Institute for Animal Breeding
Faculty of Veterinary Medicine of the Ludwig-Maximilians-University, Munich
Univ.-Prof. Dr. Dr. habil. Martin Förster

Multistage QTL mapping strategy in an advanced backcross cattle population

Thesis for the attainment of the title Doctor of Veterinary Medicine
from Faculty of Veterinary Medicine of the Ludwig-Maximilians-University, Munich

by
Stela Masle
from Sarajevo,
Bosnia and Herzegovina

Munich 2007

Gedruckt mit Genehmigung der Tierärztlichen Fakultät der
Ludwig-Maximilians-Universität München

Dekan:	Univ.-Prof. Dr. E. P. Märtlbauer
Referent:	Prof. Dr. Förster
Korreferent(en):	Prof. Dr. Reese

Tag der Promotion: 9. Februar 2007

1 Introduction

The majority of ongoing projects for detecting the underlying variance for traits of interest in the cattle are related to growth, milk and meat production. One of the objectives of the European Union research project BovMAS (N° QLK5-CT-2001-02379) was the identification of quantitative trait loci (QTL) affecting milk production in one advanced backcross Fleckvieh x Red Holstein population that are identical by descent (IBD), according to origin and effect.

Because of the strong founder influence and the optimal time since the introgression of his alleles into the population, the advanced backcross population provides a great opportunity for IBD mapping. The IBD mapping method takes advantage of historical recombinations in the region of interest. The mutation in a gene affecting some quantitative trait will occur within the context of a specific chromosomal haplotype so the QTL alleles that are IBD will tend to share the same marker haplotype in the vicinity of the mutation. During the time, the recombination process will cut the pieces of the haplotype region so the resulting haplotype in the actual population can be very small.

For the IBD method the QTL should already be mapped in a distinct chromosomal region and the QTL status of family-sires should be known. Chromosomes of QTL segregating sires should be sorted by their effect on phenotype into two groups. Comparison of haplotypes between these two groups should be able to reveal the common haplotype region shared among sires.

Combination of mapping strategies and designs for identification of the shared haplotype flanking the IBD QTL in an advanced backcross population, which led to subsequent intensive study of the identified QTL region is, to the best of our knowledge, first such study.

2 Literature

2.1 Basic principles

All genetic analyses are based on the use of naturally occurring or induced variation. The variation in phenotype between members of a population was studied from Mendel onwards in order to obtain conclusions about the studied phenomenon itself and the structure(s) that underlie it. The variation in phenotype represents the raw material for detection of gene(s), i.e. genetic variants, which cause that particular phenotype to occur. Genetic variants can be rare or common in the nature. Rare genetic variants are usually abnormal and tend to be eliminated by natural selection, but they can be kept "alive" for possible future needs. Also, one gene can have two or more common alleles in a population. This creates a state called genetic polymorphism and allows the existence of several, distinct phenotypes to be common in a population. On the other hand, there can be more than one gene influencing a specific phenotype. Hence, the simple one-to-one relationships between genotype and observed phenotype are rather rare. Usually, these relationships are far more complex and depend on a number of influencing factors, e.g. developmental noise, gene interaction, mutation, environmental influences, etc. There are many biochemical, molecular and statistical methods that have been developed for the purpose of genetic dissection i.e. identification of genetic structures responsible for variation, and the identification of their function. To be able to observe a biological structure or process through genetic dissection we need a label whose behaviour can be followed – a genetic marker (Suzuki et al., 1989).

2.2 Genetic markers

In the mid '80s *Mullis et al.* (1986), using a concept previously published by *Kleppe* and *Khorana* (1971), developed polymerase chain reaction (PCR). 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. Soon DNA sequencing was revolutionised, and PCR became the foundation for genetic and molecular analyses. Also, in the early '80s, analysis of eukaryotic genomes revealed the existence of simple sequence tandem repeats (SSTRs) widely distributed over the genome (Hamada et al., 1982; Tautz and Renz,

1984; Tautz et al., 1986). SSTRs or microsatellites consist of short repeat units, usually from one to six base pairs (bp) long, which are repeated up to a maximum of 60 times. Microsatellites are usually flanked by unique DNA stretches, so it is possible to choose primers, which specifically amplify the desired microsatellite locus. This makes them highly suitable for PCR amplification (Litt and Luty, 1989; Weber and May, 1989; Tautz, 1989). Furthermore, an amplified product is normally within the range of 50-300 bp so it can be easily resolved on sequencing gels (Tautz, 1989). Microsatellites are highly polymorphic and have a high mutation rate. High polymorphism of microsatellites is due to the variation in the number of repeat units. The main cause of mutation is reported to be the replication slippage mechanism (Levinson and Gutman, 1987; Schlötterer and Tautz, 1992; review by Ellegren, 2004). Replication slippage involves mispairing of complementary bases at the location of a microsatellite, leading to the insertion or deletion of one or more repeat units. Microsatellite mutation rates have been reported to be approximately 10^{-3} – 10^{-6} (Edwards et al., 1992; Schug et al., 1998; review by Zhang and Hewitt, 2003). In human population studies they are found to have up to 20 alleles and a heterozygosity of approximately 0.85 (Bruford and Wayne, 1993). High polymorphism, easy amplification by the PCR method and new techniques for automated fragment length analyses established microsatellites as a class of valuable and widely used genetic markers. The majority of ongoing mapping projects using nuclear DNA markers involve microsatellites. Although they are widely spread through the genome, their evolutionary origin is still not clear and their biological role is unknown. Microsatellites are usually located in non-coding regions and are often (wrongly?) considered to be “junk DNA” (Nowak, 1994; Makalowski, 2000). However, there are some issues that are still not clear and deserve attention, such as the complicated evolutionary relationships among microsatellite alleles, the considerably variable mutation rates among organisms and the questionable phenotypic neutrality of microsatellites (Zhang and Hewitt, 2003).

