View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Digitale Hochschulschriften der LMU

From the Department of Veterinary Sciences

Faculty of Veterinary Medicine

Ludwig-Maximilians-Universität München

Arbeit angefertigt unter der Leitung von

Univ. Prof. Dr. Dr. habil. Martin Förster

# Mapping of Quantitative Trait Loci for Milk Yield Traits on Bovine Chromosome 5 in the Fleckvieh Cattle

Inaugural–Dissertation

For the attainment of Doctor Degree in Veterinary Medicine

From the Faculty of Veterinary Medicine of the

Ludwig-Maximilians-Universität München

by

**Ashraf Fathy Said Awad**

from

Sharkia- Egypt

München 2011

Gedruckt mit Genehmigung der Tierärztlichen Fakultät der

Ludwig–Maximilians–Universität München

| | |
|---|---|
| Dekan: | Univ. Prof. Dr. Braun |
| Berichterstatter: | Univ. Prof. Dr. Dr. habil Förster |
| Korreferent: | Univ. Prof. Dr. Mansfeld |

Tag der Promotion: 12. February 2011

*This work is dedicated to*

*My Parents, my wife and my lovely daughters; Sama, Shaza, Hana*

# CONTENTS

**ABBREVIATION**………………………………………………………………… IV

**CHAPTER 1: GENERAL INTRODUCTION**……………………………….. 1

**CHAPTER 2: REVIEW OF LITERATURE**………………………………… 3

**2.1. DNA Markers**……………………………………………………….. 3

2.1.1. Microsatellites……………………………………………………….. 3

2.1.2. Single Nucleotide Polymorphism (SNPs)…………………………… 4

**2.2. Mapping of Quantitative Trait Loci (QTL)**…………………….. 5

2.2.1. QTL Mapping Designs……………………………………………... 6

2.2.1.1. Daughter Design……………………………………………... 6

2.2.1.2. Granddaughter Design……………………………………… 7

2.2.1.3. Complex Pedigree Design………………………………….. 9

2.2.2. QTL Mapping Strategies…………………………………………… 10

2.2.2.1. Candidate Gene Approach…………………………………. 10

2.2.2.2. Genome Scan Approach……………………………………. 11

**2.3. Principles of Linkage Mapping**………………………………… 12

**2.4. QTL Fine Mapping**……………………………………………… 14

2.4.1. Linkage Disequilibrium……………………………………………. 15

2.4.2. Combined Linkage Disequilibrium and Linkage (LDL) Mapping… 17

**2.5. Identification of Candidate Genes**……………………………… 18

**2.6. Marker Assisted Selection (MAS), Gene Assisted Selection (GAS) and Genome Selection (GS)**…………………………… 20

**CHAPTER 3: Confirmation and Refinement of a QTL on BTA5 affecting milk production traits in the Fleckvieh Dual Purpose Cattle Breed**…………………………………………………………… 23

**Summary**…………………………………………………………………… 24

**Introduction**……………………………………………………………… 25

**Materials and Methods**…………………………………………………….. 26

Families, phenotypic data and DNA samples…………………………………… 26
Markers selection, genotyping and plausibility control………………………. 27
Linkage map construction…………………………………………………… 28
QTL mapping analysis……………………………………………………….. 28
Linkage analysis by classical regression interval mapping………………. 28
Linkage analysis by variance component approach……………………… 29
Combined linkage disequilibrium and linkage mapping………………….. 30
Estimation of linkage disequilibrium…………………………………………… 32
**Results……………………………………………………………………… 32**
Genotypes and linkage map construction……………………………………. 32
Linkage analysis by interval mapping using GDD…………………………… 33
Linkage analysis by variance component approach in a complex pedigree 33
Combined linkage disequilibrium and linkage analysis……………………… 35
The level of LD in the dual purpose cattle breed Fleckvieh………………… 36
**Discussion…………………………………………………………………… 36**
**Conclusion…………………………………………………………………… 39**
**Acknowledgement………………………………………………………….. 40**
**References…………………………………………………………………… 40**

**CHAPTER 4: Mapping of a milk production quantitative trait locus to a 1.056Mb region on bovine chromosome 5 in the Fleckvieh Dual Purpose Cattle Breed…………………………………………………… 51**
**Abstract…………………………………………………………………………….. 52**
**Background………………………………………………………………………. 54**
**Materials and Methods……………………………………………………….... 56**
Animals and phenotypes…………………………………………………….. 56
DNA preparation and microsatellite marker selection and genotyping…… 56
SNP selection, genotyping and haplotyping………………………………….. 57
Linkage map construction…………………………………………………… 58
QTL fine mapping…………………………………………………………….. 59
Multiple-QTL analysis using linkage disequilibrium and linkage analysis (LDLA) method……………………………………………………………….. 60
Estimation of model parameters and test statistics…………………………. 61

**Results**……………………………………………………………………………… 61

Genotypes and linkage map construction…………………………………….. 61

Haplotype analysis in a complex pedigree……………………………………. 62

Combined linkage disequilibrium and linkage analysis……………………… 63

**Discussion**…………………………………………………………………………… 64

**Conclusion**…………………………………………………………………………… 67

**Competing interests**………………………………………………………………… 67

**Authors' contributions**…………………………………………………………….. 68

**Acknowledgment**…………………………………………………………………… 68

**References**…………………………………………………………………………… 68

**CHAPTER 5: GENERAL DISCUSSION AND CONCLUSION**………….. 80

**CHAPTER 6: GENERAL SUMMARY**……………………………………… 88

**CHAPTER 7: ZUSAMMENFASSUNG**…………………………………….. 90

**CHAPTER 8: REFERENCES**……………………………………………… 93

**ACKNOWLEDGEMENT**……………………………………………………... 105

# ABBREVIATIONS

| | |
|---|---|
| AI | Artificial Insemination |
| ASR | Arbeitgemeinschaft Süddeutscher Rinderzuchtverbände e.V. |
| bp | Base pairs |
| BLAST | Basic Local Alignment Search Tool |
| BTA | Bos *Taurus* Autosome |
| CI | Confidence Interval |
| cM | centiMorgan |
| DD | Daughter Design |
| DFG | German Research Foundation |
| DNA | Deoxyribonucleic acid |
| DYD | Daughter Yield Deviation |
| EBV | Estimated Breeding Value |
| FV | Fleckvieh |
| FY | Fat yield |
| GAS | Gene Assisted Selection |
| GDD | Granddaughter Design |
| GEBV | Genomic Breeding Value |
| GRM | Genotype relationship matrix |
| GS | Genomic Selection |
| IBD | Identity by Descent |
| IM | Interval Mapping |
| LA | Linkage Analysis |
| LD | Linkage Disequilibrium |
| LDL | combined Linkage Disequilibrium and Linkage Analysis |
| LE | Linkage Equilibrium |
| LRT | Log-likelihood ratio test |
| MAS | Marker Assisted Selection |
| Mb | Mega base (one million of Base pairs) |
| MCMC | Markov Chain Monte Carlo |
| MGS | Maternal Grandsire |
| MME | Mixed Model Equation |

| | |
|---|---|
| MS | Microsatellite |
| MY | Milk yield |
| PCR | Polymerase Chain Reaction |
| PED | Pedigree |
| PY | Protein yield |
| QTL | Quantitative Trait Locus |
| AIREML | Average Information Restricted maximum likelihood |
| RH | Red Holstein |
| SNP | Single Nucleotide Polymorphism |
| SSR | Simple Sequence Repeats |
| STR | Short Tandem Repeats |
| VC | Variance component |
| VNTR | Variable Number of Tandem Repeats |

# CHAPTER 1

# GENERAL INTRODUCTION

Most of the economically important traits in dairy cattle are quantitative in nature, which means that they are affected by environmental factors and by large number of polygenes i.e., gene with small effects that in coordination with additional genes affects the same phenotypic trait, which will result in continuous distribution of the phenotypic expression (Georges et al. 1995).

The recent development in molecular biotechnology and genomic resources for various species has made it possible to unravel the genetic architecture of quantitative traits by identifying chromosomal loci affecting these traits. These chromosomal regions are generally termed quantitative trait loci (Geldermann 1975; Falconer & Mackay 1996). Quantitative trait loci (QTL) are natural genetic variations that exist in different populations which may be under natural and artificial selection. Several QTL accounting for genetically complex traits have been mapped in different dairy cattle populations with the assistance of genetic markers and application of daughter or granddaughter designs (Weller et al. 1990).

The first step in mapping QTL is usually a genome scan where the mapping population is genotyped for markers covering the whole genome or some selected chromosomes. In genome scans, QTL are typically mapped by linkage analysis (LA) methods. The mapping resolution achieved by this method is low because the distances between markers are relatively large and no more recombination events. Also, the confidence intervals for the most likely QTL positions are about 20cM. The development of both dense genetic maps (Barendse et al. 1997; Kappes et al. 1997; Ihara et al. 2004) associated with high-throughput genotyping techniques and new models for the analysis of data have improved fine-mapping techniques greatly.

Recent advances in technology, such as high-density single-nucleotide polymorphism (SNP) genotyping, have increased the feasibility of quantitative trait loci (QTL) detection and fine mapping in outbred populations using

historical population wide linkage disequilibrium (LD). Recently, LD has received considerable attention as it may be exploited to more effectively map genes underlying both simple and complex traits. These LD mapping strategies have been developed and successfully applied for QTL fine mapping in farm animals including dairy cattle (Grisart et al. 2002; Meuwissen et al. 2002; Blott et al. 2003). In dairy cattle combined linkage and LD analysis (LDL) has been used successfully to improve the QTL mapping resolution (Farnir et al. 2002; Meuwissen et al. 2002; Olsen et al. 2005; Olsen et al. 2007; Sahana et al. 2008).

Several QTL and candidates genes for milk production, reproduction, functional, and conformation traits have been described for several bovine chromosomes and most of these regions have been mapped in multiple studies (Bovenhuis & Schrooten 2002; Boichard et al. 2003; Viitala et al. 2003; Ashwell et al. 2004; Khatkar et al. 2004; Schrooten et al. 2004; Schnabel et al. 2005; Daetwyler et al. 2008; Kolbehdari et al. 2009; Mei et al. 2009). The mapping of QTL affecting milk production traits was a main objective of several studies in different dairy cattle populations. These traits are of a complex nature and the QTL are therefore difficult to explain genetically. The mapping of trait related QTL may be a possibility to understand and to explain the physiological background of quantity and quality of milk synthesis in a better way.

The purposes of this work were

1. To map a QTL affecting the economically important milk production traits in the Fleckvieh dual purpose cattle breed on bovine chromosome 5 (BTA5).
2. To refine this QTL and detect candidate gene as possible.

# CHAPTER 2

# REVIEW OF LITERATURE

## 2.1. DNA-Markers

Since the discovery of polymerase chain reaction (PCR) technology (Mullis & Faloona 1987; Mullis et al. 1994), it has been widely used in molecular biology for many applications of medical research, gene mapping, pedigree analysis, epidemiology, strain identification, introgressive hybridization and the study of genetic variation in natural populations (Griffin & Griffin 1994; Mullis et al. 1994). This, together with the discovery of a thermostable polymerase from Thermus aquaticus (Saiki et al. 1988), opened the way to automatisation of the process and the introduction of a simple, fast and flexible diagnostic tool for molecular biology. For the first time, any genomic region could be amplified and analysed in many individuals without the requirement for cloning or isolating large amounts of ultrapure genomic DNA.

### 2.1.1. Microsatellites

The first widespread markers to take full advantage of PCR technology were microsatellites (Litt & Luty 1989; Weber & May 1989). Microsatellites are DNA sequences that belong to a class of genomic sequence known as Variable Number of Tandem Repeat (VNTR) elements. Microsatellites are composed of short motifs of 1–6 nucleotides found tandemly repeated throughout the genomes of prokaryotic and eukaryotic organisms (Hancock 1999). They are also called Simple Sequence Repeats (SSR) (Tautz 1989) and Short Tandem Repeats (STRs) (Edwards et al. 1991). Microsatellites are highly polymorphic, co-dominant inheritance, abundant and distributed throughout the genomes. These properties have made microsatellites one of the most popular genetic markers for gene mapping, paternity testing and population genetics (Goldstein & Schlötterer 1999). Microsatellites are of relatively small size, and can, therefore, be easily amplified using PCR from DNA extracted from a variety of sources including blood, hair roots, skin or even faeces. Polymorphisms can be visualized on a sequencing gel, and the availability of automatic DNA

sequencers allows high-throughput analysis of a large number of samples (Goldstein & Schlötterer 1999; Jarne & Lagoda 1996). Microsatellites are often show tens of alleles at a locus that differ from each other in the numbers of the repeats. Mutation rate in microsatellites is on average high, ranging from $10^{-7}$ to $10^{-3}$ mutations per locus per generation in eukaryotes (Primmer et al. 1996; Schug et al. 1997; Vigouroux et al. 2002). Microsatellites are thought to vary in length by a stepwise mechanism of gain and loss via two mechanisms, namely replication slippage and interchromosomal exchange (review by Ellegren 2004). The former involves mispairing of complementary bases at the location of a microsatellite, leading to the insertion or deletion of one or more repeat units. The second model of mutation consists of either recombination or unequal crossing over, each of which can lead to large-scale contractions and expansions in the repeat array (Richard & Paques 2000). Although they are widely spread through the genome, their evolutionary origin is still not clear and their biological role is unknown.

### 2.1.2. Single Nucleotide Polymorphisms (SNPs)

SNPs (pronounced as *snip*) are variations at single nucleotides of the DNA sequence in the region, say C, replaces one of the three nucleotides (T, G, A).

Seq 1 ATT **C** AATCCA

Seq 2 ATT **T** AATCCA

SNPs occur throughout the genome. A locus is viewed as polymorphic when it exists in at least two variants and the allele frequency of the most common variant is <99% (Li & Grauer 1991). They are highly abundant and are present at one SNP in every 1000 bp in the human genome (International SNP Map Working Group 2001).

Most SNPs are located in non-coding regions, and have no direct impact on the phenotype of an individual. However, some introduce mutations in expressed sequences or regions influencing gene expression (promoters, enhancers), and may induce changes in protein structure or regulation. These SNPs have the potential to detect functional genetic variation. SNPs are becoming preferred over other genetic markers because of their distribution and mode of occurrence, relatively low mutation rate as well as the ease and low cost of genotyping (Hinds et al. 2005; Snelling et al. 2005). SNP have been used for

the detection and localization of QTL for complex traits in many species (Daw et al. 2005). In cattle a 7.1 X sequence assembly has been produced with accompanying information on over 2.300.000 SNP genome wide (Bovine Genome Project: http://www.hgsc.bcm.tmc.edu/projects/bovine). From this data, a set of SNP markers spanning the whole bovine genome can be used in the large scale identification, validation and analysis of genotypic variation in cattle (Kolbehdari et al. 2009).

## 2.2. Mapping of Quantitative Trait Loci (QTL)

Genome research in farm animals differs in several respects from that in humans or experimental organisms. The identification of simple monogenic disease loci in farm animals is less important, because animals with inherited disorders (and their parents) tend to be eliminated from breeding. Most traits of interest, such as growth, milk production and meat quality, show continuous distribution of phenotypic values and have polygenic backgrounds.

Advances in molecular and quantitative genetics allow the dissection of genetic variability underlying complex traits into discrete QTL effects. Classic quantitative genetic theory assumes that there are an infinite number of genes affecting a trait, each with a small effect. In practice, QTL are found with substantial, intermediate and small effects. The presence of a QTL is detected by mapping studies that show significant differences in phenotype between individuals receiving different QTL alleles (Andersson 2001). The rationale of QTL mapping in domestic animals is based not only on the biological interest to understand the complex genetic architecture of trait variation but also on applying genomic information to practical breeding schemes in order to enhance selection programs (Andersson 2001; Dekkers & Hospital 2002; Gibson 2003). In order to understand the molecular nature of quantitative trait variation several successful efforts to map loci that affect economically important, quantitative traits in dairy cattle have been reported (reviewed by Khatkar et al. 2004). The basic resources needed for QTL mapping are appropriate pedigrees of populations with samples of genomic DNA, records for the traits of interest, selection of molecular markers and statistical methods that

utilize the preceding information in order to identify QTL and to estimate their positions and effects (Viitala 2008).

### 2.2.1. QTL Mapping Designs

QTL have been identified both from experimental and existing populations. The most powerful way to map QTL is to use experimental crossings of inbred strains or lines that are genetically different for the traits of interest (Lynch & Walsh 1998). Experimental crosses have been implemented in pigs and poultry as mapping designs, but they are very rare in cattle. Apart from the fact that inbred lines are commonly not available, genome mapping in livestock faces additional challenges such as expenses of maintaining experimental populations, limited reproduction capacity and long generation intervals (de Koning et al. 2003).

A more common mapping approach in dairy cattle is to exploit existing large paternal halfsib families, produced through the use of artificial insemination in breeding programs. Half-sib designs are beneficial because more animals can be obtained from the multiple mating, which will therefore lead to higher statistical power to detect QTL (Soller 1998). A further advantage of this approach lies in the possibility of using already recorded phenotypic values (de Koning et al. 2003). For detection of marker-QTL linkage in dairy cattle, the most common mapping designs in cattle are Daughter Design, Granddaughter Design and Complex Pedigree.

#### 2.2.1.1. Daughter Design

In the daughter design (DD, Fig. 1), marker genotypes and trait values are assessed on daughters of heterozygous sire. Progenies are grouped according to a marker allele received from the heterozygous sire. If the marker is linked to QTL, the presence of alternative alleles at QTL will tend to make a phenotypic difference between two progeny groups. In a case where the sire is heterozygous for a marker but homozygous for QTL there will be no difference in quantitative trait value between the progeny groups (Weller et al. 1990). The daughter design is more useful in situations where phenotypic data collection is difficult and/or expensive.

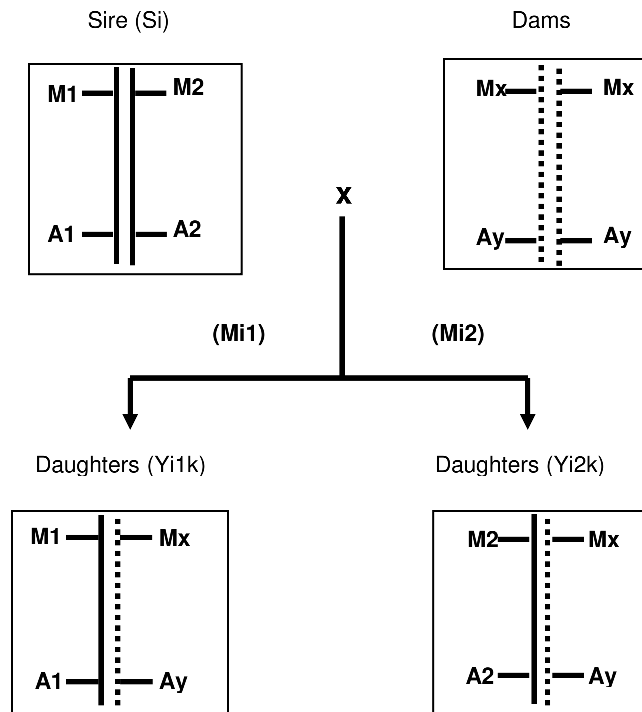



**Figure 1** Diagram of the daughter design (Weller et al. 1990), M1 and M2*,* marker alleles from the sire, Mx. marker alleles from the dams; A1 and A2*,* QTL alleles from the sire; Ay, QTL alleles from the dams,—— chromosomes from the sire ·······: chromosome from the dams.

2.2.1.2. Granddaughter Design

In the granddaughter design (GDD, Fig. 2), marker genotypes are determined for sons of heterozygous sires and trait values for the daughters of these sons. The heterozygous sire in the design is named “grandsire”, his sons are termed “sons” and daughters of the sons are termed “granddaughters”. The sons would form two subgroups per sire according to the received grandsire allele, while the trait value would be measured on granddaughters for each subgroup (Weller et al. 1990).

In order to increase the mapping power in DD, it is preferable to have fewer sires with many daughters per sire. For the GDD the power increases with number of grandsires, sons per grandsire and daughters per son. In both cases, the magnitude of the QTL effect has the greatest effect on mapping power.

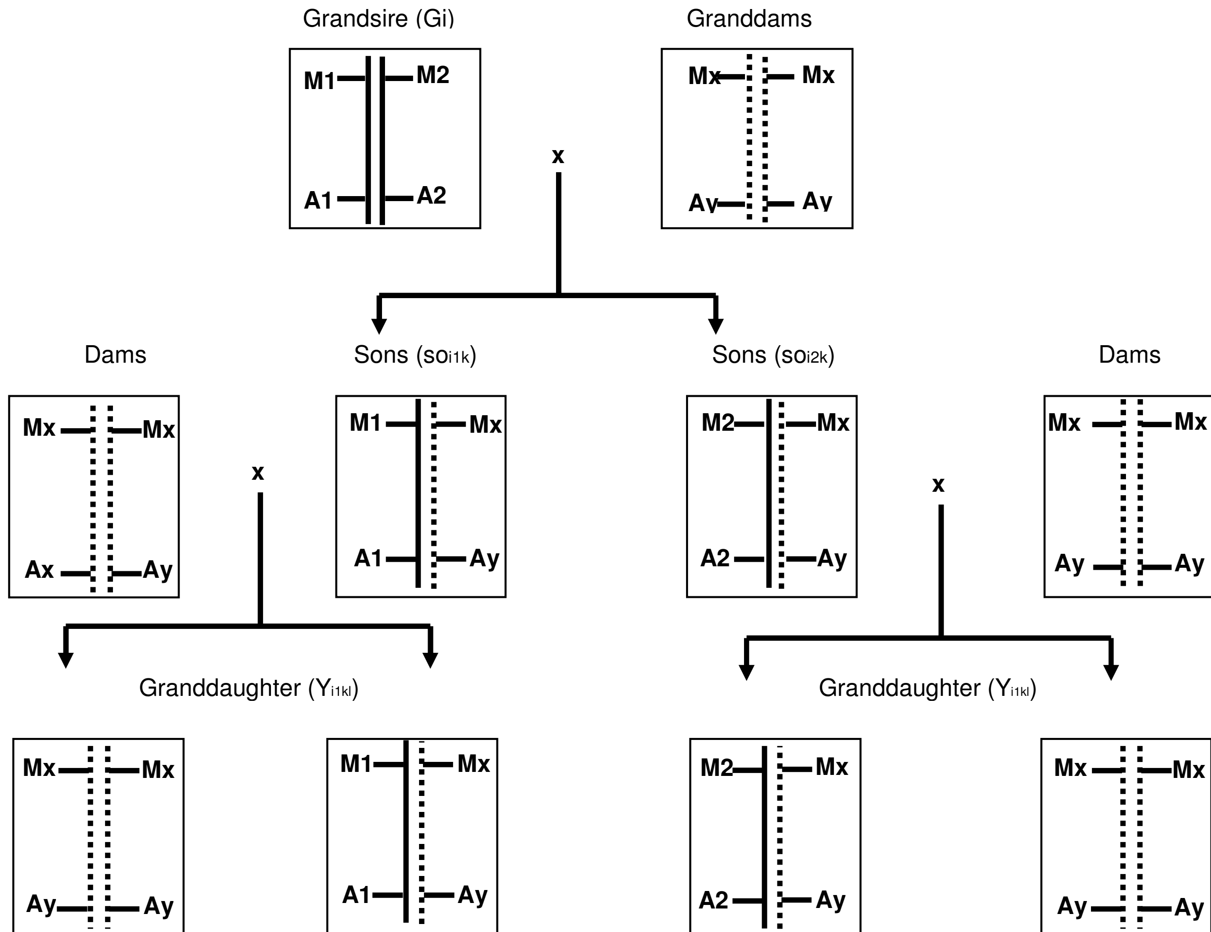



**Figure 2** Diagram of the granddaughter design, M1 and M2*,* marker alleles from the grandsire, Mx*.* marker alleles from the granddams and dams; A1 and A2*,* QTL alleles from the grandsire; Ay, QTL alleles from the granddams and dams, : ——— chromosomes from the sire, ······· :chromosome from the granddams and dams.

In general, the advantage of GDD over DD is that fewer marker assays are needed for equivalent power, it may be easier to collect blood or semen samples from sons of sires, concentrated in AI centers, than from their daughters, scattered over many farms, also The genotyping costs will be higher in DD than GDD because there are fewer animals to genotype in GDD (Weller et al.1990).

There are a number of published QTL mapping studies in cattle using GDD (e.g. Coppieters et al. 1998; Olsen et al. 2002; Freyer et al. 2003; Schrooten et al. 2004; Kolbehdari et al. 2005; Georges 2007; Druet et al. 2008 etc.) or using DD (e.g. Lipkin et al. 1998, Mosig et al. 2001; Chen et al. 2006 etc.).

#### 2.2.1.3. Complex Pedigree Design

A large complex pedigree can provide a powerful design for mapping complex traits; compared with a collection of independent nuclear families, a single pedigree may contain more linkage information and will provide greater opportunities for identifying genotyping mistakes. Large pedigrees from recently founded populations may be especially valuable, as the individuals who demonstrate a specific common characteristic are more likely to share common ancestry than those in admixed populations (Garner et al. 2001).

QTL mapping in complex pedigrees is challenging, because the number of alleles segregating at the QTL are unknown, the marker phases may be unknown or partially known, the marker and QTL allele frequencies must be estimated from the data, inbreeding loops that can exist in pedigree and markers may be uninformative or not genotyped (George et al. 2000). The implication of QTL genotypes from marker genotypes is noticeably more complicated in complex pedigrees than simple half-sib designs, as marker alleles need to be tracked over many generations. This can lead to a large number of missing genotypes. Considerable effort has been invested in creating strategies to infer genotypes with missing marker information and complex pedigrees (e.g. Kerr & Kinghorn 1996; Heath 1997). Most strategies now use simulation- based methods, predominantly Markov Chain Monte Carlo (MCMC) approaches. A review of such methods is available in George et al. (2000). Statistical methods that can fully account for the complex relationships between individuals are expected to provide greater power to detect QTL (Almasy & Blangero 1998). Regression methods, Maximum-Likelihood methods and variance component methods have been proposed for the analysis of complex pedigrees (Hoeschele et al. 1997; George et al. 2000; Visscher 2003; Hayes et al. 2005). The variance component approach may be the most useful method of analysis due to its flexibility (George et al. 2000).

### 2.2.2. QTL Mapping Strategies

There are two main approaches have been used to identify genes affecting traits of interest: candidate gene approach and genome scan approach based on linkage mapping with anonymous DNA markers (Haley 1999; Andersson 2001).

#### 2.2.2.1. Candidate Gene Approach

The candidate gene approach assumes that a gene involved in the physiology of the trait could harbour a mutation causing variation in that trait. The candidate gene approach ideally identifies the causative genes behind QTL, but more likely, it identifies markers that are close enough to the causative mutation that they are in linkage disequilibrium across the population (Dekkers 2004). The candidate gene approach has been applied to different genes in cattle. Lien et al. (1995) studied the casein gene in a granddaughter design and showed an association of casein haplotypes with yields of milk and milk protein. Lagziel et al. (1996) used a single strand conformation polymorphism method for detection of polymorphisms and definition of intragenic haplotypes in the bovine growth hormone gene in the Israeli Holstein cattle population. One haplotype was found to have a highly significant positive effect on milk protein percentage. Grisart et al. (2002) and Winter et al. (2002) reported the first positional cloning of a QTL in cattle that is associated with a significant increase in milk fat yield and a decrease in milk protein yield.

The difficulty in using the candidate gene approach is there must be an understanding of the biological mechanism controlling the trait before being able to select the potential candidate gene (Kwon & Goate 2000). The physiology underlying a phenotypic trait is often complicated and specific mechanisms are sometimes unknown. There are two problems with the candidate gene approach; firstly, there are usually a large number of candidate genes affecting a trait, so many genes must be sequenced in several animals and many association studies carried out in a large sample of animals (the likelihood that the mutation may occur in non-coding DNA further increases the amount of sequencing required and the cost). Secondly, the causative mutation may lie in

a gene that would not have been regarded *a priori* as an obvious candidate for this particular trait.

#### 2.2.2.2. Genome Scan Approach

An alternative is the genome scan approach, which does not require any prior knowledge of the underlying mechanism controlling the trait and is advantageous when the actual physiology of the trait is not yet clarified. In a genome scan, markers evenly spaced throughout the genome are selected and genotyped in a segregating population to detect an association between inheritance of the markers and expression of a phenotypic trait. The goal of the genome scan is to discover a region of the chromosome that has an effect on a phenotypic trait (Dekkers & Hospital 2002). This approach is powerful because even though the physiology is unknown, there is still the potential to detect a locus with a large effect on the trait of interest (Haley & Archibald 1998).Using statistical methods, QTL can then be identified and their position and effect estimated by associating marker data to phenotypic records. The precision of, in particular, estimates of QTL position that can be obtained from these approaches is, however, limited, and large population sizes are needed . A review of publications from the last decade demonstrates that a genome scan using anonymous markers has been used as the primary approach for QTL mapping in dairy cattle (Mosig et al. 2001; Bovenhuis & Schrooten 2002; Khatkar et al. 2004).

The main weakness of a genome scan is that, after the analysis, the specific gene that controls the quantitative trait is still unknown; only a region of the chromosome affecting the trait can be found and it may take a long time to identify the causative gene (Streelman & Kocher 2000). Another disadvantage of the genome scan is the large number of markers required to cover the entire genome (Haley & Archibald 1998). Genotyping huge numbers of animals with many markers can be fairly time consuming and expensive. Although the candidate gene and the genome scan approach are often viewed as alternate approaches for identifying genes of interest, it is clear that they can be complementary, with a genome scan identifying chromosomal regions that harbour potential QTL, followed by further investigation of genes known to be located in that region using the candidate gene approach.

### 2.3. Principles of Linkage Mapping

Linkage equilibrium and linkage disequilibrium, are terms used for the chance of coinheritance of alleles at different loci. Alleles that are in random association are said to be in linkage equilibrium. The chance of finding one allele at one locus is then independent of finding another allele at another locus. Linkage is used in mapping of genes, a concept that has been used since the beginning of the 20th century through the work of Thomas Morgan. Toward the end of the same century, abundant genetic markers were developed, mainly as microsatellites, and these could be used for systematic scanning of the whole genome for gene mapping (Van der Werf et al. 2007).

