* 116 (1999) 425-435.
- RODRIGUEZ-ZAS, S., SOUTHEY, B.; HEYEN, D. AND LEWIN, H.: Interval and Composite Interval Mapping of Somatic Cell Score, Yield, and Components of Milk in Dairy Cattle. *J. Dairy Sci.* 85 (2002) 3081-3091.

RUPP, R. AND BOICHARD, D.: Genetics of resistance to mastitis in dairy cattle. *Vet. Res.* 34 (2003) 671-688.

SCHROOTEN C.; BOVENHUIS, H.; COPPIETERS, W. AND VAN ARENDONK, J.A.M.: Whole Genome Scan to Detect Quantitative Trait Loci for Conformation and Functional Traits in Dairy Cattle. *J. Dairy Sci.* 83 (2000) 795-806.

SCHULMAN, N.S.; VIITALA, A.; DE KONING, D.; VIRTA, J.; MÄKI-TANILA, A. AND VILKKI, J.: Quantitative Trait Loci for Health Traits in Finnish Ayrshire Cattle. *J. Dairy Sci.* 87 (2004) 443-449.

SEYFERT, H. M. AND KÜHN, CH.: Characterization of a first bovine lactoferrin gene variant based on an EcoRI polymorphism. *Animal Genetics* 25 (1994) 54.

SNELLING, W.M., CASAS, E.; STONE R.T.; KEELE, J.W.; HARHAY, G.P.; BENNETT, G.L. AND SMITH, T.P.L.: Linkage mapping bovine EST-based SNP. *BMC Genomics* 6 (2005) 74.

TUCHSCHERER, A.; MAYER M. AND REINSCH, N.: Identification of gametes and treatment of linear dependencies in the gametic QTL-relationship matrix and its inverse. *Genet. Sel. Evol.* 36 (2004) 621-642.

XU. N., PAUL, S.; BENNEWITZ, J.; REINSCH, N.; THALLER, G.; REINHARDT, F.; KÜHN, CH.; SCHWERIN, M.; ERHARDT, G.; WEIMANN, C.; THOMSEN, H.; MISHRA, S. AND KALM, E.: Confirmation of quantitative trait loci for somatic cell score on bovine chromosome 18 in the German Holstein. *Arch. Tierz.* 49 (2006) 111-119.

XU, S. AND ATCHLEY, W.: A random model approach to interval mapping of quantitative genes. *Genetics* 141 (1995) 1189-1197.

CHAPTER FIVE:
LINKAGE DISEQUILIBRIUM AT A QTL FOR SOMATIC CELL SCORE ON BTA27 IN
THE GERMAN HOLSTEIN

Christine Baes¹, Manfred Mayer¹, Jens Tetens², Jörn Bennewitz², Zengting Liu³, Friedrich Reinhardt³, Georg Thaller², Norbert Reinsch¹

¹*Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere, Forschungsbereich
Genetik und Biometrie*

²*Christian-Albrechts-Universität zu Kiel, Institut für Tierzucht und Tierhaltung*

³*Vereinigte Informationssysteme Tierhaltung w.V.*

Submitted

ABSTRACT

Marker density was increased 5-fold and combined linkage and linkage disequilibrium analyses were employed to narrow the chromosomal area of interest of a previously identified QTL region affecting somatic cell score on bos Taurus autosome 27 in the German Holstein. A granddaughter design consisting of 6 grandsire families with 492 progeny tested sons was investigated. Nineteen microsatellite markers with an average marker spacing of 1.88 cM were genotyped along a chromosomal segment of 33.8 cM. The maximal restricted likelihood test ratio statistic was identified between markers DIK2879 and KIBS272; the ratio of QTL to polygenic variance in the first parity was estimated at 18% at this position. The confidence interval was narrowed down to approximately 9.7 cM. Application of a 2-QTL model did not result in higher test statistics. Estimated sums and differences of haplotype effects, relative to their standard errors, allowed the identification of animals likely to be either heterozygous or homozygous at the QTL. This study provides not only a basis for the selection of further markers in more distinct disequilibrium with the QTL but will also be useful in the search for candidate genes affecting somatic cell score.

KEYWORDS

fine-mapping / mastitis / BTA27 / somatic cell score

5.1 INTRODUCTION

Subclinical or clinical inflammation of the mammary gland is the principal manifestation of poor udder health in dairy cattle. Usually caused by microbial or fungal infection, mastitis remains one of the most expensive and persistent health problems in the dairy sector. Although management practices play the largest role in the fight against mastitis, some animals are more genetically disposed to infection than others. Furthermore, it is generally accepted that a positive genetic correlation between mastitis susceptibility and milk yield exists. The inclusion of udder health traits in breeding programmes is therefore imperative to counteract the undesirable correlated response resulting from selection for milk production alone.

As in many other countries, genetic selection for udder health in Germany is achieved through the indicator trait somatic cell score (SCS), which is highly correlated with clinical mastitis and has a moderate heritability of 16-17% (Liu et al., 2004). Through the use of genetic markers, it is possible to find specific quantitative trait loci (QTL) distributed throughout the genome which underlie the genetic variance of SCS.

A number of genome scans have been performed in the search for QTL affecting mastitis susceptibility (see Khatkar et al., 2004, for a review). Klungland et al. (2001) were the first to detect a QTL directly affecting udder health on *Bos taurus* autosome 27 (BTA27). They used a multiple marker regression model (MMRM) to analyse both clinical mastitis and lactation means of somatic cell count (SCC) in Norwegian cattle. Although a QTL affecting clinical mastitis was found between the markers IOBT313 and BM1857 ($p_{\text{chromosome-wide}} = 0.015$) there was no obvious effect on SCC. A QTL affecting SCS was found on BTA27 ($p_{\text{genome-wide}} < 0.1$) in the North American Holstein population using composite interval mapping (Rodriguez-Zas et al., 2002). Kühn et al. (2003) found a QTL affecting SCS on BTA27 between the markers BM3507 and TGLA179 ($p_{\text{chromosome-wide}} = 0.004$) in the German Holstein population using a weighted MMRM. A further QTL affecting SCS was reported close to the marker BMS641 ($p_{\text{chromosome-wide}} = 0.02$) in the Finnish Ayrshire population using a MMRM (Schulman et al., 2004). Although a 1-to-1 comparison of QTL regions across studies is difficult due to differences in the marker maps, it can be concluded that the SCS QTL regions identified in the North American, German and Finnish populations are all located within 30 cM of the centromeric end of the chromosome (Khatkar et al., 2004). Confidence intervals (CI) in such genome scans are typically of the order of 20-40 cM and may span over the entire chromosome (Kühn, personal communication). True association between specific markers and the SCS phenotype cannot be established with such sparse maps. Detailed investigation of

each individual QTL identified in the afore-mentioned studies is therefore required to validate results, to further narrow down chromosomal areas of interest and to identify true candidate genes which may be used in marker assisted selection programmes (MAS).

Marker density was increased 5-fold around the QTL affecting udder health on BTA27 previously reported by Kühn et al. (2003). The results of the linkage disequilibrium (LD) based approaches employed to narrow the localization of this QTL are reported, and a discussion with respect to further areas of interest for future fine mapping studies is given.

5.2 MATERIALS AND METHODS

5.2.1 FAMILIES

Six paternal half-sib families containing a total of 492 genotyped bulls were analysed. These animals are a subset of the German Holstein population. All 6 families originate from the granddaughter design of Thomsen et al. (2000) and were part of the national QTL mapping project of German breeding organisations, several German animal breeding institutes and animal computing centers initiated by the German cattle breeding federation (ADR). This data was also used in the genome scan of Kühn et al. (2003). Families had an average of 82 sires ranging from 28 to 274, where sire denotes the male parent common to each half-sibship. A full pedigree including non-genotyped ancestors of genotyped animals was also available for estimating random polygenic effects (4,622 animals). Pedigree information was provided by the German genetic evaluation center (VIT) in Verden, Germany.

5.2.2 GENOTYPES AND LINKAGE MAP CONSTRUCTION

Eighteen markers were chosen from the United States Department of Agriculture's Meat Animal Research Center (MARC USDA) bovine linkage map (<http://www.marc.usda.gov/genome/genome.html>). Five of these markers (BM3507, BMS2168, TGLA179, RM209 and CSSM43) were genotyped as described by Thomsen et al. (2000). Information on these 5 markers was obtained from the genome analysis project of the Federation of German Cattle Breeders (ADR) and is currently used for the marker assisted best linear unbiased prediction (MA BLUP) evaluation of the SCS trait in Germany. The remaining markers were genotyped by PCR amplification and subsequent fragment length analysis using a MegaBace1000 capillary sequencer (GE Healthcare, Munich, Germany). The PCR reaction with a final volume of 10 μ l was performed using 25 ng of genomic bovine DNA, 0.2 mM dNTPs, 0.2 μ M of each primer, and 0.5 U / Taq / polymerase in the reaction

buffer supplied by the manufacturer (Invitrogen, Karlsruhe, Germany). The forward primers were labelled with HEX, TET or FAM. The thermocycler profile was 94°C for 3 minutes; 35 cycles of 94°C for 30 seconds; marker-specific annealing for 30 seconds and 72°C for 30 seconds; and a final cooling step at 4°C for 10 minutes. Additionally, the marker KIBS272 was newly developed within this study, resulting in 19 microsatellite markers spanning over 33 cM (see Figure 1). The newly developed marker KIBS272 represents a TG repeat of variable length starting at position 525,794 of the bovine reference sequence NW_0014943954.1. PCR primers for this marker (Forward 5'-GCT ACA GTC AGT CAA TGT GCA GAG-3' and Reverse 5'-GTG GTG TGT AAC AGC AAC CTC AA-3') were designed using Primer 3.0 software

(http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi).

The marker order was obtained from the MARC USDA bovine linkage map (<http://www.marc.usda.gov/genome/genome.html>). To obtain the order of markers with

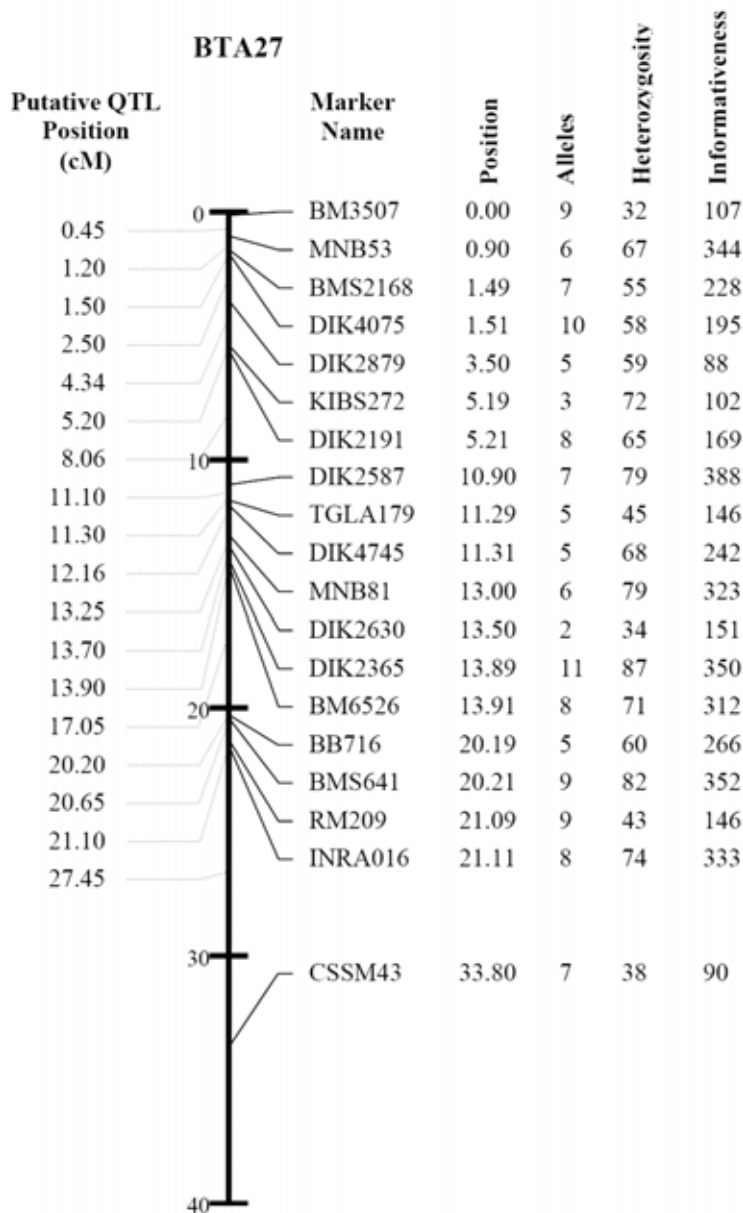


FIGURE 1. Marker and putative QTL (quantitative trait locus) map spanning over 33.8 cM. Putative QTL positions are given to the left (18) and marker names, positions, number of alleles per marker, marker heterozygosity and the number of informative meioses are given to the right of the index. The midpoint of each marker interval is considered a putative QTL position.

identical linkage map positions a BLAST (Basic Local Alignment and Search Tool) search against the bovine genome sequence build 3.1 was performed using an expectation value of 0.1 and a word size of 11. Repetitive elements were masked using RepeatMasker (<http://www.repeatmasker.org/>). Where no flanking sequence was available the primer sequences were subjected to a BLAST search applying an expectation value of 10 and a word size of 7. The map distances were calculated in Kosambi cM based on this marker order using the /BUILD/ option of the CRIMAP software (Lander and Green, 1987). A recombination rate of 0 was found between 6 pairs of adjacent markers (BMS2168-DIK4075, KIBS272-DIK2191, TGLA179-DIK4745, DIK2365-BM6526, BB716-BMS641, RM209-INRA016); here the BLAST derived marker order was used with a marker spacing of 0.02 cM.