2.3 Linkage

If two loci are located close to each other on the same chromosome, they tend to be inherited together and these two loci are said to be linked. The linkage is gets smaller when the distance between loci gets larger. The reason for this decrease in linkage is the process of crossing-over between non-sister chromatids during meiosis. This causes the recombination of the parental genotype and produces new combinations called

recombinants. The frequency of recombination is used as a quantitative index of the distance between two loci on a linkage map. One map unit on the linkage map is termed as a centiMorgan, in honour of Thomas Hunt Morgan. Various mapping functions are used to relate the observed recombination fractions to the map distance expressed in centiMorgans, among which those developed by *Haldane* (1919) and *Kosambi* (1944) are the most common. Studying the linkage between genetic markers and traits is used to estimate the relative position of the genes affecting those traits. A good example of mapping using the linkage analysis is the mapping of the bovine spinal muscular atrophy (SMA) disease locus, a disease that is caused by an autosomal recessive gene on the telomeric part of chromosome 24 (Medugorac et al., 2003).

2.4 Mapping of quantitative trait loci (QTL)

In contrast to the case in humans, the identification of simple monogenic disease loci in farm animals is generally of low importance, since affected animals are usually eliminated from breeding. The traits of interest in farm animals, e.g. milk production and quality in cattle, show continuous distribution of phenotypic values and have polygenic backgrounds. They are controlled by an unknown number of QTL and influenced by environmental factors. A QTL is defined as a chromosome region that contains one or more genes affecting a quantitative trait. 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 logic of QTL mapping is simple. In a mapping population (e.g. backcross or F_2) coming from two lines that are fixed for different alleles at loci affecting a trait of interest, animals are phenotyped and genotyped. Tests to determine the presence of significant differences in phenotype between marker genotype classes are accomplished through the application of statistical methods. If a significant difference is determined, there is a marker-QTL linkage (Mackay, 2001).

2.4.1 The QTL mapping population

Cross-breeding is used in animals, but, in contrast to plants, highly inbred lines of farm animals are not available because of a severe loss in fertility that occurs due to inbreeding depression. Thus, farm animal crosses are made from mildly inbred lines or between different breeds. Usually parental populations are crossed to produce a F_1 generation, which is then backcrossed to one or both parental lines (BC design) or crossed *inter se* to produce the F_2 generation (F_2 design; Falconer and Mackay, 1996). 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).

The advanced backcross QTL analysis (AB-QTL) is a method, proposed by *Tanksley and Nelson* (1996), which was successfully applied in plant genetics (Yamamoto et al., 2000; Huang et al., 2004; Marri et al., 2005; Blair et al., 2006). It is based on crossing a wild or indigenous plant as a founder parent and a cultivated strain as a recurrent parent. The progeny is then backcrossed over generations. The AB-QTL analysis uses the information about QTL map locations and the effects of the favourable QTL alleles from the founder i.e. genotypic selection (Tanksley and Nelson, 1996). The advantage of this method lies in the fact that it makes the introduction of favourable alleles in elite breeding lines possible, avoiding at the same time the epistatic effects of deleterious genes found in the wild (Blair et al., 2006). The parallel could be drawn to the backcrosses between different breeds used in cattle. The crosses are made in order to improve the specific performances of one breed by introduction of desirable alleles from a different breed. Preferably, the emphasis is placed on keeping the recognisable characteristic of a recurrent breed as minimally changed as possible. When the best progeny from this cross is repeatedly backcrossed, only on the recurrent breed, the obtained population could be conditionally termed "advanced backcross" QTL design (even though the marker assisted selection is not applied in this case). Also, as was mentioned before, classical backcrossing uses the inbred lines which are not available in cattle, so the AB population is actually an outbreed population. Nevertheless, the influence of the founder in an AB population can be substantial. A "founder effect" occurs when a single genetic factor obtained from a group of individuals can be traced back to the one of their common ancestors i.e. the founder

(Schork et al., 1998). A good example of the "founder effect" is the human population of Finland, which expanded from a small group with a very little immigration rate some 100 generations ago into a population of about five million people today. Such circumstances present an ideal opportunity for high-resolution mapping. By now, numerous, mainly autosomal recessive disorders, have been mapped in the Finnish population (reviewed by de la Chapelle and Wright, 1998; Norio, 2003a; Norio, 2003b; Norio, 2003c).

A more common mapping approach in dairy cattle is to exploit existing large paternal half-sib families, produced through the use of artificial insemination. A further advantage of this approach lies in the possibility of using already recorded phenotypic values (de Koning et al., 2003).

2.4.2 The QTL mapping designs

For determining linkage between marker loci and QTL, the most common mapping designs in cattle are daughter design, granddaughter design and complex pedigree.
In the daughter design (DD, Fig. 1A), marker genotypes and trait values are assessed on daughters of heterozygous fathers. Progenies are grouped according to a marker allele received from the heterozygous father. 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).
In the granddaughter design (GDD, Fig. 1B), 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 DD it is preferable to have fewer sires with many daughters per sire, in order to increase the mapping power. 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.
In general, the advantage of GDD over DD is that fewer marker assays are needed for equivalent power. Also, the semen and blood samples from sires are easier to collect and

the breeding values of the sires are highly accurate (Weller et al., 1990). There are a number of published QTL mapping studies in cattle using GDD e.g. by *Coppieters et al.* (1998), *Freyer et al.* (2003) etc. or using DD e.g. *Lipkin et al.* (1998), *Mosig et al.* (2001) etc. A review on the subject is given by *Khatkar et al.* (2004).