When DNA markers are available, they can be used to determine if variation at the molecular level (allelic variation at marker loci along the linkage map) is linked to variation in the quantitative trait. If this is the case, then the marker is linked to, or on the same chromosome as, a quantitative trait locus which has allelic variants causing variation in the quantitative trait. The main disadvantages of this method are: first, we do not receive separate estimates of QTL location and QTL effect; second, we must discard individuals whose genotypes are missing at the marker; third, when the markers are widely spaced, the QTL may be quite far from all markers, and so the power for QTL detection will decrease (Broman 2001). A solution to both detect QTL and estimate effect and location simultaneously is to apply information about marker intervals in an approach called interval mapping (Lander & Botstein 1989).

Interval mapping is the most popular approach for QTL mapping in experimental crosses. In interval mapping a genetic linkage map is used as the framework for testing the presence of QTL in fixed intervals between marker pairs whose position are known. The position that best explains the phenotypic difference between genotypic classes pinpoints the most likely QTL position. Compared to methods which consider only a single marker at a time, Interval mapping has several advantages: first, it provides a curve which indicates the evidence for QTL location; second, it allows for detection of QTLs to positions between markers; third, it provides improved estimates of QTL effects; fourth, and

perhaps most important, appropriately performed interval mapping makes proper allowance for incomplete marker genotype data.

Linkage analysis is often used for the mapping of quantitative trait loci (QTL), where the inheritance of chromosomal regions within the data set is traced by markers (Hoeschele et al. 1997). The region whose inheritance explains most of the variance of the phenotypic records indicates the most likely position of the QTL. To position the QTL, linkage mapping uses only the recombinations that occurred within the data set, which typically contains two to three generations. With closely linked markers, there will be few recombinations between adjacent markers during these two to three generations and hence a dense marker map will provide little extra information about the position of the QTL, unless the number of individuals per generation is very large (Darvasi et al. 1993).

In outbred populations only a proportion of individuals will be heterozygous for a given marker (or QTL) and the probability that an individual is heterozygous for both a marker and QTL can be small. Thus the information content varies from interval to interval causing biased QTL location estimates because these tend towards the most informative marker rather than the correct one (Haley et al. 1994). One solution to overcome this problem is to use information from multiple markers simultaneously (Georges et al. 1995; Knott et al. 1996). In principle, if one marker of a marker pair is uninformative it can be replaced with another linked and informative marker. In dairy cattle the most commonly used QTL mapping methods are based on interval mapping with multiple marker information (Georges et al. 1995; Knott et al. 1996). In a half sib pedigree the conditional probability of an offspring inheriting the alternative alleles of a sire's homologues is calculated in every position along the chromosome at fixed intervals. These probabilities are then used usually in maximum likelihood (Georges et al. 1995) or linear regression (Knott et al. 1996) methods to statistically test the presence of QTL under a null hypothesis of "no QTL". QTL analysis in a half sib pedigree is nested within families because the linkage relationship between marker and QTL alleles can differ from one family to another (Haley & Andersson 1997). This means that the same marker allele can be linked to a QTL allele of positive effect in one family and to a QTL allele of negative effect in another family. If the analysis is carried out across all sires

instead of within families there is a risk that the segregation of QTL is masked by opposing linkage relationships between QTL and marker alleles in different families.

The often used approach to derive significance thresholds for QTL mapping methods is the permutation test by Churchill & Doerge (1994) and Doerge & Churchill (1996). The permutation test can be used to create empirical distributions of the test statistics for the data under study under the null hypothesis of “no QTL”. The empirical distribution is obtained from a collection of simulated data sets that are created from the real data by randomly shuffling the phenotypes of the individuals. In a half sib design this is done within families. QTL linkage analyses are then performed with simulated data and the highest test statistics are stored and ranked. The resulting empirical distribution of the test statistics is then used to determine the chromosome-wise or experiment-wise significance level of the observed QTL signal.

The statistical certainty of QTL position can be expressed as confidence intervals (CI). A bootstrap method can be applied to estimate the CI of QTL positions (Visscher et al. 1996). Here the real data from $N$ individuals are used to create new data sets of $N$ individuals (bootstrap samples) by sampling so that some individuals can be randomly represented multiple times. The process is repeated $N$ times to create $N$ bootstrap samples. Then the interval mapping is used to detect QTL from the bootstrap samples and estimates of the most likely QTL positions are ordered. The 95% CI is the chromosomal segment that comprises 95% of the observations around the empirical centre of the most likely positions of the bootstrapped samples.

### 2.4. QTL Fine Mapping

In cattle the confidence intervals for the most likely QTL positions are usually tens of centimorgans (Georges 2007), which are too large to efficiently implement technologies such as marker assisted selection, marker assisted introgression, positional cloning or positional candidate gene identification. Fine mapping will refine the size of QTL harbouring regions, which will open the way for more efficient implementations of above technologies and will allow proceeding with high levels of accuracy and precision. The main factor limiting

mapping resolution in linkage analysis is the frequency of observable recombination in the genotyped progeny. The mapping resolution can be improved by increasing the number of recombinations using larger families or advanced generations (Darvasi & Soller 1995). Thousands of offspring are required to reduce the QTL interval to such a level that positional cloning or gene identification is possible (Darvasi 1998). In outbred populations, the large number of progeny needed for fine mapping may be achievable but is time and cost intensive effort and therefore an alternative strategy is needed. One possibility is to use the recombination events that have occurred in the history of the population (e.g. Hästbacka et al. 1992). This is done by exploiting the population level LD between QTL and closely linked markers.

### 2.4.1. Linkage Disequilibrium

LD is also known as gametic phase disequilibrium, gametic disequilibrium, and allelic association. Simply stated, LD refers to the non-random association of alleles at different loci (Hedrick 2000). LD can be a result of migration, mutation, selection, small finite population size or other genetic events. LD can also be created in livestock populations; in an F2 QTL mapping experiment LD is created between marker and QTL alleles by crossing two inbred lines (Lander & Schork 1994; Hedrick 2000). Steady-state levels of LD are typically higher for tightly linked loci.

The most commonly used LD measures are the multiallelic $D'$ (Lewontin, 1964), $r^2$ (Hill & Robertson, 1968), and $\chi^2$ (Yamazaki 1977; Zhao et al. 2005). The $D'$ measure has been commonly used in LD studies. Using $D'$, extensive LD over long distances was observed in dairy cattle, sheep and pigs (Farnir et al. 2000; McRae et al. 2002; Tenesa et al. 2003; Nsengimana et al. 2004). However, it is known that LD measured by $D'$ tends to be inflated with small sample sizes and/or with low allele frequencies (Ardlie et al. 2002; Du et al. 2007). In addition, there is no clear interpretation for intermediate values of $D'$, because the magnitude of $D'$ strongly depends on sample size, especially for SNP with rare alleles (Ardlie et al. 2002; Du et al. 2007; Sargolzaei et al. 2008). The square of the correlation coefficient between markers, $r^2$, is undisputed LD measure for diallelic markers like SNPs. The $r^2$ preferred to detect markers that might correlate with the QTL of interest, because $r^2$ quantifies the amount of

information about one locus provided by the other (Ardlie et al. 2002; McRae et al. 2002; Flint-Garcia et al. 2003; Zhao et al. 2005; Lipkin et al. 2009). Zhao et al. (2005; 2007) found that standardized $\chi^2$ (henceforth denoted $\chi^2$; Yamazaki 1977) closely tracked the regression of the allelic state at a QTL on the allelic state at a multi-allelic marker and hence conveys the same information for multi-allelic markers as $r^2$ does for diallelic markers.

The existence of LD implies there are small segments of chromosome in the current population which are descended from the same common ancestor. These identical by descent (IBD) chromosome segments will not only carry identical marker haplotypes, if there is a QTL somewhere within the chromosome segment; the IBD chromosome segments will also carry identical QTL alleles. Therefore if two animals carry chromosomes which are likely to be IBD at a point on the chromosome carrying a QTL, then their phenotypes will be correlated (Hayes et al. 2005)

Meuwissen & Goddard (2000) postulated that a base population in linkage equilibrium undergoes a mutation at the QTL, creating a novel QTL-allele embedded in one specific marker haplotype. Due to recombinations in the following generations, the original haplotype will remain only for markers close to the QTL. Thus, in the current generations, these marker alleles will be in linkage disequilibrium with the QTL alleles. The LD can be detected by estimating the effects of the marker haplotypes on the quantitative trait. Haplotypes with identical marker alleles are expected to have a similar effect on the trait because the identical marker alleles imply that the chromosomal region is inherited in a manner that is IBD from an ancient common ancestor, and the haplotypes are therefore expected to carry the same QTL allele.

Meuwissen & Goddard (2001) described a method to calculate the IBD matrix based on deterministic predictions which took into account the number of markers flanking the putative QTL position that are identical by state, the extent of LD in the population based on the expectation under finite population size, and the number of generations ago that the mutation is presumed to have occurred.

The knowledge of the extent and the pattern of LD throughout the bovine genome plays an important role in gene mapping and genome-wide association

studies and is a fundamental tool for: exploring the degree of diversity among breeds of cattle, inferring the distribution of crossing-over, and identifying regions of genome that have been subject to selective sweep (Bohmanova et al. 2010). LD mapping has been used extensively to identify genes for monogenic diseases in humans (Peltonen 2000). Contrary to the situation in humans, extensive LD over a long range was observed in dairy cattle, sheep, and pigs (Farnir et al. 2000; McRae et al. 2002; Tenesa et al. 2003; Nsengimana et al. 2004; Harmegnies et al. 2006; Khatkar et al. 2006; Taberlet et al. 2008) because of limited effective population sizes. Above studies used *D'* as a measure of LD between multi-allelic markers which overestimate true LD (Lipkin et al. 2009). Nevertheless, LD estimated by $\chi^2$ for multi-allelic marker and $r^2$ for diallelic markers is clearly lower than this estimated by *D'* (Lipkin et al. 2009) but still higher in farm animals than in humans (Bovine HapMap Consortium 2009).

Fine mapping of QTL exploiting LD was carried out and described in several studies (e.g. Riquet et al. 1999; Meuwissen & Goddard 2000; Meuwissen & Goddard 2001; Farnir et al. 2002; Meuwissen et al. 2002; Meuwissen & Goddard 2004). The occurrence of gametic-phase disequilibrium between nonsyntenic loci raises the concern about false positive result when using LD as the only means to locate genes underlying complex traits in these populations. So the preference should be given to the combined linkage and LD methods (Farnir et al. 2000).

### 2.4.2. Combined Linkage Disequilibrium and Linkage (LDL) mapping

Linkage analysis and linkage disequilibrium analysis are complementary methods. Farnir et al. (2002) and Meuwissen et al. (2002) describe for the first time the benefits from the use of combined LD and linkage mapping (LDL). The method of Meuwissen et al. (2002) allows simultaneous estimation of variance components associated with the marked QTL, with background genes, and with a residual or error variance due to environmental effects on the trait under study. This approach allows the utilization of recombinations occurring both within and outside the pedigreed and genotyped generations (i.e., linkage analysis and linkage disequilibrium analysis, respectively) and also accounts for unknown background genes. Additionally, by combining LD and linkage

analysis spurious long distance associations can be avoided because the linkage analysis information will not confirm such associations (Meuwissen et al. 2002; Meuwissen & Goddard 2004). In comparison with a classical linkage analysis, the method can make use of additional relationships within and between families, increasing the power of QTL detection. Furthermore, utilizing historical recombination events may result in a more accurate mapping of QTL, depending on marker density and the pattern of linkage disequilibrium in the population under study.

The efficiency of combined linkage disequilibrium and linkage (LDL) method was tested by means of simulation studies (Lee & van der Werf, 2004, 2005, 2006). Several authors were carried out and described combined LDL in fine mapping of QTL (Farnir et al. 2002; Meuwissen et al. 2002; Blott et al. 2003; Olsen et al. 2005; Schnabel et al. 2005; Gautier et al. 2006; Awad et al. 2010). The efficiency of LDL mapping is demonstrated by the enormous reduction of the confidence interval of a mapped QTL. For example, Meuwissen et al. (2002) used this method to reduce the CI for a QTL affecting twinning rate on BTA5 to <1 cM. the CI for the milk QTL on BTA14 was reduced to 3 cM by this method (Farnir et al. 2002), and Olsen et al. (2005) using LDL mapping, reduced the confidence interval for a QTL, affecting milk production traits, from 7.5 cM to 420 kb on BTA06 in Norwegian dairy cattle. Furthermore, they recently contributed to the characterization of a causal mutation in genes (Grisart et al. 2002; 2004; Winter et al. 2002; Blott et al. 2003; Schnabel et al. 2005)

**2.5. Identification of Candidate Genes**

The high-resolution mapping will map QTL to intervals that contain several genes and numerous DNA sequence variants. One of the greatest challenges is to determine which gene(s) and nucleotide variants (quantitative trait nucleotide(s), QTN(s)) are causing the QTL effect (Mackay 2001a). Genes that lie within the CI of the QTL and that have physiological relevance to the trait of interest should be considered as primary candidates for the QTL. The criteria of gene candidates for QTL:

- The gene has a known physiological role in the phenotype of interest;

- The gene affects the trait in question based on studies of knock-outs, mutations or transgenic in other species;
- The gene is preferentially expressed in organs related to the quantitative trait and
- The gene is preferentially expressed during developmental stages related to the phenotype (Ron & Weller 2007).

It is essential to find the relevant gene, however, this doesn't give information about the molecular nature of the QTL and therefore the QTNs that are the actual cause of the observed effect in the trait phenotype need to be identified (Mackay 2001a). The difficulty is that each gene may include numerous DNA sequence variants, some of which are located in coding region and others in the flanking genomic regions (Glazier et al. 2002). Another complicating factor is that the QTN does not have to be locate in close proximity to the gene it influences, it might be tens of kilobases away in an intragenic region (Freking et al. 2002; Smit et al. 2003) or even in another, functionally unrelated gene (Higgs et al. 1990).

There is no simple approach to facilitate the identification of functional candidates based on sequence information only because our knowledge about sequence characteristics, especially in regulatory regions, is poor. It has been speculated that the variation underlying complex traits is more often regulatory than coding e.g. (Mackay 2001b). Therefore the optimal strategy to search for causal QTN(s) is to consider each nucleotide variant as well as their combination in one or several genes (Glazier et al. 2002). One way to search for causal QTNs (or genes) is to systematically test the association of detected sequence polymorphisms and the phenotype of interest, preferably in different populations (Flint & Mott 2001). Association mapping can be used to systematically screen candidate loci in an interval defined by linkage mapping, or to evaluate associations at candidate loci even in the absence of linkage information. The simplest designs for evaluating association between markers at a candidate gene and a quantitative trait only require a sample of individuals from the population of interest, each of whom has been genotyped for the marker loci and evaluated for the trait phenotype (Mackay 2001a). In the case-control design for dichotomous traits, such as disease susceptibility, the

population sample is stratified according to disease status, and LD between a marker and the trait is revealed by a significant difference in marker allele frequency between cases and controls (Cardon & Bell 2001). For continuously distributed traits, the population sample is stratified by marker genotype, and marker-trait LD is inferred if there is a significant difference in trait mean between marker genotype classes. Testing for trait associations of multiple markers again requires that an appropriate downward adjustment of the significance threshold be made. Permutation tests developed in the context of other single-marker genome scans are ideal in this regard (Churchill & Doerge 1994; Doerge & Churchill 1996; Long et al. 1998; Lyman et al. 1999).

A review of recent publications shows that many QTL have been mapped for traits of economic importance in dairy cattle (Khatkar et al. 2004; www.animalgenome.org). However, despite the large number of QTL studies in cattle and other species, little progress has been made on the identification of major genes affecting milk production and health traits in dairy cattle: identification of causal mutations of DGAT1 on BTA14 (Grisart et al. 2002; Winter et al. 2002; Kühn et al. 2004), GHR on BTA20 (Blott et al. 2003), ABCG2 (Cohen-Zinder et al. 2005) or SPP1 (Osteopontin) gene (Cohen-Zinder et al. 2004; Schnabel et al. 2005) on BTA6.

One major limitation when choosing a candidate gene is the large number of provisional genes present in most QTL regions. Considering the current achievements QTN identification seems to be a challenging task. However, the ongoing increase of biological information and the rapid technological development of functional genomics might enable some of the limitations of QTN identification to be overcome in the future (Viitala 2008).

## 2.6. Marker Assisted Selection (MAS), Gene Assisted Selection (GAS) and Genomic Selection (GS)

The application of molecular genetic information has become an important issue in animal breeding. Breeding strategies for livestock that utilize molecular genetic information, genes or genomic regions, are broadly referred to as marker-assisted selection (MAS; reviewed by Dekkers & Hospital 2002; Dekkers 2004). Until recently, the use of molecular genetics in commercial

applications of marker-assisted selection (MAS) has focused on the use of individual genes or quantitative trait loci (QTL) linked to markers (Dekkers & Hospital 2002; Dekkers 2004). With the exception of a few genes with relatively large effects such as DGAT1 (Grisart et al. 2004), most candidate genes or QTL capture only a very small proportion of the total genetic variance. Recent empirical genome-wide association (GWAS) studies using a high-density SNP technology in humans (*e.g.* Burton et al. 2007; Weedon et al. 2008); mice (Valdar et al. 2006) and cattle (Cole et al 2009) suggest that complex traits are most likely affected by many genes with a small effect.

The success of MAS is influenced by the relationship between the markers and the genes of interest. Dekkers (2004) distinguished three kinds of relationship:

1. Direct markers: The marker is located within the gene of interest (GAS). This is the most favourable situation for MAS since, by following inheritance of the marker alleles, inheritance of the QTL alleles is followed directly. On the other hand, these kinds of markers are the most uncommon and are thus the most difficult to detect because causality is difficult to prove and, as a result, a limited number of examples are available, except for single-gene traits (Andersson 2001).
2. LD-markers: The marker is in linkage disequilibrium (LD) with functional mutation throughout the whole population. Population wide LD can be found when markers and genes of interest are physically very close to each other and/or when lines or breeds have been crossed in recent generations.
3. LE-markers: The markers that are in population wide linkage equilibrium with functional mutation in outbred populations. The LE markers can be detected on a genome-wide basis by using breed crosses or analysis of large half-sib families within the breed. Many examples of successful applications of this methodology for detection of QTL regions are available in the literature (Andersson 2001).

The three types of marker differ in their application in selection programs. Direct markers and, to a lesser extent, LD markers, allow for selection on genotype across the population because of the consistent association between genotype and phenotype, use of LE markers must allow for different linkage phases between markers and QTL from family to family. Thus, the ease and ability to

use markers in selection is opposite to their ease of detection and increases from direct markers to LD markers and LE markers. In what follows, selection on these three types of markers will be referred to as gene-assisted selection (GAS), LD markers-assisted selection (LD-MAS), and LE marker-assisted selection (LE-MAS) (Dekkers 2004).

A new form of marker assisted selection called genomic selection (Meuwissen et al. 2001), which refers to selection decisions based on genomic breeding values (GEBV). The GEBV are calculated as the sum of the effects of dense genetic markers, or haplotypes of these markers, across the entire genome, thereby potentially capturing all the quantitative trait loci (QTL) that contribute to variation in a trait. The QTL effects, inferred from either haplotypes or individual single nucleotide polymorphism markers, are first estimated in a large reference population with phenotypic information. In subsequent generations, only marker information is required to calculate GEBV (Hayes et al. 2009). Simulation studies have shown that genomic selection can lead to high correlations between predicted and true breeding value over several generations without repeated phenotyping (Meuwissen et al. 2001; Habier et al. 2007). Therefore, genomic selection can result in lower costs and increased rates of genetic gain.

In cattle, an assay for simultaneous genotyping of more than 50,000 SNP markers is commercially available from the beginning of 2008. This opens an opportunity for effective selection using dense markers through the whole genome (i.e., genomic selection). Genomic selection is based on breeding values that are directly estimated from genome-wide dense marker panels. Therefore, genetic evaluation can be performed as soon as DNA is obtained, which allows accurate selection in both genders early in life (Su et al. 2010). It is expected that by using genomic selection in dairy cattle breeding, the genetic progress would be doubled whereas the cost for proving bulls would be reduced by 92% (Schaeffer 2006). Currently there are tens of different research project evaluating capability and suitability for continuous use of genomic selection in cattle.

# CHAPTER 3

## Confirmation and Refinement of a QTL on BTA5 affecting Milk Production Traits in the Fleckvieh Dual Purpose Cattle Breed

A. Awad[1], I. Russ[2], R. Emmerling[3], M. Förster[1,2], and I. Medugorac[1]

[1] Chair of Animal Genetics and Husbandry, Faculty of Veterinary Medicine, The Ludwig-Maximilians-University Munich, Veterinärstr. 13, 80539 Munich, Germany

[2] Tierzuchtforschung e.V. München, Senator-Gerauer-Straße 23, D-85586 Grub, Germany

[3] Bavarian State Research Center for Agriculture, Institute of Animal Breeding, Prof.-Dürrwaechter-Platz 1, 85586 Grub, Germany

**Published in Animal Genetics 2010, 41 (1): 1-11**

## Summary

We analysed a QTL affecting milk yield (MY), milk protein yield (PY) and milk fat yield (FY) in the dual purpose cattle breed Fleckvieh on BTA5. Twenty six microsatellite markers covering 135 cM were selected to analyze nine half-sib families containing a total of 605 sons in a granddaughter design. Thereby we assigned two new markers to the public linkage map using CRI-MAP program. Phenotypic records were daughter yield deviations (DYD) originating from the routinely performed genetic evaluations of breeding animals. To determine the position of the QTL, three different approaches were applied: interval mapping (IM), linkage analysis by variance component analysis (LAVC) and combined linkage disequilibrium (LD) and linkage (LDL) analysis. All three methods mapped the QTL in the same marker interval (*BM2830-ETH152*) with the greatest test-statistic value at 118, 119.33 and 119.33 cM respectively. The positive QTL allele simultaneously increases DYD in the first lactation by 272 kg milk, 7.1 kg milk protein and 7.0 kg milk fat. Although the mapping accuracy and the significance of a QTL effect increased from IM over LAVC to LDL, the confidence interval was large (13, 20 and 24 cM for FY, MY and PY, respectively) for the positional cloning of the causal gene. The estimated averages of pair-wise marker LD with a distance < 5 cM was low (0.107) and reflects the large effective population size of the analysed Fleckvieh subpopulation. This low level of LD suggests a need for increase in marker density in following fine mapping steps.

**Key words:** Cattle chromosome 5, linkage disequilibrium, milk production traits, QTL mapping

## Introduction

Milk production traits in dairy cattle are controlled by large numbers of quantitative trait loci (QTL) and environmental factors (Zhang et al. 1998). The presence of a QTL is detected by mapping studies, which show significant differences in phenotypes between individuals receiving different marker alleles in appropriately designed mapping populations (Andersson 2001). In dairy cattle, QTL mapping utilizes existing half-sib breeding populations, which are routinely produced by artificial insemination. To use the existing outbred half-sib data structure Weller et al. (1990) proposed the granddaughter design (GDD) to detect linkage between a single marker and a QTL. There are numerous studies detecting QTL in cattle using GDD. Many of these were recently reviewed in a meta-analysis performed by Khatkar et al. (2004). Mapped QTL can be accessed in the QTL databases (http://www.animalgenome.org/QTLdb/cattle.html, http://bovineqtl.tamu.edu/).

Associations between marker polymorphism and trait variation can be assessed using single or multiple marker genotypes. The usual methods are based on regression to detect linkage between adjacent informative markers and QTL (Lander & Botstein 1989; Knott et al. 1996). The interval mapping (IM) approach applied to the most common mapping designs in cattle (GDD and DD) uses only linkage information and recognise only recombination events in the paternal gametes. Only a small number of such events are expected in a short chromosomal region, resulting in large confidence intervals for the QTL. Increasing the number of markers within the same IM design will only detect a small number of previously undetected recombination events, resulting in minimal increase in mapping resolution (Darvasi et al. 1993). In contrast, linkage disequilibrium (LD) mapping methods are based on historical recombination events between genetic markers (Riquet et al. 1999; Meuwissen & Goddard 2000). The classical linkage and LD analysis are complementary methods. Farnir et al. (2002) and Meuwissen et al. (2002) proposed that these two analyses should be applied simultaneously in one model. The method of Meuwissen et al. (2002) allows simultaneous estimation of variance components associated with the marked QTL, with background genes, and with a residual or error variance due to environmental effects on the trait under

study. In comparison with a classical regression analysis, the method can make use of additional relationships within and between families, increasing the power of QTL detection. Furthermore, utilizing historical recombination events may result in a more accurate mapping of QTL, depending on marker density and the pattern of linkage disequilibrium in the population under study. The simultaneous development of high-throughput genotyping and fine-mapping techniques has improved the development of dense marker and QTL maps (e.g. Ihara et al. 2004; Gautier et al. 2006). The efficiency of combined linkage disequilibrium and linkage (LDL) mapping is demonstrated by the enormous reduction of the confidence interval for location of a mapped QTL. For example, Olsen et al. (2005) using LDL mapping, reduced the confidence interval for a QTL, affecting milk production traits, from 7.5 cM to 420 kb on BTA06 in Norwegian dairy cattle.

Bovine chromosome 5 (BTA5) has been shown to harbour QTL that influence milk production traits (e.g. Olsen et al. 2002; Bennewitz et al. 2003; Ashwell et al. 2004), reproduction (Kappes et al. 2000; Lien et al. 2000) and growth and carcass traits (e.g. Casas et al. 2000; Gutierrez-Gil et al. 2009). In an initial study in our laboratory (unpublished), ten markers and 14 GDD families of the Fleckvieh breed were used for QTL mapping on BTA5. Two QTL affecting milk production traits were identified, one at the proximal and the second at the distal regions of BTA5, respectively. The aim of the present study is to confirm and to refine map location of the distal BTA5 QTL. To achieve this, we designed a set of 18 microsatellite markers spanning the distal region of BTA5 and used this to map nine GDD families with respect to milk yield (MY), milk fat yield (FY), and milk protein yield (PY).

## Materials and methods

### Families, Phenotypic Data and DNA Samples

Nine paternal half-sib families of the Fleckvieh dual purpose cattle breed were analysed in a GDD (Weller et al. 1990). The total number of sons in the study was 605, ranging from 42 to 140 sons per grandsire, with an average of 67 sons. Genetic evaluations were routinely performed by a multiple lactation random regression test-day model for the joint Fleckvieh population in Germany

and Austria (Emmerling et al. 2002). For each of the sons, daughter-yield deviations (DYD) in the first, second and third lactations for milk yield (MY1; MY2; MY3), milk fat yield (FY1; FY2; FY3), and milk protein yield (PY1; PY2; PY3) were derived from the November 2008 routine genetic evaluations.

In addition, we sampled all available male ancestors of GDD sires, some important maternal grandsires (MGS) of the sons, and two very important recent founders of the Fleckvieh population. The first is a pure bred Fleckvieh bull born in 1966 which was used intensively in artificial insemination over a long period. The pedigree of a sample of 500 bull-dams born from 1998 to 2001 (dams of the current bull generation) was traced back and this founder was detected in 98% of pedigrees. The second, more recent founder is a 50% Red-Holstein (RH) bull used in the early 1980's to improve the milk yield and udder quality of the Fleckvieh population. This founder was present in 60% of the above mentioned 500 bull-dam pedigrees. To trace the haplotypes of both founders, we sampled and genotyped both founders, six very important sons for each founder, the sire of the 50% RH founder (which was 100% RH), and 69 other ancestors connecting the nine GDD families with founders into a complex pedigree. Pedigree information was important, as systematic ancestry differences can cause spurious associations in mapping studies using LD information (e.g. Lander & Schork 1994). The possible influence of 50% RH bull in our results will be discussed later.

Genomic DNA was prepared from semen using standard methods based on ethanol precipitation and from whole blood samples using the QIAamp Blood-Kits (Qiagen) following the manufacturer's protocol.

**Marker selection, genotyping and plausibility control**

This study includes ten evenly distributed markers (Set0, Table 1) genotyped in an earlier project for initial QTL mapping (unpublished results). Six of the ten markers were not genotyped in all nine GDD families. For confirmation and fine mapping of the QTL at the distal end of BTA5 18 evenly distributed microsatellite markers were added, covering the region from 40 cM to the end of the chromosome. For 26 of the 28 markers, relevant information was obtained from the MARC-ARS-USDA public database at http://www.marc.usda.gov

(Ihara et al. 2004). Primers were optimized (Program Primer3 v.0.4.0) according to the current available bovine genome sequence data and the appropriate fragment size in the multiplex marker set. The remaining two markers (*LMU0501* and *LMU0502*) were newly designed. The quality of the new markers was tested by genotyping a small set of animals. The eighteen added markers were divided into two PCR multiplex sets (Table 1, Set1 and Set2) that were combined after PCR for electrophoresis and fragment analysis. PCR reactions were performed in a 15 µl final volume using Primus 96$^{plus}$ (MWG-Biotech) and PTC-100 (MJ Research). The fragment analysis of the resulting PCR products was performed on ABI377 Sequencers. Genotypes were assigned using *GENESCAN* and *GENOTYPER* (Applied Biosystems) software programs.

We performed independent double genotyping of GDD families and ancestors. In case of questionable genotypes, the raw data were re-evaluated and if necessary all animals involved were genotyped again. As a further quality control, haplotypes derived by *QTL EXPRESS* (based on half-sib design) as well as *CRI-MAP* (based on the complex pedigree) were examined for double recombination within a relatively short chromosomal region (~20 cM). Genotypes presenting a double recombination signal were re-evaluated and if necessary genotyped again.

**Linkage map construction**

The relative positions of the 26 public markers were re-evaluated by the *CRI-MAP* program (Green et al. 1990). In addition, a physical map was constructed according to the sequence data of all markers (Table 1), using the basic alignment search tool (BLAST) and the latest cattle genome sequence (http://genome.ucsc.edu/cgibin/hgGateway). The linkage and physical map were used as the framework for insertion of the two newly designed markers (*LMU0501* and *LMU0502*) by *CRI-MAP* option build. The resulting final map (Table 1) was used for all following analyses.

**QTL mapping Analysis**

**Linkage Analysis (LA) by classical regression interval mapping (IM)**

Linkage analysis was done by regression interval mapping (Knott et al. 1996) with the half-sib option of *QTL EXPRESS* program (Seaton et al. 2002,

http://qtl.cap.ed.ac.uk). One QTL model with an additive effect was fitted along BTA5 by regressing the phenotype (DYD) on the conditional probability of inheriting the sire's alternative haplotype. An across-family analysis was conducted using all 9 GDD families. The regression coefficient was calculated at each cM along the BTA5 and the position with the maximum F-ratio was considered to be the most likely QTL location. Each family was individually analyzed by using *QTL EXPRESS* to determine the sire's QTL segregation status for each trait. The families detected as segregating for the distal QTL by family-wise analysis were used for an additional across-family analysis as described above, based on these families only. Chromosome-wise significance levels (*Pchr*) for the across-family and within family analyses were obtained by carrying out 10 000 permutations (Churchill & Doerge 1994). To determine the 95% confidence interval (CI) for QTL position we generated and analyzed 10 000 bootstrap replicates by *QTL EXPRESS*.