5.2.3 PHENOTYPIC DATA

Phenotype information was obtained from the VIT in Verden, Germany. The phenotypic unit of measurement for genotyped offspring was the daughter yield deviation (DYD), which is the average phenotype of daughters corrected for fixed effects such as herd, season, and calving interval, and corrected for genetic contributions of the daughters' dams (Liu et al., 2004). The DYD were obtained from the official release of the January 2008 genetic evaluation and were based on a random regression test-day model (Liu et al., 2001).

The trait SCS was calculated from the somatic cell count (SCC) using $SCS = \log_2(SCC/100000) + 3$, where lower values represent lower SCC and therefore favourable effects on udder health. The DYD for SCS of the first lactation were used for analysis, as the number of observations was highest in this parity. The reliabilities associated with the DYD were also obtained from the January 2008 evaluation and was expressed by the number of effective daughter contributions (EDC) following Liu et al. (2004). The DYD for SCS of the 492 genotyped sires ranged from -0.052 to 1.63 (average = -0.002, standard deviation = 0.41), with lower values representing a lower SCC and therefore favourable effects on udder health. The average EDC was 188.96 with a standard deviation of 749.46.

5.2.4 STATISTICAL ANALYSIS

The most probable linkage phases of genotyped sires were calculated in order to determine which haplotypes were paternally and maternally inherited. The paternal transmission probability (the probability that the sire of an animal passed his paternal gamete to his offspring) was also calculated for every sire. Both transmission probabilities and marker allele

frequencies were determined using BIGMAP (Reinsch, 1999). The midpoint of each marker bracket was considered a putative QTL position, which resulted in evaluation of 18 putative QTL positions.

5.2.4.1 LINKAGE ANALYSIS

The model applied to the data for across-family linkage analysis (LA) contained a random polygenic effect as well as a random additive QTL effect (Xu and Atchley, 1995; Hoeschele et al., 1997):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{v} + \mathbf{e} \quad [1]$$

where $\mathbf{y}_{(m \times k)}$ is a vector of phenotypic observations (DYD) for sires, $\mathbf{X}_{(m \times k)}$ is a design matrix, $\boldsymbol{\beta}_{(k \times 1)}$ is a vector of fixed effects common to all animals, $\mathbf{Z}_{(m \times n)}$ is an incidence matrix relating animals to phenotypes, $\mathbf{u}_{(n \times 1)}$ is a vector of random polygenic effects, $\mathbf{W}_{(m \times gam)}$ is an incidence matrix, $\mathbf{v}_{(gam \times 1)}$ is a vector of random QTL effects and $\mathbf{e}_{(m \times 1)}$ is a residual vector.

Expectations of the random effects \mathbf{u} , \mathbf{v} and \mathbf{e} and their covariances were assumed to be zero. Their distribution is assumed with the multivariate normal densities $\mathbf{u} \sim N(0, \mathbf{A}_p \sigma_p^2)$, $\mathbf{v} \sim N(0, \mathbf{G}_q \sigma_q^2)$ and $\mathbf{e} \sim N(0, \mathbf{D} \sigma_e^2)$, where variances σ_p^2 , σ_q^2 and σ_e^2 describe the polygenic variance, the additive variance of the QTL and the residual variance, respectively; \mathbf{A}_p is the additive genetic relationship matrix; \mathbf{G}_q is the (co)variance matrix for the additive effects of the QTL conditional on marker haplotypes at a given position and the diagonal matrix \mathbf{D} containing 1 over the EDC (Liu et al., 2004). Subscripts in parenthesis of the vectors and matrices denote their dimensions; m is the number of animals in the subpedigree of genotyped animals, k equals 1 in our study since only a common mean effect was fitted in the analysis, n is the number of animals in the full pedigree and gam is the number of allelic effects with 1 effect per sire plus 2 times the number of grandsires.

The restricted log likelihood ratio test statistic ($RLRT^{Pos} = -2[\ln(L_0) - \ln(L_1^{Pos})]$) was calculated for each putative QTL position ($Pos=1-18$), whereby L_1^{Pos} is the RLRT of model 1 for position Pos and L_0 is that of the same model without random gametic effects for a given QTL. Test statistics were calculated using ASReML (Gilmour et al., 2006) within the TIGER software system (Baes and Reinsch, 2008).

5.2.4.2 INCORPORATING LD USING VARIANCE COMPONENTS

Combined LA and LD analysis (LALD) was conducted using the method described by Meuwissen et al. (2002), in which similar marker haplotypes are more likely to have identical QTL alleles. The method models the covariance between individuals by deriving identical by descent (IBD) probabilities for QTL alleles carried by different marker haplotypes. If a common ancestor occurred within the known pedigree, IBD probabilities were calculated from the marker information (i.e. as in LA). However, if no common ancestor existed, the probability of being IBD at the QTL was calculated from the similarity between the founder marker haplotypes (i.e. LD). In this manner, IBD sub-matrices containing the covariances between founder haplotypes were calculated and used to supplement the previously mentioned \mathbf{G}_q matrix with LD information.

Haplotypes were defined as all marker combinations within a 5 cM window (2.5 cM left and right of the putative QTL position) surrounding the given position. The window was widened in the case of larger marker intervals. The assumed effective population size was held constant at 100 and coalescence was calculated over 100 generations. This process was repeated 10,000 times. One IBD-submatrix for every putative QTL position was formed, resulting in 18 different IBD submatrices. These matrices were tested for positive definiteness; matrices with negative eigenvalues were appropriately adjusted using the BENDPDF program (EBI, 1991). The \mathbf{G}_q matrices in model 1 were supplemented with the IBD sub-matrices containing the covariances between founder marker haplotypes with no known ancestors. The dimension gam was then equal to $2(m) + 2$ times the number of grandsires. In this way, the LA model described above became an LALD when the IBD sub-matrices were included. The LD analysis was carried out similarly to the LALD analysis, however transmitting probabilities between all non-founder haplotypes were set to 0.5.

5.2.4.3 2-QTL MODEL

A 2-QTL model was calculated using a model similar to model 1:

$$\mathbf{y} = \mathbf{XB} + \mathbf{Zu} + \mathbf{W}_1\mathbf{v}_1 + \mathbf{W}_2\mathbf{v}_2 + \mathbf{e} \quad [2]$$

where $\mathbf{W}_{i(m \times gam)}$ is an incidence matrix and $\mathbf{v}_{i(gam \times 1)}$ is a vector of random QTL effects for the putative QTL i ($i = 1,2$). Dimensions and other effects are defined as in model 1. The

restricted log likelihood ratio test statistic ($RLRT^{Pos} = -2[\ln(L_0) - \ln(L_1^{Pos})]$) was calculated for each pair of putative QTL positions using ASReMl (Gilmour et al., 2006).

The covariance matrices (\mathbf{G}_q) for the additive effects of the QTL conditional on marker haplotypes in all analyses were calculated using the condensing algorithm of Tuchscherer et al. (2004), which has an effect similar to that of haplotype clustering (e.g. Calus et al., 2008). If correlation between haplotypes very close to one or zero occurred, a predefined threshold (e.g. $\varepsilon = 0.03$) was given and all haplotypes with correlations above 0.97 were considered identical. Haplotypes with such high correlations were considered 'non-unique'. The resulting condensed gametic relationship matrix has between n and $2n$ rows and columns, whereby n is the number of genotyped animals in the pedigree plus 2 times the number of grandsires. This method prevents numerical difficulties due to quasi-linear dependencies which may occur due to the inclusion of almost identical rows and columns in the QTL (co)variance matrices. The matrices were computed using a modified version of the Fortran 90 software COBRA (Baes and Reinsch, 2007).

For the sake of simplicity, significance thresholds were calculated as RLRT-equivalents of logarithm of odds (LOD) units, where a LOD score > 2 indicated chromosome-wide significance (Lander and Botstein, 1989). For chromosome-wide linkage, the significance level required is traditionally set at a LOD score of 2, which is equivalent to a RLRT value of 9.2 (Ott, 1991, pp. 65). Likewise, the LOD drop-off method of Lander and Botstein (1989) was employed to estimate approximate confidence intervals, which were calculated by identifying the highest point in the RLRT profile and subtracting 1 LOD unit (equivalent RLRT difference = 4.6); all positions within 1 LOD unit from the maximum were considered within the confidence interval, and, asymptotically, these were approximately equivalent to 96.8% (Mangin et al., 1994).

5.3. RESULTS AND DISCUSSION

5.3.1 MARKER CHARACTERISTICS

The average marker interval was 1.88 cM with a maximum interval of 12.69 cM and a minimum interval of 0.02 cM between the markers. Markers had an average of 6.84 alleles; a maximum of 11 alleles was observed for the marker DIK2365 and a minimum of 2 alleles was observed for the marker DIK2630. Heterozygosity, measured as the percentage of informative markers per grandsire, ranged from 52.6% in grandsires 1 and 2 to 84.2% in grandsire 5, with an average heterozygosity of 62.3% over all grandsires. The average

information content for linkage was measured as the mean of $|1 - 2p_{ijk}|$ over all progeny at each marker interval, where p_{ijk} is the transmitting probability of the putative QTL from grandsire i to sire j at marker interval k . The mean average information content over all marker intervals was 95.6% and ranged from 84.7% to 97.8%. The number of informative meioses for the markers used ranged from 88 to 388 and averaged at 228.

5.3.2 RESTRICTED LIKELIHOOD RATIO TESTS

The RLRT profiles for LA, LD and LALD for a single QTL are depicted in Figure 2. Neither the LA profile nor the LD profile reached the chromosome-wide significance threshold. The LA curve reached a maximum between positions 4.34 and 8.06. A distinct maximum value at position 4.34 and one smaller secondary local maximum at position 20.20 was found in the LD analysis. The combined LALD analysis resulted in the highest peak at marker interval 5 (position 4.34, $RLRT = 9.79$, $\sigma_e^2 = 0.000013$) between the markers DIK2879 (3.50 cM) and KIBS272 (5.19 cM). This peak clearly exceeded the chromosome-wide significance threshold as well as the RLRT-curve of the LA analysis. An approximate 1 LOD drop confidence interval includes markers in the interval from DIK4075 to DIK4745.