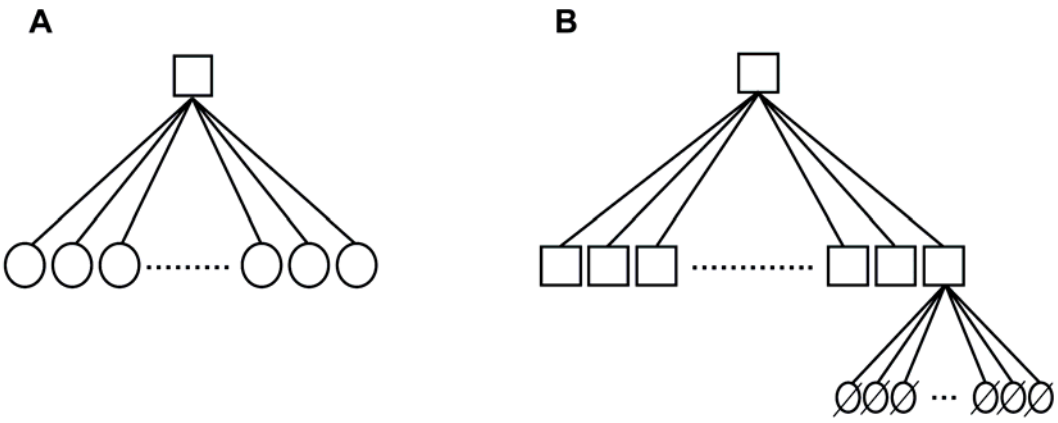

Figure 1. The common mapping designs in cattle: daughter design, where the daughters are genotyped and phenotyped (A) and granddaughter design, where the granddaughters of genotyped sons are only phenotyped (B). Squares represent male and circles female animals. Symbols for non-genotyped animals have a diagonal line through them.

A complex pedigree can provide a powerful design for mapping complex traits. It will 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 then those in admixed populations (Garner et al., 2001). QTL mapping in complex pedigrees is challenging, because the number of segregating 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). Statistical methods that can fully account for the complex relationships between individuals are expected to provide greater power to detect QTL (Almasy and Blangero, 1998).

2.4.3 The QTL mapping strategies

There are two main strategies for finding trait loci: association tests which use candidate genes and genome scans which are based on linkage mapping with anonymous DNA markers. The candidate gene approach can be very powerful, in cases where the candidate gene is a true causative gene, even in detecting loci with small effects. But this approach is also time-consuming and can fail because of current insufficient knowledge about gene function. Further dangers lie in the presence of linkage disequilibrium between loosely linked loci or even loci on different chromosomes (Farnir et al., 2000) and in the setting of proper statistical thresholds (Schaid, 2004) when testing with this approach. In contrast, a genome scan will always map a trait locus with a major effect if an accurate genetic model is postulated, reasonable sample size is used and the marker set fully covers the genome. However, it will fail to detect a trait locus with smaller effects, because of the stringent significance threshold applied (Andersson, 2001). Since the first genome wide scan experiment by *Georges et al.* (1995) a number of full or partial genome scans have been published which were able to detect QTL in dairy cattle. For the review, see *Khatkar et al.* (2004).

2.4.3.1 QTL mapping by means of "selective DNA pooling"

"Selective DNA pooling" has proven statistical power for detecting marker-QTL linkage. The method is based on combining selective genotyping and DNA pooling. In selective genotyping, only the individuals with extreme quantitative trait values are genotyped. Through DNA pooling, DNA of the individuals of two phenotypic extremes are pooled, creating "high" pool out of individuals with the highest values for the given trait and "low" pool out of individuals with the lowest values for the given trait. The determination of linkage is based on differences in the frequency distribution of parental alleles among pooled DNA samples of the extreme phenotypic groups of the offspring. The relative allele frequencies can be estimated on allelic band intensity, determined by densitometry. "Selective DNA pooling" reduces genotyping costs, because independent of the number of animals constituting one pool, each marker should be genotyped just twice, once in each extreme phenotypic group (Darvasi and Soller, 1994). Additional reduction of costs in dairy cattle comes with usage of milk as a source of DNA. As proposed by *Lipkin et al.* (1993), milk represents a good source of DNA for PCR amplification and consequently for

direct sequencing when an efficient protocol for DNA extraction is used and there are a sufficient number of cells in the sample. Direct PCR on milk samples provides identical results to the ones from DNA extracted from milk and blood. Milk with added preservatives, refrigerated or frozen, also proved to be a reliable source of DNA for a longer period of time. Milk samples are routinely collected as part of milk recording schemes and can be made available for QTL mapping (Mosig et al., 2001). Studies conducted using "selective DNA pooling" as QTL mapping method in daughter design (DD) and using milk as a source of DNA were able to detect large numbers of QTL segregating in a given population (Lipkin et al., 1998, Mosig et al., 2001). Despite its advances in reducing the amounts and costs of genotyping, "selective DNA pooling" has some disadvantages over individual genotyping. For example, it doesn't provide individual genotypes, but only estimates of allele frequencies. Allele frequencies are usually inaccurately estimated and have some degree of technical error. The QTL value of each individual cannot be individually assessed for a particular marker genotype (Pareek et al., 2002).

2.4.3.2 Interval mapping

Associations between marker and trait can be assessed using one-, two-, or multiple-marker genotypes. Usage of just one marker can detect a marker-QTL linkage, but the distance of the QTL from the marker cannot be estimated. This problem is overcome with the use of two flanking markers and determining the QTL position within the flanked interval. Interval mapping considers a pair of adjacent informative loci when determining the marker-QTL linkage (Paterson et al., 1988; Lander and Botstein, 1989). For n number of informative markers there is going to be $n - 1$ number of marker-QTL association tests. This increases the power of detection and offers more precise estimates of QTL effect and position, except when there is multiple QTL linked to the interval (Lynch and Walsh, 1998). To reduce or remove such bias the interval mapping method is further developed to use three or more marker loci simultaneously. So-called "composite interval mapping" considers a marker interval plus a few other well-chosen markers in each analysis (Zeng, 1994; Jansen and Stam, 1994).