**Linkage Analysis by variance component approach (LAVC approach)**

*Mixed linear model*: The genetic model that is used in the LAVC approach is relatively general. A vector of phenotypic observations is modelled as a linear function of the effects of QTL, a polygenic term representing the sum of other unidentified genetic effects, fixed effects and residuals:

$$y = X\beta + Zu + \sum_{i=1}^{NQ} Zq_i + e$$

Where **y** is a vector of *N=605* DYDs of the trait of interest, **β** is a vector of fixed effects, **u** is a vector of *n* random polygenic effects for each animal, *NQ=n* is the number of QTL effects, $\mathbf{q}_i$ is a vector of *n* random effects due to the putative QTL at position *i* and **e** is a vector of residuals. Analyses of un-weighted DYD give results of high similarity with weighted and un-weighted EBV by LAVC and LDL. Using of weighted DYD in LAVC and LDL analysis show the convergence problems especially for FY. To omit additional figures and tables which finally led to the same conclusions we prefer to mention only results based on un-weighted DYD here. The random effects (**u**, $\mathbf{q}_i$ and **e**) are assumed to be normally distributed with mean zero and variances $\mathbf{A}\sigma_u^2$, $\mathbf{G}_i\sigma_{q_i}^2$ and **R**, respectively, where **A** is the numerator relationship matrix based on recorded pedigree, $\mathbf{G}_i$ is the additive genotype relationship matrix whose elements are

the identity by descent (IBD) probabilities at putative QTL position *i* conditional on marker information, and **R** is the covariance matrix among residual effects, assumed diagonal in this study (**R=I**$\sigma_e^2$). **X** and **Z** are incidence matrices for the effects **β** and **u** and $\mathbf{q}_i$, respectively. We used the program *LOKI* version 2.4.5 (http://www.stat.washington.edu/thompson/Genepi/Loki.shtml) to estimate IBD probabilities. The data set of 605 phenotyped and genotyped sons from nine GDD families interwoven into a complex pedigree by 78 genotyped and 828 not genotyped (mostly dams) ancestors (total of 1511 individuals) was included in a variance component (VC) analysis. To implement this, *LOKI* applies a Bayesian Markov Chain Monte Carlo (MCMC) algorithm (Heath 1997) and uses all linkage information in the known part of the complex pedigree to estimate IBD probabilities. An initial burn-in of 50 000 iterations was followed by 500 000 iterations; with parameter estimates collected for each iteration. We set the total genome length to 2900 cM to fit unlinked QTL and to estimate IBD in the middle position *p* of each marker interval. Then the estimated IBD matrices ($\mathbf{G}_p$) were used for VC analysis by the external program *ASREML* (Gilmour et al. 2000) as well as by average information REML (AIREML; Johnson & Thompson 1995) as implemented in the program *LDL* version 1.46 (Lee & Van der Werf 2006). AIREML uses the variance covariance matrix directly, which is more efficient and robust against dense structures and singular **G**-matrices than the approach based on Mixed Model Equation (Lee & Van der Werf 2006). Here, we used the mixed linear model described above to estimate variance components and likelihood (L) of a model containing a QTL as well as background genes at position *p* ($L_p$). The likelihood of a model without QTL ($L_0$) was calculated on the basis of a model containing only background genes. The log-likelihood ratio was calculated as the double difference in *logL* between the models with and without a QTL, i.e., LRT=-2 ($logL_0$-$logL_p$). The LRT test statistic is distributed approximately as chi-square with 1 degree of freedom (Olsen et al. 2004). The marker bracket with the greatest LRT value was taken as the most likely QTL position.

### Combined linkage disequilibrium and linkage mapping

The combined linkage disequilibrium and linkage (LDL) analysis is also a variance component approach, which uses the same data set and the same

mixed linear model as described for LAVC. Here, we used the program *LDLRAMS* version 1.76 (http://www-personal.une.edu.au/~slee7/program/program.html) to estimate IBD probabilities by a random walk approach and meiosis sampler (Lee et al. 2005). The *LDLRAMS* as well as the program *LDL* version 1.46 (http://www-personal.une.edu.au/~slee7/program/program.html) also performs VC analysis by AIREML (Johnson & Thompson 1995). The differences between LAVC accomplished by *LOKI* and LDL analysis by *LDLRAMS* are in the method by which the IBD probabilities are estimated and used: *LOKI* uses locus Gibbs sampling (i.e. the genotypes for markers and QTL's are simultaneously updated for all individuals, one locus conditional on all other loci, although only one locus at a time) whereas *LDLRAMS* combines meiosis Gibbs sampling and random walk in estimating IBD. This combined method could remedy the reducibility problems in the meiosis Gibbs sampler. Due to more comprehensive coverage of the sampling space it contributes to more accurate estimates and enhances the probability to climb the global maxima. Furthermore, *LOKI* uses linkage information whereas *LDLRAMS* combines LD and linkage information. To estimate LD-based IBD probabilities we assumed the mutation age and past effective population size was 100, initial homozygosity at each marker in the base population was set to 0.35 and allele frequencies were estimated by allele counting within the complex pedigree. We counted both alleles at genotyped unrelated individuals (based on the partly known pedigree) and only the maternal allele of descendents not having genotyped maternal grandsires (MGS). The same complex pedigree of 1511 individuals described above was analysed by *LOKI* and *LDLRAMS*. For LDL analysis by the program *LDLRAMS* we used an initial burn-in of 500 iterations followed by 5000 iterations, with parameter estimates collected at each five iterations. Two independent sampling procedures (i.e. two *LDLRAMS* runs with different random number seed) ensured convergence at a global maximum. The appropriate variance components and LRT along BTA5 were estimated by AIREML (Johnson & Thompson 1995; Lee & Van der Werf 2006). The program *LDL* uses IBD matrices estimated by external programs. The confidence interval (CI) for the QTL position was determined as 1-LOD support interval. In theory, the 1-LOD support interval approximately corresponds to a 97% CI, because 1 LOD is

equivalent to a LRT statistic of 4.605 ($2 \times ln$ (10)), which corresponds to an asymptotic $\chi^2$ *P*-value of 0.032 (Mangin et al. 1994, see also Visscher & Goddard 2004). 1-LOD drop-off support interval was constructed as interval surrounding the QTL peak where the LRT exceeds $LRT_{max} - 2 \times ln$ (10), where $LRT_{max}$ is the maximum LRT-value for the tested QTL.

**Estimation of linkage disequilibrium**

Most of the livestock studies based on microsatellite data used Hedrick's multi-allelic *D'* value (Hedrick 1987) as the measure of LD, while Zhao et al. (2005) found that the standardized $\chi^2$ (Yamazaki 1977; henceforth denoted $\chi^{2\prime}$) is the best predictor of useable LD of multi-allelic markers with QTL. Consequently, in according with Lipkin et al. (2009) we used both of these measures to evaluate LD in our population. To estimate both *D'* and $\chi^{2\prime}$ we considered only maternal haplotypes of the sons. To avoid effects of population structure on estimated LD values we excluded 38 haplotypes originating from common dams and 104 haplotype presumably originating from common maternal grand sires. At the end, 477 independent maternal haplotypes were considered for estimation of the level of LD in the mapping population.

## Results

**Genotypes and Linkage map construction**

Genotypes for 28 microsatellite markers were available to build the BTA5 genetic map. During plausibility control of the genotype and haplotype data two markers (*BM6026* and *BMS610*) genotyped in the initial mapping study, showed extensive double recombinations with markers included by this project (18 added markers). To reduce possible mapping errors we decided to exclude these two most proximal markers from all subsequent analyses. Using the *build* option of the *CRI-MAP* program we re-estimated marker distances and order. The resulting linkage map based on our data was consistent with the position of the markers in the USDA linkage map and the physical map (see Table 1). Two new markers developed in this study (*LMU0501* and *LMU0502*) are highly informative for linkage analysis and have 16 and 17 alleles, respectively. The relative positions of these two markers (Table 1) were confirmed by linkage

analysis (*CRI-MAP*) and physical position according to the current available information on bovine genome sequence data (Baylor release Btau_4.0, http://genome.ucsc.edu/cgi-bin/hgGateway).

**Linkage analysis by interval mapping using GDD**

Initial interval mapping on BTA5 was conducted by *QTL EXPRESS* using nine GDD families genotyped at 26 markers and using DYD as phenotypes. The across-family analysis did not confirm the QTL detected by the initial mapping study (unpublished results) and showed a flat F-statistic profile with a nominal F-value between 1 and 2 (Fig. 1A). The results of the permutation test across all families indicated the null hypothesis (QTL not present) as correct.

The analysis of individual GDD families, however, indicated three families segregating for a QTL affecting MY1, PY1 and FY1 at the distal region of BTA5. Figure 1B illustrates these results for MY1. The across-family analyses including only the three families with the putative QTL at the same distal position and affecting the same set of traits, revealed a significant QTL effect ($P$=0.03) at 118 cM (Fig. 1C). The least square estimators of the QTL allele substitution effects on the DYDs for the three segregating sire families were 272 kg milk, 7.0 kg milk fat and 7.1 kg milk protein in the first lactation. It should be noted that these effects are likely to be overestimated as they were obtained from preselected material, and the detection power of our GDD is limited (Georges et al. 1995). Despite evident grouping of bootstrap estimates around the position of 118 cM (two third within 5 cM) the bootstrapping analysis estimated a very broad 95% confidence interval of QTL position from 36 to 133 cM for MY1 (see Fig. 1C).

**Linkage analysis by variance component (LAVC) approach in a complex pedigree**

In many cases, the complex pedigree with 1511 animals (683 genotyped) allowed tracing the paternal haplotype without interruption along five generations. Very few dams were genotyped, so that for most of the dams' haplotypes had to be inferred for IBD estimation. The LDL analysis derives the most probable haplotypes for dams that were not genotyped, and uses these for IBD estimation. In contrast, for dams that were not genotyped, LAVC as

implemented by the *LOKI* program considers expected IBD-values based on additive genetic relationship, i.e. 0.5. In practice, this enabled the LAVC relatively good use of linkage information comprised in maternal haplotypes only when the MGS were genotyped. For each of 25 marker brackets, **G** matrices were generated by *LOKI* and variance components estimated with each **G**, using ASREML and AIREML as implemented in the programs *ASREML* and *LDL,* respectively.

The patterns of the LRT values indicated the presence of a QTL affecting MY, PY and FY in the first lactation at the distal region of BTA5. The LRT reached a maximum value of 4.42 within marker bracket *BM2830-ETH152* at the relative position of 119.33 cM (Fig. 2). This LRT value corresponded to $P$=0.035 and was significant for MY1. The LRT curve showed a possible second QTL at the middle part of the chromosome from 65-75 cM affecting MY1, PY1 and FY1 (Fig. 2) corresponding to a significant P value of 0.045 for FY1 (LRT=4.01, at position 73.31 cM). Despite numerous independent estimates of IBD matrices by *LOKI* the variance component estimation by ASREML did not converge for FY1. The AIREML converged for all three traits, consequently the presented LRT curves (Fig. 2) and variance components were estimated by AIREML in both LAVC and combined LDL analyses. Table 2 shows the estimates of variance components for all three traits by LAVC and LDL. For the most significant trait (MY1), we obtained estimates of the following components of variance for a QTL assumed located at 119.33 cM: QTL variance, $\sigma_q^2$ = 12948 kg; polygenic variance, $\sigma_u^2$=221753 kg; and phenotypic variance, $\sigma_y^2$ = 251793 kg. $\sigma_q^2$ and $\sigma_u^2$ sum to additive-genetic variance $\sigma_A^2$ = 234701 kg. Note that above components based on DYD pre-corrected for most important environmental factors. The estimation of breeding values in the Fleckvieh population assumed $\sigma_A^2$ = 260437 kg. Therefore our relatively small mapping population underestimated $\sigma_A^2$ only by about 9.9%. According to our estimates of $\sigma_q^2$ and $\sigma_A^2$ from the linkage analysis, the QTL mapped in this study explained 5.5% of $\sigma_A^2$.

**Combined Linkage Disequilibrium and Linkage Analysis (LDL)**

The estimation of IBD probabilities in a complex pedigree of 1511 individuals with 55% missing data (i.e. 55% not genotyped animals) by MCMC as implemented in *LDLRAMS* program needs an enormous computing effort. Therefore, with our available computing power, we were only able to perform an analysis with 5500 MCMC iterations. To ensure the convergence at global maximum, we used two MCMC runs with the same parameters but with different random number seeds. The LDL analysis mapped a QTL affecting MY1 and PY1 at the same position (119.33 cM; Fig. 3) as LAVC (119.33 cM; Fig. 2) and nearly at the similar position estimated by IM in the GDD mapping (118 cM; Fig. 1C). The LRT rejected the null hypothesis (no QTL) in favour of the alternative hypothesis at level *P*=0.002 for MY1 (Fig. 3). Additionally, LDL analysis mapped a significant QTL effect (*P*=0.006) for FY1, with nearly identical LRT-value at 114.67 cM and 119.33 cM. In general, the LRT curve for all three traits was significantly sharper when using the LDL method compared to the results of the IM and LAVC approach, respectively. The 97% confidence interval was still large and covered the chromosomal segment from 113-133 cM for MY1, 109-133 for PY1 and 110-123 for FY1. For the most significant trait MY1 (Table 2), we estimated at position 119.33 cM a QTL variance, $\sigma_q^2$ = 48037 kg; polygenic variance, $\sigma_u^2$ = 203652 kg; and phenotypic variance, $\sigma_y^2$ = 269059 kg. Consequently, the estimated additive-genetic variance $\sigma_A^2$ is 251689 kg. Again, $\sigma_A^2$ was underestimated but only by 3.4%. According to our estimates of $\sigma_q^2$ and $\sigma_A^2$ from the LDL analysis, the putative QTL mapped in this study explained 19.1% of the $\sigma_A^2$. The LDL analyses did not confirm the possible second QTL at the middle part of the chromosome (65-75 cM) which was mapped by LAVC (Fig. 2). On the other hand, at the beginning of the investigated chromosomal segment (36.27 cM) the LDL analysis pointed to a putative QTL affecting FY1 (*P*=0.007) with indications for effects on PY1 and MY1 also mapped at this position.

The analyses of all three traits in the second and the third lactation with IM, LAVC and LDL show non-significant test-statistic with mostly parallel course to the appropriate curve in the first lactation (data not shown).

**The Level of the LD in the dual purpose Cattle Breed Fleckvieh**

Figure 4 shows the distribution of LD between microsatellite markers according to LD statistic *D′* or $\chi^{2\prime}$. Average LD values were given for marker pairs sorted in 5-cM bins (0–50 cM) and for separation distances > 50 cM (Table 3). At the <5 cM separation distance, the mean *D′* was 0.336 and the corresponding $\chi^{2\prime}$ values were 0.107. For the separation distance > 50 cM, mean values for *D′* and for $\chi^{2\prime}$ were 0.168 and 0.029, respectively. *D′* and $\chi^{2\prime}$ values rapidly decreased to 0.18 and 0.035 for separation distances of >20 cM, and then reached a respective plateau slightly below 0.17 and 0.03 for greater separation distances.

## Discussion

Numerous studies reported segregation of QTL affecting milk production traits on BTA5 (e.g. Olsen et al. 2002; Bennewitz et al. 2003; Khatkar et al. 2004), some of these studies mapped two QTL affecting milk yield, milk protein yield and milk fat yield on this chromosome, one at the proximal and another at the distal region of the chromosome. The present study confirmed a QTL affecting first lactation milk yield, milk protein yield and milk fat yield at the distal region of BTA5.

*Comparison of different mapping methods*; IM, LAVC and LDL, mapped this QTL at the same marker interval (*BM2830-ETH152*) and nearly at the same position: 118.00, 119.33 and 119.33 cM, respectively. As expected on the basis of theoretical considerations (Lee & Van der Werf, 2005) and experimental results (Meuwissen et al. 2002; Olsen et al. 2005), the shape of the curve is significantly sharper with the LDL method compared to IM and LAVC, and the peak obtained by the LDL method was higher and narrower than that of the LAVC approach. The test statistics used for IM, LAVC and LDL are not directly comparable, but the null hypothesis was rejected with higher probability as more information was used. In particular, IM analysis across nine families was not able to detect a QTL segregating in a set of nine GDD families, while connecting the same nine families by numerator relationship matrix (LAVC approach) did result in significant effects of mapped QTL. However, the shape

of test statistics estimated by IM based on all nine GDD families (Fig. 1A) was nearly comparable with that obtained by the LAVC approach (Fig. 2): both approaches showed a peak for FY1 at the middle region of the chromosome and another peak for MY1 and PY1 at the distal region of the chromosome. The crucial difference was in the significance of the results: the LAVC approach estimated significant QTL effects for FY1 at 73.3 cM and for MY1 and PY1 at 119.33 cM whereas IM based on these nine families showed a fairly flat non-significant test statistic along the chromosome. However, analysis of the individual GDD families indicated three families segregating for a QTL affecting milk production traits in the distal region of BTA5. This is a rather unusual result because experience shows that even a single significant family is usually sufficient to render a multi-family GDD analysis significant. Be that as it may, IM analysis of these selected families did reveal a substantial QTL effect in each of them, but since these families were selected, a formal significance threshold could not be easily determined. However, these families mapped the "detected" QTL to the same general location as the LAVC analysis indicating that they were indeed segregating for this QTL, and provided estimates of QTL effects which could not be obtained from the overall LAVC or LDL analyses.

As shown in Figure 3, LDL indicated a possible QTL affecting FY1 at position 36.27 cM. The most proximal marker in the region (*AGLA293*) with relative position at 32.25 cM stems from the initial mapping study (Set 0 in Table 1) and was genotyped in only five GDD families. In addition, our marker map shows the largest distance between this and the next most proximal marker at this region (8.04 cM, Table 1). Considering this, and the necessity to reach high marker informativity at the ends of the chromosomal segments under investigation (Krebs et al. 2007), we were not confident about this putative QTL and therefore did not follow up this result.

As mentioned in the introduction section, there is some introgression of Red-Holstein (RH) genes into Fleckvieh population. We sampled and genotyped a 50% RH bull and six of his sons with the largest impact on the current Fleckvieh population as well as his 100% RH sire. This enabled us to clearly trace RH haplotypes derived from this 50% RH bull in our mapping population. In the total of 605 sons phenotyped, 37 sons (~6%) inherited the RH haplotype for the

distal half of BTA5. Of these, 27 were part of GDD family nine (Fig. 1B) and we were able to undoubtedly trace this haplotype over family sires to the 100% RH bull. The IM in outbred half-sib design is thoroughly capable of mapping QTL allele originating from different subpopulations (Lander & Botstein 1989; Knott et al 1996). As clearly shown in Figure 1B, Family 9 is not even indicative for the QTL affecting MY1 at the distal half of the BTA5. The same is true for FY1 and PY1 (data not shown). Therefore, we concluded that the introgression of the RH genes into the Fleckvieh population did not contribute to our mapping results for the three traits at the distal region of BTA5.

*Linkage disequilibrium in the Fleckvieh population*: The results of the present study confirm and enlarge results previously reported from our laboratory comparing LD in the Fleckvieh population with LD in Israel and Italian Holstein and in Italian-German-Austrian Brown Swiss cattle (Lipkin et al. 2009). In particular, the average *D'* value for all marker pairs with a distance < 5 cM (0.34) was very close to that estimated in the Lipkin et al. 2009 study (0.35), and is considerably less than estimated for the Holstein breed (Farnir et al. 2000; Vallejo et al. 2003; Tenesa et al. 2003; Lipkin et al. 2009). This is consistent with results revealing Fleckvieh as the breed with the highest genetic diversity among seven Alpine and three North-western European cattle breeds (unpublished results I. Medugorac), confirming a large effective population size (Sölkner et al. 1998). Large effective population size would be expected to result in a lower level of LD in comparison to breeds, such as the Holstein, with much smaller effective population size.

It is well known for diallelic markers, that estimates of *D'* are strongly inflated in small samples, even for SNPs with common alleles, but especially for SNPs with rare alleles (e.g. Shifman et al. 2003). This tendency is undoubtedly exacerbated for microsatellite markers because of the general presence of one or more alleles at low frequency. Our data (Fig. 4), as well as similar comparative results for *D'* and $\chi^{2\prime}$ reported for sheep (Meadows et al. 2008) and previously found by us for cattle (Lipkin et al. 2009) confirmed this tendency. As shown by Zhao et al. (2005), $R^2=\chi^{2\prime}$ for LD between a multi-allelic marker and a biallelic QTL, and $R^2=\chi^{2\prime}=r^2$ when both loci are biallelic, where $R^2$ is coefficient of determination of the regression of QTL allele *Q* on alleles ($A_i$) at a single

marker. Therefore we used *D'* to compare the level of LD with other populations but $\chi^{2\prime}$ to estimate the actual level of useable LD in the mapping population. In this study, we estimated $\chi^{2\prime}$=0.107 for all marker pairs with a distance <5 cM in Fleckvieh. This corresponds to the value found in our previous study of this population (Lipkin et al. 2009) and is less than half than estimated for the Italian and Israel Holstein cattle population (Lipkin et al. 2009) at the same marker distance. The $\chi^{2\prime}$ values dropped off rapidly with increasing separation distance, and were very low (<0.045) for separation distances > 10 cM. Therefore, our results indicate the absence of a useful long-range intra-chromosomal LD in the Fleckvieh population and suggest the presence of only moderate LD at the < 5 cM range. In the present study, only 15% of marker pairs at this separation distance showed LD at a level >0.2 which would be useful for fine-mapping applications. For comparison, the fraction of marker pairs with $\chi^{2\prime}$ >0.2 at the < 5 cM separation distance in Fleckvieh, Israel and Italian Holstein is at 10%, 44% and 67% respectively (Lipkin et al. 2009). For LDL mapping LD between markers and QTL alleles is of primary interest. At the distal part of BTA5 we used markers with an average distance of 4.6 cM and estimated IBD at the mid point of each interval. Therefore, the maximum distance between the closest markers and the QTL is 2.3 cM, and average distance between the closest marker and the QTL is 1.15 cM, which could provide higher levels of LD useful for fine-mapping, than the LD between markers estimated above.

## Conclusion

By successive mapping steps we concluded that the distal part of BTA5 harbours a QTL affecting first lactation milk yield, milk protein yield and milk fat yield. The mapping accuracy and the significance of a QTL effect increased from IM over LAVC to LDL. However, the confidence interval was large for the positional candidate gene approach. The relatively low level of LD in the Fleckvieh breed suggests a need for a denser marker map to achieve fine LDL mapping of QTL in this population. Therefore, we conclude that an additional marker set of 12 to 16 markers covering BTA5 from cM 107 to 133 should be genotyped in the same families. These new genotypes should lead to finer mapping and promising positional candidate genes. It should be noted,

however, that the values obtained for $\chi^{2'}$ at the <5 cM separation distance were based on microsatellite markers. Thus, although much less than those given by *D'*, they are still considerably greater than those reported in cattle for biallelic SNP markers at this separation distance, using the comparable $r^2$ measure (McKay et al. 2007; Khatkar et al. 2008). Consequently, the new markers added to this region for high resolution mapping should preferably be microsatellite markers, to the possible extent.

## Acknowledgements

A. Awad was supported by the Ministry of Higher Education, Egypt. We thank all breeders and AI Centres for providing samples. A part of this study (10 markers of an earlier project i.e. Set0) was genotyped by Tierzuchtforschung e.V. München. The authors thank M. Soller for useful comments on an earlier version of this manuscript and editing of final version. The authors thank C. Edel for fruitful discussion. We also thank the two anonymous reviewers whose comments resulted in substantial improvement of the final manuscript version.

## References

Andersson L. (2001) Genetic dissection of phenotypic diversity in farm animals. *Nature Reviews Genetics* **2**: 130-138.

Ashwell M.S., Heyen D.W., Sonstegard T.S. & Van Tassell C.P. (2004) Detection of quantitative trait Loci affecting milk production, health, and reproductive traits in Holstein cattle. *Journal of Dairy Science* **87**: 468-475.

Bennewitz J., Reinsch N., Grohs C., et al. (2003) Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle. *Genetics Selection Evolution* **35**: 319-338.

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

Churchill G.A. & Doerge R.W. (1994) Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–971.

Darvasi A., Weinreb A., Minke V., Weller J.I. & Soller M. (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* **134**: 943-951.

Emmerling R., Lidauer M. & Mäntysaari E.A. (2002) Multiple lactation random regression test-day model for Simmental and Brown Swiss in Germany and Austria. *Interbull Bulletin* **29**: 111–117.

Farnir F., Coppieters W., Arranz J.J., Berzi P. & Cambisano N. (2000) Extensive genome-wide linkage disequilibrium in cattle. *Genome Research* **10**: 220–227.

Farnir F., Grisart B., Coppieters W. et al. (2002) Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: Revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. *Genetics* **161**: 275–287.

Gautier M., Barcelona R.R., Fritz S., Grohs C., Druet T., Boichard D., Eggen A. & Meuwissen T.H. (2006) Fine mapping and physical characterization of two linked quantitative trait loci affecting milk fat yield in dairy cattle on BTA26. *Genetics* **172**: 425-436

Georges M., Nielsen D., Mackinnon M., et al. (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* **139**: 907-920.

Gilmour A.R., Cullis B.R., Welham S.J. & Thompson R. (2000) ASREML reference manual. Available: ftp.res.bbsrc.ac.uk/pub/aar

Green P., Falls K. & Crooks S. (1990) Documentation for CRI- MAP, Version 2.4.Washington University School of Medicine, St. Louis.

Gutierrez-Gil B., Williams J.L., Homer D., Burton D., Haley C.S. & Wiener P. (2009) Search for QTL affecting growth and carcass traits in a cross population of beef and dairy cattle. *Journal of Animal Science* **87**:24-36

Heath S.C. (1997) Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *American Journal of Human Genetics* **61**:748–760.

Hedrick P.W. (1987) Gametic disequilibrium measures: Proceed with caution. *Genetics* **117**: 331-341

Ihara N., Takasuga A., Mizoshita K. et al. (2004) A comprehensive genetic map of the cattle genome based on 3802 microsatellites. *Genome Research* **14**:1987-98.

Johnson D. L. & Thompson R. (1995) Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. *Journal of Dairy Science* **78**: 449–456.

Kappes S.M., Bennett J.W., Echternkamp S.E., Gregory K.E. & Thallman R.M. (2000) Initial results of genomic scans for ovulation rate in a cattle population selected for increased twinning rate. *Journal of Animal Science* **78**: 3053-3059.

Khatkar M.S., Thomson P.C., Tammen I. & Raadsma H.W. (2004) Quantitative trait loci mapping in dairy cattle: Review and meta-analysis. *Genetics Selection Evolution* **36**: 163–190.

Khatkar M.S., Nicholas F.W., Collins A.R. et al. (2008) Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics* **24**: (9):187.

Knott S.A., Elsen J.M. & Haley C.S. (1996) Methods for multiple markers mapping of quantitative trait loci in half-sib populations. *Theoretical and applied genetics* **93**: 71–80.

Krebs S., Medugorac I., Röther S., Strässer K. & Förster M. (2007) A missense mutation in the 3-ketodihydrosphingosine reductase FVT1 as candidate causal mutation for bovine spinal muscular atrophy. Proc Natl Acad Sci, *U S A*, 104: 6746-6751.

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

Lander E.S. & Schork N.J. (1994) Genetic dissection of complex traits. *Science* **265**: 2037-2048

Lee S.H. & Van der Werf J.H. (2005) The role of pedigree information in combined linkage disequilibrium and linkage mapping of quantitative trait loci in a general complex pedigree. *Genetics* **169**: 455-66.

Lee S.H. & Van der Werf J.H. (2006) An efficient variance component approach implementing an average information REML suitable for combined LD and linkage mapping with a general complex pedigree. *Genetics Selection Evolution* **38**: 25-43.

Lee S.H., Van der Werf J.H. & Tier B. (2005) Combining the Meiosis Gibbs Sampler with the random walk approach for linkage and association studies with a general complex pedigree and multimarker Loci. *Genetics* **171**: 2063-2072

Lien S., Karlsen A., Klemetsdal G. et al. (2000) A primary screen of the bovine genome for quantitative trait loci affecting twinning rate. *Mammalian Genome* **11**: 877-882.

Lipkin E., Straus K., Tal Stein R. et al. (2009) Extensive long-range and non-syntenic linkage disequilibrium in livestock populations: deconstruction of a conundrum. *Genetics* **181**: 691–699

Mangin B., Goffinet B. & Rebai A. (1994) Constructing confidence intervals for QTL location. *Genetics* **138**: 1301–1308.

McKay S.D., Schnabel R.D., Murdoch B.M. et al. (2007) Whole genome linkage disequilibrium maps in cattle. *BMC Genetics* **8**: 74.

Meadows J.R.S., Chan E.K.F. & Kijas J.W. (2008) Linkage disequilibrium compared between five populations of domestic sheep. *BMC Genetics* **9**: 61

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

MeuwissenT.H.E., Karlsen A., Lien S., Olsaker I. & Goddard M.E. (2002) Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* **161**: 373–379.

Olsen H.G., Gomez-Raya L., Vage D.I. et al. (2002) A Genome Scan for Quantitative Trait Loci Affecting Milk Production in Norwegian Dairy Cattle. *Journal of Dairy Science* **85**: 3124-3130.

Olsen H.G., Lien S., Gautier M. et al. (2005) Mapping of a milk production quantitative trait locus to a 420-kb region on bovine chromosome 6. *Genetics* **169**: 275-83.

Olsen H.G., Lien S., Svendsen M. et al. (2004) Fine Mapping of milk production QTL on BTA6 by Combined Linkage and Linkage Disequilibrium Analysis. *Journal of Dairy Science* **87**: 690-698

Riquet J., Coppieters W., Cambisano N. et al. (1999) Fine mapping of quantitative trait loci by identity by descent in outbred populations: Application to milk production in dairy cattle. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 9252–9257.

Seaton G., Haley C.S., Knott S. A., Kearsey M. & Visscher P.M. (2002) QTL Express: mapping quantitative trait loci in simple and complex pedigrees, *Bioinformatics* **18**: 339-340.

Shifman S., Kuypers J., Kokoris M., Yakir B. & Darvasi A. (2003) Linkage disequilibrium patterns of the human genome across populations. *Human molecular genetics* **12**: 771-776.