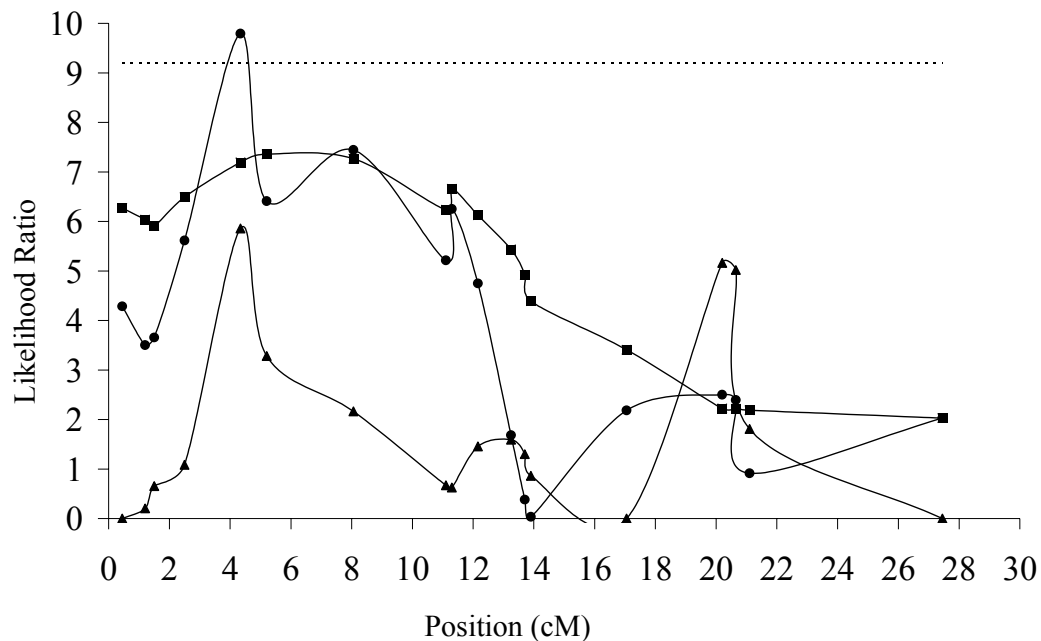


FIGURE 2. Restricted likelihood ratio curves of linkage equilibrium (▲), linkage disequilibrium (■) and combined linkage disequilibrium/equilibrium analysis (●) in lactation 1. Line markings denote interval midpoints / putative QTL positions. The dashed horizontal line shows the chromosome-wide significance threshold.

Marker interval 5 is 1.69 cM long and the associated founder haplotype consists of 1 marker to the left of the QTL and 2 markers to the right of the QTL. The use of an equivalent number of SNP as opposed to microsatellites would have enabled a more even marker spacing, which would likely be beneficial for the characterisation of haplotypes in the variance component-based LALD method. Varilo et al. (2002) stated that one microsatellite may roughly be equivalent to 5 SNP and 95 SNP can be regarded equivalent to our microsatellite map, however our markers were not evenly distributed along the chromosomal segment under investigation.

The estimated random polygenic variance σ_p^2 at marker interval 5 was 0.17 in the combined LALD analysis. The QTL variance σ_q^2 at this position (calculated as twice the gametic variance) was 0.03 and the ratio of QTL to polygenic variance $\frac{\sigma_q^2}{\sigma_p^2}$ was estimated at 0.18. The

QTL located in this chromosomal area is therefore responsible for approximately 18% of the genetic variation in SCS in the first lactation.

Although the LD analysis using a 2-QTL model showed no significant results (data not shown), the 2-QTL analysis using LALD pinpointed similar chromosomal areas as those of the single QTL analyses (Figure 3). Significant RLRT values ($\text{LOD} > 2$) were achieved in the LALD analysis at positions 4.34 and 13.9 (RLRT = 9.88), 8.06 and 20.20 (RLRT = 9.79), 4.34 and 11.30 (RLRT = 9.58), 4.34 and 8.06 (RLRT = 9.52) and 4.34 and 20.20 (RLRT = 9.50) with $\hat{\sigma}_e^2$ between 0.000022 and 0.000724. In contrast to the distinct peak found using a single QTL LALD analysis, all significant test ratios of QTL combinations were very similar. Furthermore the RLRT-values of the 2-QTL analysis did not differ markedly from the single-QTL model. With very dense marker maps the discovery of multiple QTL within coarsely-mapped QTL regions may become a frequently observed phenomenon. One very well investigated example is a recent study regarding female fertility in French dairy cattle (Druet et al., 2008), in which one QTL region disaggregated into 8 individual RLRT peaks when a combined LALD model similar to the one used here was applied. In our less informative case, however, no conclusive evidence for a second QTL was found.

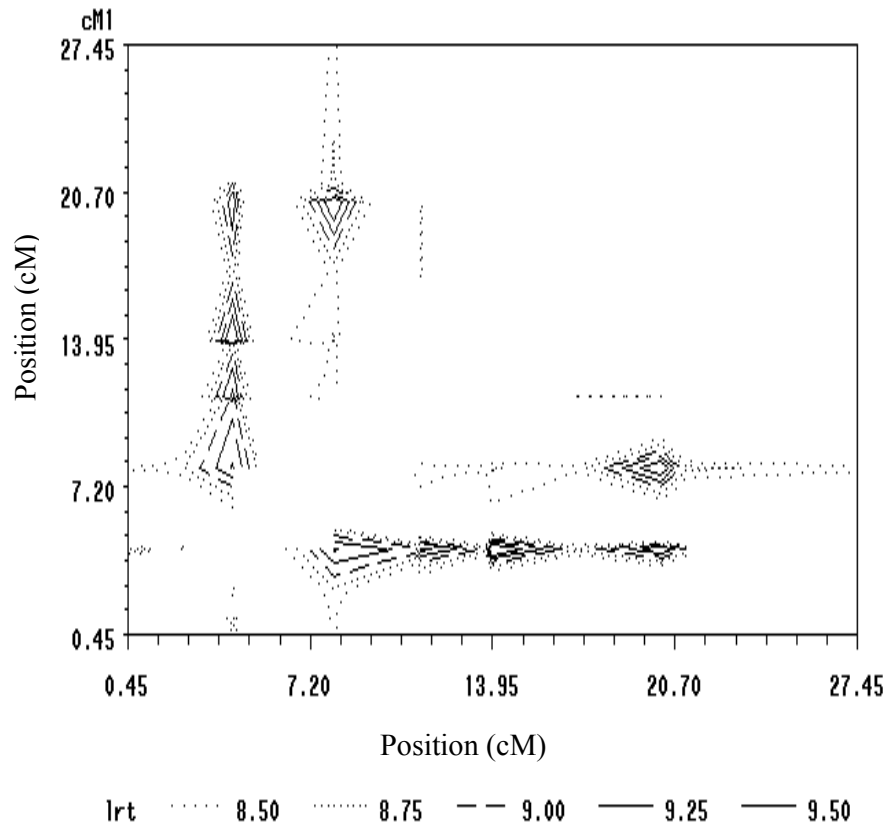


FIGURE 3.

Contour plot of the Restricted Log Likelihood Ratio profile using combined linkage and linkage disequilibrium analysis for a 2-QTL model affecting somatic cell score on *Bos taurus* autosome 27. QTL-positions in cM are given on the x- and y-axis, each point on the plane represents a particular combination of QTL-positions (each point is mirrored across the diagonal). Combinations with equal height of the test-statistics are connected by contour lines, where dark lines are used to indicate the top levels of the surface.

It is important to note that the chromosomal region examined in this study includes the bovine β -defensin cluster on BTA27 (roughly between the markers BMS2168 and TGLA179, distal to marker interval 5). Defensins are important components of the innate immune system, which play a crucial role as key effectors of the immune system and act as endogenous antibiotics (Boman, 2003; Ganz, 2003). Beta-defensins are expressed in bovine mammary gland tissue (Roosen et al., 2004); the expression of the bovine lingual antimicrobial peptide (LAP) and neutrophil β -defensin 5 (BNBD5) are induced during clinical mastitis (Goldammer et al., 2004; Swanson et al., 2004). Furthermore, associations between polymorphisms in the bovine neutrophil β -defensin gene (BNBD4) and other β -defensins and the SCC in Holstein-Friesian and Jersey cows have been reported (Wojdak-Maksymiec et al., 2006; Bagnicka et al,

2007). Since these results clearly indicate that β -defensins play a role in local host defense during udder infections, they have to be regarded as strong candidate genes for underlying the QTL affecting SCS on BTA27.

5.3.3 HAPLOTYPE ANALYSIS

A second goal of the study was to identify haplotypes associated with an increase or decrease in mastitis resistance in the population. Due to the condensing algorithm used to calculate \mathbf{G}_q , the number of unique haplotypes per marker interval was less than twice the number of individuals; there were 549 unique haplotypes in marker interval 5. The use of a denser map with SNP could make many of the unique haplotypes in our experiment ‘non-unique’ through the increased information content which would result from the use of evenly spaced dense SNP. The average haplotype estimate for a putative QTL at position 4.34 in lactation 1 was -0.020 with a maximum of 0.086, a minimum of -0.16 and one standard deviation equal to 0.047. The standard errors for haplotype estimates ranged from 0.069 to 0.10.

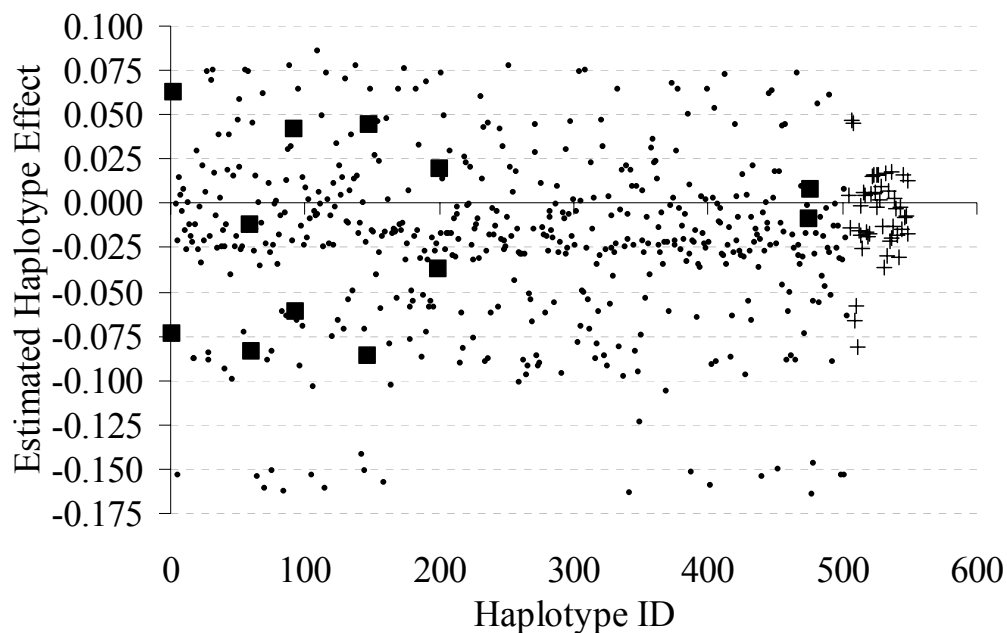


FIGURE 4. Estimated haplotype effects for 12 unique grandsire haplotypes and non-unique paternal haplotypes of sires (■), 492 maternal haplotypes of sires (●) and 45 unique paternal haplotypes of sires (+) at the fifth marker interval (4.34 cM). Haplotype ID numbers are plotted along the X-axis and the estimated haplotype effects on SCS as deviations from 0 on the Y-axis. The effects of the 447 non-unique paternal haplotypes of sires are identical to those of their respective grandsires

The estimated haplotype effects for unique haplotypes in interval 5 for the first parity are plotted in Figure 4, with haplotype effect on the Y-axis and haplotype ID number on the X-axis. Haplotypes consisted of 1 marker to the left of the putative QTL and 2 markers to the right. Twelve effects were estimated for maternal and paternal grandsire haplotypes. Additionally, 492 effects were estimated for unique founder maternal haplotypes of sires and a further 45 effects were estimated for unique paternal haplotype effects of sires; these values can be seen to the right of Figure 4. It can be observed that the values for the latter effects deviate less from the mean than those of maternal or grandsire haplotypes. This is because in order to remain “unique” the corresponding transmitting probabilities - or, equivalently, information contents - are low for these haplotypes, causing a larger shrinkage towards the mean. A total of 447 haplotype effects of more informative sires with transmitting probabilities very close to 1 or 0 were assumed to share the same QTL effect as the first (second) chromosome of their sire and therefore have no unique effect of their own.

Non-unique gametes have a potentially large effect on the population, especially if they occur frequently. Haplotypes occurred 1.83 times on average, with a maximum occurrence of 125 (i.e. 125 animals had exactly the same marker haplotype surrounding interval 5 as their sire). All non-unique haplotypes were considered identical to those of the respective grandsires; all maternal haplotypes were considered unique. Table 1 shows the origin of the non-unique gametes in this study at marker interval 5.