2.4.3.3 "Approximate interval mapping" (AIM)

"Approximate interval mapping" is a novel method, developed for interval mapping, using "selective DNA pooling" data (Dolezal et al., 2005). The test was developed to predict test statistics for markers for which the sire is homozygous, or for any other location on chromosome, when the test statistic for markers for which the sire was heterozygous is available. In contrast with previous two methods developed by *Dekkers* (2000) and *Wang et al.* (2002), AIM does not need haplotype information (which is often missing) but it has less power as a consequence (Dolezal et al., 2005).

2.5 High-resolution mapping

2.5.1 The principles of identity by descent (IBD) mapping

Alleles that are identical by descent are direct descendants of a specific variant carried by an ancestral individual. If two alleles have identical nucleotide sequences but come from different origins in the reference population, they are identical by state but not by descent. On the other hand, genes that are identical by descent are always identical by state as well (Lynch and Walsh, 1998). Due to recombinations in the area surrounding a QTL allele, the inherited segment will tend to get smaller through generations. Rather than generating new recombinants by producing more offspring, IBD mapping takes advantage of historical recombinations in a region carrying a functional mutation in a gene which affects some quantitative trait (Schork et al., 1998). If a QTL allele located in a specific chromosomal region is segregating in the population and represents a mutation in the gene, we would expect it to derive historically from the same original mutation rather than to represent a novel mutation in the same gene (de Koning et al., 2003). Thus, we would say that the mutant QTL allele is IBD. Since any mutation will occur within the context of a specific chromosomal haplotype, mutant alleles that are IBD will tend to share the same marker haplotype in the vicinity of the mutation. By comparison of marker haplotypes over QTL-containing regions we can define a small chromosomal region within the QTL lies. The longer the period from the occurrence of the mutation, the smaller the shared haplotype between different QTL-carrying chromosomes will tend to be (Peltonen et al., 2000). However, too grate a time can cause the appearance of independent novel mutations which result in different marker haplotypes and the IBD approach will not work. General principles of IBD mapping include initial QTL mapping, identification of QTL-segregating

sires, genotyping of these sires for set of markers and the establishment of the linkage phase, sorting out sire chromosomes according to their effect on phenotype and the identification of the shared haplotype flanking the IBD QTL (Riquet et al., 1999).

2.5.2 Linkage disequilibrium (LD)

Linkage disequilibrium describes a situation in which particular alleles occur in a specific haplotype more frequently than it is expected by chance. There are many factors influencing LD including genetic drift, migration, mutation and selection. Due to the widespread use of artificial insemination and the intense selection for increased milk production, most of the cattle subpopulations, i.e. breeds, are structured as very large half-sib families but have low effective population size (N_e), i.e. low numbers of independent chromosomes. This produces extensive LD that could be used for fine mapping in dairy cattle. The occurrence of LD between nonsyntenic loci raises the concern about false-positive result when applying only LD mapping in effectively small cattle populations so the preference should be given to the combined linkage and LD methods (Farnir et al., 2000).

A combined linkage disequilibrium and linkage (LDL) mapping method was first proposed *Meuwissen and Goddard* (2001). *Farnir et al.* (2002) described the method for combined LDL as an extension of a multipoint association method and applied it to map a QTL with a major effect on milk production on BTA14. The combination of both methods was used in linkage disequilibrium variance component mapping (LDVCM) by *Blott et al.* (2003). *Lee and van der Werf* (2004; 2005, 2006) also described combined LDL method. The method is established in three simulation studies, all testing the variance component (VC) method using combined LDL mapping for fine mapping of a QTL. The first one (Lee and van der Werf, 2004) was investigating the efficiency of the experimental designs, the second one (Lee and van der Werf, 2005) the role of pedigree information in a general complex pedigree and the third one (Lee and van der Werf, 2006) the implementation of restricted maximum likelihood (REML) into the VC approach. All simulation studies were based on the following: In the first part of the simulation, the population was developed in a historical sense, beyond recorded pedigree. In each generation the number of male and female parents was $N_e/2$, and unique numbers of mutant alleles to QTL were provided. Parents were randomly mated with a total of two offspring for each mating pair. Parental alleles were passed to their offspring using the gene dropping method (MacCluer et al.,

1986). In the last generation (t), sires and dams were selected randomly. Surviving mutant allele was also randomly selected, with a frequency of >0.1 and < 0.9, and was then treated as a favourable QTL allele with effect α compared to other QTL alleles. The marker alleles mutated at a rate of $4x10^{-4}$ per generation. The population in the last generation was designed in the second part of the simulation. The descendants in the t + 1 generation were given phenotypic data and pedigree information. For animals in this generation, as well as for animals in the generation t, the marker genotypes were known and the phase was assumed as known. For a multiallelic marker model, the number of alleles was four and base allele frequencies at 0.25. To evaluate effects of marker densities and computational stability and efficiency, eleven markers were positioned at 10, 1 and 0.1 cM intervals. Studies showed that the half-sib design of few sires and a large number of dams, which is common in dairy cattle populations, could be efficiently used for fine mapping of QTL by combined LDL mapping. Also, they show that the combined LDL mapping had generally higher power in positioning the QTL than the linkage analysis alone. When there is substantial LD in the population pedigree, information is not so important, but its importance increases with decreasing the LD. We applied the software program developed by *Lee and van der Werf* for the fine mapping of QTL, but analyses by other programs, based on combined LDL method, have already begun. There are reports of the successful application of combined LDL in fine mapping of QTL in dairy cattle, e.g. the QTL for twinning rate at chromosome 5 (Meuwissen et al., 2002), the QTL affecting milk yield on chromosome 14 (Farnir et al., 2002), the QTL affecting milk yield and composition on chromosome 20 (Blott et al., 2003), two linked QTL affecting the milk fat yield on chromosome 26 (Gautier et al., 2005), the QTL affecting protein percent on chromosome 6 (Schnabel et al., 2005), the QTL affecting milk yield on chromosome 6 (Olsen et al., 2005).