Sölkner J., Filipcic L. & Hampshire N. (1998) Genetic variability of populations and similarity of subpopulations in Austrian cattle breeds determined by analysis of pedigrees. *Animal Science* **67**: 249-256

Tenesa A., Knott S.A., Ward D., Smith D., Williams J.L. & Visscher P.M. (2003) Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. *Journal of Animal Science* **81**: 617-623.

Vallejo R.L., Li Y.L., Rogers G.W. & Ashwell M.S. (2003) Genetic diversity and background linkage disequilibrium in the North American Holstein cattle population. *Journal of Dairy Science* **86**: 4137-4147.

Visscher P.M. & Goddard M.E. (2004) Prediction of the confidence interval of quantitative trait Loci location. *Behaviour Genetics* **34**: 477-82

Weller J.I., Kashi Y. & Soller M. (1990) Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *Journal of Dairy Science* **73**: 2525–2537.

Yamazaki T. (1977) The effects of over dominance on linkage in a multi locus system. *Genetics* **86**: 227–236.

Zhang Q., Boichard D., Hoeschele I. et al. (1998) Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbreed pedigree. *Genetics* **149**: 1959–1973.

Zhao H., Nettleton D., Soller M. & Dekkers J.C.M. (2005) Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genetic Research* **86**: 77–87

**Table 1** All markers used on BTA5 with marker name, relative position on the USDA map (cM), physical position (bp), forward and reverse primers sequences and marker set description (set: Set0 previous Project; Set1 this project multiplex 1; Set2 project multiplex 2).

| No | Marker ID | cM | bp | Forward primer | Reverse primer | Set |
|---|---|---|---|---|---|---|
| 1 | *BM6026* | 6.05 | 7023092-7023270 | GCAACTAAGACCCAACCAAC | ACTGATGTGCTCAGGTATGACG | Set0 |
| 2 | *BMS610* | 12.02 | 13167579-13167698 | TTTCACTGTCATCTCCCTAGCA | ATGTATTCATGCACACCACACA | Set0 |
| 3 | *AGLA293* | 32.25 | 26701514-26702173 | GAAACTCAACCCAAGACAACTCAAG | GACACTAGTAGATTTGAAACCCA | Set0 |
| 4 | *DIK4759* | 40.29 | 30345678-30346337 | CTGAAGTTGGACCTGCCATT | GGCAGAGCCAGAACTACACC | Set2 |
| 5 | *BMS1898* | 44.48 | 33042590-33043249 | TCCACCGTGTGACATCACTT | GAAGCAGGTCTCCACTCAGG | Set1 |
| 6 | *CSSM34* | 45.51 | 33862045-33862704 | GAAACAAAGAATCAGGCCATAA | TGATCAATGGGTAAGTGAACAAA | Set0 |
| 7 | *DIK4782* | 50.52 | 40849000-40849659 | GCATCTGAGAGCCTCTTTGG | TCCTGGGTGTATAGGGCATC | Set1 |
| 8 | *BL37* | 52.09 | 45122278-45122937 | GCAATCCCACTCTCCAGGTG | CATTCATGTTGCTGTAAATGGC | Set0 |
| 9 | *RM084* | 56.63 | 51937288-51937947 | TAGGAACTGATGGCCCTTTG | GGCAGAGTTCGTGACTGGAG | Set1 |
| 10 | *DIK2732* | 63.92 | 55833522-55834181 | GGGGAACTTTAATGGGAGGA | GCTTGGCAATCCATAGAGGA | Set1 |
| 11 | *LMU0501* | 65.00 | 57808736-57809395 | CATGTTGGTAATGAATGGGCTA | TAATCAACGGCATCATCAGG | Set1 |
| 12 | *BMS490* | 66.21 | 57285385-57286044 | CCATTTTGGTAAATGGCTCAA | TTTGGAAGCTTAAGGGAAACTT | Set2 |
| 13 | *ETH10* | 71.76 | 60836415-60836774 | GTTCAGGACTGGCCCTGCTAACA | CCTCCAGCCCACTTTCTCTTCTC | Set0 |
| 14 | *DIK5248* | 74.86 | 66083848-66084507 | TGGATAGAGCCTTGGAATGG | TTCCAATGATGCAGGTCAAA | Set2 |
| 15 | *BMS1216* | 78.21 | 76369042-76369701 | TTCCATCCAAGCTACCCAAC | CTTGAATCCGCATGCTTGAC | Set1 |
| 16 | *TEXAN15* | 80.96 | 78557768-78558407 | CGCAAACAGTCAGAGACCAC | CAGCAGTTCCTGGATGAGAA | Set0 |
| 17 | *DIK4356* | 82.88 | 80778499-80779158 | TGTCCCATCATATCCCATCTC | TTGCAAACACAGAACTGAAAGAA | Set1 |
| 18 | *DIK545* | 85.53 | 88255774-88256433 | AAAGTTTGCAAGGGGCTTTT | GGTTAGGATTGGGAGGGAAT | Set1 |
| 19 | *BMS1248* | 90.85 | Unknown | GTAATGTAGCCTTTTGTGCCG | TCACCAACATGAGATAGTGTGC | Set0 |
| 20 | *LMU0502* | 95.00 | 98418609-98419268 | TGGAAGAATATGCAGGTAACTCT | GTCGCTCTTTGTGGCTTCAC | Set1 |
| 21 | *DIK2336* | 99.79 | 101071987-101072659 | ATGTGGAATGTAGGGCAAGG | TCCCTCACCTTTCGAACAAA | Set1 |
| 22 | *BM315* | 103.17 | 104045839-104046013 | TGGTTTAGCAGAGAGCACATG | GCTCCTAGCCCTGCACAC | Set0 |
| 23 | *DIK1135* (a | 108.22 | 10181475-10181838 | GTCTGCCATCTAGCCAAAAA | GTTTTTCAGTGGGCATTTGG | Set1 |
| 24 | *ETH2* | 112.43 | 112903664-112904283 | ATTTGCCCTGCTAGCTTTGA | AAGACTCTGGGCTTCAAAAGG | Set1 |
| 25 | *BM2830* | 116.91 | 115261807-115262486 | AATGGGCGTATAAACACAGATG | TGAGTCCTGTCACCATCAGC | Set0 |
| 26 | *ETH152* | 121.75 | Unknown | GTTCTCAGGCTTCAGCTTCG | TGATCAGAGGGCACCTGTCT | Set1 |
| 27 | *URB060* | 127.55 | 122472467-122473126 | TTGTCATTTCTGGACTCCACTG | CAGGTCCAACCCTGTTTAGC | Set1 |
| 28 | *MNB71* | 133.09 | unknown | CATCTAAGGCAGAGCCAACC | TTCTTGGTGCCTCTCTCTCC | Set1 |

[(a] The whole genome sequence map *DIK1135* at the beginning of chromosome but our linkage analysis confirm the USDA position.

**Table 2** The estimates of the variance components at the most likely QTL position (119.33 cM) using linkage analysis by variance components (LAVC) and combined linkage disequilibrium and linkage analysis (LDL). The estimates of QTL variance ($\sigma_q^2$),Polygenic variance ($\sigma_u^2$), phenotypic variance ($\sigma_y^2$), additive-genetic variance ($\sigma_A^2$), $\sigma_A^2$ used for breeding values estimation ($\sigma_A^2$ BV), % of additive-genetic variance explained by QTL (% of $\sigma_A^2$) for all three traits: milk yield (MY1) milk fat yield (FY1) and milk protein yield (PY1) in the first lactation.

| Variance components | LAVC | | | LDL | | |
|---|---|---|---|---|---|---|
| | MY1 | FY1 | PY1 | MY1 | FY1 | PY1 |
| $\sigma_q^2$ | 12948 kg | 14 kg | 10 kg | 48037 kg | 59 kg | 30 kg |
| $\sigma_u^2$ | 221753 kg | 233 kg | 179 kg | 203652 kg | 197 kg | 169 kg |
| $\sigma_y^2$ | 251793 kg | 312 kg | 217 kg | 269059 kg | 331 kg | 227 kg |
| $\sigma_A^2$ | 234701 kg | 247 kg | 189 kg | 251689 kg | 256 kg | 199 kg |
| $\sigma_A^2$ BV | 260437 kg | 358 kg | 201 kg | 260437 kg | 358 kg | 201 kg |
| $\sigma_A^2/\sigma_A^2$ BV | 0.90 | 0.69 | 0.94 | 0.97 | 0.72 | 0.99 |
| % of $\sigma_A^2$ | 5.52 | 5.77 | 5.24 | 19.09 | 22.98 | 15.31 |

**Table 3** Mean LD values estimated as *D'* and $\chi^{2}$' for marker pairs sorted in 5-cM bins (0–50 cM) and for separation distances (Dist) > 50 cM.

| cM | *D'* | $\chi^{2}$' |
|---|---|---|
| <5 | 0.336 | 0.107 |
| 5< Dist <10 | 0.245 | 0.059 |
| 10< Dist <15 | 0.224 | 0.045 |
| 15< Dist <20 | 0.195 | 0.044 |
| 20< Dist <25 | 0.178 | 0.036 |
| 25< Dist <30 | 0.178 | 0.036 |
| 30< Dist <35 | 0.194 | 0,035 |
| 35< Dist <40 | 0.172 | 0.034 |
| 40< Dist <45 | 0.193 | 0.028 |
| 45< Dist <50 | 0.203 | 0.043 |
| Dist >50 | 0.168 | 0.029 |

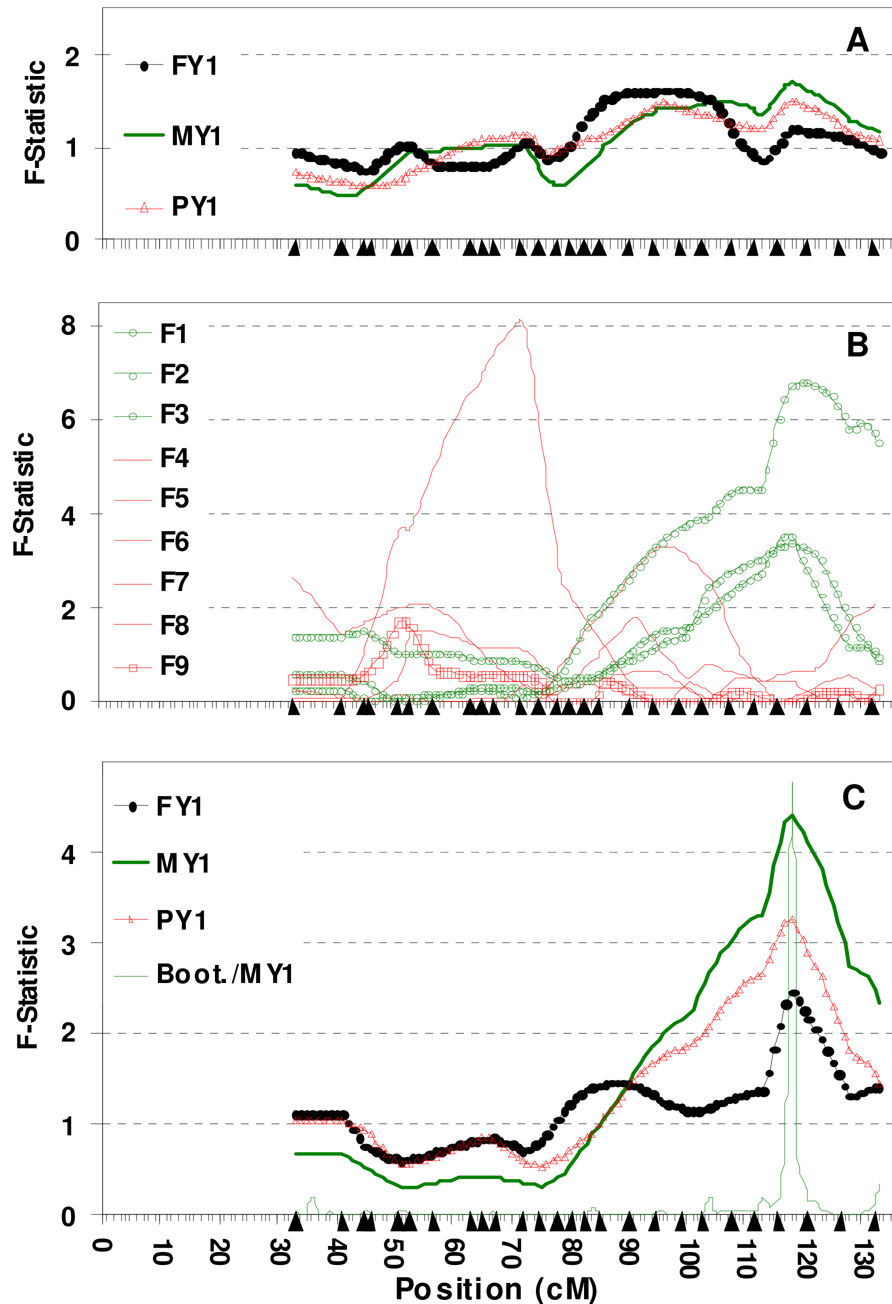



**Figure 1** Interval mapping analysis by half-sib option of *QTL EXPRESS* using grand daughter design (GDD) and DYD as phenotype. Solid triangles on the X-axis represent positions of markers used for analysis. (A) Interval mapping across all 9 GDD families. (B) Family-wise analysis for MY1, the F-statistic of three families (F1, F2 and F3) indicative for QTL at the distal part of BTA5 and Family 9 which has a distal haplotype from 50% RH are highlighted. (C) Interval mapping across 3 GDD families, F-statistic curve for MY1, PY1 and FY1 and Bootstrap samples for MY1 (divided by 1000 to fit scale).

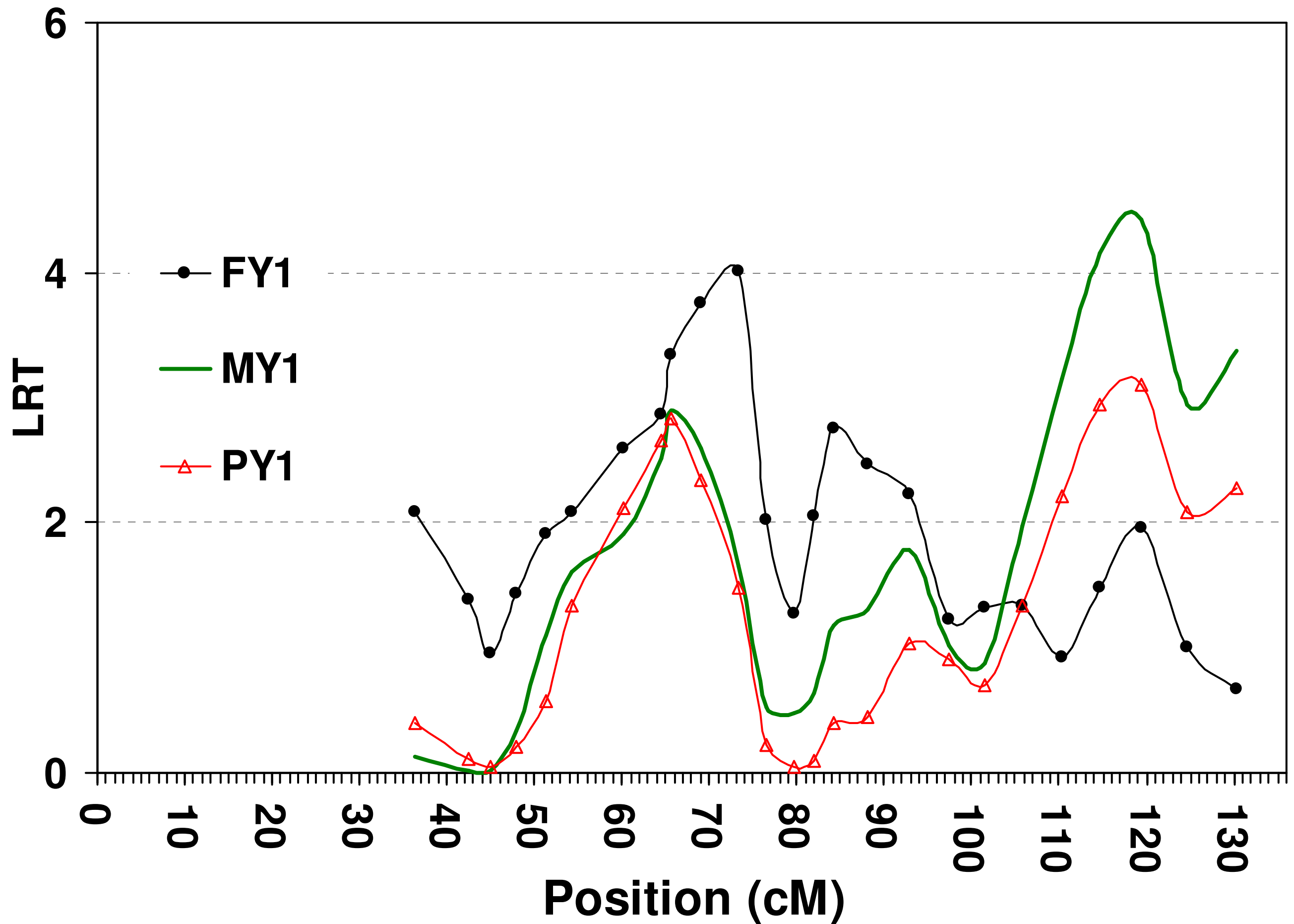



**Figure 2** Linkage analyses by variance component approach (LAVC) for MY1, PY1 and FY1 using complex pedigree as implemented in *LOKI,* DYD as phenotype and variance components estimated by AIREML as implemented in program *LDL*. Solid triangles on the X-axis represent positions of markers used for analysis

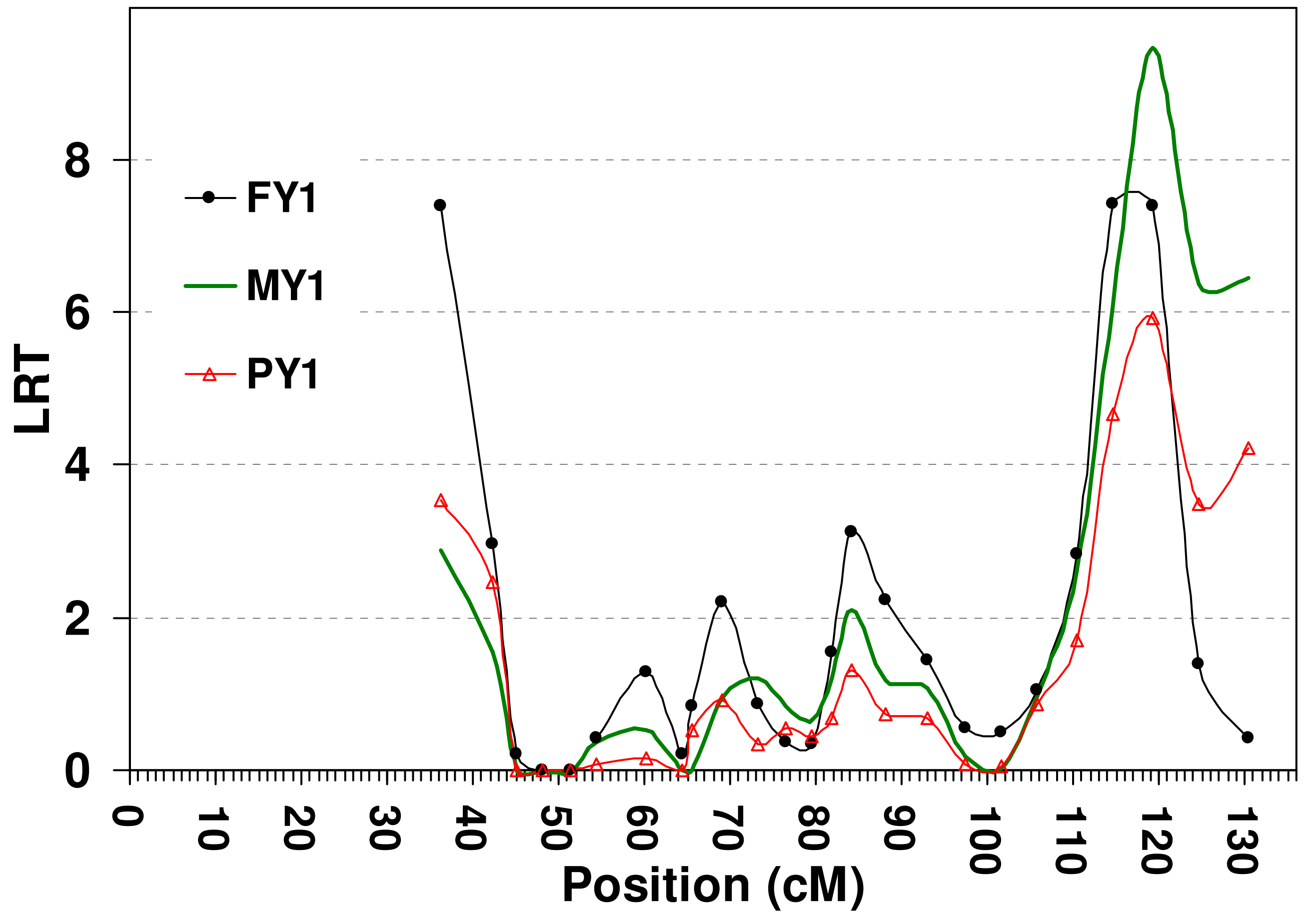



**Figure 3** Combined Linkage Disequilibrium and Linkage (LDL) analysis by variance component approach using complex pedigree, DYD as phenotype and AIREML as implemented in *LDLRAMS* and *LDL* program. Solid triangles on the X-axis represent positions of markers used for analysis

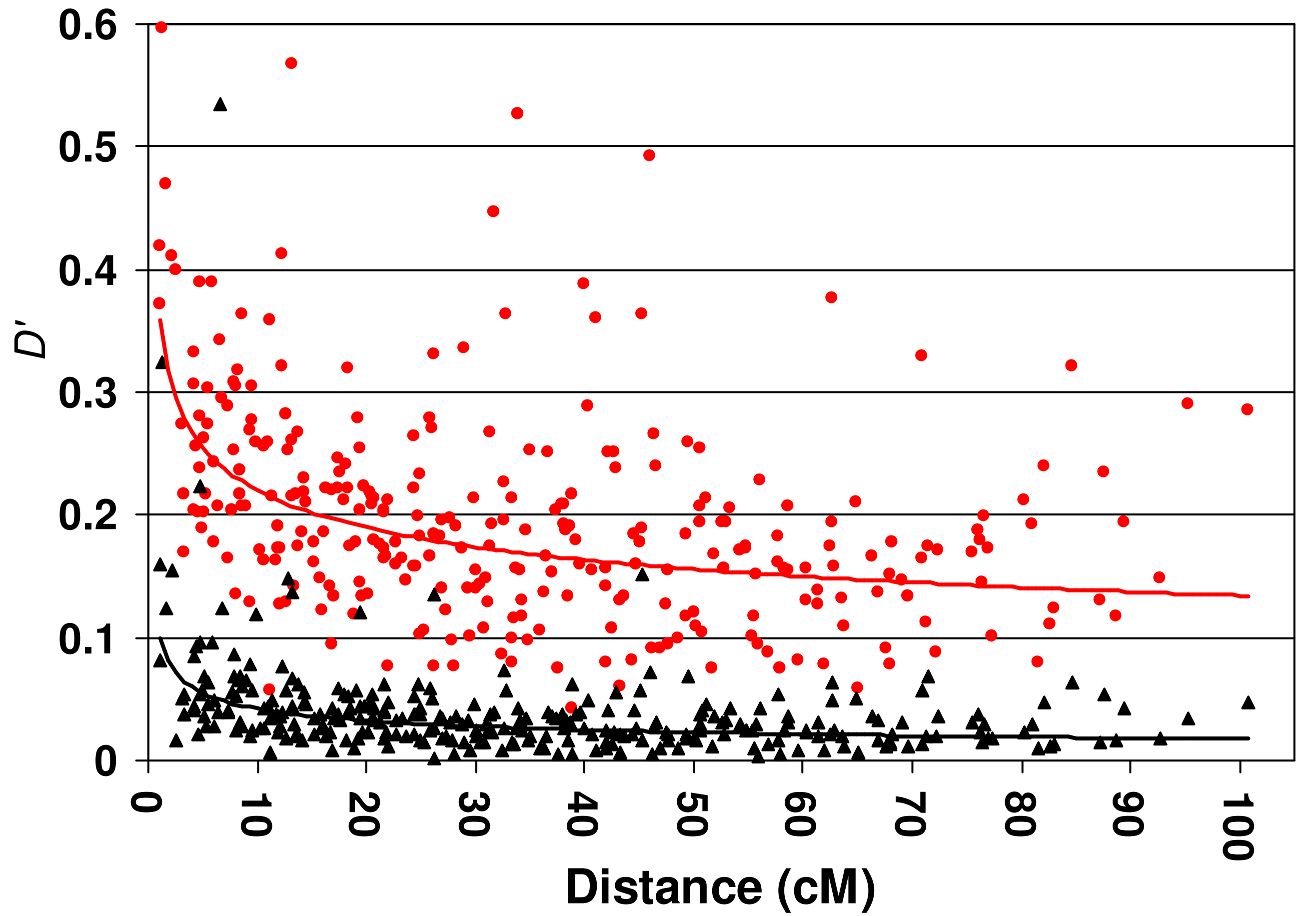



**Figure 4** Distribution of *D*' (dots) and $\chi^{2'}$ (triangles) values observed between marker pairs of BTA5 as a function of genetic distance in centiMorgan (cM). The curves correspond to trends of *D*' and $\chi^{2'}$ values

# CHAPTER 4

# Mapping of a milk production quantitative trait locus to a 1.056 Mb region on bovine chromosome 5 in the Fleckvieh Dual Purpose Cattle Breed

A. Awad[1], I. Russ[2], M. Förster[1,2], and I. Medugorac[1]

[1] Chair of Animal Genetics and Husbandry, Faculty of Veterinary Medicine, The Ludwig-Maximilians-University Munich, Veterinärstr. 13, 80539 Munich, Germany

[2] Tierzuchtforschung e.V. München, Senator-Gerauer-Straße 23, D-85586 Grub, Germany

## Abstract

### Background

In a previous study in the Fleckvieh dual purpose cattle breed, we mapped a quantitative trait locus (QTL) affecting milk yield (MY1), milk protein yield (PY1) and milk fat yield (FY1) during first lactation to the distal part of bovine chromosome 5 (BTA5) but the confidence interval was too large for positional cloning of the causal gene. Our objective here was to refine the position of this QTL and to define the candidate region for high-throughput sequencing.

### Material and methods

For genotyping, 12 new microsatellite and 240 SNP markers covering the most likely QTL region on BTA5 were added. New Fleckvieh families were also genotyped in order to increase the number of recombination events around the putative QTL. Based on haplotype analysis performed in this complex pedigree, families segregating for the low frequency allele of this QTL (minor allele) were selected. Single- and multiple-QTL analyses using combined linkage and linkage disequilibrium methods were performed.

### Results

Single Nucleotide Polymorphism (SNP) haplotype analysis on representative family sires and their ancestors revealed that the haplotype carrying the minor QTL allele is rare and most probably originates from a unique ancestor in the mapping population. Analyses of different subsets of families, created according to the result of haplotype analysis and availability of SNP and microsatellite data, refined previously detected QTL affecting MY1 and PY1 to a region ranging from 117.962 Mb to 119.018 Mb (1.056 Mb) on BTA5 However, the possibility of a second QTL affecting only PY1 at 122.115 Mb was not ruled out.

### Conclusion

This study demonstrates that targeting families segregating for a less frequent QTL allele is a useful method that improves mapping resolution of the QTL due to the division of the mapping population according to haplotype analysis and

the increased frequency of minor allele in the families. Consequently, we succeeded in refining the region containing the previously detected QTL to 1 Mb on BTA5. This candidate region contains 27 genes with unknown or partially known function(s) and is small enough for high-throughput sequencing, which will allow future detailed analyses of candidate genes..

## Background

Recent developments in molecular biology and statistical methodologies for quantitative trait loci (QTL) mapping have made it possible to identify the genetic factors affecting economically important traits. Such developments have the potential to significantly increase the rate of genetic improvement of livestock species through marker-assisted selection of specific loci, genome-wide selection, gene introgression and positional cloning [1]. However, after an initial exaggerated enthusiasm, animal geneticists like their colleagues in human genetics e.g. [2] have faced somewhat unexpected challenges.

The first step in QTL mapping usually involves a complete or partial genome scan where the mapping population is genotyped for markers covering the entire genome or some selected chromosomes respectively The QTLs are then mapped using linkage analysis (LA) methods. The resolution of this mapping approach is low because relatively few new recombination events are generated in the single generation separating parents and progeny. Typically, the size of confidence intervals for the most likely QTL positions ranges between 20 and 40 cM.

Fine-mapping approaches have been developed to reduce these confidence intervals e.g. [3-5], leading in some instances to the identification of the underlying causal mutation [6-9]. These approaches are mostly based on the addition of new families, new markers and the use of statistical methods combining linkage-disequilibrium and linkage (LDL) analysis. In general, the marker density is increased by adding a few tens of new markers (microsatellite markers or SNP) identified within the QTL region or candidate genes.

At present, high-throughput SNP analysis provides the opportunity to genotype many animals for hundreds or even thousands of SNP per bovine chromosome [10-12]. Thus, the limiting factors in QTL fine-mapping studies have now switched partly from marker density to applied methods and designs. Use of linkage-disequilibrium (LD) information increases the precision of the QTL mapping because it exploits the entire number of recombinations accumulated since the original mutation generating the new QTL allele occurred [13].

The degree of LD in livestock populations has attracted much attention because it provides useful information regarding the possibility of fine-mapping QTL and the potential to use marker-assisted selection. In cattle, previous reports using a low density of microsatellite map (10 cM interval on average) and Hedrick's normalized measure of LD [14] *D'* have shown that LD extends over several tens of centimorgans [10,15,16]. However, an exceedingly low long-range and non-syntenic LD has been estimated [17] when evaluated by the standardized chi-square measure of LD, which is related to the predictive ability of LD. Nevertheless, the extent of LD in cattle [18] is greater than in humans [19] but smaller than in dog [20].

Combined linkage disequilibrium and linkage (LDL) analysis [3] makes it possible to exploit recombinations occurring both within and outside the pedigree and genotyped population. It also gives a clearer signal for QTL positions compared with LA or LD mapping alone [3]. Additionally, the LDL approach reduces the risk of false-positive QTL identification caused by accidental marker-phenotype associations when LA and LD are used separately and it also increases the power and resolution of QTL mapping by combining all available information [21].