TABLE 1.

Grandsire number, frequency, and effect estimated for haplotypes occurring more than once in marker interval 5

Grandsire	Frequency ¹	Effect ²
1	24	-0.06784
	26	0.07650
2	19	-0.01258
	9	-0.09144
3	32	0.06236
	18	-0.05117
4	27	-0.09137
	20	0.05431
5	125	-0.05024
	112	0.008301
6	10	-0.01517
	13	0.01297

¹ Number of identical haplotype copies in sire population; total number of haplotypes = 984

² Estimated haplotype effect; standard error ranged from 0.069 to 0.10.

Each unique haplotype effect was estimated in our study; the paternal and maternal effects of a particular animal can be summed to obtain the total estimated effect for the animals' genotype. Figure 5(a) displays the distribution of genotype (QTL) effects at marker interval 5 as the sum of both haplotypes of an individual divided by the corresponding standard error of the sum. Values ranged from -1.88 to 1.15 (average = -0.26, standard deviation = 0.51). Animals with negative values (left side of graph) are likely to be QTL homozygous with a positive effect on SCS, whereby animals with positive values show an increase in SCS (right side of graph) and can be considered likely homozygous with an unfavourable effect. The animals from the extreme left and right of the graph (with the lowest and highest genotype effects) could be chosen for comparative sequencing analyses. Similarly, the absolute difference between estimated haplotype effects were analysed in order to determine likely QTL heterozygotes (on the far right of Figure 5(b)).

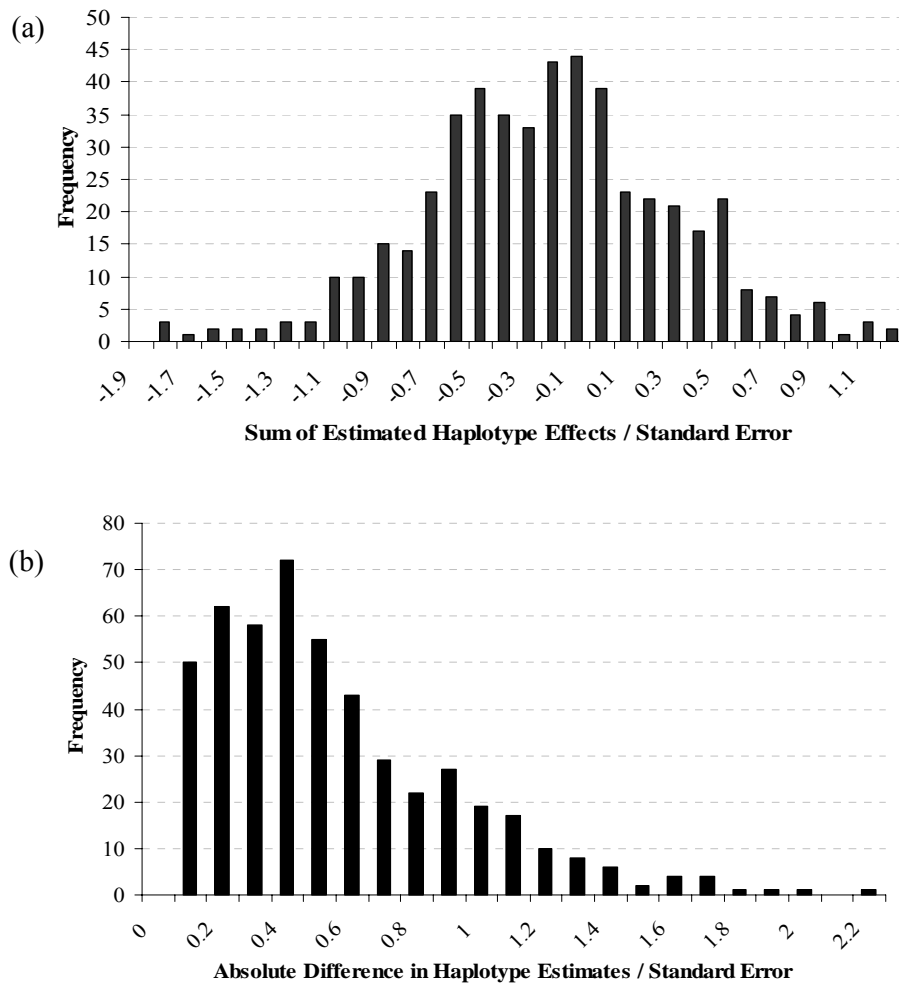


FIGURE 5. Frequency distribution of the sums (a) and differences (b) of estimated genotype effects at the fifth marker interval (4.34 cM) divided by the standard error.

Figure 6 depicts the estimated haplotype effects for the most frequent maternal 2-marker allele combinations flanking the putative QTL at marker interval 5. It is apparent that the majority of estimated effects for the haplotypes 187-131, 189-131 and 193-131 lie below the average of -0.020 and are therefore favourable. It can be concluded that these markers are very likely in strong linkage disequilibrium with the putative QTL. On the other hand, 59.2% of the effects estimated for the haplotype 189-131 are favourable and 40.7% unfavourable, indicating almost perfect linkage equilibrium. The haplotypes 189-129 and 191-131 can be considered unfavourable.

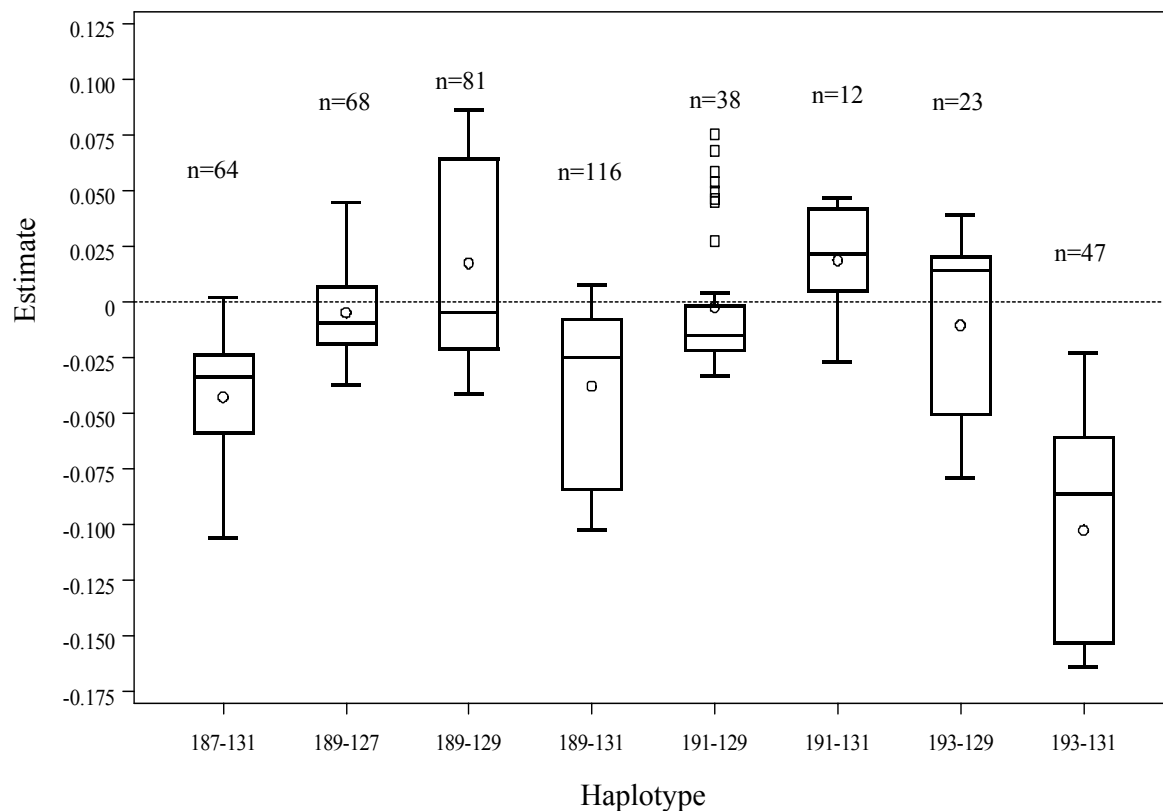


FIGURE 6.

Box-whisker plot containing estimated haplotype effects for maternal and grandsire flanking allele combinations with frequency > 2% at marker interval 5 in the first parity. The frequency (n) of the given haplotype is given above the box-whisker; markers flanking the putative QTL are written on the X-axis. The mean of each haplotype is represented by (+), the median is represented by a horizontal line through the box (—). The boxes contain 50% of all values, with the third quartile represented by the upper edge of the box and the first quartile represented by the lower edge. Whiskers extend from the first and third quartiles ± 1.5 times the interquartile range. Observations outside of the given range are denoted with \square .

The information gained from estimated genotypic values at the QTL and haplotype effects may serve as a further aid for comparative sequencing. One may capitalize on probable QTL heterozygosity and homozygosity - as well as on haplotype effects - as derived above by

comparing homozygosity and heterozygosity of potential SNP markers. This can be done in order to preselect those SNP likely to be in LD with the QTL for further fine mapping.

5.4. CONCLUSIONS

This study was undertaken with the goal of further investigating the chromosomal area affecting SCS on BTA27 identified by Kühn et al. (2003) using a much denser marker map and employing linkage disequilibrium methods. Compared to the point of departure, the region of interest for future fine mapping experiments could be decisively narrowed to the marker interval DIK2879- KIBS272, with an approximate confidence interval spanning from DIK4075 to DIK4745 (9.7 cM). A 2-QTL analysis did not give convincing evidence for a second QTL in the chromosomal region under investigation. Specific haplotypes associated with both increases and decreases in SCS could be identified. The results of this study provide the basis for the comparative sequencing of animals with extreme QTL effects and will aid in the selection of appropriate candidate genes.

5.5. ACKNOWLEDGEMENTS

The authors would like to thank the Federation of German Cattle Breeders (Arbeitsgemeinschaft Deutscher Rinderzüchter, ADR) for their supply of data. Financial support from the FUGATO.M.A.S.-Net project is gratefully acknowledged.

5.6. REFERENCES

- BAES, C. AND REINSCH, N.: Computing the condensed conditional gametic QTL relationship matrix and its inverse. *Arch. Tierz.* 50 (2007) 294-308.
- BAES, C. AND REINSCH, N.: TIGER: A software system for fine-mapping quantitative trait loci. *Arch. Tierz.* 51 (2008) 402-412.
- BAGNICKA, E.; STRZALKOWSKA, N.; FLISIKOWSKI, K.; SZREDER, T.; JÓŻWIK, A.; PRUSAK, B.; J. KRZYZEWSKI, J. AND ZWIERZCHOWSKI, L.: The polymorphism in the beta 4-defensin gene and its association with production and somatic cell count in Holstein-Friesian cows. *J. Anim. Breed. Genet.* 124 (2007) 150-6.
- BOMAN, H. G.: Antibacterial peptides: basic facts and emerging concepts. *J. Intern. Med.* 254 (2003) 197-215.
- CALUS, M.; MEUWISSEN, T.; DE ROOS, A. AND VEERKAMP, R.: Accuracy of Genomic Selection Using Different Methods to Define Haplotypes. *Genetics* 178 (2008) 553-561.
- DRUET, T.; FRITZ, S.; BOUSSAHA, M.; BEN-JEMAA, S.; GUILLAUME, F.; DERBALA, D.; ZELENKA, D.; LECHNER, D.; CHARON, C.; BOICHARD, D.; GUT, I.; EGGEN, A. AND GAUTIER, M.: Fine Mapping of Quantitative Trait Loci Affecting Female Fertility in Dairy Cattle on BTA03 Using a Dense Single-Nucleotide Polymorphism Map. *Genetics* 178 (2008) 2227-2235.
- EBL, A.: Choice of an appropriate bending factor using prior knowledge of the parameters. *J. Anim. Breed. Genet.* 108 (1991) 89-101.
- GANZ, T.: Defensins: antimicrobial peptides of innate immunity. *Nat. Rev. Immunol.* 40 (2003) 710-720.
- GILMOUR, A.; GOGEL, B.; CULLIS, M. AND THOMPSON, R., 2006. ASReml User Guide Release 2.0. VSN International Ltd., Hemel Hempstead, HP1 1ES, UK.
- GOLDAMMER, T.; ZERBE, H.; MOLENAAR, A.; SCHUBERTH, H. J.; BRUNNER, R. M.; KATA, S. R. AND SEYFERT, H. M.: Mastitis increases mammary mRNA abundance of beta-defensin 5, toll-like-receptor 2 (TLR2), and TLR4 but not TLR9 in cattle. *Clin. Diagn. Lab. Immunol.* 11 (2004) 174-85.
- HOESCHELE, I.; UIMARI, P.; GRIGNOLA, F.; ZHANG, Q. AND GAGE, K.: Advances in statistical methods to map quantitative trait loci in outbred populations. *Genetics* 147 (1997) 1445-1457.
- KHATKAR, M.S.; THOMSON, P.C.; TAMMEN, I. AND RAADSMA, H.W.: Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet. Sel. Evol.* 36 (2004) 163-190.
- KLUNGLAND, H.; SABRYIU, A.; HERINGSTAD, B.; GRO OLSEN H.; FOMEZ-RAYA, L.; INGE VÅGE, D.; OLSAKER, I.; ØDEGÅRD, J.; KLEMETSDAL, G.; SCHULMAN, N.; VILKKI, J.; RUANE, J.; AASLAND, M.; RØNNINGEN, K. AND LIEN, S.: Quantitative trait loci affecting clinical mastitis and somatic cell count in dairy cattle. *Mammalian Genet.* 12 (2001) 837-842.