2.6 Ultimate goals

2.6.1 Finding causative gene/mutation

The main strategy for finding a causative mutation is positional cloning of candidate gene(s). High-resolution mapping is the step which restricts the region of interest and reduces the number of potential candidate genes. Combining the information on map location and gene function leads to identification of positional and functional candidate

genes for further analysis. Pure positional cloning is used rarely in animals; this procedure primarily derives from comparative data from humans and experimental animals (Andersson, 2001). Positional cloning led, for example, to the identification of the bovine *DGAT1 (diacylglycerol acyltransferase)* gene which affects milk yield and composition, when the QTL was already mapped to the 3-5 cM interval on the chromosome 14 (Grisart et al., 2002; Winter et al., 2002). One further study reports that the phenylalanine to tyrosine substitution in growth hormone receptor (GHR) is associated with a major effect on milk yield and composition on bovine chromosome 20 (Blott et al., 2003). The analyses on bovine chromosome 6 indicate the possibility of osteopontin (OPN) as a candidate gene with effects on protein percentage (Schnabel et al., 2005).

2.6.2 Marker-/gene-assisted selection (MAS/GAS)

The identified region that contains a locus affecting a trait of importance in livestock or the gene(s) itself can be used for indirect or direct selection of genomic regions or gene(s) by marker-assisted selection and gene introgression. According to *Dekkers* (2004), in regards to the application of molecular information in selection programs, three types of genetic loci can be distinguished:

- *direct markers* i.e. genes– (GAS)
- *LD markers* i.e. loci that are in population-wide linkage disequilibrium with functional mutation – LD-MAS
- *LE markers* i.e. loci that are in population-wide linkage equilibrium (LE) with functional mutation in outbred populations – LE-MAS

Direct markers are the most difficult to detect, followed by *LD markers* and *LE markers*. On the other hand, *direct markers* are the easiest for application in selection programs, followed by *LD markers* and *LE markers*. Whereas *direct markers* and *LD markers* can be used across populations, *LE markers* can only be used within families. There are already substantial examples of utilisation of genetic information in commercial breeding programs e.g. in Germany described by *Liu et al.* (2004) and *Szyda et al.* (2005). As for milk yield and composition, there are gene tests based on direct markers, e.g. DGAT1, and growth hormone receptor (GHR), on LD markers e.g. LD marker near to prolactin gene (PRL; Cowan et al., 1990) and some LE markers in so-called “in-house” breeding programs. Successful implementation of MAS depends not only on markers but also on the careful considerations of selection strategies and approaches (Dekkers, 2004).

3 Material

3.1 Animals

The majority of the animal material used for the analyses presented here was collected in the context of the EU-project BovMAS (N° QLK5-CT-2001-02379). One of the project's tasks was mapping of QTL affecting milk production traits in a chosen population by means of selective DNA pooling in a daughter design (DD; Mosig et al., 2001). For this purpose we collected milk samples from a total of 18 half-sib daughter families (DD-18). These 18 DD families can be divided into two groups. The first group, consisting of 10 DD families, comes from the purebred Bavarian and Austrian Fleckvieh population (FV). This is a large dual-purpose population with about 1.4 million recorded cows and with reproduction based on artificial insemination (A.I.), using progeny tested sires. There are numerous large half-sib daughter families within the population. Family sires of the 10 chosen DD families are some of the most influencing Fleckvieh sires. Their most important founder was demonstrated to be the bull "HAXL" (born 1966).

The second group, consisting of eight DD families, represents a unique, advanced backcross population Fleckvieh x Red Holstein (ABFV). It was included in this study as a dual-purpose subpopulation within Fleckvieh. The aim of the crossing was to increase milk production and to improve udder quality. The initial crossing of Swiss Simmental females to a male from a dairy breed Red Holstein produced a generation of potential parents for backcrossing with Fleckvieh. Out of this F_1 generation the bull "REDAD" (born 1973; Fig. 2) produced, in backcross with Fleckvieh females, nine sons with favourable breeding values and more than 5000 daughters each, among five with numbers of daughters varying from approximately 10,500 to 28,500 (BC_1). Backcrossing of "REDAD" sons and their sons was repeated producing backcross generations two and three (BC_2 and BC_3). This process is continued with chosen BC_3 sires creating the current generation of tested bulls (BC_4 – BC_6). A set of eight families from this Fleckvieh x Red Holstein backcross was taken into the study. Six of the half–sib daughter families used come from the BC_3 generation. Further, two half-sib families are sampled from the BC_4 generation (Fig.2).