In dairy cattle, several studies have reported the presence of one or more QTL affecting milk production traits on BTA5 e.g. [22-25], but the results differ among studies with respect to the number of QTL detected, their positions, and the extent to which the milk traits are affected by the QTL.

The present study aimed to refine the previously detected QTL affecting milk yield (MY1), milk protein yield (PY1) and milk fat yield (FY1) during first lactation in the distal part of BTA5 in the Fleckvieh dual-purpose cattle breed [24] and to define the candidate region for high-throughput sequencing. To achieve this, we sampled additional families carrying the low frequency allele of the putative QTL (minor QTL allele) and genotyped additional markers covering the most likely QTL region on BTA5. These new families were identified by combining results from QTL-mapping based on microsatellites and haplotype analyses based on SNP in a complex pedigree. Single- and multiple-QTL analyses based on the LDL method were performed in different sample-sets in order to allocate the

minor QTL allele to specific families and to use the increased frequency of the minor QTL allele for refined mapping.

## Materials and Methods

### Animals and phenotype

In this study, we analysed the same nine granddaughter (GD) families used in our previous study [24], in which we identified three GD families (G01, G02 and G03) as heterozygous for a QTL located in the distal region of BTA5. The grandsires of these three GD families are designated as G01, G02 and G03, respectively. Grandsires G01 and G02 are half-sibs and have inherited the same haplotype in the distal region of BTA5 from their common ancestor A0 [24]. By target sampling (see haplotyping section below), we introduced two additional GD families, family G10 with 85 sons and family G11 with 47 sons. Grandsire G10 (grandsire of family G10), was connected through his dam to A0. Grandsire G11, (grandsire of family G11), is a son of grandsire G02. In addition, we identified all available, progeny-tested maternal grandsons of grandsires G01, G02, G10 and G11 to add more, possibly recombinant, A0 haplotypes into the mapping population. In this way, we created three maternal grandsire (MGS) families, M02 with 21 grandsons, M10 with 32 grandsons and M11 with 33 grandsons, descendants of grandsires G02, G10 ad G11, respectively. Samples of maternal grandsons were not available for grandsire G01. Thus, the analysis included 11 GD families: G01 to G11 and three MGS families (M02, M10 and M11). Figure 1 shows the relationships of all families included in this study. In some cases, mapping analyses were carried out on 173 additional animals available from other projects that are not descended from ancestor A0. Estimated breeding values (EBV) of the Fleckvieh bulls for milk production traits MY1, PY1, and FY1, (along with their reliability values) were obtained from the 2009 joint Austria-Germany genetic evaluation of the Fleckvieh population [26].

### DNA preparation and microsatellite marker selection and genotyping

Genomic DNA was prepared from semen using standard methods and from whole blood samples with QIAamp Blood-Kits (Qiagen), according to the manufacturer’s protocol.

Twelve evenly distributed microsatellite markers were added to the 28 microsatellite markers used in the previous study [24]. Twenty-one of these 40 microsatellite markers covered the most likely region containing the QTL in the distal part of BTA5 (Table 1) and were used in the present study. Previously analysed animals were genotyped only for the new markers, but the five new families (G10, G11, M02, M10 and M11) were genotyped for all marker sets [24]. For 11 of the 12 markers, relevant information was obtained from the MARC-ARS-USDA public database at http://www.marc.usda.gov [27]. The new marker LMU0505 was obtained by targeted search for dinucleotide repeats in genomic regions with a low marker density. The unique sequences flanking the newly identified dinucleotide repeats were tested for informativity by genotyping a small set of animals first. Primers for the 12 new microsatellite markers were optimized using Primer3 (v.0.4.0) according to the bovine genome sequence data current available (i.e. Baylor release Btau_4.0, http://genome.ucsc.edu/cgi-bin/hgGateway) and the appropriate fragment size in the currently designed marker set. The 12 new markers were divided into two PCR multiplex sets (Table 1) that were combined again after PCR for electrophoresis and fragment analysis. The fragment analysis of the PCR products was performed on ABI377 and ABI Prism 310 sequencers. Genotypes were assigned using *GENESCAN* and *GENOTYPER* (Applied Biosystems) software programs. We performed double genotyping of all families and ancestors using two independent runs. For ambiguous genotypes, the raw data were re-evaluated and animals were re-genotyped if necessary.

**SNP selection, genotyping and haplotyping**

SNP genotyping was carried out by Tierzuchtforschung e. V. München using the commercial Illumina Bovine SNP50 Bead chip featuring 54.001 SNPs (http://www.illumina.com/; Illumina, San Diego) that span the bovine genome, excluding Y-chromosome (BTAY). The genotype calling was performed with the GenCall application, as implemented in Illumina Bead chip Genotyping analysis software. This application computes a Gencall score for each locus, which evaluates the quality of genotypes. We included only animals with confirmed paternity and with a call rate above 0.98. Furthermore, we only used markers with a call rate above 0.90. We excluded all markers producing more than 1%

paternity problems in pairs with confirmed paternity, and also excluded all markers that were non-informative in the Fleckvieh population or with an unknown chromosomal position. This yielded 43.806 informative SNPs available for the whole-genome analysis in the Fleckvieh population. Two hundred and forty of these informative SNPs covered the region most likely containing the QTL in the distal part of BTA5 and were used in the present study.

We performed SNP genotyping in two stages. First, 75 animals i.e. the grandsires of the nine initial GD families and their ancestors, and also a number of potential GD-family sires and their ancestors, were genotyped with the SNP chip [24] and their haplotypes were reconstructed with the *BEAGLE* program [28]. These 75 animals constitute a complex pedigree (Figure 1) in which it is possible to trace the segregating haplotypes five generations back to some important ancestors of the Fleckvieh population, born in the 1960's and 1970's. This pedigree represents almost all of the important bull lines originating from a wide range of dams. Considering this, and the fact that a large proportion of the included bulls and dams are unrelated (no common grand-parents), these 75 animals provide a good representation of the haplotype diversity in the breeding Fleckvieh population. Second, the new families (G10, G11, M02, M10 and M11) containing the target haplotype segment of ancestor A0 were genotyped with microsatellite markers and with the genome-wide SNP chip. These animals and 173 additional Fleckvieh animals not closely related to ancestor A0 (but genotyped with the SNP chip in other projects running in our laboratory) were also haplotyped using the *BEAGLE* program.

**Linkage map construction**

The relative positions of microsatellite markers were re-evaluated by the *CRI-MAP* program [29]. A physical map was constructed according to the sequence data of all the markers (Table 1), using the basic alignment search tool (BLAST) and the latest cattle genome sequence (http://genome.ucsc.edu/cgibin/hgGateway). Our genetic data was used to resolve cases where more than one marker order was obtained from published linkage and physical maps. When our genetic data supported a marker order different from that of the public linkage map but in accordance to the physical

map, we modified the relative position (cM) of the markers along with the corresponding sequence. The linkage and physical maps were used as a framework to insert the newly designed marker (*LMU0505*) with the build option of the *CRI-MAP* program. The resulting final map (Table 1) was used for all the following analyses.

## QTL fine mapping

### *LDL mapping by microsatellite markers*

Joint linkage disequilibrium and linkage (LDL) analysis is a variance component approach. We applied a mixed linear model and variance component estimation as described previously [24]. Thereby we used the Markov chain Monte Carlo (MCMC) to estimate IBD probabilities in general complex pedigrees [30-32]. To estimate LD-based IBD probabilities we assumed the number of generation since the base population (mutation age) and the past effective population size to be 100, and the initial homozygosity at each microsatellite marker in the base population was set to 0.35. In addition, the program *LDLRAMS* version 1.76 [30-32] exploits allele frequencies in the population. To calculate an unbiased estimation of allele frequencies in the Fleckvieh population, we performed allele counting within the complex pedigree. We counted both alleles at all genotyped founder individuals and only the maternal allele of descendents in the pedigree. Two complex pedigrees consisting of 2089 (*MSPED2089*) and 1038 (*MSPED1038*) animals, respectively, were analysed by *LDLRAMS.* The *MSPED2089* pedigree included nine GD families from the previous study (G01 to G09), two additional GD families (G10 and G11), three maternal grandsire families (M02, M10 and M11), some highly related animals and some important ancestors (paternal and maternal grandsires of phenotyped sons and of family sires). The *MSPED1038* pedigree included two GD families (G01 and G02) founded to be segregating for QTL in the previous study, two additional GD (G10 and G11) families and three MGS families (M02, M10 and M11) sampled according to the results of the haplotype analysis. For both LDL analyses, as implemented in the MCMC approach of the program *LDLRAMS*, we used an initial burn-in of 500 iterations followed by 2500 iterations, with parameter estimates collected for each iteration. To avoid entrapment in a local maximum,

we performed two independent sampling procedures (i.e. two *LDLRAMS* runs with different random number seeds).

***LDL mapping by SNP***

Here we used three complex pedigrees for LDL mapping by SNP. The first pedigree, *SNPPED723,* was based on all progeny-tested Fleckvieh animals genotyped with the SNP chip, and consisted of 325 genotyped and phenotyped sons, and 16 genotyped and 382 non genotyped ancestors. The second pedigree, *SNPPED421,* was based on progeny-tested animals that could be traced back to ancestor A0, and consisted of 175 genotyped and phenotyped sons, eight genotyped and 238 non genotyped ancestors. The third pedigree, *SNPPED308,* was based on animals not related to ancestor A0 according to the known pedigree, and consisted of 144 genotyped and phenotyped animals, 12 genotyped and 152 non genotyped ancestors. These pedigrees were analysed with *LDLRAMS* using a dense map of 240 SNPs covering the region from 112.650 to 124.780 Mb on BTA5, i.e. a region larger than the 97% confidence interval as determined by 1-LOD support interval [24]. Due to computing constraints the total marker set was divided into five overlapping sets of 80 SNPs each. Since IBD estimates are most accurate in the middle of an investigated marker set, we present log-likelihood ratio (LRT) values only for the internal 40 marker intervals within these windows (that is, excluding the most proximal and most distal 20 markers). We used the model described above, setting the initial homozygosity at each SNP in the base population to 0.75 and using an initial burn-in of 500 iterations followed by 2500 iterations. The parameter estimates were collected after each iteration. Two independent MCMC sampling procedures (i.e. two *LDLRAMS* runs with different random number seeds) indicated convergence to a global maximum.

**Multiple-QTL analysis using linkage disequilibrium and linkage (LDL) analysis method**

We used the analysis method of Olsen et al. [33], i.e. the same model as for single-QTL analysis, but included a random QTL effect in another specified marker bracket. That is, bracket that showed a highest LRT in the single-QTL analysis was included as a random effect in the QTL model in turn, and the

analysis was repeated. These analyses searched for an additional QTL, given that the QTL in the bracket is accounted for, and is similar to the fitting of cofactors [34].

### Estimation of model parameters and test statistics

The variance components and the logarithm of the likelihood (*L*) of a model containing a QTL as well as residual polygenic effects at position *p* ($logL_p$) were estimated by AIREML [32,35], which is an integral part of the *LDLRAMS* and *LDL* programs. The likelihood of a model without QTL effect ($logL_0$) was calculated on the basis of a polygenic model. The log-likelihood ratio (LRT) was calculated as double difference in *logL* between models with and without a QTL, i.e. LRT = -2 ($logL_0$-$logL_p$). The LRT test statistic is distributed approximately as chi-square with 1 degree of freedom [36]. The confidence interval (CI) for the QTL position was determined as 1-LOD support interval which was constructed as the interval surrounding the QTL peak where the LRT exceeds $LRT_{max}$ - 2 × *ln* (10), where $LRT_{max}$ is the maximum LRT-value for the tested QTL [37].

## Results

### Genotypes and linkage map construction

Genotypes for 40 microsatellite markers were available to build the BTA5 genetic map. In most LDL analyses, only the 21 most distal markers (Table 1) covering the 97% confidence interval were considered. When we controlled if the genotype and haplotype data were plausible, the distal marker (MNB71) which was genotyped in the previous projects [24], showed extensive double recombinations with the 12 markers added in the present project. To reduce possible mapping errors, we excluded this marker from all subsequent analyses. Using the build option of the *CRI-MAP* program we re-estimated the marker distances and order. In addition to mapping results based on own data we also considered the physical map results (bp position along BTA5 Baylor release Btau_4.0) to separate markers that were at the same position on the USDA linkage map and to confirm the orders of the markers on BTA5 that we used.

The following changes with respect to the public USDA linkage map were made: (i) according to the physical map results and confirmed by applying the build option of the *CRI-MAP* program to our own data, the positions of markers BM49 and BM733 are inverted (Table 1); (ii) markers DIK2035 and DIK5277 are both at the same position (120.85 cM) on the USDA linkage map but, according to our genotypes and the physical map results, they are separated, placing DIK2035 (120.38 cM) upstream of DIK5277 (120.82 cM); (iii) the new marker developed in this study (LMU0505) is highly informative for linkage analysis and its relative position between DIK5106 and ETH152 was estimated by applying the build option of the *CRI-MAP* program. The positions of both flanking markers DIK5106 and ETH152 also changed (Table 1).

**Haplotype analysis in a complex pedigree**

Using the algorithm implemented into the program *BEAGLE*, we haplotyped the 75 animals of the complex pedigree in Figure 1 with 1976 SNP markers on BTA5 that are informative in the Fleckvieh population. Thus reconstructed haplotypes were used to identify families segregating for QTL detected in the initial study [24]. As already shown by the microsatellite analysis, the grandsires of families G01 and G02 which are heterozygous at the QTL, inherited the same haplotype in the distal region of BTA5 from their ancestor A0 (Figure 1). This was confirmed by the haplotypes reconstructed with the 240 SNP. This haplotype of ancestor A0 is named “haplotype 1” or ($A0_{H1}$) and the alternative haplotype of A0 “haplotype 2” or ($A0_{H2}$). Family G03, previously declared as heterozygous for the target QTL [24] but not here, has inherited haplotypes not related to $A0_{H1}$ (Figure 1). All animals with haplotype $A0_{H1}$ can be traced back to A0. Two of these, grandsires G10 and G11 are paternal and maternal grandsons of A0 and are very important Fleckvieh bull sires. We have collected samples of all available progeny-tested sons of these two grandsires and all available progeny-tested maternal grandsons of grandsires G01, G02, G10 and G11 to add more recombinant A0 haplotypes into the mapping population. In a total of 485 animals were genotyped by the SNP chip and haplotyped for BTA5. By calculating the independent haplotypes in the complex pedigrees and considering the traceability of all $A0_{H1}$ haplotypes to A0, we estimated a very low frequency (<0.005) of $A0_{H1}$ in the Fleckvieh population. Consequently,

throughout the rest of this paper the less frequent putative QTL allele embedded in this less frequent haplotype is referred to as the minor QTL allele.

**Combined linkage disequilibrium and linkage analysis**

Thirty-seven microsatellite markers and the complex pedigree *MSPED2089* were used for initial LDL mapping analyses. As shown in Figure 2, we observed a highly significant QTL effect (LRT = 20 to 22, i.e. $P$ = 0.0000077 to 0.0000027) but because of the presence of two or three peaks we were not able to improve the mapping accuracy.
According to previous results [24] and to results obtained in the first part of this study, we have assumed that only one QTL has been introduced by haplotype $A0_{H1}$ into the mapping population. Therefore, we performed a second LDL analysis using the 21 most distal markers, and limited to GD and MGS families descending from A0 and known to carry $A0_{H1}$, i.e. pedigree *MSPED1038* (Figure 3). Unlike the analysis of pedigree *MSPED2089*, Figure 3 illustrates a single rather broad peak between positions 119.005 cM and 120.166 cM. However, this highly significant QTL ($P$ = 0.000062 to 0.000021) is still mapped with a low accuracy, i.e. 1-LOD drop-off support intervals are 4.7 cM for FY1, 10.4 cM for PY1 and 11.5 cM for MY1.

Since the confidence interval achieved by LDL analyses using pedigree *MSPED1038* was still too large for a positional candidate gene approach, we analysed pedigree *SNPPED723* using the LDL approach. The results were similar to those obtained with microsatellite markers and pedigree *MSPED2089*, namely, multiple peaks suggesting multiple QTL or no QTL (Figure 4).

To resolve this dilemma, we divided pedigree *SNPPED723* into pedigree *SNPPED421* consisting of all progeny-tested animals descending from ancestor A0, and pedigree *SNPPED308* consisting of the remaining progeny-tested animals. The LDL analyses of *SNPPED308* pedigree showed a moderately flat non significant test statistic along the investigated chromosomal segment (Figure 5). Only LRT values for FY1 reached an indicative level of 3.99 (P=0.046). Conversely, it was possible to map a QTL with pedigree *SNPPED421* whose minor allele is most probably originating from ancestor A0 (Figure 6). There were two distinct peaks, one with LRT values over 17 ($P$ < 0.000037) for

both MY1 and PY1 in a region of 0.5 Mb (from 118.107 to 118.606 Mb) and one with a very high LRT value for only PY1 (LRT = 20.72, *P* = 0.0000053) at position 122.115 Mb. Considering 1-LOD drop-off support intervals, the 97% confidence intervals were located between 117.962 Mb and 119.018 Mb (i.e. 1.056 Mb) for the QTL affecting MY1 and PY1 and between 121.800 Mb and 122.200 Mb (i.e. 0.400 Mb) for the QTL affecting only PY1. There were two additional peaks with LRT values over 15 in regions around the positions 115.650 and 116.300 Mb but they were not included in the 97% confidence interval for PY1 and were not supported by the highly correlated MY1 trait.

The two identified peaks (located between 118.107 Mb and 118.606 Mb and at 122.115 Mb, respectively) may be due to either the presence of more than one QTL or to the presence of one QTL with carryover effects to another region. Thus, a multiple-QTL analysis was performed. The two-QTL analyses using pedigree *SNPPED421* for MY1 and PY1 fitting a QTL at position 118.202 revealed a single QTL affecting MY1 at this location only and an additional QTL affecting PY1 at position 122.115 Mb (*P* = 0.019). However, two-QTL analyses accounting for the QTL at position 122.115 Mb did not rule out a possible second QTL affecting PY1 at position 118.202 Mb (*P* = 0.019).

## DISCUSSION

The aim of this study was to refine the position of a previously mapped QTL by increasing the marker density in the region, by target sampling of additional families and by adapting fine mapping methods. According to our previous results [24] and to results from the initial part of this study, we hypothesized the presence of a minor QTL allele with a strong effect but at a very low frequency in the Fleckvieh dual-purpose cattle breed. In such a situation, a random sampling of additional families for confirmation and fine-mapping purposes can result in an increased frequency of the common QTL allele in the mapping design. Thus, the capacity to differentiate between genetic background noise and initially targeted QTL will be decreased. The reduced accuracy of QTL position estimates when using all genotyped animals (pedigrees *MSPED2089* or *SNPPED723*) compared to a subset of animals (pedigrees *MSPED1038* or *SNPPED421*) is counterintuitive to the general notion that the use of more

information should result in better estimates. To further explore this unexpected result, we have investigated several possible explanations, including the effects of the haplotype distribution and the possibility of additional QTL. To study the haplotype distribution in the Fleckvieh population, 485 animals were genotyped with the Illumina 50K SNP chip. Of these, a subset of 144 animals was not progeny tested and was not relevant for QTL mapping but very informative to study haplotype distribution. In particular, considering the putative QTL affecting MY1 and PY1 located within the 97% CI (between 117.962 Mb and 119.018 Mb), a haplotype of 25 markers ($A0_{H1}$) covering this region was detected in 89 of 485 animals. This haplotype $A0_{H1}$, most probably carrying the minor QTL allele, could be traced back to the ancestor A0 in all 89 cases (Figure 1). The alternative haplotype $A0_{H2}$, most probably carrying the common QTL allele, was found in 13 cases but it was traced back to the ancestor A0 only in three. A perfect LD between the minor QTL allele and $A0_{H1}$ (and only $A0_{H1}$) would result in a relatively low allele frequency (0.137) of the derived QTL allele in phenotyped animals of pedigree *SNPPED723* and in a frequency about double (0.254) in pedigree *SNPPED421.* The mapping results did reflect this difference too. In contrast, consider the six markers located within the 97% CI (between 121.800 Mb and 122.200 Mb) of the putative QTL region affecting only PY1. Ancestor A0 is homozygous for a very long segment of this region i.e. from positions 118.266 Mb to 123.347 Mb (three SNP telomeric to the main peak of QTL affecting MY1 and PY1). This segment of 5.080 Mb includes 109 informative markers in the Fleckvieh population. Comparison of mapping results from pedigrees *SNPPED723* (Figure 4), *SNPPED421* (Figure 6), and *SNPPED308* (Figure 5) revealed a highly significant QTL allele affecting PY1 only when the pedigree included families segregating for haplotype $A0_{H1}$ (see comparison between Figures 4 and 6). Excluding these families yielded LRT values below 3.99 ($P > 0.045$) for all three milk yield traits and for the complete investigated region (Figure 5, between 113.500 Mb and 123.700 Mb). We therefore used the linkage information in the *SNPPED421* pedigree ($A0_{H1}$ always traceable to A0), to map a QTL affecting both MY1 and PY1 in a 97% CI of 1 Mb.

Haplotype and LDL analyses by microsatellite designs (Figures 2 and 3) and SNP designs (Figures 4 and 6) clearly suggest that the minor QTL allele associated with the putative QTL around the physical position 118.00 Mb (97%

CI between 117.962 Mb to 119.018 Mb) has been introduced by ancestor A0 into the mapping population. The explanation of the second possible QTL that maps to the physical position 122.115 Mb and affects only PY1 is different. First, this QTL should also be associated with ancestor A0 haplotypes, i.e. absence of effect in the smaller *SNPPED308* pedigree (Figure 5). Second, both ancestor haplotypes at the physical position 122.115 Mb are most probably identical by descent (i.e. homozygous for a 5.080 Mb segment with 109 informative SNP). Therefore, ancestor A0 is most probably homozygous for the putative QTL at this position too. Third, this part of the haplotype is not unique to A0, but also segregates in other families, i.e. there is LD information for mapping, too. The relatively sharp LRT peak at position 122.115 Mb and homozygosity of A0 suggest an essential contribution of LD to this mapping result. Fourth, analyses with the two-QTL model did not rule out the possibility of a second QTL affecting PY1 within the candidate region on BTA5. And finally, despite the overall presence of haplotypes with a high IBD to ancestor haplotypes around position 122.115 Mb, the complete absence of this peak in *SNPPED308* pedigree can be explained by either a novel mutation in ancestor A0 or by the incapacity of the method and design used here to map it in a relatively small pedigree like *SNPPED308*. more reasonable explanation may be the lower statistical power of the pedigree *SNPPED308* possible local inconsistencies in the map order (which was based on map release Btau_4.0), the presence of a strong QTL at position 118.000 Mb with carryover effects to other regions, or a combination of all these explanations.

The LDL analysis using SNP and pedigree *SNPPED723* indicate several peaks affecting MY1 and PY1 in the region investigated here. In principle, these results (Figure 4) are comparable to the fine-mapping results reported on BTA3 by Druet et al. [38]. In this study, the authors also first carried out mapping by linkage analysis and finally ended up with LDL analyses and multiple LRT peaks. We used larger overlapping marker windows (80 SNP) than Druet et al. [38]. By dividing the data set according to the results of linkage and haplotype analyses, most of the multiple peaks were explained as genetic background noise in a larger family set. The multiple peak profile could be explained by the heterogeneous LD structure within the QTL region or by the use of LD in the model when there is no LD information at all [38]. This might be increased by

possible local inconsistencies in the map order, which was based on the draft assembly or on comparative map information. Moreover, the method and the data structure may not make it possible to discard some regions even though they do not harbour the QTL [38].

To check for a possible effects of the data structure on the reported mapping results we tested regression of EBV on genetic distance from ancestor A0 for all carriers of haplotype 1 ($A0_{H1}$). The apparent lack of this regression suggests that we are looking at a real QTL effect, and not an artifact of pedigree-tracking.

Searching the region between 117.900 and 119.100 Mb for candidate genes revealed 27 genes, 13 of which had no known function. Based on current biological information, genes with partly known function could only be indirectly related to milk yield traits.

## Conclusions

In the present study, we have performed a haplotype-assisted extension of the mapping design and thus increased the allele frequency of the minor QTL allele in mapping families. Alternative analyses with family subsets resulted in a substantial reduction of the genetic background noise and an increased frequency of the minor QTL allele. Using these subsets, we succeeded in refining the map position of the previously detected QTL for milk production traits on BTA5 to a 1 Mb interval. In spite of implementing a two-QTL analysis, the possibility of a second QTL affecting only PY1 could not be ruled out. All in all, the results of both this study and the previous study by Awad et al. [24] support the presence of a QTL affecting both, MY1 and PY1 that close to the centromeric part of the long homozygous region (~5 Mb) in ancestor A0. Therefore, positional cloning and high-throughput sequencing of the candidate region located between 117.900 Mb and 119.100 Mb should now be considered but should also not neglect the second possible QTL around position 122.115 Mb.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

AA carried out DNA extraction, microsatellite genotyping; AA and IM performed all data analysis and wrote the paper; IM and MF designed the study; IR performed SNP genotyping and partly performed sampling. All authors read and approved the final manuscript.

## Acknowledgements

Ashraf Awad was supported by the Ministry of Higher Education, Egypt. We thank Tierzuchtforschung e.V. München for providing genome-wide SNP genotypes of animals related to ancestor A0 and for providing a part of the DNA samples analysed here. One hundred and seventy-three Fleckvieh animals not connected with A0 were genome-wide genotyped within the project ME3404/1-1 gratefully funded by German Research Foundation (DFG). We thank all breeders and breeding associations who sent us remaining samples free of charge to support this study. In particular, we thank Arbeitsgemeinschaft Süddeutscher Rinderzucht- und Besamungsorganisationen e. V. (ASR) and the Bavarian Gene Reserves at LfL, Grub, Germany. We also thank the associate editor in charge and two reviewers whose comments resulted in substantial improvement of the final manuscript version.

## References

[1] Andersson L: **Genetic dissection of phenotypic diversity in farm animals.** *Nat Rev Genet* 2001, **2**:130-138.

[2] Weiss KM, Terwilliger JD: **How many diseases does it take to map a gene with SNPs?** *Nat Genet* 2000, **26**:151-157.

[3] Meuwissen THE, Karlsen A, Lien S, Olsaker I, Goddard ME: **Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping.** *Genetics* 2002, **161**:373-379.

[4] Olsen HG, Lien S, Gautier M, Nilsen H, Roseth A, Berg PR, Sundsaasen KK, Svendsen M, Meuwissen TH: **Mapping of a milk production quantitative trait locus to a 420-kb region on bovine chromosome 6.** *Genetics* 2005, **169**:275-283.

[5] Gautier M, Barcelona RR, Fritz S, Grohs C, Druet T, Boichard D, Eggen A, Meuwissen TH: **Fine mapping and physical characterization of two**

**linked quantitative trait loci affecting milk fat yield in dairy cattle on BTA26.** *Genetics* 2006, **172**:425-436.

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

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

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

[9] Cohen-Zinder M, Seroussi E, Larkin DM., Loor JJ, Everts-van der Wind A, Heon-Lee J, Drackley JK, Band MR, Hernandez AG, Shani M, Lewin HA, Weller JI, Ron M: **Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle.** *Genome Res* 2005, **15**:936-944.

[10] Khatkar MS, Thomson PC, Tammen I, Cavanagh JA, Nicholas FW, Raadsma HW: **Linkage disequilibrium on chromosome 6 in Australian Holstein-Friesian cattle.** *Genet Sel Evol* 2006, **38**:463-477.

[11] Khatkar MS, Zenger KR, Hobbs M, Hawken RJ , Cavanagh JAL , Barris W, McClintock AE, McClintock S, Thomson PC , Tier B, Nicholas FW, Raadsma HW: **A primary assembly of a bovine haplotype block map based on a 15,036-single-nucleotide polymorphism panel genotyped in Holstein–Friesian cattle.** *Genetics* 2007, **176**:763-772.

[12] Gautier M, Faraut T, Moazami-Goudarzi K, Navratil V, Foglio M, Grohs C, Boland A, Garnier JG, Boichard D, Lathrop GM, Gut IG, Eggen A: **Genetic and haplotypic structure in 14 European and African cattle breeds.** *Genetics* 2007, **177**:1059-1070.

[13] Meuwissen THE, Goddard ME: **Fine mapping of quantitative trait loci using linkage disequilibrium with closely linked marker loci.** *Genetics* 2000, **155**:421-430.

[14] Hedrick PW: **Gametic disequilibrium measures: Proceed with caution.** *Genetics* 1987, **117**:331-341.

[15] Farnir F, Coppieters W, Arranz JJ, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M: **Extensive genome-wide linkage disequilibrium in cattle.** *Genome Res* 2000, **10**:220-227.

[16] Tenesa A, Knott SA, Ward D, Smith D, Williams JL, Visscher PM: **Estimation of linkage disequilibrium in a sample of the United**

**Kingdom dairy cattle population using unphased genotypes.** *Anim Sci* 2003, **81**:617-623.

[17] Lipkin E, Straus K, Tal Stein R, Bagnato A, Schiavini F, Fontanesi L, Russo V, Medugorac I, Foerster M, Sölkner J, Dolezal M, Medrano JF, Friedmann A, Soller M: **Extensive long-range and non-syntenic linkage disequilibrium in livestock populations: deconstruction of a conundrum.** *Genetics* 2009, **181**:691-699.

[18] Bovine HapMap Consortium: **Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds**. *Science* 2009, **324**:528-531.

[19] Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME, Visscher PM: **Recent human effective population size estimated from linkage disequilibrium.** *Genome Res* 2007, **17**:520-526.

[20] Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL: **Genome sequence, comparative analysis and haplotype structure of the domestic dog.** *Nature* 2005, **438**:803-819.

[21] Goddard ME, Meuwissen THE: **The use of linkage disequilibrium to map quantitative trait loci.** *Aust J Exp Agric* 2005, **45**:837-845.

[22] Bennewitz J, Reinsch N, Grohs C, Leveziel H, Malafosse A, Thomsen H, Xu N, Looft C, Kuhn C, Brockmann GA, Schwerin M, Weimann C, Hiendleder S, Erhardt G, Medjugorac I, Russ I, Forster M, Brenig B, Reinhardt F, Reents R, Averdunk G, Blumel J, Boichard D, 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* 2003, **35**:319-338.