KÜHN, C.; BENNEWITZ, J.; REINSCH, N.; XU, N.; THOMSEN, H.; LOOFT, C.; BROCKMANN, G. A.; SCHWERIN, M.; WEIMANN, C.; HIENDLEDER, S.; ERHARDT, G.; MEDJUGORAC, I.; FÖRSTER, M.; BREINIG, B.; REINHARDT, F.; REENTS, R.; RUSS, I.; AVERDUNK, G.; BLÜMEL, J. AND KALM, E.: Quantitative trait loci mapping functional traits in the German Holstein Cattle Population. *J. Dairy Sci.* 86 (2003) 360-368.

LANDER, E. AND BOTSTEIN, D.: Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps. *Genetics* 121 (1989) 185-199.

LANDER, E. AND GREEN, P.: Construction of multilocus genetic linkage maps of humans. *Proceedings of the National Academy of Sciences USA* 84 (1987) 2363–2367

LIU, Z.; REINHARDT, F.; BÜNGER, A. AND REENTS, R.: Application of a random regression model to genetic evaluation test-day yields and somatic cell scores in dairy cattle. *Interbull Bull.* 27 (2001) 159-166.

LIU, Z.; REINHARDT, F.; BÜNGER, A. AND REENTS, R.: Derivation and Calculation of Approximate Reliabilities and Daughter Yield-Deviations of a Random Regression Test-Day Model for Genetic Evaluation of Dairy Cattle. *J. Dairy Sci.* 87 (2004) 1896-1907.

MANGIN, B.; GOFFINET, B. AND REBAÏ, A.: Constructing confidence intervals for QTL location. *Genetics* 138 (1994) 1301–1308.

MEUWISSEN, T.; KARLSEN, A.; LIEN, S.; OLSAKER, I. AND GODDARD, M.: Fine Mapping of a Quantitative Trait Locus for Twinning Rate using Combined Linkage and Linkage Disequilibrium Mapping. *Genetics* 161 (2002) 373-379.

OTT, J.: *Analysis of Human Linkage*. John Hopkins University Press. (1991) Baltimore Maryland.

REINSCH, N.: A multiple-species, multiple-project database for genotypes at codominant loci. *J. Anim. Breed. Genet.* 116 (1999) 425-435.

RODRIGUEZ-ZAS, S.; SOUTHEY, B.; HEYEN, D. AND LEWIN, H.: Interval and Composite Interval Mapping of Somatic Cell Score, Yield, and Components of Milk in Dairy Cattle. *J. Dairy Sci.* 85 (2002) 3081-3091.

ROOSEN, S.; EXNER, K.; PAUL, S.; SCHRÖDER, J. M.; KALM, E. AND LOOFT, C.: Bovine beta-defensins: identification and characterization of novel bovine beta-defensin genes and their expression in mammary gland tissue. *Mamm. Genome.* 15 (2004) 834-42.

SCHULMAN, N.; VIITALA, S.; DE KONING, D.; VIRTA, J.; MÄKI-TANILA, A. AND VILKKI, J.: Quantitative Trait Loci for Health Traits in Finnish Ayrshire Cattle. *J. Dairy Sci.* 87 (2004) 443-449.

SWANSON, K.; GORODETSKY, S.; GOOD, L.; DAVIS, S.; MUSGRAVE, D.; STELWAGEN, K.; FARR, V. AND MOLENAAR, A.: Expression of a beta-defensin mRNA, lingual antimicrobial peptide, in bovine mammary epithelial tissue is induced by mastitis. *Infect Immun.* 72 (2004) 7311-4.

THOMSEN, H.; REINSCH, N.; XU, N.; LOOFT, C.; GRUPE, S.; KÜHN, CH.; BROCKMANN, G.; SCHWERIN, M.; LEYHE-HORN, B.; HIENDLEDER, S.; ERHARDT, G.; MEDJUGORAC, I.; RUSS, I.

FÖRSTER, M.; BRENIG, B.; REINHARDT, F.; REENTS, R.; BLÜMEL, J.; AVERDUNK, G. AND KALM, E.: A male bovine linkage map for the ADR granddaughter design. *J. of Anim. Breed. Genet.* 117 (2000) 289-306.

TUCHSCHERER, A.; MAYER, M. AND REINSCH, N.: Identification of gametes and treatment of linear dependencies in the gametic QTL-relationship matrix and its inverse. *Genet. Sel. Evol.* 36 (2004) 621-642.

VARILO, T.; PAUNIO, T.; PARKER, A.; PEROLA, M.; MEYER, J.; TERWILLIGER, J. AND PELTONEN, L.: *Hum. Mol. Genet.* 12 (2003) 51-59.

WOJDAK-MAKSYMIEC, K.; KMIEĆ, M. AND ZUKIEWICZ, A.: Associations between defensin polymorphism and somatic cell count in milk and milk utility traits in Jersey dairy cows. *J. Vet. Med. A. Physiol. Pathol. Clin. Med.* 53 (2006) 495-500.

XU, S. AND ATCHLEY, W.: A random model approach to interval mapping of quantitative genes. *Genetics* 141 (1995) 1189-1197.

GENERAL DISCUSSION

The overall aim of this thesis was twofold: 1) to develop software for fine mapping quantitative trait loci (QTL) using haplotype-based variance-component methods and 2) to use this software for reducing confidence intervals on QTL regions previously identified on *bos taurus* autosomes (BTA) 2, 18 and 27 in the German Holstein population. Technical considerations and computer software for calculation of the condensed conditional gametic relationship matrix were discussed in chapter 1. Practical validation of the software using data on a single highly polymorphic marker on BTA18 for a sample of 12,008 German Holstein dairy cattle was also described. In chapter 2 the software system TIGER was presented. TIGER was designed for comprehensive variance component analysis of fine mapping data and calculates a residual maximum likelihood-based test for each putative quantitative trait locus using raw genotypes, pedigree and marker map information. The software is capable of linkage analysis, linkage disequilibrium analysis and combined linkage / linkage disequilibrium analysis. Chapter 3 described the exploitation of linkage disequilibrium using different methods at a previously identified QTL region on BTA02. Chapter 4 described a fine mapping study done on BTA18 using a single QTL model and a 2-QTL model. Finally, chapter 5 presented fine mapping results for BTA27, including haplotype analysis for the identification of possible candidates for comparative sequencing analysis. In the present chapter, the methods and results presented in chapters 3, 4 and 5 are compared and summarised.

EXPERIMENTAL DESIGN OF FINE MAPPING STUDIES

The Holstein Friesian dairy cattle population structure is ideally suited for QTL mapping. The implementation of artificial insemination has led to very large groups of paternal half-sibs and paternal ancestry can often be traced back to only a handful of bulls. Due to the systematic and routine recording of phenotypic data on milk-producing females, the performance of sire's progeny can be included in the calculation of the sire's breeding value. This 'granddaughter design' (Georges et al., 1995; Weller et al., 1990) particularly lends itself to use in QTL mapping studies.

The structure of granddaughter designs analysed in this thesis are presented in Table 1. Designs consisted of between 6 and 10 grandsire families with between 492 and 1121 genotyped sires per chromosome.

TABLE 1.
Structure of granddaughter designs used in analyses on BTA02, BTA18 and BTA27

BOS TAURUS CHROMOSOME	02	18	27
Number of grandsire families	10	6	6
Number of genotyped sires	1121	1054	492
Average number of sons per sire	112	176	82
Average number of effective daughter contributions / sire	350	468	208
Median	51.93	55.59	54.69
Standard deviation	1569.19	1857.65	822.09
Size of full pedigree	8073	7627	4622

Although the total number of genotyped sires was highest for BTA02, the average number of sons per sire and the average number of effective daughter contributions per sire were highest for BTA18. The average number of effective daughter contributions per sire ranged from 208 to 468, although standard deviations were very high. Calculation of the respective median values shows that the distributions of effective daughter contributions are highly skewed, which is the result of very strong implementation of a few bulls.

CALCULATION OF MARKER MAPS

An essential component in fine mapping analysis using haplotype approaches is calculation of the marker map. Although many linkage maps have been published and are widely consistent, discrepancies in marker order and distance between markers are not barred. Map distance is not additive; it is based on recombination rates between markers of a given population. There are two main problems with existing linkage maps: the recombination rate between two markers may be zero, and the marker order may differ from that in the population being studied. If the recombination rate is zero (or operationally near zero), it is necessary to find the correct marker order using sequence information. Wrong marker order in haplotype analyses can lead to unexplainable drops in the likelihood profiles between very close markers. It is therefore important to calculate marker maps specific to the population being analysed. When possible, the additional use of radiation hybrid maps (Itoh et al. 2005) and human and bovine sequence assemblies (NCBI Map Viewer; <http://www.ncbi.nlm.nih.gov/mapview/>) prove useful.

In contrast to haplotype-based methods, fine mapping based on regression is independent from marker map information. In the case of BTA02, not all sequence information was available

for testing marker order. Both variance component and regression methods were therefore employed. On BTA18 and BTA27, recombination rates of zero were calculated for a number of adjacent marker pairs; in the case of BTA27, the BLAST-derived marker order was used, whereby radiation-hybrid maps and both human and bovine sequence assemblies as well as own linkage mapping results were used to define the marker order for BTA18. In these cases, a fixed marker interval (0.1 cM for BTA18 and 0.02 cM for BTA27) was set due to the technical difficulties in calculated transmitting probabilities with recombination rates of zero.

TABLE 2.
Marker map information used in analyses on BTA02, BTA18 and BTA27

BOS TAURUS CHROMOSOME	02	18	27
Length of chromosomal segment in cM ¹	81.6	50.8	33.8
Average number of markers per cM ¹	0.3	0.5	0.6
Maximum marker interval in cM ¹	8.3	7.9	12.7
Average marker interval in cM ¹	3.1	2.0	1.8

¹ Map distances were calculated in Kosambi centi Morgans

Table 2 shows general information on the marker maps used. The shortest chromosomal segment under investigation was that on BTA27 (33.8 cM), whereby that on BTA02 was the longest (81.6 cM). On average, marker intervals on BTA27 were smallest (1.8 cM), whereby intervals on BTA02 were largest (3.1 cM).

MOLECULAR MARKERS: MICROSATELLITES VS. SNP

Molecular markers are easily assayed phenotypes that have a direct correspondence with DNA sequence variation at a specific location in the genome and have proven to be an invaluable tool in the search for QTL. Molecular genetic DNA markers can be described as belonging to one of three main groups: bi-allelic dominant (random amplification of polymorphic DNA and amplified fragment length polymorphisms), bi-allelic co-dominant (restriction fragment length polymorphisms and single-stranded conformation polymorphisms) and multi-allelic co-dominant (microsatellites) (Vignal et al. 2002). In this thesis, only the microsatellite and the single nucleotide polymorphism were analysed.