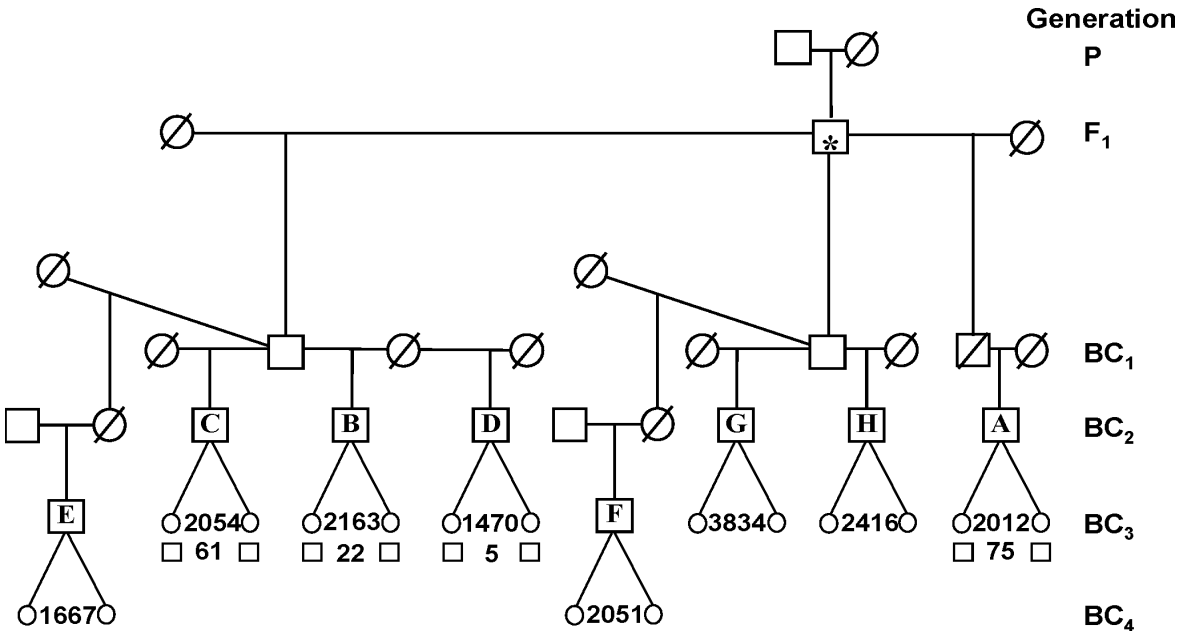

Figure 2. The advanced backcross Fleckvieh x Red Holstein population. This population is produced when one Red Holstein male (P) was crossed to a Swiss Simmental female and produced the important F1 bull "REDAD" (*). His sons with the favourable breeding values comprise the first backcross generation (BC1). Repeated backcrossing of chosen males on the Fleckvieh females produced BC2, BC3 and BC4. Eight family sires are marked A-H. Male animals are presented with squares and female animals with circles. Symbols for non-genotyped animals have a diagonal line through them. The number of sampled daughters in each family is shown, as well as the number of sons, if they were sampled.

We sampled a total of 48,190 daughters in the year 2002 to create DD-18. The number of daughters varied from 1470 to 6057, with an average of 2677 daughters per family. We also sampled sons in nine DD families (Table 1) to provide an independent sample for confirmation of the mapping results. Eleven DD-18 family sires were, at the same time, present as sons in a granddaughter design (Table 1).

Figure 3 presents 33 connected GDD families, along with DD families. The subset of 20 GDD families (GDD-20; Table 2) was considered in this study. For seventeen DD-18 families and 24 GDD families, respectively, we were able to sample the sire of the sire (Fig. 3). In addition, we sampled all available male ancestors up to important founders. This allowed us to build up a complex five-generation pedigree (FV-ROOT) comprised of 69 animals (Fig. 3). For the intensive study, a total of 11 families were selected in accordance with the results of the analysis performed in GDD-20 and DD-18. For these 11 families the granddaughter design (GDD-11) was chosen (Table 2). For the families that were already in the GDD, only the additionally available sons, if any, were collected. For the families from DD, all available sons were collected.

Table 1. Daughter design (DD-18) comprised of 18 half-sib families used for "selective DNA pooling".

No	Sire	Year of birth	RH[1] (%)	Number of sampled		Country of sampling
				Daughters	Sons	
01	STEFFEN*	1990	-	2033	-	G
02	SAMURAI*	1992	-	2554	54	G
03	SPORT	1992	-	2063	-	G
04	ZITAT*	1987	-	2275	1	G
05	ZEUSOR*	1992	-	1635	-	G
06	WINZER*	1992	6[b]	6057	59	G
07	HONER*	1992	-	4904	28	G
08	MALF	1988	3[b]	4228	144	A&G
09	LANDON	1989	-	1969	-	A
10	DONES	1990	-	2806	-	A
11	REXON	1989	12[a]	2012	74	G
12	REDER*	1989	12[a]	2163	22	G
13	RENGER*	1989	12[a]	2054	61	G
14	RENNO*	1992	12[a]	1470	5	G
15	HONNEF*	1993	6[a]	1667	-	G
16	UTNACH*	1991	6[a]	2051	-	G
17	RAUDI	1987	12[a]	3834	-	A
18	RUMSI	1990	12[a]	2416	-	A

[1] Red Holstein breed proportion

[a] breed proportion coming from REDAD

[b] breed proportion coming from other Red Holstein sires

* Eleven family sires that are at the same time sons in a granddaughter design (GDD-20)

Table 2. Granddaughter design families used in initial granddaughter design (GDD-20) and for fine mapping of QTL affecting PP on bovine chromosome 19 (GDD-11).

No	Grandsire	GDD-20	GDD-11	Year of birth	RH[1] (%)	Number of sampled	
						Sons	Daughters[2]
1	HORROR	X	X	1979	-	55	157,684
2	PROPELLER	X		1981	-	39	15,761
3	RENNER	X	X	1981	25[a]	57	118,214
4	ZEUS	X		1981	-	55	88,896
5	HODSCHA	X		1983	-	60	27,217
6	HORLER	X		1983	-	78	13,079
7	UTERINO	X	X	1983	-	60	26,434
8	STREITL	X		1984	-	84	72,178
9	HORB	X		1986	-	54	68,646
10	HORWEIN	X		1986	-	98	70,937
11	HORST	X		1987	-	64	44,636
12	HUMBERG	X		1987	-	44	15,015
13	RADON	X		1987	12[a]	41	7925
14	RALBO	X	X	1987	12[a]	96	77,423
15	ZAX	X	X	1987	-	80	49,336
16	MALF	X		1988	3[b]	145	33,687
17	ROMEN	X		1988	-	62	36,197
18	STREUF	X		1988	-	50	10,730
19	RENOLD	X	X	1989	12[a]	77	8126
20	REPORT	X		1989	12[a]	93	20,883
21	RENGER		X	1989	12[a]	61	6223
22	REDER		X	1989	12[a]	22	2929
23	REXON		X	1989	12[a]	74	7554
24	WINZER		X	1992	6[b]	59	5259
25	SAMURAI		X	1992	-	54	5463

[1] Red Holstein breed proportion
[a] breed proportion coming from REDAD
[b] breed proportion coming from other Red Holstein sires
[2] Numbers of all daughters from all collected sons at the time of sampling