[23] Ashwell MS, Heyen DW, Sonstegard TS, Van Tassell CP: **Detection of quantitative trait Loci affecting milk production, health, and reproductive traits in Holstein cattle.** *J Dairy Sci* 2004, **87**:468-475.

[24] Awad A, Russ I, Emmerling R, Förster M, Medugorac I: **Confirmation and refinement of a QTL on BTA5 affecting milk production traits in the Fleckvieh dual purpose cattle breed.** *Anim Genet* 2010*,* **41**:1-11

[25] Kolbehdari D, Wang Z, Grant JR, Murdoch B, Prasad A, Xiu Z, Marques E, Stothard P, Moore SS: **A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls.** *J Anim Breed Genet* 2009, **126**:216-227.

[26] Emmerling R, Lidauer M, Mäntysaari EA: **Multiple lactation random regression test-day model for Simmental and Brown Swiss in Germany and Austria**. *Interbull Bull* 2002, **29**:111-117.

[27] Ihara N, Takasuga A, Mizoshita K, Takeda H, Sugimoto M, Mizoguchi Y, Hirano T, Itoh T, Watanabe T, Reed KM, Snelling WM, Kappes SM, Beattie CW, Bennett GL, Sugimoto Y: **A comprehensive genetic map of the cattle genome based on 3802 microsatellites.** *Genome Res* 2004, **14**:1987-1998.

[28] Browning BL, Browning SR: **A unified approach to genotype imputation and haplotype phase inference for large data sets of trios and unrelated individuals.** *Am J Hum Genet* 2009, **84**:210-223.

[29] Green P, Falls K, Crooks S: **Documentation for CRI- MAP, Version 2.4.,** Washington University School of Medicine, St. Louis, 1990.

[30] Lee SH, van der Werf JH: **The efficiency of designs for fine-mapping of quantitative trait loci using combined linkage disequilibrium and linkage.** *Genet Sel Evol* 2004**, 36**:145-161**.**

[31] Lee SH, Van der Werf JH: **The role of pedigree information in combined linkage disequilibrium and linkage mapping of quantitative trait loci in a general complex pedigree.** *Genetics* 2005, **169**:455-466.

[32] Lee SH, Van der Werf JH: **An efficient variance component approach implementing average information REML suitable for combined LD and linkage mapping with a general complex pedigree.** *Genet Sel Evol* 2006, **38**:25-43.

[33] Olsen HG, Meuwissen THE, Nilsen H, Svendsen M, Lien S: **Fine mapping of quantitative trait loci on bovine chromosome 6 affecting calving difficulty.** *J Dairy Sci* 2008, **91**:4312-4322.

[34] Jansen RC: **Interval mapping of multiple quantitative trait loci.** *Genetics* 1993, **135**:205-211.

[35] Johnson DL, Thompson R: **Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information.** *J Dairy Sci* 1995, **78**:449-456.

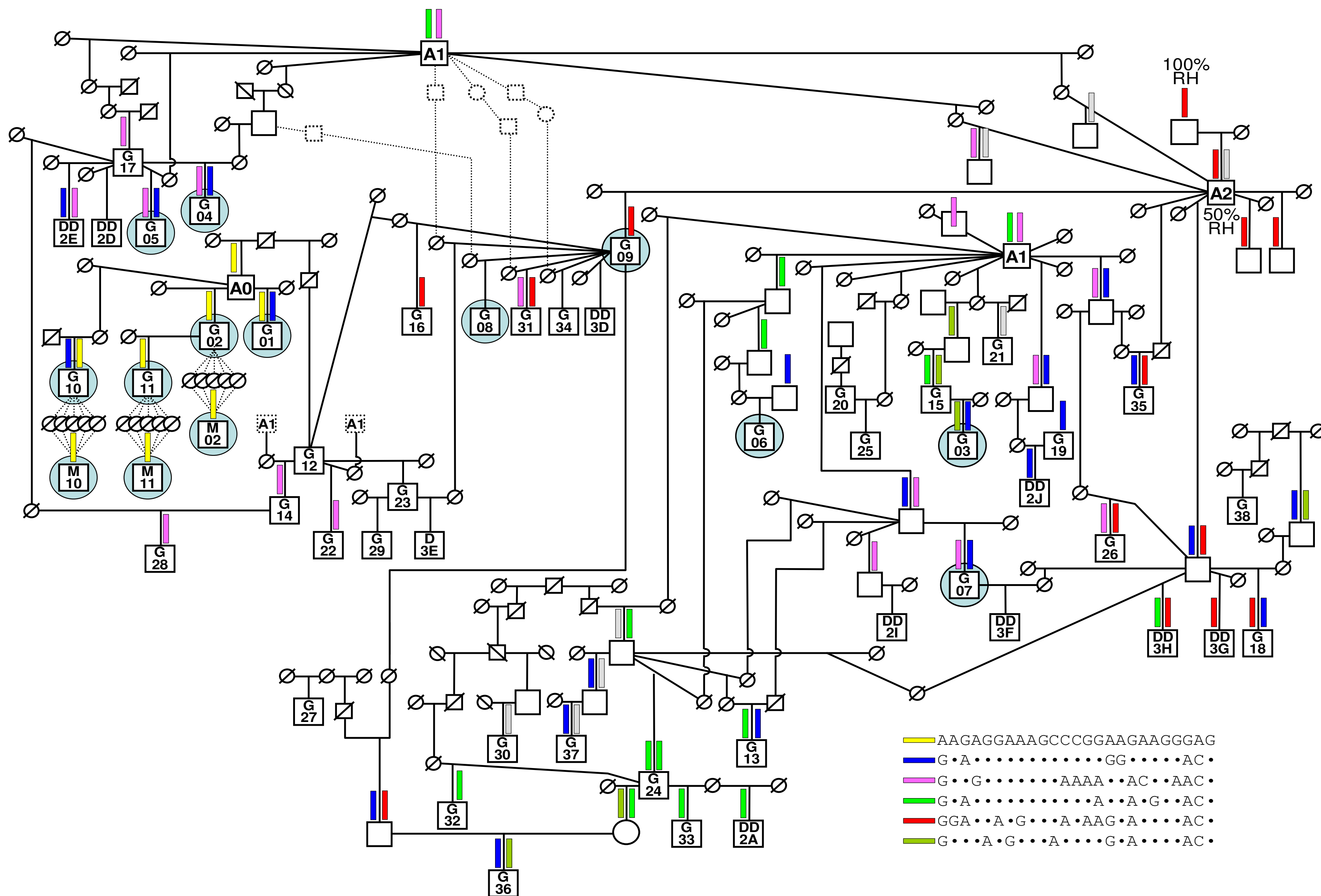
[36] Olsen HG, Lien S, Svendsen M, Nilsen H, Roseth A, Aasland OM, Meuwissen TH: **Fine mapping of milk production QTL on BTA6 by combined linkage and linkage disequilibrium analysis**. *J Dairy Sci* 2004, **87**:690-698.

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

[38] Druet T, Fritz S, Boussaha M, Ben-Jemaa S, Guillaume F, Derbala D, Zelenika D, Lechner D, Charon C, Boichard D, Gut IG, Eggen A, Gautier M**: Fine mapping of quantitative trait loci affecting female fertility in dairy cattle on BTA03 using a dense single-nucleotide polymorphism map.** *Genetics* 2008, **178**:2227-2235.

**Table 1 Microsatellite markers used for QTL mapping.** marker name, relative position (cM), physical position (bp), forward and reverse primer sequences and marker set (set: Set0 & Set1 as in previous study; Set3 & Set4 multiplex 1&2 in this study )

| No | Marker ID | cM | bp | Forward primer | Reverse primer | Remarks |
|---|---|---|---|---|---|---|
| 1 | LMU0502 | 95.00 | 98418609-98419268 | TGGAAGAATATGCAGGTAACTCT | GTCGCTCTTTGTGGCTTCAC | Set1 |
| 2 | DIK2336 | 99.79 | 101071987-101072659 | ATGTGGAATGTAGGGCAAGG | TCCCTCACCTTTCGAACAAA | Set1 |
| 3 | BM315 | 103.17 | 104045839-104046013 | TGGTTTAGCAGAGAGCACATG | GCTCCTAGCCCTGCACAC | Set0 |
| 4 | DIK4843 | 107.02 | 107077504-107078179 | CATGCAAGCTTTCAAGAATGA | TGCAGAGATAAGCCGAGGAC | Set4 |
| 5 | DIK1135 | 108.22 | 10181410-10182069 | GTCTGCCATCTAGCCAAAAA | GTTTTTCAGTGGGCATTTGG | Set1 |
| 6 | DIK5238 | 110.97 | 111864734-111865363 | TGGAACCAGTGAAGTTTAGGG | GAAATGCCCACTGAAGCTCT | Set3 |
| 7 | ETH2 | 112.43 | 112903902-112909263 | ATTTGCCCTGCTAGCTTTGA | AAGACTCTGGGCTTCAAAAGG | Set1 |
| 8 | DIK2122 | 114.68 | 113216193-113216706 | CAACAAACTGTGCGTTGTGA | ACTCAGCAGTTGCCCTCAGT | Set3 |
| 9 | BM2830 | 116.91 | 115262054-115262075 | AATGGGCGTATAAACACAGATG | TGAGTCCTGTCACCATCAGC | Set0 |
| 10 | BM49 | 118.06 | 116205343-116205972 | CACCATATTTGCCAGGATCA | GCGGGATCTCACTAAACCAG | Set3 |
| 11 | BM733 | 119.95 | 117125799-117126005 | CTGGAGTCTCCTCCGTTGAG | AGAGAGGGCCCTTGTGAGAT | Set4 |
| 12 | DIK2035 | 120.85 | 119370626-119371127 | CAGTCAATGCAGGAAAAGCA | GCTGCTAGAGGGAGACAGGA | Set3 |
| 13 | DIK5277 | 121.53 | 120099447-120100247 | ACCCAAACTTAGCGTGGATG | GTCTCCAAGGCTGCTCACTC | Set3 |
| 14 | DIK5106 | 121.47 | 118461214-118461602 | GCATGTGTGCAGAAGAAGGA | TGTTCAGTGGTTCCCTGTGA | Set3 |
| 15 | LMU0505 | 123.64 | 121423920-121424520 | TGCAAGGAGAAGCGGTAGAT | TGCACACTTACCCCATGTTC | Set3 |
| 16 | ETH152 | 124.95 | Unknown | GTTCTCAGGCTTCAGCTTCG | TGATCAGAGGGCACCTGTCT | Set1 |
| 17 | URB060 | 127.55 | 122472602-122473177 | TTGTCATTTCTGGACTCCACTG | CAGGTCCAACCCTGTTTAGC | Set1 |
| 18 | DIK5212 | 129.17 | 123262266-123262905 | GGCTGGAACAGTGACTCTGG | GGACCCAGATTTCAATGGAG | Set3 |
| 19 | DIK5247 | 129.80 | 123619504-123619855 | GGGTCTGTAGGGAGAAGCTG | GCTTTCGAGAAGCATCCACT | Set3 |
| 20 | MNB71 | 133.09 | Unknown | CATCTAAGGCAGAGCCAACC | TTCTTGGTGCCTCTCTCTCC | Set1 |
| 21 | NOR44 | 133.98 | 125340968-125341598 | ACCCACCCGTACACATTCAA | GGGGAGGAGATGGACTGTTC | Set3 |



A1
A2
A0
100% RH
50% RH
G 01
G 02
G 03
G 04
G 05
G 06
G 07
G 08
G 09
G 10
G 11
G 12
G 13
G 14
G 15
G 16
G 17
G 18
G 19
G 20
G 21
G 22
G 23
G 24
G 25
G 26
G 27
G 28
G 29
G 30
G 31
G 32
G 33
G 34
G 35
G 36
G 37
G 38
M 02
M 10
M 11
DD 2A
DD 2D
DD 2E
DD 2I
DD 2J
DD 3D
DD 3F
DD 3G
DD 3H
D 3E
AAGAGGAAAGCCCGGAAGAAGGGAG
G•A••••••••••••GG•••••AC•
G••G•••••••AAAA••AC••AAC•
G•A•••••••••••A••A•G••AC•
GGA••A•G•••A•AAG•A••••AC•
G•••A•G•••A••••G•A••••AC•

**Figure 1 - Familial relationships considered in this study and segregation of most important haplotypes**

A complex pedigree of 38 sires (squares) of GD families (G), ten sires of daughter design (DD) families, three maternal grandsire (M) families and 26 sampled and genotyped relevant ancestors; the pedigree has been simplified by showing only ancestors who made it possible to trace haplotypes from family-sires to the most important ancestors (A0, A1, A2); furthermore, to reduce the complexity of the figure, ancestor A1 is represented more than one; correspondingly, letters and numbers within squares of family-sires represent the internal family ID; non- genotyped individuals are represented by smaller circles (females) and squares (males) marked with a diagonal line; the estimated haplotype of 25 markers ($A0_{H1}$) comprising a derived QTL allele affecting MY1 and PY1 with 97% CI between 117.962 Mb and 119.018 Mb is graphically presented by yellow bars above the individual's symbol; five other most frequent haplotypes are represented by five different coloured bars; introgression of Red-Holstein genes into the mapping populations is represented by ancestor A2 and the corresponding haplotype presented by a red bar; to reduce the complexity of the figure, 77 low frequency haplotypes are omitted; the allelic composition of the respective haplotypes is presented within the figure; the pedigree *MSPED2089* is a subset of the total material which can be constructed by keeping the families marked by a grey circle around squares and associated ancestors; pedigrees *MSPED1038* and *SNPPED421* are subsets of *MSPED2089* which can be constructed by removing appropriate families as described in material and methods; the pedigree *SNPPED308* consists of GD family G36 and animals across the entire mapping population but not descending from A0; the pedigree *SNPPED723* is a sum of pedigrees *SNPPED308* and *SNPPED421*

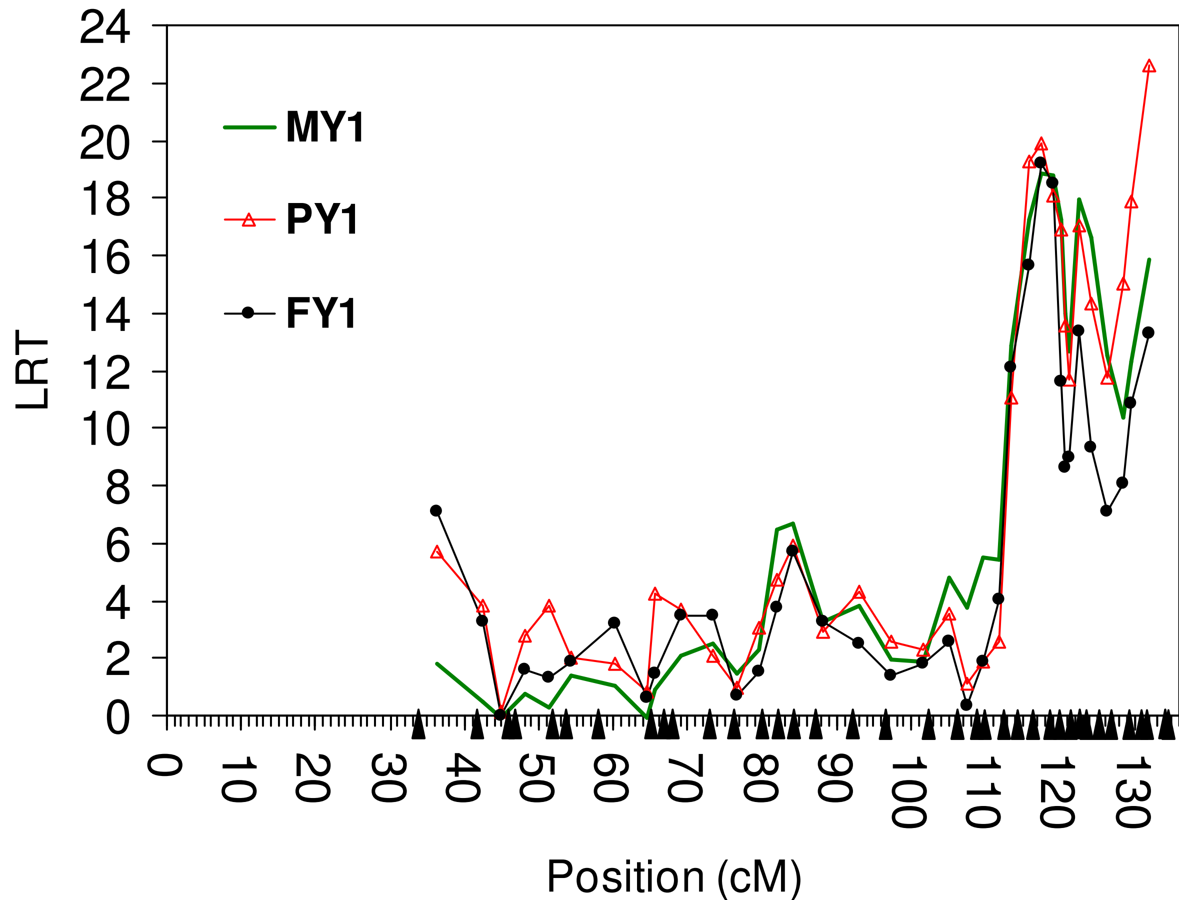



**Figure 2 LDL analysis by variance component approach using microsatellites in a complex pedigree of 2089 animals**
Joint linkage disequilibrium and linkage (LDL) analysis for three milk yield traits; Milk Yield (MY1), Milk Protein Yield (PY1) and Milk Fat Yield (FY1) during first lactation using thirty-seven microsatellites, a complex pedigree of 2089 animals, EBV as phenotype and AIREML as implemented in *LDLRAMS* and *LDL* program. Chromosome length in centiMorgan (cM) on the X-axis, log-likelihood ratio test (LRT) values on the Y-axis. Solid triangles on the X-axis represent positions of markers included in the analysis.

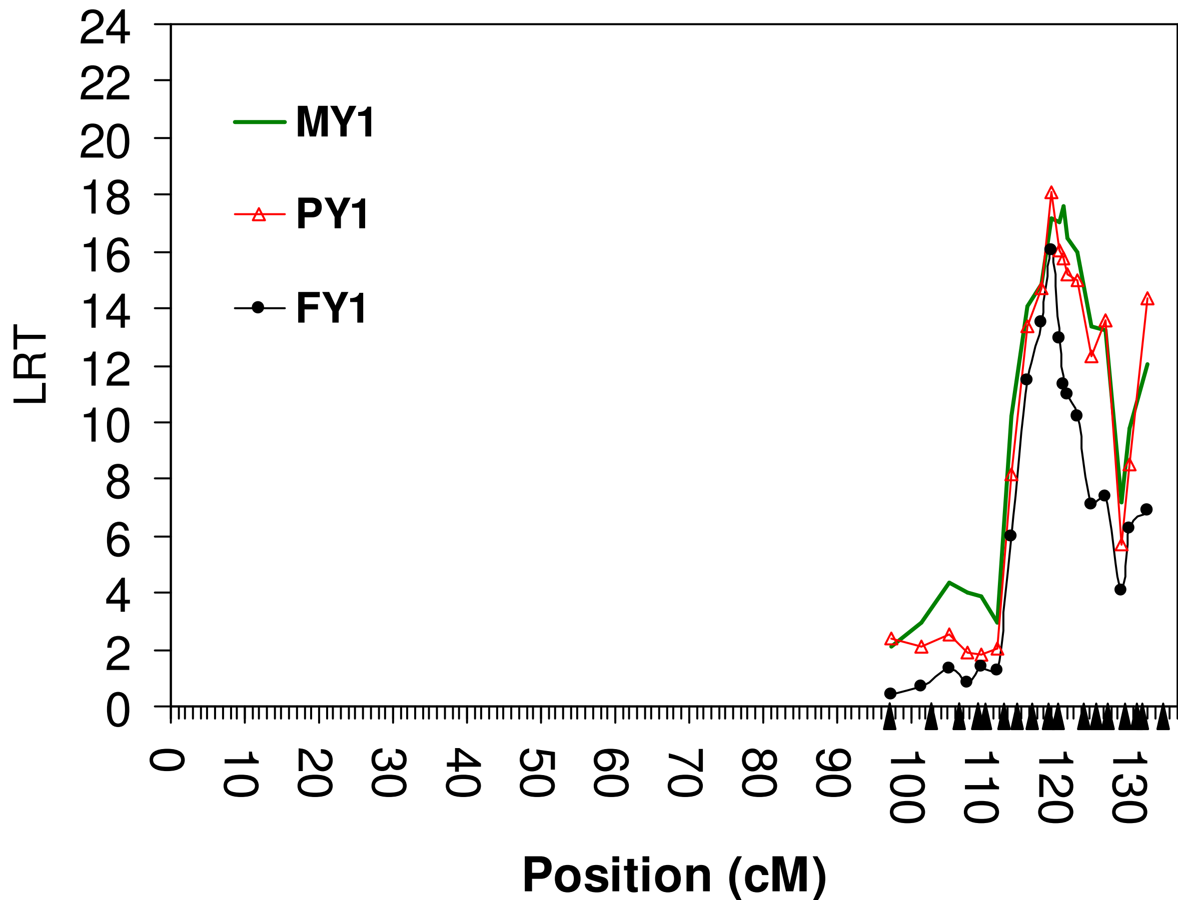



**Figure 3 LDL analysis by variance component approach using microsatellites in a complex pedigree of 1038 animals**
Joint linkage disequilibrium and linkage (LDL) analysis for three milk yield traits; Milk Yield (MY1), Milk Protein Yield (PY1) and Milk Fat Yield (FY1) during first lactation using twenty-one microsatellites covered the most likely region containing the QTL in the distal part of bovine chromosome 5 (BTA5), a complex pedigree of 1038 animals, EBV as phenotype and AIREML as implemented in *LDLRAMS* and *LDL* program. Chromosome length in centiMorgan (cM) on the X-axis, log-likelihood ratio test (LRT) values on the Y-axis. Solid triangles on the X-axis represent positions of markers included in the analysis

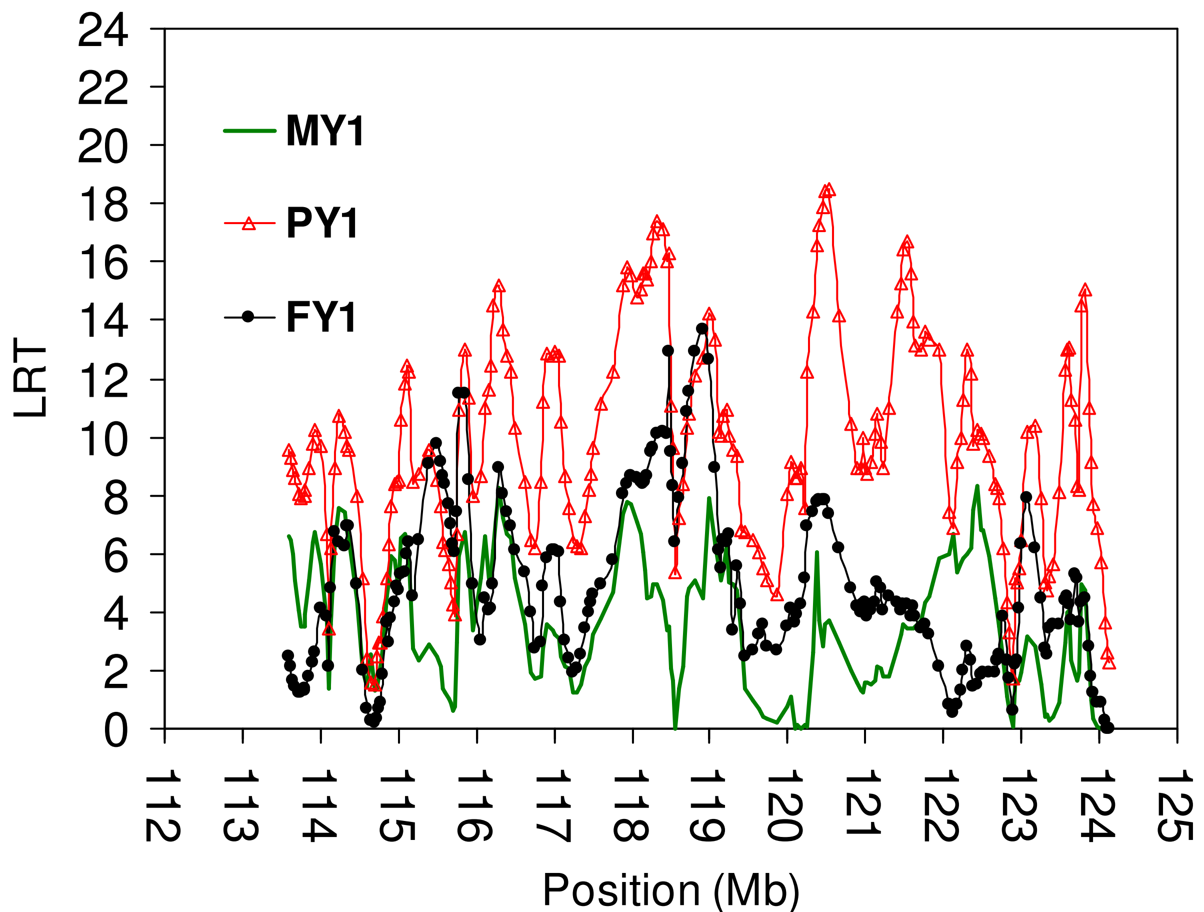



**Figure 4 - LDL analysis by variance component approach using SNP in a complex pedigree of 723 animals**

Joint linkage disequilibrium and linkage (LDL) analysis for three milk yield traits; Milk Yield (MY1), Milk Protein Yield (PY1) and Milk Fat Yield (FY1) during first lactation using Two hundred and forty SNPs covering the most likely region containing the QTL in the distal part of bovine chromosome 5 (BTA5), a complex pedigree of 723 animals, EBV as phenotype and AIREML as implemented in *LDLRAMS* and *LDL* program. Chromosome length in Megabase (Mb) on the X-axis, log-likelihood ratio test (LRT) values on the Y-axis.

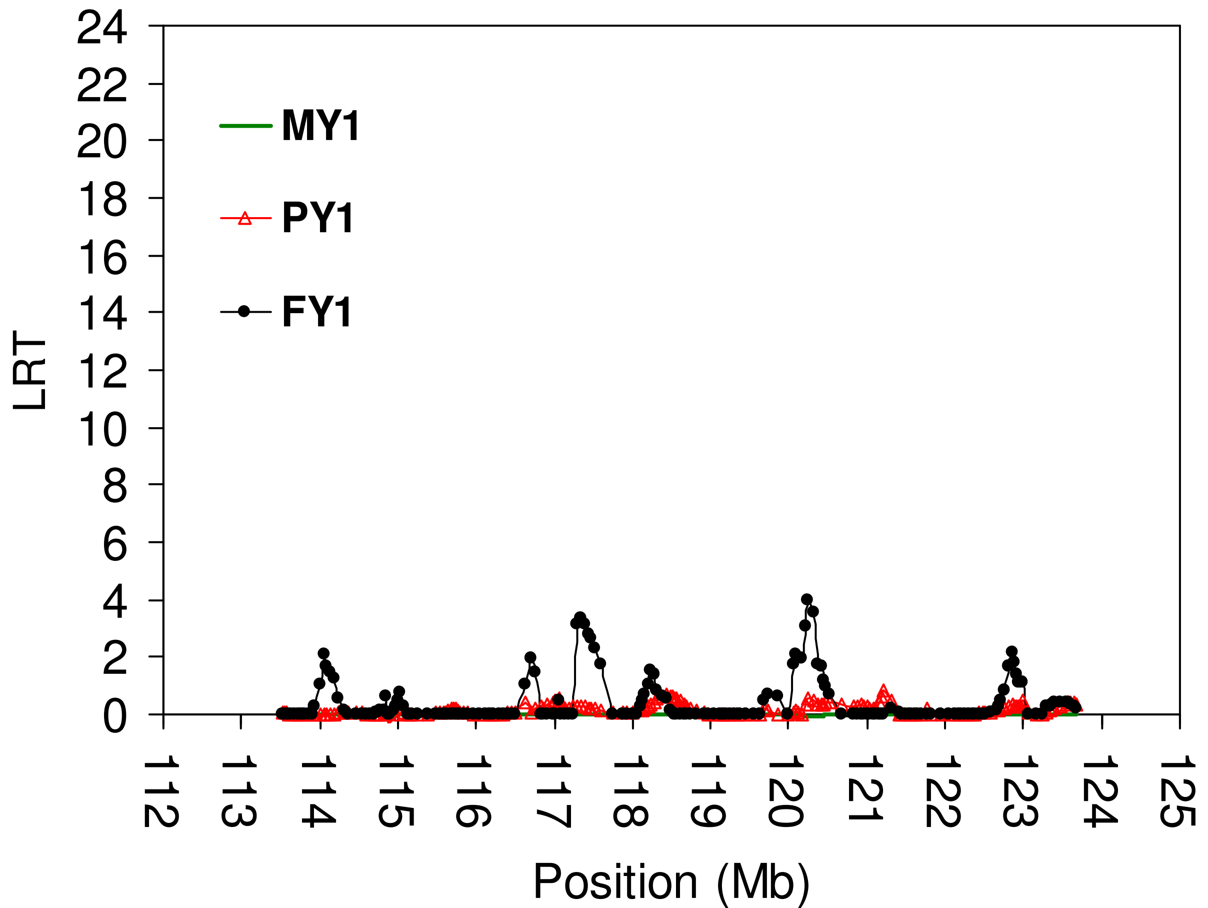



**Figure 5 - LDL analysis by variance component approach using SNP in a complex pedigree of 308 animals**
Joint linkage disequilibrium and linkage (LDL) analysis for three milk yield traits; Milk Yield (MY1), Milk Protein Yield (PY1) and Milk Fat Yield (FY1) during first lactation using Two hundred and forty SNPs covering the most likely region containing the QTL in the distal part of bovine chromosome 5 (BTA5), a complex pedigree of 308 animals, EBV as phenotype and AIREML as implemented in *LDLRAMS* and *LDL* program. Chromosome length in Megabase (Mb) on the X-axis, log-likelihood ratio test (LRT) values on the Y-axis.