Until recently, microsatellite markers played the most important role in genome analyses. Initially described in 1989 by 3 independent research groups (Litt and Luty, 1989; Tautz, 1989;

Weber and May, 1989) microsatellites are genomic regions containing repetitions of a simple sequence. Microsatellites may be highly polymorphic (up to 70 or 80 alleles), whereby the number of alleles is dependant on variation in the number of sequence repetitions. Specific microsatellite alleles can be detected using amplification with PCR followed by the separation of different fragment lengths on a polyacrylamide gel. Discrepancy between alleles is based on electrophoretic mobility, which is highly dependant on electrophoretic conditions. The detection of a particular allele is therefore impossible to standardise, meaning alleles may be identified differently in different labs, or even from gel to gel in a single lab. This makes the process of genotyping difficult and expensive.

A more recent development in molecular genetics is the implementation of single nucleotide polymorphisms (SNP), which are sites in the genome where the DNA sequences of many individuals differ by a single nucleotide. SNP are becoming the method of choice for fine-mapping experiments and genomic selection (Meuwissen et al., 2001), as they are extremely abundant, have a low mutation rate and are well-suited for automated large-scale genotyping (Khatkar et al., 2004). Furthermore, it is much easier to evenly and densely cover a chromosomal region with SNP than with microsatellites. Nevertheless, a single polymorph microsatellite can be considered equivalent to 3-5 SNP (Varilo et al., 2002), which validates the use of microsatellites for fine mapping studies when marker density is sufficient.

TABLE 3.
Overview of marker information used in analyses on BTA02, BTA18 and BTA27

BOS TAURUS CHROMOSOME	02	18	27
Number of microsatellites	24	20	19
Number of SNP ¹	2	5	0
Number of other markers	0	1	0
Total number of markers	26	26	19
Maximum number of alleles	11	34	11
Minimum number of alleles	2	2	2
Average number of alleles	6.38	8.19	6.84
Maximum number of informative meioses	803	707	388
Average number of informative meioses	430	365	228

¹Single nucleotide polymorphism

An overview of the marker information used for analyses in this thesis is given in Table 3. The number of microsatellites used ranged from 19 to 24, and considering 3-5 SNP per microsatellite (Varilo et al., 2002), the equivalent number of SNP markers would range from 57 to 120. Although no SNP were genotyped on BTA27, SNP on both BTA02 and BTA18 were included in the analysis. On BTA18, a highly polymorph erythrocyte antigen with a total of 34 alleles was also used as a marker.

GENOTYPIC INFORMATION

Genotypic information was gathered in laboratories in Giessen, Dummerstorf and Kiel. Some genotypic information used was collected in previous studies on the German Holstein population (i.e. BTA02: Thomsen et al., 2000 and Bennewitz et al., 2003; BTA18: Thomsen et al., 2000, Kühn et al., 2003, Brink, 2003 and Xu et al., 2006; BTA27: Thomsen et al., 2000 and Kühn et al., 2003). Table 4 shows the number of actual genotypes used in these analyses. The most complete genotypes are those of BTA02, in which only 3% of genotype information was missing.

TABLE 4.
Genotypes used in analyses on BTA02, BTA18 and BTA27

BOS TAURUS CHROMOSOME	02	18	27
Partner laboratory	Giessen	Dummerstorf	Kiel
Total possible number of genotypes	29,146	27,404	9,348
Actual number of genotypes	28,275	22,559	8,574

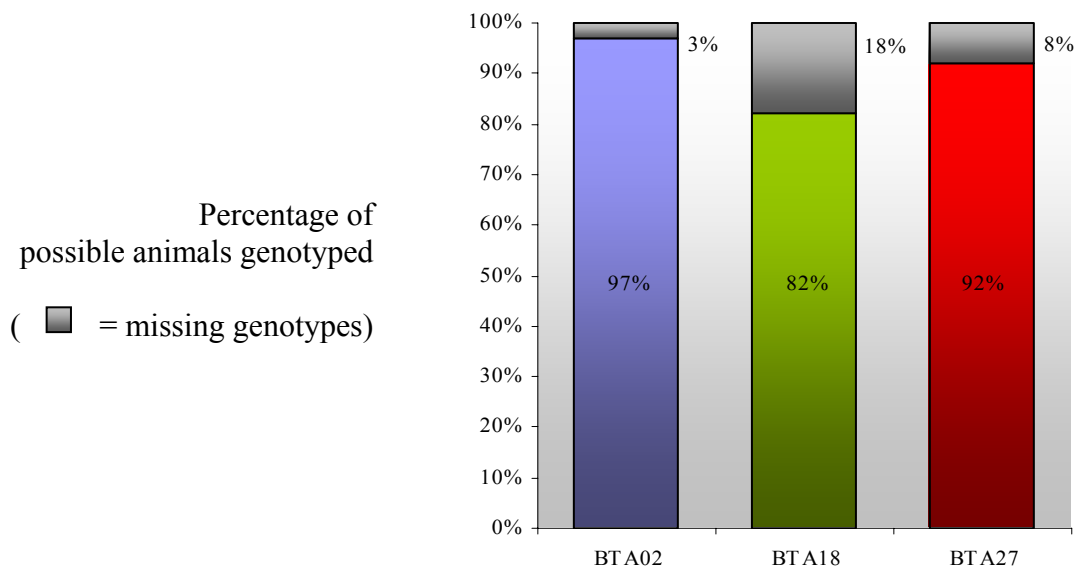


Table 4. shows that almost 20% of genotype information was missing for BTA18. This was due to the lack of DNA samples of animals used in previous studies. The animals were only genotyped for a few of the markers, and no new genotyping could take place. This problem is common; unknown linkage phases, non-informative markers and missing marker genotypes complicate the calculation of transmitting probabilities, which are required for calculation of the gametic relationship matrix, which is in turn required for solving the mixed model equations in variance component analysis for linkage analysis and linkage disequilibrium studies. Nevertheless, the attempt to incorporate as much genotype information from previous studies is worthwhile.

PHENOTYPIC INFORMATION

Phenotypic information for all analyses was obtained from the German genetic evaluation centre (VIT) in Verden, Germany. Udder health is often incorporated in breeding programs through the indicator trait SCS, which is calculated from the somatic cell count (SCC) per ml of milk as $SCS = \log_2 (SCC / 100,000) + 3$. The phenotypic unit of measurement for SCS in genotyped offspring was the daughter yield deviation (DYD), which is the average phenotype of daughters corrected for fixed effects such as herd, season and calving interval and corrected for genetic contributions of daughters' dams. An overview of DYD information used is displayed in Table 5. The frequency distributions of DYD for BTA02, BTA18 and BTA27 are presented in Figure 1.

TABLE 5.
Phenotypic information used in analyses on BTA02, BTA18 and BTA27

BOS TAURUS CHROMOSOME	02	18	27
Average daughter yield deviation	-0.085	-0.102	-0.002
Standard deviation	0.378	0.383	0.412
Maximum	1.391	1.072	1.625
Minimum	-1.211	-1.211	-1.051

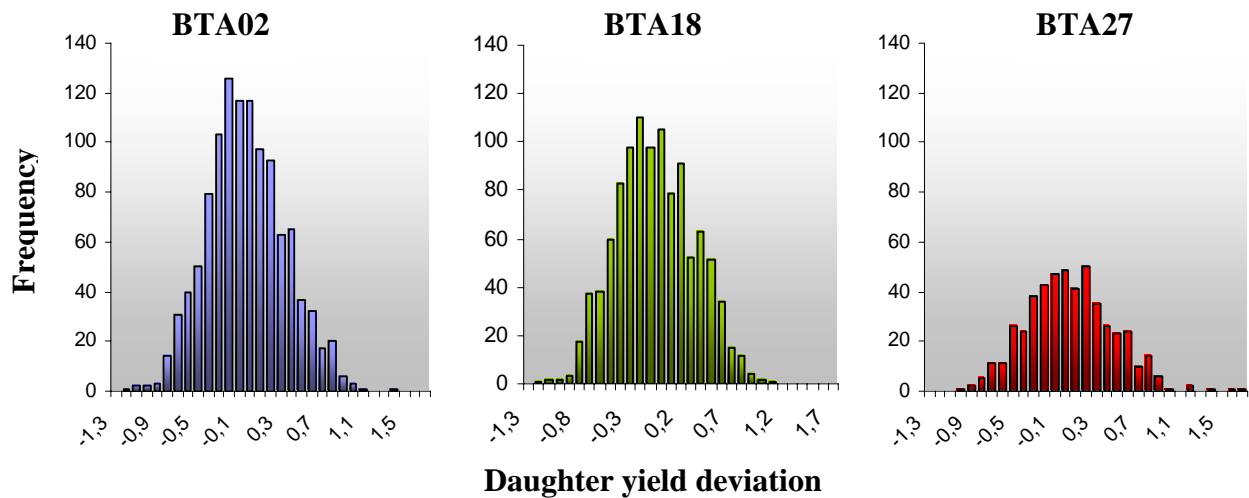


FIGURE 1.
 Frequency distributions of daughter yield deviations for somatic cell score used in analyses on BTA02, BTA18 and BTA27

LINKAGE AND LINKAGE DISEQUILIBRIUM

The terms linkage and linkage disequilibrium are, though central concepts of mapping and fine mapping experiments, somewhat misleading. Linkage only considers the disequilibrium that exists within families. When observed on a population-wide basis, there may be no association between any specific haplotypes and the QTL in question, i.e. the genotypes at different loci can be independent of one another. This disequilibrium extends over relatively large chromosomal regions and is therefore broken by recombination relatively quickly. Linkage disequilibrium, on the other hand, requires a marker to be in disequilibrium with a QTL on a population-wide basis, i.e. genotypes at different loci are not independent of one another. In order to fulfill this property, the association must have persisted over many generations and the marker must be very tightly linked to the QTL. The terms are misleading for two reasons: firstly, non-random association of alleles at two loci can occur even if the genes are unlinked. Secondly, two linked loci are not necessarily in linkage disequilibrium. It therefore becomes apparent that markers may be in strong or weak disequilibrium with a given QTL.

Linkage disequilibrium can result from various processes, including epistatic natural selection, mutation, random drift, genetic hitchhiking and gene flow. Natural selection can create disequilibrium by favouring specific allelic combinations at different loci. Particular allele combination may function better as a group, in which case natural selection acts to “push” the

different loci together physically into a so-called “supergene” (Joron et al., 2006); this tight linkage reduces the recombination rate between individual genes within the supergene. Although mutation alone does not lead to substantial disequilibrium, recurrent mutation produces non-random association between alleles at different loci. Mutation can, however, form linkage disequilibrium between loci in non-recombining areas of the genome. This disequilibrium does not decay and can increase in frequency through drift or selection. Random drift causes random fluctuation of gamete frequencies in the population and thereby increases the variance of the linkage disequilibrium coefficient (Ohta, 1982). Genetic hitchhiking is the process by which gene frequency changes because of selection of a linked gene (Smith and Haigh, 1974). This, of course, can also be a cause of linkage disequilibrium. Gene flow, or the transfer of alleles from one population to another, can also produce significant disequilibrium in a population. This will, however, only occur when the allele frequencies of both loci differ between the two populations.

ANALYSES AND RESULTS

Our goal was to exploit the linkage disequilibrium in the German Holstein as described by Meuwissen et al. (2000) and Meuwissen et al. (2002) in order to narrow the confidence intervals of previously reported putative QTL regions. Regression was only used in the special case of BTA02, in which the marker order could not be established with certainty. Table 6 displays an overview of analyses done within the context of this thesis. All results were calculated using the TIGER system and modifications thereof. Table 7 presents the main results of these analyses.