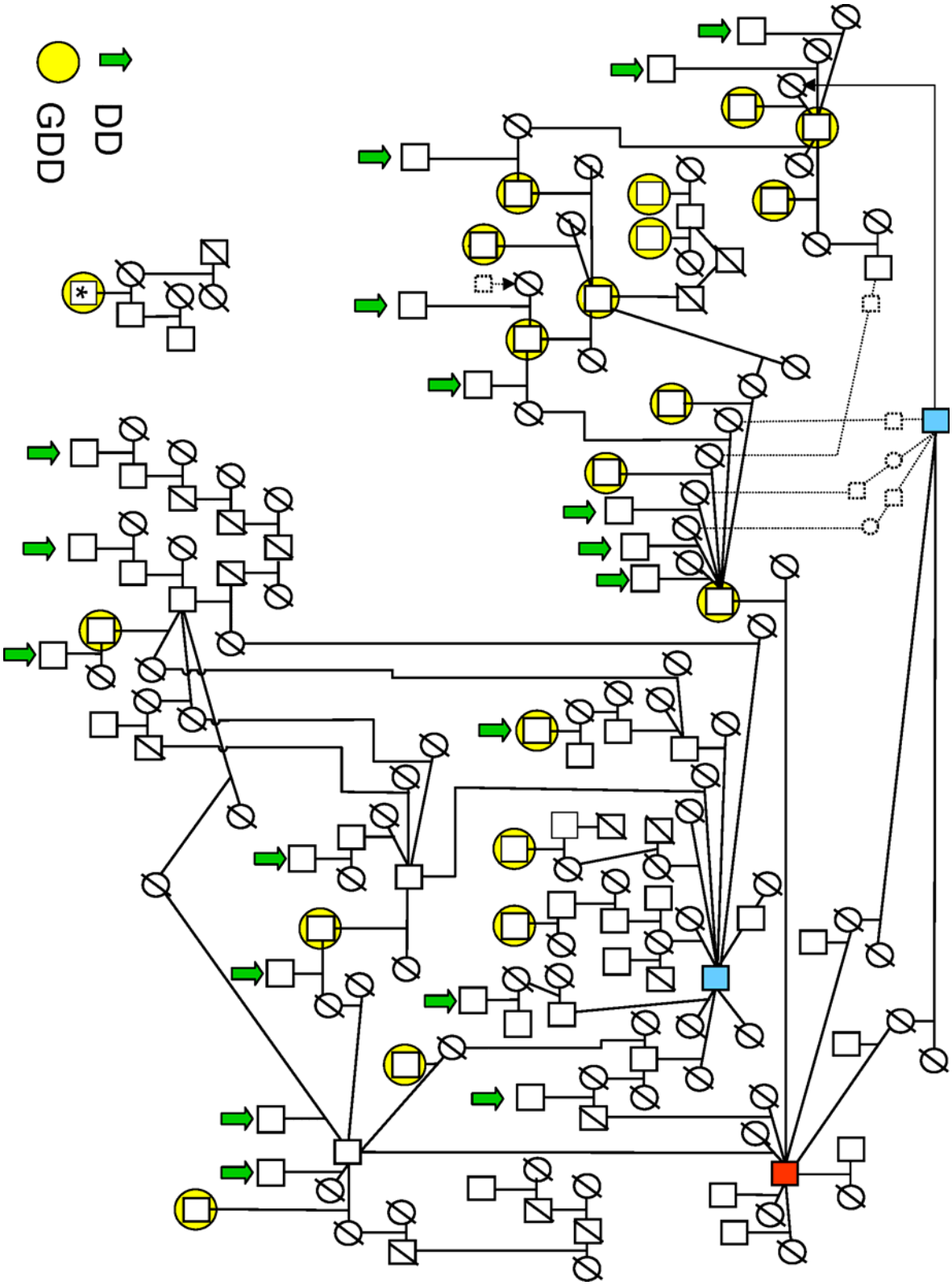

Figure 3. In the complex pedigree (FV-ROOT) the 18 daughter design sires (DD-18, arrows) and 33 granddaughter design sires (yellow circles) are shown with all available ancestors up to the important founders ("REDAD" marked red, "HAXL" marked blue). Only one granddaughter design sire comes from an independent family that connects neither to both REDAD and HAXL nor to the remaining family sires (*). In the pedigree the squares represent males and circles represent female animals. Symbols for non-genotyped animals have a diagonal line through them. In order to reduce the complexity of the picture, one founder ("HAXL" marked blue) is shown twice.

3.1.1 Complex pedigree based on GDD-11

Eleven families were selected for the intensive study on BTA19 in the granddaughter design, marked as GDD-11 and comprised of 694 animals. Since we had already used some of these families in earlier mapping projects, there was not enough DNA available to provide samples from 13 animals. For this reason, only 681 were actually genotyped. These animals were, together with their sires and mothers, connected through ancestors to the FV-ROOT for the final haplotype analysis and the analysis with linkage disequilibrium and linkage (LDL) method, thus building a so-called complex pedigree based on GDD-11. This pedigree was then filtered on animals genotyped for 12 to 21 markers. The applied filter left 593 genotyped animals in the pedigree, which then comprised of totally 1460 animals. The threshold of 12 markers for rating the success of genotyping process was established empirically.

3.2 Sampling

3.2.1 Semen samples

Semen samples were collected from A.I. centres, from a *Bavarian Genreserve* or directly from breeders. Animal information was entered into the computer database under a unique laboratory number for each animal. Every sample got its own unique coordinate under which it was stored. A bar code printer was used to print labels with animal laboratory numbers, box names and positions. Samples with adequate bar code labels were then stored in corresponding boxes under corresponding coordinates.

3.2.2 Blood samples

All blood samples were obtained from a blood reserve bank, established by routine paternity testing in Tierzuchtforschung e.V. Müenchen, or directly from breeders. They were systematically stored in the same manner as the semen samples.