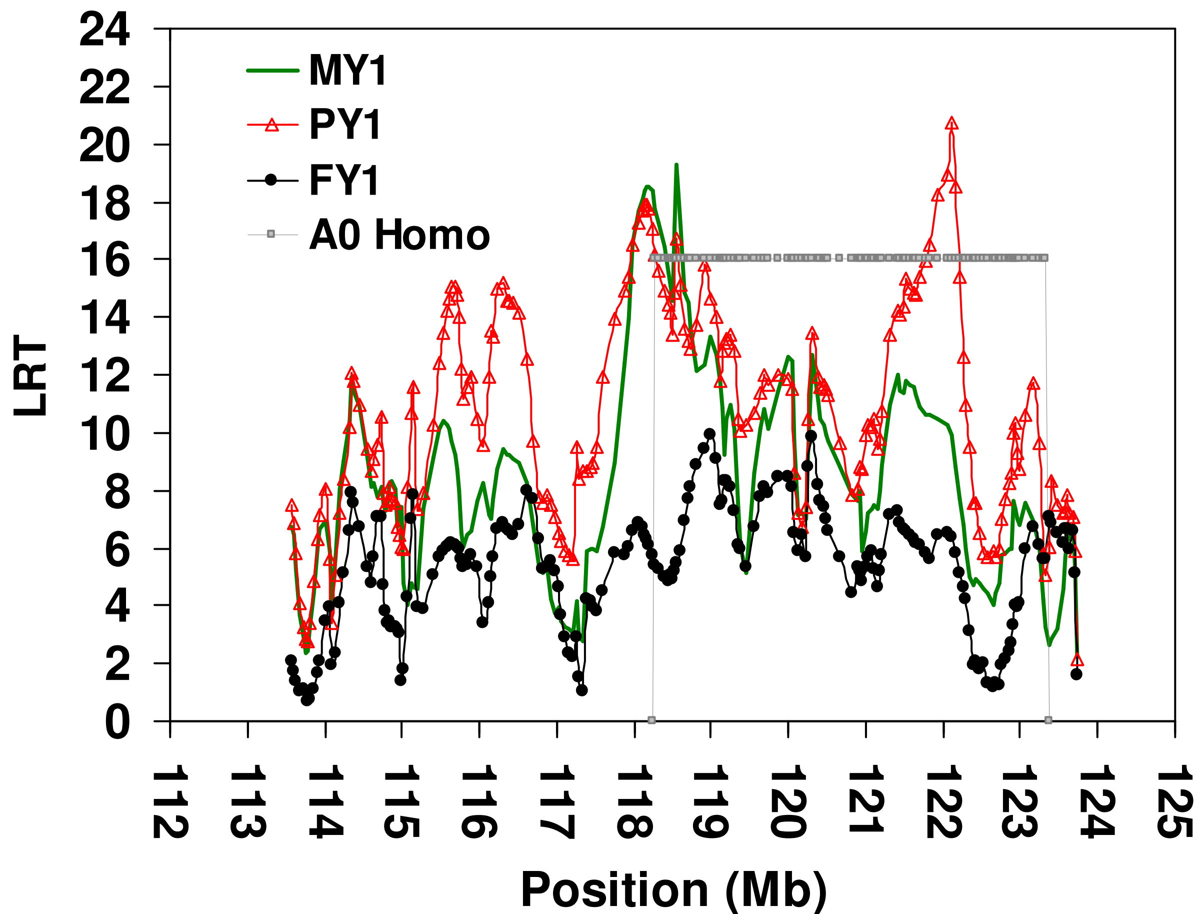



**Figure 6 - LDL analysis by variance component approach using SNP in a complex pedigree of 421 animals**

Joint linkage disequilibrium and linkage (LDL) analysis for three milk yield traits; Milk Yield (MY1), Milk Protein Yield (PY1) and Milk Fat Yield (FY1) during first lactation using Two hundred and forty SNPs covering the most likely region containing the QTL in the distal part of bovine chromosome 5 (BTA5), a complex pedigree of 421 animals, EBV as phenotype and AIREML as implemented in LDLRAMS and LDL program. Chromosome length in Megabase (Mb) on the X-axis, log-likelihood ratio test (LRT) values on the Y-axis. The long homozygous region (~5 Mb) in ancestor A0 was showed (A0 Homo)

# CHAPTER 5

# GENERAL DISCUSSION AND CONCLUSION

One of the primary goals of modern genetics is to understand the genetic basis of complex traits. What are the genes and alleles that contribute to the trait of interest, and how do they interact with each other and with environmental factors to produce phenotypes? The availability of genetic linkage maps mainly composed of highly polymorphic microsatellite markers allows the genetic dissection of complex traits into QTL. Identification of QTL is a first step towards the identification of the genes involved in the regulation of a quantitative trait. However, even if the actual genes are not known it is possible to enhance the selection efficiency by use of genetic markers that are closely linked to a QTL of interest.

Several QTL for milk production traits have been identified on all autosomes of the bovine genome. Khatkar et al. (2004), in a meta-analysis, considered 55 studies in dairy cattle populations and provided an overview of QTL reported in literature. Some of these QTL were repeatedly detected and mapped in numerous studies (e.g. Boichard et al. 2003; Ashwell et al. 2004; Schnabel et al. 2005; Chen et al. 2006; Daetwyler et al. 2008; Kolbehdari et al. 2009; Mei et al. 2009). The results of previous QTL studies differ somewhat with respect to the number of QTL detected on a chromosome, their positions, and the extent to which the milk production traits are affected (Khatkar et al. 2004).

The first objective of this study was to map QTL for milk production traits in the dual purpose Fleckvieh cattle breed on BTA5. In view of that, twenty-eight evenly distributed microsatellite markers covering bovine chromosome 5 were selected to genotype nine paternal half-sib families constructed in a granddaughter design. Three different mapping methods: interval mapping (IM), linkage analysis by variance component analysis (LAVC) and combined linkage disequilibrium (LD) and linkage (LDL) mapping method were used for analysis.
By successive mapping steps we illustrated that the distal part of chromosome five harbour a QTL affecting MY, PY and FY in the first lactation. Different

mapping methods –IM, LA by VC approach and LDL– mapped this QTL with highest test-statistic value to almost the same marker interval (BM2830-ETH152) and nearly at the same position: 118, 119.33 and 119.33 cM respectively. This results is in coordination with a number of studies which mapped QTL for milk production traits at the distal part of BTA5 e.g. Olsen et al. (2002) reported a QTL affecting fat yield at a position 115 cM from the centromere. De Koning et al. (2001) mapped a QTL for milk yield at 107 cM in the Finnish Ayrshire population. Using the same population, Viitala et al. (2003) detected another QTL for milk yield at 98 cM. A QTL for milk yield was also reported at position 109 cM by Bennewitz et al. (2003). Rodriguez-Zas et al. (2002) reported that a QTL affecting protein yield was located at position 91 cM. The three applied methods show differences in the shape and in the peak of test statistic curve, the shape of the curve is significantly sharper by LA-VC than by IM and more sharper by LDL than by the LA-VC approach, and the peak obtained by the LDL method was higher and narrower than that of the LAVC approach. The test statistics used for IM and VC-LA as well as for LDL approach are not directly comparable, but we reject the null hypothesis with higher probability as we use more information.

The improved accuracy of QTL position estimates when using additional information on relationships (from IM to VC-LA) or additional information on relationships and historical recombination events (from IM over VC-LA to LDL) is plausible to the general notion that use of more information should result in better estimates. In our case, this general notion does not apply for across-family interval mapping, where analysis across nine families was not able to detect QTL segregating in subset of families. On the other hand, connecting the same nine families by numerator relationship matrix (LA-VC approach) did result in significant effects of mapped QTL. However, the shape of test statistics estimated by IM based on all nine GDD families (Fig. 1a) was nearly comparable with that obtained by the LAVC approach (Fig. 2): both approaches showed a peak for FY1 at the middle region of the chromosome and another peak for MY1 and PY1 at the distal region of the chromosome 5. The crucial difference was in the significance of the results: the LAVC approach estimated significant QTL effects for FY1 at 73.3 cM and for MY1 and PY1 at 119.33 cM

whereas IM based on these nine families showed a fairly flat non-significant test statistic along the chromosome.

Few studies have reported on a comparison of data analyzed with the different methods (de Koning et al. 2003; Nagamine et al. 2004; Grapes et al. 2004; Kolbehdari et al. 2005). The regression and variance component method have similar power in a simple pedigree structure. In complex pedigrees, the variance component method is thought to achieve greater power to detect QTL. Furthermore, the variance component method uses potential information from segregation on the maternal side (Nagamine et al. 2004). Other advantages of the VC method are simultaneous estimation of (none) genetic effects, less sensitivity for small family sizes or less informative markers, or both (Kolbehdari et al. 2005). Also, the VC method provides estimates of the effect of each haplotype, which links up with breeding value estimation. According to the theoretical expectations (e.g. Lee & Van der Werf 2005) and practical observation (Meuwissen et al. 2002; Olsen et al. 2005) the LDL mapping compared to the LA mapping as performed by *LOKI* should give not only sharper mapping but also higher LRT value at the most probable QTL position. This is the case in our study, although increase of test-statistic value and mapping accuracy depend on present level of LD in the mapping population and marker density.

Comparison of allelic richness (El Mousadik & Petit 1996) and other diversity parameters among seven Alpine and three north-west European cattle breeds reveal Fleckvieh as breed with highest genetic variability (Medugorac et al. 2009). The highest allelic richness among compared breeds reflects a large effective population size of Fleckvieh (Sölkner et al. 1998, Pirchner 2002). According to large effective population size we expect lower level of LD comparing to one-purpose breeds like Holstein-Friesian cattle that have smaller effective population sizes. Indeed, here estimate average LD, as measured by Hedrick's multi-allelic *D'* value (Hedrick 1987), for all marker pairs with distance $< 5$ cM amounts to 0.336. This is close to that estimated in the study of Lipkin et al. (2009) (0.35), and is considerably less then estimated for dairy breed Holstein-Friesian (0.50 in Farnir et al. 2000 and Vallejo et al. 2003; 0.45 in

Tenesa et al. 2003;) or Japanese beef cattle breeds (0.55 and 0.45 in Odani et al. 2006).

However, it is difficult to compare the level and extent of LD obtained in different studies because the estimates of LD strongly depends on various factors such as: history and structure of the studied population (evolutionary forces that affected the population), sample size, marker type (microsatellites or SNPs), density and distribution of markers, type of method used for haplotype reconstruction, strictness of SNP filtering (threshold of minor allele frequencies and Hardy-Weinberg equilibrium), use of maternal haplotypes only or both maternal and paternal haplotypes (Bohmanova et al. 2010). Each of these factors can affect the estimates of LD. For example, with the $D'$ measure, a high level of LD extending over more than 10 cM has been reported (e.g., Farnir et al., 2000; Khatkar et al., 2006; Odani et al., 2006), but this measure can be inflated at large distances, for SNP with low allele frequencies, and by small sample sizes (Zhao et al. 2005; Du et al. 2007; Zhao et al. 2007). This tendency is undoubtedly exacerbated for microsatellite markers because of the general presence of one or more alleles at low frequency. The rare alleles and not observed haplotypes inflated $D'$ but not $\chi^{2\prime}$ (Heifetz et al. 2005; Thévenon et al. 2007). Therefore we use $D'$ value only to compare level of LD with other populations but $\chi^{2\prime}$ to estimate actual level of useable LD in mapping population. Here we estimated 10.7% LD for separation distance <5 cM in Fleckvieh. This is only the half of appropriate estimates for Italian and Israel Holstein cattle population (Lipkin et al. 2009). The $\chi^{2\prime}$ values dropped off rapidly with increasing separation distance, and were very low (<0.045) for separation distances > 10 cM.

Therefore, our results suggest the presence of only moderate LD at the < 5 cM range. The values obtained for $\chi^{2\prime}$ at this separation distance are still considerably greater than those reported at this separation distance in cattle for diallelic SNP markers, using the comparable $r^2$ measure (Mckay et al. 2007; Khatkar et al. 2008; Sargolzaei et al. 2008).

Relatively low level of LD in Fleckvieh suggest a need for more dense marker map for fine mapping of QTL in this population. The confidence interval as determined by 1-LOD support interval was large (13, 20 and 24 cM for FY, MY and PY respectively) for the positional candidate gene approach. Consequently,

the second objective was to refine the position of a previously detected QTL on BTA5. To achieve this, we increased the marker density at the most likely QTL region by adding twelve new microsatellite markers to the map used in previous study (Awad et al. 2010) as well 240 SNPs corresponded to a region larger than 1-LOD support interval, selection of new animals according to haplotype analyses in the complex pedigree. Single and multiple QTL analysis using combined linkage and linkage disequilibrium mapping method were used to analyse the data.

Due to substantial improvement of the material and methods, we were able to refine QTL position considerably. We analysed different subset of families according to haplotype analysis and availability of SNP and microsatellite data. The reduced accuracy of QTL position estimates when using all genotyped animals (*MSPED2089* or *SNPPED723*) compared to a subset of animals (*MSPED1038* or *SNPPED421*) is counterintuitive to the general notion that use of more information should result in better estimates. To explore this unexpected result further, we investigated several possible explanations; including the effects of the haplotype distribution and the possibility of additional QTLs. A total of 485 animals were genotyped with SNP-chip and are used for study of haplotype distribution in Fleckvieh population. The haplotype of 25 markers (*H0P*) comprising a derived QTL allele affecting MY1 and PY1 with 97% CI (117.962 to 119.018) is in all 89 cases traceable back to the founder F0 (Fig. 1). The alternative founder haplotype is in only 3 of 13 cases traceable back to the founder F0. Considering the possible QTL affecting only PY1 at position 122Mb there are six markers covering 97% CI (121.8 to 122.2Mb). The founder F0 is homozygous in this six markers and large surrounding region from 118.266 – 123.347Mb. This segment of 5.080Mb includes 109 informative markers in Fleckvieh. Prediction of identity by descent probabilities (IBD) from marker-haplotypes according to Meuwissen and Goddard (2001) suggest IBD>0.999 for both haplotypes in this homozygous segment in founder F0.

A highly significant QTL allele affecting PY1 was detected only if families segregating for founder haplotypes included in analysis. Excluding these families produce the LRT values below 3.99 (*P*>0.045) for all three milk yield traits and for the complete investigated region (Fig. 5, 113.5-123.7 Mb). This

and LDLA analyses by microsatellite designs (Fig. 2 & 3) clearly suggest, derived QTL allele is introduced by founder F0 in to used mapping population. Haplotype analyse show very seldom haplotype associated with derived QTL around physical position 118.00Mb. Taking together, there is no LD information for mapping of this derived QTL in *SNPPED723* and *SNPPED421* designs. Therefore, using virtually only linkage information in *SNPPED421* design we mapped a QTL affecting both MY1 and PY1 in a 97% CI of 1 Mb. The explanation of possible second QTL affecting only PY1 and mapping to the physical position 122.115 Mb is different. First, also this QTL is associated with founder haplotype. Second, both founder haplotypes at this position are most probably (P>0.999) identical by descent and therefore founder is most probably homozygous for QTL alleles at this position too. Third, this part of the haplotype is not unique for founder but segregate also in other families, i.e. there is LD information for mapping too. The relatively sharp LRT peak at position 122.115 Mb and Homozygosity of F0 suggests essential contribution of LD to this mapping. Fourth, two-QTL analyse ruled not out the possibility of two QTL affecting PY1 within candidate region on BTA5.

LDLA analysis using SNPs in a complex pedigree of 2089 animals indicated several peaks affecting MY1 and PY1 in our investigated region on BTA5 (Fig. 4). These results are comparable with fine-mapping results on BTA03 presented by Druet et al. (2008), where they mapped QTL for female fertility to a small set of narrow peaks on BTA3 by genotyping 17 Holstein half-sib families for a set of 437 SNPs, linkage analysis and LDLA were performed. The multiple peak profile would be explained by the heterogeneous LD structure within the QTL region or by use of LD in the model where there is no LD information at all. This might be increased by possible local inconsistencies in the map order, which was based on draft assembly or on comparative map information. Moreover, the method and the data structure might not allow the discarding of some regions even though they do not harbour the QTL (Druet et al. 2008). This study demonstrates way of target sampling of families segregating for derived QTL allele. Furthermore, the results of this study illustrates substantially improve in mapping resolution by preselecting of families and reduction of the noise in the test-statistic curve by dividing the mapping population according to haplotype analyses.

In the search for candidate gene in the region from117.9 to 119.1Mb, there are totally 27 genes. Thirteen of those encoding for proteins with unknown function. Genes with known function in this region (with descriptions taken out of Online Mendelian Inheritance in Man (OMIM) database) include for example:

- TAB1 (Tak-Binding Protein 1): found in both yeast and mammalian cells, TAB1 activated the kinase activity of TAK1 by direct interaction, may be an important signalling intermediate between TGFB receptors and TAK1.
- MGAT3: It is involved in the regulation of the biosynthesis and biological function of glycoprotein oligosaccharides. Catalyzes the addition of N-acetylglucosamine in beta 1-4 linkage to the beta-linked mannose of the trimannosyl core of N-linked sugar chains. It is one of the most important enzymes involved in the regulation of the biosynthesis of glycoprotein oligosaccharides.
- ATF4 (Activating Transcription Factor 4): transcription activator activity, protein dimerization activity, sequence-specific DNA binding.
- RPS19BP1 (Ribosomal Protein S19 Binding Protein 1): Direct regulator of SIRT1. Enhances SIRT1-mediated deacetylation of p53/TP53, thereby participating in inhibition of p53/TP53-mediated transcriptional activity.
- GRAP2 (Growth Factor Receptor-Bound Protein 2): is a divergent member of the GRB2/Sem5/Drk family and suggested that it is an adaptor-like protein involved in leukocyte-specific protein-tyrosine kinase signalling.
- CACNA1I (Voltage-dependent T-type calcium channel subunit alpha-1I): calcium ion transport, signal transduction, transmembrane transport.
- TNRC6B (Trinucleotide Repeat-Containing Gene 6B): Plays a role in RNA-mediated gene silencing by both micro-RNAs (miRNAs) and short interfering RNAs (siRNAs). Required for miRNA-dependent translational repression and siRNA-dependent endonucleolytic cleavage of complementary mRNAs by argonaute family proteins.

- SGSM3 (small G protein signaling modulator 3): acts in cooperation with merlin to enhance suppression of cell growth, regulation of Rab GTPase activity, regulation of Rab protein signal transduction.
- RUTBC3 (RUN and TBC1 domain-containing protein 3): similar to SGSM3.
- MKL1 (Megakaryoblastic leukemia 1 protein): Transcriptional coactivator of serum response factor (SRF) with the potential to modulate SRF target genes. Suppresses TNF-induced cell death by inhibiting activation of caspases; its transcriptional activity is indispensable for the antiapoptotic function. It may up-regulate antiapoptotic molecules, which in turn inhibit caspase activation.
- MAP3K7IP1 (Mitogen-activated protein kinase kinase kinase 7 interacting protein 1): similar to TAB1.

Also, the possible second QTL affecting only PY1 and mapping to 122.115 Mb (from 121.8 to 122.2 Mb) includes about five genes (LDOC1L, PRR5, LOC553158, ARHGAP8 and PHF2). None of them could be associated with milk protein in dairy cattle. Therefore we will prefer positional cloning and high-throughput sequencing in the candidate regions.

## CONCLUSION

The first aim of this study was to map quantitative trait loci (QTL) affecting traits that are important to the milk producers and dairy industry. These are the milk production traits that directly affect the income of the dairy farmer and are also important for the dairies, the functional health and fertility traits that affect the income through effects on the milk quality and costs of production in general. By applying different mapping methods we showed that the distal part of BTA5 harbours a QTL affecting first lactation milk yield, milk protein yield and milk fat yield in the dual purpose Fleckvieh cattle breed. The confidence interval of this QTL was large for positional candidate gene approach. So, the second aim was to refine the position of previously detected QTL. By increasing material and improving methods, we refined the position of QTL and reduce the confidence interval to 1Mb. Several genes with unknown function may be identified in our candidate region but this region is small enough for high-throughput sequencing and future detailed analyses of candidate genes.

# CHAPTER 6

# GENERAL SUMMARY

### Mapping of Quantitative Trait Loci for Milk Yield Traits on Bovine Chromosome 5 in the Fleckvieh Cattle

The aim of this thesis was to map quantitative trait loci (QTLs) affecting milk yield traits on bovine chromosome 5 (BTA5) in the Fleckvieh dual purpose cattle breed. These QTLs are of high interest in dairy and dual purpose cattle breeds because of their high economic weight in selection index. Segregation between a genetic marker and a locus affecting quantitative trait is the basis for success in mapping of QTLs.

Twenty eight microsatellite markers with a coverage of 135 cM on BTA5 were selected and nine half-sib families containing a total of 607 sons in a granddaughter design (GDD) were analysed for mapping QTLs affecting Milk Yield (MY1), Milk Protein Yield (PY1) and Milk Fat Yield (FY1) during first lactation in the dual purpose cattle breed Fleckvieh. Phenotypic records were daughter yield deviations for these traits and corresponding reliabilities originated from the routinely performed genetic evaluations of breeding animals. The build option of the *CRI-MAP* program was applied to our genotype and pedigree date to find the marker order. As quality control of data, the constructed linkage map was compared to the published linkage map at USDA and to physical map of the bovine genome (Genome sequence Btau 4.0) and to separate markers that were at the same linkage position but with different physical position. Three different QTL mapping approaches were implemented: interval mapping (IM), variance component linkage analysis (LA-VC) and combined linkage disequilibrium and linkage analysis (LDL).

The QTL analysis across nine half-sib families using one-QTL model interval mapping found that there might be two QTL affecting the three yield traits (MY1, PY1 and FY1), one QTL near the middle part of BTA5 and the other QTL at the distal part of chromosome. We confirmed a QTL at the distal part of BTA5 with

both variance component approaches (LA-VC and LDL), which mapped a QTL affecting MY1, PY1 and FY1 in the same direction and located it to the same marker interval (BM2830-ETH152) which has the highest test-statistic value at 119.33 cM. The estimated average LD for marker pairs with distance < 5 cM was low (0.107) and reflected the large effective population size of the Fleckvieh breed. The relatively low level of LD in Fleckvieh suggests a need for a denser marker map for fine mapping of QTL in this population. Consequently, the achieved confidence interval was too large for the positional candidate gene approach.

Therefore, we increased the marker density at the most likely QTL position by adding twelve new microsatellite markers to the map used in previous study as well as 240 SNPs corresponding to candidate region larger than the 97% confidence interval. Furthermore we added some new animals to get more recombination events around the candidate region. Families in which the derived QTL allele segregated were selected according to haplotype analyses of 41 GDD family sires and their available ancestors. These animals comprise one complex pedigree in which the segregating haplotypes can trace along five generations back to some important founders of the Fleckvieh population. Single and multiple QTL analysis using combined linkage and linkage disequilibrium method were performed. Analysis of different subset of families according to haplotype analysis and availability of SNP and microsatellite data, succeeded in refining the map position of the previously detected QTL for milk yield traits on BTA5. Considering the 97% confidence interval, we mapped a QTL affecting MY and PY to a region of only 1.056 mega base (Mb), i.e. from physical position 117.962 Mb to 119.018 Mb on BTA5. Additionally, there were some indications of a second QTL affecting only PY at physical position 122.0 Mb. According to haplotype analyses of most important family sires and founders, this second QTL at position 122.115 Mb is less plausible. We identified several genes in our candidate region from 117.962 to 119.018 Mb but these are either with unknown function or with known function which can be only indirectly related to milk yield traits. Therefore, we suggest and initiated positional cloning and high-throughput sequencing in candidate region from 117.9 to 119.1Mb.

# CHAPTER 7

# ZUSAMMENFASSUNG

## QTL-Kartierung für Milchleistungsmerkmale im Deutschen Fleckvieh auf dem Rinderchromosom 5

Ziel dieser Studie war es, „Quantitative Trait Loci“ (QTLs) mit Effekt auf drei Milchleistungsmerkmale auf dem Rinderchromosom 5 (BTA5) der Zweinutzungsrasse Deutsches Fleckvieh (DFV) zu kartieren. Diese QTLs sind aufgrund ihres hohen ökonomischen Gewichts in dem Selektionsindex bei Milch- und Doppelnutzungsrassen von sehr hohem Interesse. Dabei ist die Erkennung einer Ko-Segregation von Marker- und QTL-Allele, die das Merkmal von Interesse beeinflussen, die Basis für eine erfolgreiche QTL-Kartierung.

Es wurden 28 Mikrosattelitenmarker, welche 135 cM des BTA5 abdecken, ausgewählt und in 607 Söhnen aus neun Halbgeschwisterfamilien (Granddaughter Design (GDD) Familien) genotypisiert. Dieses GDD wurde zur Kartierung von QTL mit Effekt auf die Milchleistungsmerkmale: Milchmenge (MY), Milchproteinmenge (PY) und Milchfettmenge (FY) im Doppelnutzungsrind DFV verwendet. Dabei wurden Leistungsabweichungen der Töchter, stammend aus quartalsmäßiger Zuchtwertschätzung, als Phänotypen verwendet. Die *build* Option des Programmes *CRI-MAP* wurde zur Anordnung von Markern basierend auf deren Genotypen und Abstammungsdaten verwendet. Die auf diese Weise entstandene Markerkarte wurde zur Qualitätskontrolle mit veröffentlichten USDA Kopplungskarten sowie mit der physikalischen Karte des Rindergenoms (genomische Sequenz *Btau* 4.0) verglichen und bei Bedarf (gleiche Kopplungs- aber verschiedene physikalische Position zweier enggekoppelter Marker) angepasst. Es wurden drei verschiedene QTL-Kartierungsansätze angewendet: Intervallkartierung (IM), Kartierung mit Hilfe einer Varianzkomponenten-Analyse (LA-VC) und eine kombinierte Kopplungsungleichgewicht- (LD) und Kopplungsanalyse (LDL).

Die Intervallkartierung in neun GDD Familien zeigte, dass zwei QTLs einen Einfluss auf die drei Milchleistungsmerkmale (MY, PY und FY) haben könnten. Einer befindet sich im mittleren Teil, der andere im distalen Bereich des BTA5. Durch die beiden Varianzkomponentenansätze (LA-VC and LDL) konnte der QTL auf dem distalen Teil von BTA5 bestätigt werden. Dieser QTL zeigte für alle drei Leistungsmerkmale gleichgerichtete Effekte und kartierte im selben Markerintervall (BM2830-ETH152) mit höchstem statistischem Wert bei Position 119.33 cM. Alle Markerpaare mit einer Distanz kleiner 5 cM zeigten einen geringen durchschnittlichen LD-Wert (0,107) und reflektieren damit eine große effektive Populationsgröße in der Rasse Fleckvieh. Der relativ niedrige LD-Wert macht deutlich, dass für eine QTL-Feinkartierung in der Rasse DFV eine dichtere Markerkarte nötig ist. Deswegen ist die erzielte Auflösung, in dieser ersten Kartierungsphase, nicht hoch genug für ein erfolgversprechendes positionelles Klonen und weiterführende Analysen der möglichen Kandidatengene.

Deshalb wurde die Markerdichte in der Kandidatenregion erhöht. Es wurden zwölf neue Mikrosattelitenmarker und 240 SNPs zu der Karte, die in der vorangegangenen Kartierung verwendet wurde, hinzugefügt. Diese 252 neuen Marker decken eine Region ab, die mehr als 97% des Vertrauensintervalls umfasst. Außerdem wurden einige neue Tiere hinzugefügt, um mehr Rekombinationsereignisse zu beobachten. Dieses zusätzliche Tiermaterial wurde anhand einer Haplotypenanalyse von 41 GDD Familienväter und deren 34 verfügbaren Vorfahren ausgewählt. Anhand der SNP-Haplotypen in diesen 75 Tieren wurde es möglich QTL-Assozierte Haplotypen zu bestimmen und gezielt GDD-Familien für weiterführenden Analysen auszuwählen und für die oben genannten 252 Marker zu genotypisieren. Es wurden 1-QTL und 2-QTL Analysen unter Verwendung der LDL Methode durchgeführt. Die Analysen der verschiedenen Subgruppen in den Familien gemäß der Haplotypenanalyse und der Verfügbarkeit von SNP- und Mikrosattelitendaten konnten schließlich die Position des zuvor entdeckten QTLs im distalen Bereich des BTA5 genauer eingrenzen. Unter Berücksichtigung des 97%-igen Vertrauensintervalls konnten wir einen QTL mit Effekt auf MY und PY in einer relativ kleinen Region von 1.056 Megabase (Mb) kartieren. Diese Region befindet sich auf BTA5 mit der

physikalischen Position von 117.962 MB bis 119.018 MB. Zusätzlich wurde ein Hinweis auf einen zweiten QTL, der lediglich PY beeinflusst, bei der

physikalischen Position 122.0 MM gefunden. Dieser zweite QTL konnte nicht durch eine Analyse mit der QTL segregierenden Haplotypen in GDD-Familien und wichtige Gründertieren bestätigt werden. In unserer Kandidatenregion befinden sich mehrere Gene, bei einigen davon ist jedoch die Funktion bisher unbekannt, bei anderen mit bekannter Funktion konnte keine direkte funktionelle Verbindung mit Milchleistungsmerkmalen hergestellt werden. Deswegen wurde ein positionelles Klonen und eine Hochdurchsatzsequenzierung in der Kandidatenregion von 117.9 bis 119.1 Mb auf BTA5 empfohlen und eingeleitet.

# CHAPTER 8

# REFERENCES


Almasy L. & Blangero J. (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *American Journal of Human Genetics* **62**:1198-1211.

Andersson L. (2001) Genetic dissection of phenotypic diversity in farm animals. *Nature Reviews Genetics* **2**: 130-138.

Ardlie K.G., Kruglyak L. & Seielstad M. (2002) Patterns of linkage disequilibrium in the human genome. *Nat. Rev. Genet*. **3**: 299–309.

Ashwell M.S., Heyen D.W., Sonstegard T.S., & Van Tassell C.P. (2004) Detection of quantitative trait Loci affecting milk production, health, and reproductive traits in Holstein cattle. *Journal of Dairy Science* **87**: 468-475.

Awad A. Russ I., Emmerling R., Förster M., Medugorac I. (2010) Confirmation and refinement of a QTL on BTA5 affecting milk production traits in the Fleckvieh dual purpose cattle breed. *Animal Genetics* **41**:1-11.

Barendse W., Vaiman D., Kemp S.J., Sugimoto Y., Armitage S.M., Williams J.L., Sun H.S., Eggen A., & et al. (1997) A medium-density genetic linkage map of the bovine genome. *Mammalian Genome* **8**: 21–28.

Bennewitz J., Reinsch N., Grohs C., Leveziel H., Malafosse A., Thomsen H., Xu N., Looft C., Kuhn C., Brockmann G.A., Schwerin M., Weimann C., Hiendleder S., Erhardt G., Medjugorac I., Russ I., Forster M., Brenig B., Reinhardt F., Reents R., Averdunk G., Blumel J., Boichard D. & Kalm E. (2003) Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle. *Genetics Selection Evolution* **35**: 319-338.