TABLE 6.
General overview of analyses implemented on BTA02, BTA18 and BTA27

BOS TAURUS CHROMOSOME	02	18	27
Linkage analysis	x	x	x
Linkage disequilibrium analysis	x	x	x
Combined Linkage / Linkage disequilibrium analysis	x	x	x
2-QTL combined Linkage / Linkage disequilibrium analysis	-	x	x
Linkage Analysis (Regression)	x	-	-
Linkage disequilibrium analysis (Regression)	x	-	-
Haplotype Analysis	-	-	x

TABLE 7.
Results of analyses on BTA02, BTA18 and BTA27

BOS TAURUS CHROMOSOME	02	18	27
Number of peaks exceeding chromosome-wide significance ¹ (Linkage analysis)	1	1	0
Number of peaks exceeding genome-wide significance ² (Linkage analysis)	1	2	0
Number of peaks exceeding chromosome-wide significance ¹ (Combined linkage / linkage disequilibrium analysis)	4	2	1
Number of peaks exceeding genome-wide significance ² (Combined linkage / linkage disequilibrium analysis)	1	3	0
Confidence interval of previous studies (cM) ³	128 ⁵	>40 ⁶	>40 ⁶
Approximate confidence intervals (1-LOD drop) (cM) ³	38 [*]	10.5	9.8
Number of markers identified using regression analysis (Combined linkage / linkage disequilibrium analysis)	2	-	-
Number of QTL regions identified	2	1	1
Maximum test statistic ⁴	20.29 [*]	21.89	9.79
Maximum 2-QTL test statistic ⁴	-	23.51	9.88

¹ Significance thresholds based on a LOD score ≥ 2 , equating to a restricted maximum likelihood test statistic of 9.2

² Significance thresholds based on a LOD score ≥ 3 , equating to a restricted maximum likelihood test statistic of 13.8

³ Map distances were calculated in Kosambi centi Morgans

⁴ Restricted maximum likelihood test statistics

⁵ Bennewitz et al., 2004

⁶ Kühn, personal communication

* Linkage analysis

An important goal of this thesis was to reduce confidence intervals on QTL regions previously identified on bos taurus autosomes BTA02, BTA18 and BTA27 in the German Holstein population. The confidence interval on BTA02 was narrowed to approximately 38 cM between the markers DIK2862 and BMS778 by linkage analysis. Using regression on single markers, 2 putative QTL regions were found close to the markers ILSTS098 and BMS778. The chromosomal region investigated on BTA02 therefore may harbour more than one QTL region affecting somatic cell score. On BTA18, 1 main peak between the markers ILSTS002 and BMS833 was found with a confidence interval of approximately 10.5 cM. Although likelihood values were generally higher in a 2-QTL model than in the single QTL model, no conclusive evidence of a second QTL was found. The region of interest for future fine mapping experiments on BTA27 could be decisively narrowed to the marker interval DIK2879- KIBS272, with an approximate confidence interval spanning from DIK4075 – DIK4745 (9.8 cM). Information for

the comparative sequencing of animals with extreme QTL effects was presented for BTA27 and will aid in the selection of appropriate candidate genes.

FUTURE PERSPECTIVES

The use of SNP markers in future mapping and fine mapping experiments will certainly play a major role in understanding and unravelling the genetic architecture of quantitative traits. Due to their extreme abundance in the genome, their low mutation rate and their propensity for automated large-scale genotyping, the ability to create very dense marker maps with evenly-spaced SNP markers can be realized. New techniques and cost-effective genotyping methods of these markers will hopefully alleviate some of the difficulties which have been encountered in the past. However, the use of conventional microsatellite markers and 'rough' maps has and will continue to prove useful in finding causative mutations. Many projects have been initiated with the goal of narrowing down confidence intervals for future fine mapping experiments using various statistical models. In this thesis, a software solution was developed which is capable of linkage analysis, linkage disequilibrium analysis and combined linkage / linkage disequilibrium analysis, as well as association analyses based on marker regression. Two independent chromosomal regions (BTA18 and BTA27) have been narrowed to < 11 cM using the combined linkage / linkage disequilibrium software, and a third chromosomal region on BTA02 was notably narrowed using linkage analysis. Although only a small step in the long journey towards discovering causal genes, the information gained from the analyses conducted in this thesis provides an excellent starting point for implementation of the new generation of markers.

LITERATURE

BENNEWITZ, J.; REINSCH, N.; GROHS, C.; LEVÉZIEL, H.; MALAFOSSE, A.; THOMSEN, H.; XU, N.; LOOFT, C.; KÜHN, C.; BROCKMANN, G.; SCHWERIN, M.; WEIMANN, C.; HIENDLEDER, S.; ERHARDT, G.; MEDJUGORAC, I.; RUSS, I.; FÖRSTER, M.; BRENIG, B.; REINHARDT, F.; REENTS, R.; AVERDUNK, G.; BLÜMEL, J.; BOICHARD, D. AND KALM, E.: 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 (2003) 319-338.

BRINK, M.: Ein Beitrag zur Feinkartierung von QTL-Regionen für Eutergesundheit beim Rind. Schriftenreihe des Instituts für Tierzucht und Tierhaltung der Christian-Albrechts-Universität zu Kiel (2003) 142.

GEORGES, M.; NIELSEN, D.; MACKINNON, M.; MISHRA, A.; OKIMOTO, R.; PASQUINO, A.; SARGEANT, L.S.; SORENSEN, A.; STEELE, M.R.; ZHAO, X.; WOMACK, J.E. AND HOESCHELE, I.: Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 139 (1995) 907-920

ITOH, T.; WATANABE, T.; IHARA, N.; MARIANI, P.; BEATTIE, C.W.; SUGIMOTO, Y. AND TAKASUGA, A.: A comprehensive radiation hybrid map of the bovine genome comprising 5593 loci. *Genomics* 85 (2005) 413-424.

JORON, M.; PAPA, R.; BELTRAN, M.; CHAMERLAIN, N.; MAVAREZ, J.; BAXTER, S.; ABANTO M.; BIRMINGHAM, E.; HUMPHRAY, S.; ROGERS, J.; BEASLEY, H.; BARLOW, K.; FRENCH-CONSTANT, R.; MALLET, J.; McMILLAN, W. AND JIGGINS, C.: A Conserved Supergene Locus Controls Colour Pattern Diversity in *Heliconius* Butterflies. *PLoS Biology* (2006) 4(19):e303 doi:10.1371/journal.pbio.0040303

KHATKAR, M.S.; THOMSON, P.C.; TAMMEN, I. AND RAADSMA, H.W.: Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet. Sel. Evol.* 36 (2004) 163–190.

KÜHN, C.; BENNEWITZ, J.; REINSCH, N.; XU, N.; THOMSEN, H.; LOOFT, C.; BROCKMANN, G. A.; SCHWERIN, M.; WEIMANN, C.; HIENDLEDER, S.; ERHARDT, G.; MEDJUGORAC, I.; FÖRSTER, M.; BRENIG, B.; REINHARDT, F.; REENTS, R.; RUSS, I.; AVERDUNK, G.; BLÜMEL, J. AND KALM, E.: Quantitative trait loci mapping functional traits in the German Holstein Cattle Population. *J. Dairy Sci.* 86 (2003) 360-368.

LITT, M. AND LUTY, J.A.: A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* 44 (1989) 397-401.

MEUWISSEN, T. AND GODDARD, M.: Fine Mapping of Quantitative Trait Loci Using Linkage Disequilibrium With Closely Linked Marker Loci. *Genetics* 155 (2000) 421-430.

MEUWISSEN, T.; HAYES, B.J. AND GODDARD, M.E.: Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157 (2001) 1819-1829.

MEUWISSEN, T.; KARLSEN, A.; LIEN, S.; OLSAKER, I. AND GODDARD, M.: Fine Mapping of a Quantitative Trait Locus for Twinning Rate using Combined Linkage and Linkage Disequilibrium Mapping. *Genetics* 161 (2002) 373-379.

OHTA, T.: Linkage disequilibrium with the island model. *Genetics* 101 (1982) 139-155.

SMITH, M. AND HAIGH, J.: The hitch-hiking effect of a favourable gene. *Genetical Research* 23 (1974) 23-35.

TAUTZ, D.: Hypervariability of simple sequenz repeats as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 17 (1989) 6463-6471.

THOMSEN, H.; REINSCH, N.; XU, N.; LOOFT, C.; GRUPE, S.; KÜHN, CH.; BROCKMANN, G.; SCHWERIN, M.; LEYHE-HORN, B.; HIENDLEDER, S.; ERHARDT, G.; MEDJUGORAC, I.; RUSS, I.; FÖRSTER, M.; BREINIG, B.; REINHARDT, F.; REENTS, R.; BLÜMEL, J.; AVERDUNK, G. AND KALM, E.: A male bovine linkage map for the ADR granddaughter design. *J. of Anim. Breed. Genet.* 117 (2000) 289-306.

VARILO, T.; PAUNIO, T.; PARKER, A.; PEROLA, M.; MEYER, J.; TERWILLIGER, J. AND PELTONEN, L.: *Hum. Mol. Genet.* 12 (2003) 51-59.

VIGNAL, A.; MILAN, D.; SAN CRISTOBAL, M. AND EGGEN, A.: A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.* 34 (2002) 275-305

WEBER, J.L. AND MAY, P.E.: Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44 (1989) 388-396.

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

XU, N., PAUL, S.; BENNEWITZ, J.; REINSCH, N.; THALLER, G.; REINHARDT, F.; KÜHN, CH.; SCHWERIN, M.; ERHARDT, G.; WEIMANN, C.; THOMSEN, H.; MISHRA, S. AND KALM, E.: Confirmation of quantitative trait loci for somatic cell score on bovine chromosome 18 in the German Holstein. *Arch. Tierz.* 49 (2006) 111-119.

GENERAL SUMMARY (ENGLISH)

The objectives of this thesis were two-fold. Firstly, software was developed to properly analyse data collected in the context of a national genome project (FUGATO M.A.S.-Net). Secondly, analyses were carried out for data collected from M.A.S.-Net partner laboratories in Giessen, Kiel and Dummerstorf. Phenotypes were contributed by the animal computing center (VIT) in Verden.

Calculation of the “condensed” gametic relationship matrix \mathbf{G}^* - a version of \mathbf{G} where linear dependencies have been removed - and its inverse \mathbf{G}^{*-1} is described and discussed in **Chapter 1**. **Chapter 2** presents and describes the software system TIGER, which was specifically developed for comprehensive variance component analysis of fine mapping data. Starting with raw genotype data, pedigree and marker map information and ending with a residual maximum likelihood-based test for each putative quantitative trait locus position, the software provides the user with an ‘all in one’ package capable of linkage analysis, linkage disequilibrium analysis and combined linkage / linkage disequilibrium analysis. The developed programs also proved useful for association analyses where polygenic effects are included in the model and linkage disequilibrium is depicted by a fixed regression on marker alleles.

Further examination of a previously identified chromosomal area affecting SCS on BTA02 is presented in **Chapter 3** and is based on marker-data from six half-sib families genotyped in Giessen. Linkage analysis revealed 15 contiguous positions with genome-wide significance, whereby variance component methodology could not detect linkage disequilibrium in the haplotypes under investigation. Regression on marker alleles resulted in the identification of 2 markers (ILSTS098 and BMS779) with a significant effect on somatic cell score. It could be concluded that multiple QTL are likely harboured in the chromosomal area studied; in contrast to previous reports, localizations were substantially narrowed.

The distal end of BTA18 has been identified by various groups as likely harbouring a QTL for SCS. **Chapter 4** deals with further investigation of this chromosomal region, in which single and 2-QTL models were used to investigate the linkage disequilibrium and to narrow down the confidence interval. Marker data and the underlying map were prepared by the project partners in Dummerstorf. A single-QTL linkage analysis model resulted in several putative QTL positions which exceeded the significance threshold of LOD 3. Combined linkage and linkage disequilibrium analysis was successful in assigning 1 QTL with a distinct significant peak to the marker interval ILSTS002-BMS833. Results of a 2-QTL model confirmed these results without strong evidence for the presence of a second QTL. The QTL position was

narrowed considerably; future analysis can concentrate on an interval of approximately 10.5 cM.

In **Chapter 5**, the chromosomal area affecting SCS on BTA27 around the *defensin*-cluster was investigated using a much denser marker map and employing linkage disequilibrium methods with both single-QTL and 2-QTL models. Marker data and the underlying map were compiled by the project partners in Kiel. A single QTL estimated to be responsible for 18 % of the genetic variance in SCS could be identified between the markers DIK2879 and KIBS272, with an approximate confidence interval of 9.8 cM. Specific haplotypes associated with both increases and decreases in SCS could be identified and will certainly be useful in further sequencing projects.

GENERAL SUMMARY (GERMAN)

Ziel dieser Arbeit war es, eine Software zu entwickeln, die die Analyse von Markerinformationen anhand von Varianzkomponentenmethoden und Haplotyp-Effekten möglich macht. Ebenfalls sollten die Daten, die im Rahmen des FUGATO M.A.S.-Net Projektes gesammelt wurden und aus den Partnerlabors in Gießen, Kiel und Dummerstorf stammen, analysiert werden. Die Phänotypen wurden vom VIT (Vereinigte Informationssysteme Tierhaltung) in Verden zur Verfügung gestellt.