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

Bohmanova J., Sargolzaei M. & Schenke F.S. (2010) Characteristics of linkage disequilibrium in North American Holsteins. *BMC Genomics*: **11**(1): 421.

Boichard D., Grohs C., Bourgeois F., Cerqueira F., Faugeras R., Neau A., Rupp R., Amigues Y., Boscher M.Y., & Leveziel H. (2003) Detection of genes influencing economic traits in three French dairy cattle breeds. *Genetics Selection Evolution* **35**:77–101.

Bovenhuis H. & Schrooten C. (2002) quantitative trait loci for milk production traits in dairy cattle. Proc. 7th World Congr. Genet. Appl. Livestock Prod. No. 09–07, Montpellier, France.

Bovine HapMap Consortium (2009) Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science* **324**: 528-531.

Broman K.W. (2001) Review of statistical methods for QTL mapping in experimental crosses. *Lab Animal* **30** (7): 44–52

Burton P.R., Clayton D.G., Cardon L.R., Craddock N., Deloukas P., Duncanson A., Kwiatkowski D.P., et al. (2007) Association scan of 14,500 non synonymous SNPs in four diseases identifies autoimmunity variants**.** *Nat Genet*. **39**:1329-1337.

Cardon L. R. & Bell J. I. (2001) Association study designs for complex diseases. *Nature Reviews Genetics* **2**: 91–99.

Chen H.Y., Zhang Q., Yin C.C., Wang C.K., Gong W.J. & Mei G. (2006) Detection of Quantitative Trait Loci Affecting Milk Production Traits on Bovine Chromosome 6 in a Chinese Holstein Population by the Daughter Design. *Journal of dairy science* **89**:782-790.

Churchill G.A. & Doerge R.W. (1994) Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963-971.

Cohen-Zinder M., Seroussi E., Band M.R., Lewin H. A., Drackley J. K., Larkin D. M., Everts-van der Wind A., Heon-Lee J., Loor J.J. & Shani M. (2004) *SPP1* is a candidate gene for the QTL affecting milk protein concentration on BTA6 in Israeli Holstein. 29th Int Conf Ani Gen, ISAG, F015, Tokyo, Japan.

Cohen-Zinder M., Seroussi E., Larkin D.M., Loor J.J., Everts-van der Wind A., Heon-Lee J., Drackley J.K., Band M.R., Hernandez A.G., Shani M., Lewin H.A., Weller J.I., & Ron M. (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Research* **15** (7): 936-944.

Cole J.B., VanRaden P.M., O'Connell J.R., Van Tassell C.P., Sonstegard T.S., Schnabel R.D., Taylor J.F. & Wiggans G.R. (2009) Distribution and location of genetic effects for dairy traits. *J. Dairy Sci*. **92**: 2931–2946.

Coppieters W., Riquet J., Arranz J.J., Berzi P., Cambisano N., Grisart B., Karim L., Marcq F., Moreau L., Nezer C., Simon P., Vanmanshoven P., Wagenaar D.,& Georges M. (1998) A QTL with major effect on milk yield and composition maps to bovine chromosome 14. *Mammalian Genome* **9** (7): 540-544.

Daetwyler H.D., Schenkel F.S., Sargolzaei M. & Robinson J.A. (2008) A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. *Journal of Dairy Science* **91** (8): 3225-3236.

Darvasi A. (1998) Experimental strategies for the genetic dissection of complex traits in animal models. *Nature Genetics* **18**: 19-24.

Darvasi A. & Soller M. (1995) Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* **141**: 1199-1207.

Darvasi A., Weinreb A., Minke V., Weller J.I. & Soller M. (1993) detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* **134**: 943-951

Daw E.W., Heath S.C. & Lu Y. (2005) Single nucleotide polymorphism versus microsatellite markers in a combined linkage and segregating analysis of a quantitative trait. *BMC Genetic*: 6 (Suppl. 1), S32.

Dekkers J. C. M. (2004) Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. *Journal of Animal Science* **82** E-Suppl: E313-328.

Dekkers J.C. & Hospital F. (2002) the use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics* **3**: 22-32.

de Koning D.J., Pong-Wong R., Varona L., Evans G.J., Giuffra E., Sanchez A., Plastow G., Noguera J.L., Andersson L., & Haley C.S. (2003) Full pedigree quantitative trait locus analysis in commercial pigs using variance components. *Journal of Animal Science* **81**: 2155–2163.

de Koning D.J., Schulman N.F., Elo K., Moisio S., Kinos R., Vilkki J., & Maki-Tanila A. (2001) Mapping of multiple quantitative trait loci by simple regression in half-sib designs. *Journal of Animal Science* **79**: 616–622.

Doerge R.W. & Churchill G.A. (1996) Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**: 285-294.

Druet T., Fritz S., Boussaha M., Ben-Jemaa S., Guillaume F., Derbala D., Zelenika D., Lechner D., Charon C., Boichard D., Gut I.G., Eggen A. & Gautier M. (2008) Fine Mapping of Quantitative Trait Loci Affecting Female Fertility in Dairy Cattle on BTA03 Using a Dense Single-Nucleotide Polymorphism Map. *Genetics* **178**: 2227–2235.

Du F., Clutter A.C. & Lohuis M.M. (2007) Characterizing Linkage Disequilibrium in Pig Populations. *International journal of biological sciences* **3** (3):166-178.

Edwards A., Civitello A., Hammond H.A. & Caskey C.T. (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American journal of human genetics* **49**: 746–756.

Ellegren H. (2004) Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics* **5**: 435–445.

El Mousadik A. & Petit R.J. (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [Argania spinosa (L.) Skeels] endemic to Morocco. *Theoretical and applied genetics* **92**: 832-839

Falconer D.S. & Mackay T.F.C. (1996) Introduction to quantitative genetics, Fourth edition, Pearson education limited.

Farnir F., Coppieters W., Arranz J.J., Berzi P., Cambisano N., Grisart B., Karim L., Marcq F., Moreau L., Mni M., Nezer C., Simon P., Vanmanshoven P., Wagenaar D., & Georges M. (2000) Extensive genome-wide linkage disequilibrium in cattle. *Genome Research* **10**: 220–227.

Farnir F., Grisart B., Coppieters W., Riquet J., Berzi P., Cambisano N., Karim L., Mni M., Moisio S., Simon P., Wagenaar D., Vilkki J., & Georges M. (2002) Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: Revisiting the

location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. *Genetics* **161**: 275–287.

Flint J., & Mott R. (2001) Finding the molecular basis of quantitative traits: Successes and pitfalls. *Nature Reviews Genetics* **2**: 437-445.

Flint-Garcia S.A., Thornsberry J.M. & Buckler E.S. (2003) Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* **54**: 357–374.

Freking B.A., Murphy S.K., Wylie A.A., Rhodes S.J., Keele J.W., Leymaster K.A., Jirtle R.L. & Smith T.P. (2002) Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar over dominance in mammals. *Genome Research* **12**: 1496-1506.

Freyer G., Sorensen P., Kuhn C., Weikard R. & Hoeschele I. (2003) Search for pleiotropic QTL on chromosome BTA6 affecting yield traits of milk production. *Journal of Dairy Science* **86** (3): 999-1008.

Garner C., McInnes L.A., Service S.K., Spesny M., Fournier E., Leon P. & Freimer N.B. (2001) Linkage analysis of a complex pedigree with severe bipolar disorder, using a Markov Chain Monte Carlo method. *American Journal of Human Genetics* **68** (4):1061-1064.

Gautier M., Barcelona R.R., Fritz S., Grohs C., Druet T., Boichard D., Eggen A., & Meuwissen T.H. (2006) Fine mapping and physical characterization of two linked quantitative trait loci affecting milk fat yield in dairy cattle on BTA26. *Genetics* **172**: 425-436.

Geldermann H. (1975) Investigations of quantitative characters in animals by gene markers. *Theoretical and applied genetics* **46**: 319-330.

George A.W., Visscher P.M. & Haley C.S. (2000) Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. *Genetics* **156**: 2081-2092.

Georges M. (2007) Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. *Annual review of genomics and human genetics* **8**:131-162.

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., & Hoeschele I. (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* **139**: 907-920.

Gibson J.P. (2003) Incorporating molecular genetic information in a global approach to genetic improvement for low and medium input systems, p. 1 in The Book of Abstracts of the 54th Annual Meeting of the European Association for Animal Production, edited by Y. VAN DER HONING. Wageningen Academic Publishers, Wageningen, The Netherlands.

Glazier A.M., Nadeau J.H. & Aitman T.J. (2002) Finding genes that underlie complex traits. *Science* **298**: 2345-2349.

Goldstein D. & Schlötterer C. (1999) Microsatellites: Evolution and Applications (Oxford Univ. Press, Oxford).

Grapes L., Dekkers J.C.M., Rothschild M.F., & Fernando R.L. (2004) Comparing linkage disequilibrium-based methods for fine mapping quantitative trait loci. *Genetics* **166**:1561–1570.

Griffin H.G. & Griffin A.M. (eds) (1994) PCR technologies: current innovations. CRC Press, Inc., Boca Raton, FL.

Grisart B., Coppieters W., Farnir F., Karim L., Ford C., Berzi P., Cambisano N., Mni M., Reid S., Simon P., Spelman R., Georges M., & Snell R. (2002) Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research* **12** (2): 222-231.

Grisart B., Farnir F., Karim L., Cambisano N., Kim J., Kvasz A., Mni M., Simon P., Frère J., Coppieters W., & Georges M. (2004) Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. *PNAS* **101** (8): 2398-2403.

Habier D., Fernando R.L. & Dekkers J.C. (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* **177**: 2389-2397.

Haley C. (1999) Advances in quantitative trait locus mapping. In: Dekkers JCM, Lamont SJ, Rothschild MF, eds. From JL Lush to genomics: visions for animal breeding and genetics. Ames, IA, USA: Iowa State University Press, 47-59.

Haley C. & Archibald A. (1998) Progress in Pig Gene Mapping. In: Progress in Pig Science. J. Wiseman, M.A. Varley, and J.P. Chadwick (eds). Nottingham University Press. pp. 5-27.

Haley C.S. & Andersson L. (1997) Linkage mapping of quantitative trait loci in plants and animals, pp. 49-71 in Genome Mapping : A Practical Approach, edited by P. H. Dear. Oxford University Press, New York.

Haley C.S., Knott S.A. & Elsen J.M. (1994) Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136**:1195-1207.

Hancock J.M. (1999) Microsatellites and other simple sequences: genomic context and mutational mechanisms. In: Goldstein D., Schlötterer C., editors. Microsatellites - Evolution and Applications. New York: Oxford University Press. pp 1–9.

Harmegnies N., Farnir F., Davin F., Buys N., Georges M. & Coppieters W. (2006) Measuring the extent of linkage disequilibrium in commercial pig populations. *Anim. Genet.* **37**: 225–231.

Hayes B.J., Kinghorn B.P. & Ruvinsky A. (2005) Genome scanning for quantitative trait loci, CAB International, UK. 507 – 537.

Hayes B.J., Bowman P.J., Chamberlain A.J. & Goddard M.E. (2009) Invited review: Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci.* **92** (2): 433-43.

Hästbacka J., de la Chapelle A., Kaitila I., Sistonen P., Weaver A. & Lander E. (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genetics* **2**: 204 – 211

Heath S. (1997) Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *American Journal of Human Genetics* **61**: 748–760.

Hedrick P.W. (1987) Gametic disequilibrium measures: Proceed with caution. *Genetics* **117**: 331-341

Hedrick P.W. (2000) Genetics of Population, 2nd edn. Jones and Bartlett. pp. 553.

Heifetz E.M., Fulton J.E., O'Sullivan N., Zhao H., Dekkers J.C.M. & Soller M. (2005) Extent and Consistency Across Generations of Linkage Disequilibrium in Commercial Layer Chicken Breeding Populations. *Genetics* **171** (3): 1173–1181

Higgs D.R., Wood W.G., Jarman A.P., Sharpe J., Lida J. Pretorius I.M. & Ayyub H. (1990) A major positive regulatory region located far upstream of the human alpha-globin gene locus. *Genes and Development* **4**:1588-1601.

Hill W.G. & Robertson A. (1968) Linkage disequilibrium in finite populations. *Theoretical and applied genetics* **38**: 226-231.

Hinds D.A., Stuve L.L., Nilsen G.B., Halperin E., Eskin E., Ballinger D.G., Frazer K.A. & Cox D.R. (2005) Whole genome patterns of common DNA variation in three human populations. *Science* **307**: 1072–1079.

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

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., & Sugimoto Y. (2004) A comprehensive genetic map of the cattle genome based on 3802 microsatellites. *Genome Research* **14**: 1987–1998.

International SNP Map Working Group (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**: 928–933.

Jarne P., & Lagoda P.J.L. (1996) Microsatellites, from molecules to populations and back. *Tree* **11**: 424–429.

Kappes S.M., Keele W.J., Stone R.T., McGraw R.A., Sonstegard T.S., et al. (1997) A second-generation linkage map of the bovine genome. *Genome Research* **7**: 235–249.

Kerr R.J. & Kinghorn B.P. (1996) An efficient algorithm for segregation analysis in large populations. *Journal of Animal Breeding and Genetics* **113**: 457–469.

Khatkar M.S., Thomson P.C., Tammen I., & Raadsma H.W. (2004) Quantitative trait loci mapping in dairy cattle: Review and meta-analysis. *Genetics Selection Evolution* **36**:163–190.

Khatkar M.S., Thomson P.C., Tammen I., Cavanagh J.A., Nicholas F.W., & Raadsma H.W. (2006) Linkage disequilibrium on chromosome 6 in australian holstein-friesian cattle. *Genetics Selection Evolution* **38**: 463-477.

Khatkar M.S., Nicholas F.W., Collins A.R., Zenger K.R., Cavanagh J.A., Barris W., Schnabel R.D., Taylor J.F. & Raadsma H.W. (2008) Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics* **9**:187.

Knott S.A., Elsen J.M. & Haley C.S. (1996) Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theoretical and applied genetics* **93**: 71-80.

Kolbehdari D., Jansen G.B., Schaeffer L.R., & Allen B.O. (2005) Power of QTL detection by either fixed or random models in half-sib designs", *Genetics Selection Evolution* **37** (6): 601-614.

Kolbehdari D., Wang Z., Grant J.R., Murdoch B., Prasad A., Xiu Z., Marques E., Stothard P. & Moore S.S. (2009) A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. *Journal of Animal Breeding and Genetics* **126** (3): 216-227.

Kühn C., Thaller G., Winter A., Bininda-Edmonds O.R.P., Kaupe B., Erhardt G., Bennewitz J., Schwerin M. & Fries B. (2004) Evidence for multiple alleles at the DGAT1 locus better explains a quantitative trait locus with major effect on milk fat content in cattle. *Genetics* **167**: 1873-1881.

Kwon J.M. & Goate A.M. (2000) The candidate gene approach. *Alcohol Research and Health* **24**: 164-168.

Lagziel A., Lipkin E., & Soller M. (1996) Association between SCCP haplotypes at the bovine growth hormone gene and milk protein percentage. *Genetics* **142**: 945–951.

Lander E.S. & Botstein D. (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**(1):185-199.

Lander E.S. & Schork N.J. (1994) Genetic dissection of complex traits. *Science* **265**: 2037-2048.

Lee S. H. & van der Werf J.H. (2004) The efficiency of designs for fine-mapping of quantitative trait loci using combined linkage disequilibrium and linkage. *Genetics Selection Evolution* **36**: 145-161.

Lee S.H. & Van der Werf J.H. (2005) The role of pedigree information in combined linkage disequilibrium and linkage mapping of quantitative trait loci in a general complex pedigree. *Genetics* **169**: 455-466.

Lee S. H. & Van der Werf J.H. (2006) Simultaneous fine mapping of multiple closely linked quantitative trait Loci using combined linkage disequilibrium and linkage with a general pedigree. *Genetics* **173**: 2329-2337.

Lewontin R. C. (1964) The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**:49-67.

Li W. H. & Grauer D. (1991) Fundamentals of molecular evolution. Sinauer Associates, Sunderland, MA.

Lien S., Gomez-Raya L., Steine T., Fimland E., & Roger S. (1995) Associations between casein haplotypes and milk yield traits. *Journal of Dairy Science* **78**:2047–2056.

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

Lipkin E., Straus K., Tal Stein R., Bagnato A., Schiavini F., Fontanesi L., Russo V., Medugorac I., Foerster M., Sölkner J., Dolezal M., Medrano

J.F., Friedmann A. & Soller M. (2009) Extensive long-range and non-syntenic linkage disequilibrium in livestock populations: deconstruction of a conundrum. *Genetics* **181**: 691–699

Litt M. & Luty J.A. (1989) a hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* **44** (3): 397-401.

Long A. D., Lyman R.F., Langley C.H. & Mackay T.F.C. (1998) Two sites in the Delta gene region contribute to naturally occurring variation in bristle number in Drosophila melanogaster. *Genetics* **149**: 999–1017.

Lyman R.F., Lai C. & Mackay T.F.C. (1999) Linkage disequilibrium mapping of molecular polymorphisms at the scabrous locus associated with naturally occurring variation in bristle number in Drosophila melanogaster. *Genetical Research* **74**: 303–311

Lynch M. & Walsh B. (1998) Genetics and Analysis of Quantitative Traits. Sinauer Associates, Sunderland, MA.

Mackay T.F. (2001a) The genetic architecture of quantitative traits. *Annual Review of Genetics* **35**: 303-339.

Mackay T.F., (2001b) Quantitative trait loci in drosophila. *Nature Review Genetics* **2**: 11-20.

McKay S.D., Schnabel R.D., Murdoch B.M., Matukumalli L.K., Aerts J., Coppieters W., Crews D., Neto E.D., Gill C.A., Gao C., Mannen H., Stothard P., Wang Z., Van Tassell C.P., Williams J.L., Taylor J.F. & Moore S.S. (2007) Whole genome linkage disequilibrium maps in cattle. *BMC Genetics* **8**: 74.

McRae A.F., McEwan J.C., Dodds K.G., Wilson T., Crawford A.M. & Slate J. (2002) Linkage disequilibrium in domestic sheep. *Genetics* **160**: 1113–1122.

Medugorac I., Medugorac A., Russ I., Veit-Kensch C.E., Taberlet P., Luntz B., Mix H.M. & Förster M. (2009) Genetic diversity of European cattle breeds highlights the conservation value of traditional unselected breeds with high effective population size. *Mol Ecol*. **18** (16): 3394-3410.

Mei G., Yin C., Ding X. & Zhang Q. (2009) Fine mapping quantitative trait loci affecting milk production traits on bovine chromosome 6 in a Chinese Holstein population. *Journal of Genetics and Genomics* **36** (11): 653-660.

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

Meuwissen T.H. & Goddard M.E. (2001) Prediction of identity by descent probabilities from marker-haplotypes. *Genetics Selection Evolution* **33** (6): 605-634.

Meuwissen T.H. & Goddard M.E. (2004) Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genetics Selection Evolution* **36** (3): 261-279.

Meuwissen T.H., Hayes B.J. & Goddard M.E. (2001) prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819-1829.

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

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

Mullis K.B. & Faloona F.A. (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology* **155**: 335-350.

Mullis K., Ferre F. & Gibbs R.A. (eds) (1994) The Polymerase chain reaction. Springer Verlag.

Nagamine Y., Visscher P.M. & Haley C.S. (2004) QTL detection and allelic effects for growth and fat traits in outbred pig populations. *Genetics Selection Evolution* **36**: 83–96.

Nsengimana J., Baret P., Haley C.S. & Visscher P.M. (2004) Linkage disequilibrium in the domesticated pig. *Genetics* **166**: 1395–1404.

Odani M., Narita A., Watanabe T., Yokouchi K., Sugimoto Y., Fujita T., Oguni T., Matsumoto M. & Sasaki Y. (2006) Genome wide linkage disequilibrium in two Japanese beef cattle breeds. *Animal Genetics* **37**:139–144.

Olsen H.G., Gomez-Raya L., Vage D.I., Olsaker I., Klungland H., Svendsen M., Adny T., Sabry A., Klemetsda G., Schulman N., Kramer W., Thaller G., Rnningen K. & Lien S. (2002) A Genome Scan for Quantitative Trait Loci Affecting Milk Production in Norwegian Dairy Cattle. *Journal of Dairy Science* **85**: 3124-3130.

Olsen H.G., Lien S., Gautier M., Nilsen H., Roseth A., Berg P.R., Sundsaasen K.K., Svendsen M., & Meuwissen T.H. (2005) Mapping of a milk production quantitative trait locus to a 420-kb region on bovine chromosome 6. *Genetics* **169**: 275-283.

Olsen H.G., Nilsen H., Hayes B., Berg P.R., Svendsen M. Lien S. & Meuwissen T. (2007) Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk composition of dairy cattle. *BMC Genetics* **8**: 32.

Peltonen L. (2000) Positional cloning of disease genes: advantages of genetic isolates. *Hum. Hered.* **50**: 66–75.

Pirchner F. (2002) Schätzung inzuchtwirksamer (effektiver ) Populationsgrößen aus Genfrequenzschwankungen bei Bayerischem Fleckvieh und Tiroler Grauvieh. *Archiv für Tierzucht* **4**: 331-339

Primmer C. R., Ellegren H., Saino N., & Moller A. P. (1996) Directional evolution in germline microsatellite mutations. *Nature Genetics* **13**: 391–393.

Richard G.F. & Paques F. (2000) Mini- and microsatellite expansions: the recombination connection. *EMBO Reports* **1**:122–126.

Riquet J., Coppieters W., Cambisano N., Arranz J.J., Berzi P., Davis S.K., Grisart B., Farnir F., Karim L., Mni M., Simon P., Taylor J.F., Vanmanshoven P., Wagenaar D., Womack J.E. & Georges M. (1999) Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle, Proc Natl Acad Sci, *USA*, **96** (16): 9252-9257.

Rodriguez-Zas S.L., Southey B.R., Heyen D.W. & Lewin H.A. (2002) Interval and composite interval mapping of somatic cell score, yield, and components of milk in dairy cattle. *Journal of Dairy Science* **85** (11): 3081-3091

Ron M. & Weller J.I. (2007) From QTL to QTN identification in livestock – winning by points rather than knock-out: a review. *Animal Genetics* **38**: 429–439

Sahana G., Lund M.S., Andersson-Eklund L., Hastings N., Fernandez A., Iso-Touru T., Thomsen B., Viitala S., Sørensen P., Williams J.L., & Vilkki J. (2008) Fine-mapping QTL for mastitis resistance on BTA9 in three nordic red cattle breeds. *Animal Genetics* **39** (4): 354-362.

Saiki R.K., Gelfand D.H., Stoffel S., Scharf S.J., Higuchi R., Horn G.T., Mullis K.B., & Erlich H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239** (4839): 487-491.

Sargolzaei M., Schenkel F.S., Jansen G.B. & Schaeffer L.R. (2008) Extent of linkage disequilibrium in Holstein cattle in North America. *Journal of Dairy Science* **91** (5): 2106-2117.

Schaeffer L.R. (2006) Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* **123**: 218–223.

Schnabel R.D., Kim J.J., Ashwell M.S., Sonstegard T.S., Van Tassell C.P., Connor E.E., & Taylor J.F. (2005) Fine-mapping milk production quantitative trait loci on BTA6: analysis of the bovine osteopontin gene, Proc Natl Acad Sci, *USA*, **102** (19): 6896-6901.

Schrooten C., Bink M.C.A.M., & Bovenhuis H. (2004) Whole genome scan to detect chromosomal regions affecting multiple traits in dairy cattle. *Journal of Dairy Science* **87**: 3550-3560.

Schug M.D., Mackay T.F., & Aquadro C.F. (1997) Low mutation rates of microsatellite loci in Drosophila melanogaster. *Nature Genetics* **15**: 99–102.

Smit M., Segers K., Carrascosa L.G., Shay T., Baraldi F., Gyapay G., Snowder G., Georges M., Cockett N., & Charlier C. (2003) Mosaicism of solid gold supports the causality of a noncoding A-to-G transition in the determinism of the callipyge phenotype. *Genetics* **163**: 453-456.

Snelling W.M., Casas E., Stone R.T., Keele J.W., Harhay G.P., Bennet G.L. & Smith T.P. (2005) Linkage mapping bovine EST-based SNP. *BMC Genome* **6**:74.

Sölkner J., Filipcic L. & Hampshire N. (1998) Genetic variability of populations and similarity of subpopulations in Austrian cattle breeds determined by analysis of pedigrees. *Animal Science* **67**: 249-256

Soller M. (1998) Marker-assisted selection in animal biotechnology. pp 387-406 In: Agriculture Biotechnology: Books in soils, plants and the environment. A. Altman (ed), New York Marcel Decker, Inc., NY.

Streelman J.T. & Kocher T.D. (2000) From phenotype to genotype. *Evolution and Development* 2/3: 166-173.

Su G., Guldbrandtsen B., Gregersen V.R. & Lund M.S. (2010) Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. *Journal of Dairy Science* **93** (3): 1175–1183.

Taberlet P., Valentini A., Rezaei H .R., Naderi S., Pompanon F., Negrini R. & Ajmone-Marsan P. (2008) Are cattle, sheep, and goats endangered species? *Molecular Ecology* **17**: 275-284.

Tautz D. (1989) Hypervariability of simple sequences as a general source of polymorphic DNA markers. *Nucleic Acids Research* **17**: 6463-6471.

Tenesa A., Knott S.A., Ward D., Smith D., Williams J.L. & Visscher P.M. (2003) Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. *Journal of Animal Science* **81**: 617-623.

Thévenon S., Dayo G.K., Sylla S., Sidibe I., Berthier D., Legros H., Boichard D., Eggen A. & Gautier M. (2007) The extent of linkage disequilibrium in a large cattle population of western Africa and its consequences for association studies. *Animal Genetics* **38**: 277-286.

Valdar W., Solberg L.C., Gauguier D., Burnett S., Klenerman P., Cookson W.O., Taylor M.S., Rawlins J.N., Mott R. & Flint J. (2006) Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat Genet* **38**:879-887.

Vallejo R.L., Li Y.L., Rogers G.W. & Ashwell M.S. (2003) Genetic diversity and background linkage disequilibrium in the North American Holstein cattle population. *Journal of Dairy Science* **86**: 4137-4147.

Van der Werf J.H., Marshall K., & Lee S.H. (2007) Methods and experimental designs for detection of QTL in sheep and goats. *Small Ruminant Research* **70**: 21–31.

Vigouroux Y., Jaqueth J.S., Matsuoka Y., Smith O.S., Beavis W.D., Smith J.S. C. & Doebley J. (2002) Rate and pattern of mutation at microsatellite loci in maize. *Molecular Biology and Evolution* **19**: 1251–1260.

Viitala S. (2008) Identification of genes controlling milk production in dairy cattle. Doctoral dissertation. MTT Agrifood Research Finland.

Viitala S.M., Schulman N.F., De Koning D.J., Elo K., Kinos R., Virta A., Virta J., Maki-Tanila A. & Vilkki J.H. (2003) Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle. *Journal of Dairy Science* **86** (5):1828-1836.

Visscher P.M. (2003): Principles of QTL mapping. QTL mapping course. Salzburg.

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

Weber J.L., & May P.E. (1989) abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44** (3): 388-396.

Weedon M.N., Lango H., Lindgren C.M., Wallace C., Evans D.M., Mangino M., Freathy R.M., et al. (2008) Genome-wide association analysis identifies 20 loci that influence adult height**.** *Nat Genet.* **40**: 575-583.

Weller J.I., Kashi Y. & Soller M. (1990) Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *Journal of Dairy Science* **73**: 2525–2537.

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

Yamazaki T. (1977) The effects of over dominance on linkage in a multi locus system*. Genetics* **86**:227–236.

Zhao H., Nettleton D. & Dekkers J.C.M. (2007) Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms. *Genetic Research* **89**: 1–6.

Zhao H., Nettleton D., Soller M. & Dekkers J.C.M. (2005) Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genetic Research* **86**: 77–87.

## ACKNOWLEDGEMENT

Ultimate thanks are to *GOD* who gives me strength to complete this work.

I am indebted to many people for their long-lasting support and encouragement which was invaluable for the successful completion of this work. In the following lines some of them are gratefully acknowledged. However, I am aware of the fact that there are many more and these words cannot express the gratitude and respect I feel for all of those.

First of all, I wish to express my sincere gratitude and deep thanks to my supervisor Prof. Dr. Dr. habil. Martin Förster for giving me the opportunity to work on this thesis in Chair of Animal Genetics and Husbandry, Faculty of Veterinary Medicine, Ludwig-Maximilians-University (LMU), Munich. I am very grateful for his support and for accepting this thesis.

I would like to express special thanks to Dr. Ivica Medugorac, Chair of Animal Genetics and Husbandry, LMU for his outstanding support and patience in answering many of my questions during these years. I appreciate his contribution, especially for reviewing this thesis and all my manuscripts. His comments have been most helpful, a heartfelt “thank you”!

I am thankful to Dr. Stela Masle, Dr. Jennifer Schär, Dr. Jan Maxa, Markus Neuditschko, Ina Vogl, Sophie Rothammer, Verena Schneider and Monika Höpfl, Chair of Animal Genetics and Husbandry, LMU for their advice and highly collaborative attitude.

I wish to thank Mr Martin Dinkel, Ms Renate Damian, Ms Tatiana Schröter, Ms Gisela Andorfer, Ms Ursula Werner, Ms Heike Zierahn and Ms Christa Nanou, Chair of Animal Genetics and Husbandry, LMU, for their support and advice during the work in the laboratory. I acknowledge also, Ms Schadt Karina, secretary in Chair of Animal Genetics and Husbandry for her help during these years.

I would like to thank my professors in Egypt, Prof. Dr. Khairy M. El-Bayoumi, Prof. Dr. Said El-Fiky and Prof. Dr. Shabaan Hemeda for their continuous support. I like also to express further thanks to all my friends and colleagues in the Faculty of Veterinary Medicine, Zagazig University, Egypt for their continuous encouragement. Special thanks to my colleague Dr. Hatem Soliman, Fish Medicine and Livestock Management, University of Veterinary Medicine, Vienna, Austria for his advice and continuous encouragement.

Special thanks to my parents, my brother and my sisters for their love and continuous support during all my life. I want to acknowledge my wife, Dr. Samah Khalil Lecturer of Forensic Medicine and Toxicology, Faculty of Veterinary Medicine, Zagazig University, Egypt for her efforts and support while I do this work. Thank you for encouraging me when the times were tough.

Sincere thanks should be also offered to the Egyptian government and Cultural Bureau in Berlin for their support of my doctor study in Germany.