Kapitel 1 beschreibt, wie die „eingedampfte“ gametische Verwandtschaftsmatrix und ihre Inverse in einer Abfolge einfacher Rechenschritte nach einem bekannten Algorithmus berechnet werden kann. **In Kapitel 2** wurde das Softwaresystem TIGER, das für Varianzkomponenten-Analysen von Feinkartierungsdaten eingesetzt wird, vorgestellt und beschrieben. Anhand von Genotypen, einer passenden Markerkarte und Abstammungsinformationen kann eine Residual-Maximum-Likelihood-Testgröße für jede putative QTL-Position ermittelt werden. Das Software-System bietet sich als umfassendes Paket an und kann sowohl für Kopplungs-, Kopplungsungleichgewichts- und für kombinierte Analysen eingesetzt werden. Die Programme sind auch für Assoziationsanalysen mit Regression auf Markerallele nützlich.

Ein bereits identifizierter Bereich auf BTA02, der auf SCS (Zellzahl) Einfluss nimmt, wurde in **Kapitel 3** näher analysiert. Die Analysen basierten auf in Gießen typisierten Markerdaten aus 6 Halbgeschwisterfamilien. Aus der Kopplungsanalyse gingen 15 Positionen hervor, die eine genomweite Signifikanz überschritten. Aus einer Regressionsanalyse auf Markerallele ergaben sich 2 weitere Marker (ILSTS098 und BMS779) mit signifikantem Effekt auf die somatische Zellzahl. Es kann daher angenommen werden, dass multiple QTL im untersuchten Bereich liegen.

Das distale Ende von BTA18 wurde bereits als eine wahrscheinliche Region für QTL mit Einfluss auf SCS identifiziert. Im **Kapitel 4** werden weitere Untersuchungen dieser Region beschrieben. Dabei wurden 1- und 2-QTL-Modelle verwendet, um die Kopplungsungleichgewichte zu untersuchen und das Konfidenzintervall zu verengen. Die Markerdaten und die zugrunde liegende Markerkarte wurden von den Projektpartnern in Dummerstorf vorbereitet. Mit einer kombinierten Kopplungsungleichgewicht- und Kopplungsungleichgewichtsanalyse konnte ein QTL eindeutig dem Markerintervall ILSTS002-BMS833 zugewiesen werden. Die Ergebnisse aus einem 2-QTL-Modell bestätigten die oben genannten Resultate ohne eindeutigen Beweis für das Vorhandensein

eines zweiten QTL. Die QTL-Position wurde deutlich eingengt, so dass sich zukünftige Analysen auf einen Abstand von ungefähr 10.5 cM konzentrieren können.

In Kapitel 5 wurde der Bereich der SCS auf BTA27 um den Defensinkomplex beeinflusst, unter Verwendung einer viel dichteren Markerkarte und unter Zuhilfenahme von Kopplungsungleichgewichtsmethoden mit 1-QTL- und 2-QTL- Modellen untersucht. Markerdaten und die zugrunde liegende Markerkarte wurden von den Projektpartnern in Kiel geliefert. Ein einzelnes QTL, das für geschätzte 18% genetische Varianz in SCS verantwortlich zu sein scheint, konnte zwischen den Markern DIK2879 und KIBS272 mit einem ungefähren Konfidenzintervall von 9.8 cM identifiziert werden. Die spezifischen Haplotypen, die sowohl mit erhöhtem als auch verringertem SCS verbunden sind, konnten aufgezeigt werden und sind zweifellos für zukünftige vergleichende Sequenzierungs-Analysen nützlich.

ACKNOWLEDGEMENTS

From experience I can tell you that the last pages of a doctoral thesis are perhaps the most widely read pages of the entire publication. Unfortunately, by the time a candidate gets to write these pages there is a disturbingly low level of sanity left - after several years of trying to understand the essence of genetic variance and wrapping your brain around DO loops, IF queries and condensed conditional gametic relationship matrices, not much can be expected. Nevertheless, it would be so nice to put hash algorithms and variance components aside and take a good long look at a nice dairy cow....

The three years spent in Dummerstorf were a crucial learning period. I would like to thank my supervisor Norbert Reinsch for guiding me through the joys and pitfalls of genetic analyses. Your patience, knowledge and time have given the fields of genetics and statistics a whole new meaning in my life. The opportunity to attend so many courses and conferences has broadened my professional, technical and geographical knowledge beyond my wildest dreams, and is quite possibly unparalleled in the history of a doctoral student at the FBN! Many thanks as well to Manfred Mayer for the hours spent discussing if the thick black curve is better than the thin black one, and for your support, knowledge, patience and never-ending willingness to send another file!

I am thoroughly grateful for the thoughtful attention and parent-like qualities of my colleagues in the Forschungsbereich für Genetik und Biometrie, especially those of Frau Inge Räder, Frau Helrun Pichmann and Herr Lothar Panicke. You have made me feel safe, warm and welcome in a place far from home. Many thanks as well to Herr Karsten Schlettwein, who managed to answer all my questions, repeatedly resurrect my computer from the dead, fix all technical problems and reconstruct any file I accidentally deleted in the three years I spent in Dummerstorf.

I am deeply indebted to the natives and non-natives with whom I had the pleasure of discovering beautiful Mecklenburg, especially Dörte, Henrike, Marika, and Theresa (note the alphabetical order!): your friendship and laughter have made the years in Dummerstorf richer than I could have ever imagined. If the phone rings, and a voice on the other line says something like “ey altäää...!” I’ll know who it is...I would also like to apologize to Azora for the hours of torture I subjected her to. You are wonderful, and even though we had our differences, you played a considerable role in my physical well-being and I enjoyed spending time with you. Azze and Lavendel, your ‘fliegende Wechseln’ were brilliant and I am grateful to Folker and Katrin for all your equestrian advice. I would also like to thank Michi and the

Dummerstorfer-Döner-man for giving me a warm place to sit and the feeling of being a real Dummerstorfer.

This thesis would certainly never have been possible without the perseverance, character and stubbornness on my paternal chromosomes inherited from my father, Leon, and the humour, dedication and love of life on the maternal chromosomes inherited from my mother Bernadette. You have given me the freedom to choose my own path, and I hope with all my heart that you can be proud of me, even if I don't have any cows! While I'm at it I can also thank my brothers and sisters Steven, Michael, Myriam and Suzanne for always giving me a licking when I deserved one and for showing me the ropes. Mulder and Suki – the fridge magnet says it best: good friends are like stars, you can't always see them, but you always know they are there.

And last but definitely not least I would like to thank Nikltje from the bottom of my heart for always being there for me, always taking care of me, and for constantly challenging me to set new limits for myself. Without you I wouldn't be where I am today.

M.SC. (AGR) CHRISTINE FRANCOISE BAES

70599 STUTTGART, HOHENHEIM,
DEN 15. SEPTEMBER 2008
GARBENSTRASSE 17
TEL.: 0172/164/8823

EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich an Eides statt, dass ich die vorlegte Dissertation mit dem Titel „FINE MAPPING OF QUANTITATIVE TRAIT LOCI AFFECTING SOMATIC CELL SCORE ON BTA02, BTA18 AND BTA27 IN THE GERMAN HOLSTEIN“ selbständig und ohne unerlaubte Hilfe angefertigt habe und dass ich die Arbeit noch keinem anderen Fachbereich bzw. noch keiner anderen Fakultät vorlegt habe.

Stuttgart-Hohenheim, den 15.09.2008

Christine Baes

Hiermit erkläre ich, dass gegen mich kein strafrechtliches Ermittlungsverfahren schwebt.

Stuttgart-Hohenheim, den 15.09.2008

Christine Baes

Persönliche Information

Christine Baes

Kirchweg 12, 74706 Osterburken-Bofsheim

Tel: +49 0711/459/22475

Mobile: +49 0172/1648823

E-mail: baes@uni-hohenheim.de



Name	Christine Francoise Baes
Geburtstag	12. September 1980
Geburtsort	Stratford, Kanada
Staatsangehörigkeit	Kanada

Ausbildung

- Sep. 1994 – Dez. 1998 St. Michael Secondary School, *Kanada*
- Sep. 1999 – Feb. 2004 University of Guelph, *Kanada*:
Bachelor of Science in Agriculture with Honours in Animal Science
- Jan. – Mai 2000 Moscow State University, *Russische Föderation*:
Austausch-Semester
- Mai – Jun. 2002 University of Victoria, *Kanada, (Kassel, Deutschland)*:
Deutschkurs
- Okt. 2002 – Jul. 2003 Universität Hohenheim, *Deutschland*:
Ontario – Baden-Württemberg Austausch
- Apr. 2004 – Sep. 2005 Universität Hohenheim, *Deutschland*:
Master of Science in Agriculture
- Okt. 2005 – Aug. 2008 Christian-Albrechts-Universität zu Kiel, *Deutschland*:
Promotion (Agrarwissenschaften: Institut für Tierzucht und Tierhaltung)

Fortbildung

- 6.-10. Jun. 2006 Georg-August-Universität Göttingen, *Deutschland*:
Bayesian Analysis in Animal Breeding
- 22.-29. Sep. 2006 Doktorandenfortbildungskurs der Gesellschaft für Tierwissenschaften,
Lehr-, Versuchs- und Prüfungsanstalt Schwarzenau, *Deutschland*:
**Statistische Methoden in Quantitativer Genetik und Tierzüchtung
unter Verwendung von Statistikpaketen**
- 11.-15. Dez. 2006 Gembloux faculté universitaire des sciences agronomiques, *Belgien*:
**Advanced Scientific Computing with FORTRAN 90 Application to
Animal Breeding and Beyond**
- 20.-21. Dez. 2006 Zentrum für Graphische Datenverarbeitung e.V., *Deutschland*:
SQL-Datenbanksprache
- 11.-15. Jun. 2007 Center for the Genetic Improvement of Livestock Guelph, *Kanada*:
Animal Breeding Applications using R
- 10.-14. Mär. 2008 University of Veterinary Medicine Hannover / Institute of Animal Genetics
Friedrich-Loeffler-Institut, *Deutschland*:
**Computation of Random and Fixed Effects in Animal Breeding with
PEST**
- 6.-7. Jun. 2008 Georg-August-Universität Göttingen, *Deutschland*:
Genomic Breeding Values and Genomic Assisted Selection
- 8.-12. Sep. 2008 Università Cattolica del Sacro Cuore, *Italien*:
Globaldiv Summer school

Praxiserfahrung

Jan. – Mai 1999	Robertsons Dairy, Neuseeland: Praktikum
Jan. – Mai 2000	Lingva.Ru, Russische Föderation: Englisch Lehrerin
Apr. – Aug. 2001, Aug. 2003 – Mär. 2004	Bayer CropScience, Kanada: Forschungsassistentin in der Abteilung Forschung und Entwicklung
Jun. – Aug. 2004, Okt. 2004 – Jun. 2007	Universität Hohenheim, Deutschland: Studentische Hilfskraft im Institut für Tierernährung (bis Sep. 2005), im Institut für Kleintierzucht und Ethologie (bis Sep. 2005) und im Akademischen Auslandsamt (bis Jun. 2007)
Sep. – Okt. 2004	Gestüt Claratal, Namibia: Praktikum
Sep. 2005 – Aug. 2008	Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere, Deutschland: Wissenschaftliche Mitarbeiterin im Forschungsbereich Genetik und Biometrie
seit Sep. 2008	Universität Hohenheim, Deutschland: Wissenschaftliche Assistentin, Institut für Tierhaltung und Tierzucht

Auszeichnungen

2002	National Sciences and Engineering Research Council of Canada Scholarship
2003 & 2004	Baden-Württemberg Landesstipendium
2005	DAAD-Preis 2005 (beste/n ausländische/n Studierende/n an der Universität Hohenheim)
2008	Förderstipendium der Europäische Vereinigung für Tierproduktion (EVT)

Kenntnisse und Interessen

Computer	MS-Office Programme, FORTRAN 90, SAS, AS Reml
Sprachen	Englisch (Muttersprache), Deutsch, Flämisch (Grundkenntnisse), Französisch (Grundkenntnisse), Russisch (Grundkenntnisse)
Mitgliedschaft	Deutsche Gesellschaft für Züchtungskunde, International Student Organisation, International Student Initiative
Zusätzliche Interessen	Lesen, Querflöte spielen, Fußball spielen, reiten, reisen