3.2.3 Milk samples

As proposed by *Lipkin et al.* (1993), milk represents a good source of DNA for the PCR amplification and consequently for the direct sequencing when an efficient protocol for

DNA extraction is used and there are a sufficient number of cells in the sample. Routine milk recording schemes and its collection in centralised laboratories for testing make milk very attractive for genetic analysis methods. In accordance with the above-mentioned proposal, we developed a very efficient logistic scheme for collection and storage of milk samples. Samples were collected in Germany and Austria. Thirteen out of 18 DD families were collected only in Germany, four DD families were exclusively collected in Austria and one DD family was collected in Germany and Austria (Table 1). The principle used for sampling in Germany was as follows: target half-sib families were chosen for sampling in accordance with the Consortium of Cattle Breeds Associations (Arbeitsgemeinschaft Süddeutscher Rinderzuchtverbände e.V.; ASR). The Bavarian milk recording organisation (Landeskuratorium der Erzeugerringe für tierische Veredelung in Bayern e.V.; LKV Bayern) used this information for actual selection of daughters. The desired milk samples from selected daughters were then marked on a list for the field workers. If a marked animal was on his list, the field worker simply placed a blue plastic ring on the milk sample bottle (this system is routinely used for special treatment in quality control programs). All bottles from LKV Bayern were sent to the Bavarian milk analysing organisation (Milchprüfring Bayern e.V.; MPR) where milk analyses were performed. After routine analysis, bottles marked with a blue ring were manually sorted out and sent to our laboratory. All sample bottles were labelled with bottle number information only. Sample bottles are reused many times in milk recording schemes and they have permanent barcodes. The temporary connections between permanent bottle numbers and animal ear-tag numbers and the results of analyses for each sample are provided by the LKV Bayern. After we received this information, we were able to connect the samples with their analysis results and to sort them to corresponding families. Special database applications were used to provide reliable identification, labelling and storage of samples in co-ordinate system boxes, as previously described.

About one million samples are routinely analysed every month in the Bavarian milk analysing organisation. With a carefully planned strategy we were able to incorporate ourselves into this cycle with minimal investment and minimal changes in the routine milk recording schemes. Also, the involved organisations had to invest only minimum additional effort. We were consequently able to create an effective and reliable sampling system.

3.2.3.1 Milk pools

"Selective DNA pooling", as proposed by *Darvasi and Soller* (1994), is a combination of "selective genotyping" and DNA pooling, which means that only phenotypic extremes of a population, pooled together into two pools, are genotyped. For each of the two main traits, milk yield (MY) and protein percent (PP), two tail pools were formed, one consisting of individuals with high phenotypic values and another one of individuals with low phenotypic values for the trait. Daughters for each tail were selected according to corrected breeding values (cBV) as follows:

cBV = daughter breeding value – half of mother's breeding value.

From any of the selected daughters we pooled 10,000 somatic cells. Two independent sub-pools, so-called "replicates", were constructed for each tail. This process was performed twice, for the first and for the second duplicates. The term "replicate" refers to two pools of the same trait and tail, consisting of different daughters. We use the term "duplicate" for two pools of the same trait and tail, consisting of the same family daughters but created through two independent pipetting processes. Therefore, there were eight pools for one trait and one family: two tail pools (high and low tail) and two replicates, both in two duplicates. The number of animals in each of eight pools was averaged 101.5 (98-102). Besides the two main traits we made pools for seven associated traits including: milk protein yield (PY), milk fat yield (FY), milk fat percent (FP), milk somatic cell count (SCC), maternal non-return rate (mNR), maternal calving ease (mCE) and maternal stillbirth (mSB). For these pools the same number of animals was selected with 5000 cells pooled per animal. We also created eight pools for associated traits. Exceptions to this were the pools made for two threshold traits, mCE and mSB. For mSB, there were only enough affected animals to create one high pool, an average of 108 animals (41-152). The low pool contained the same number of randomly chosen non-affected daughters. Thus, four pools for the trait were made: one pool for two tails in two replicates. Just three families had enough affected animals to create eight pools for mSB. For mCE, the normalised mCE observations were used as a selection criterion. All daughters were standardised for the same calving numbers. There were 102 daughters with the most difficult calving score pooled into a high pool and 102 daughters with the easiest calving score pooled into a low pool (Table 3). The entire pooling procedure was controlled by a database application and performed by a liquid handling robot station.

Table 3. Pools made from milk samples of 18 half-sib daughter families for main and associated traits

Trait	Abbr.	Number of				
		Pooled cells per daughter	Tails	Replicates	Duplicates	Pools
Milk yield[a]	MY	10,000	2	2	2	8
Protein percent[a]	PP	10,000	2	2	2	8
Protein yield[b]	PY	5000	2	2	2	8
Fat yield[b]	FY	5000	2	2	2	8
Fat percent[b]	FP	5000	2	2	2	8
Somatic cell count[b]	SCC	5000	2	2	2	8
Maternal non-return[b]	mNR	5000	2	2	2	8
Maternal calving ease[b]	mCE	5000	2	1	2	4
Maternal stillbirth[b]	mSB	5000	2	1(2)*	2	4(8)*

[a] main traits
[b] associated traits
* just three families had enough affected animals to construct eight pools

3.3 Microsatellite markers

For the genome wide scan we chose 209 microsatellite markers covering all 29 autosomes from the public database (http://www.marc.usda.gov/genome/genome). After the genome-wide scan, we added 28 markers to seven regions of interest. In total, 237 microsatellite markers were considered in the study. Out of them, 18 were not included in the analyses (Results). Totally nine markers were used in genome-wide scan (GWS) on chromosome 19 (Fig. 4; Table 4). Two more marker sets (set-1 and set-2) were chosen for intensive study and fine mapping on chromosome 19 (BTA19). There were totally 24 selected markers, 12 in set-1 and 12 in set-2 (Fig. 4; Table 4). Additional information about markers used on BTA19 and the reasons for discarding four markers are given in Table 4.

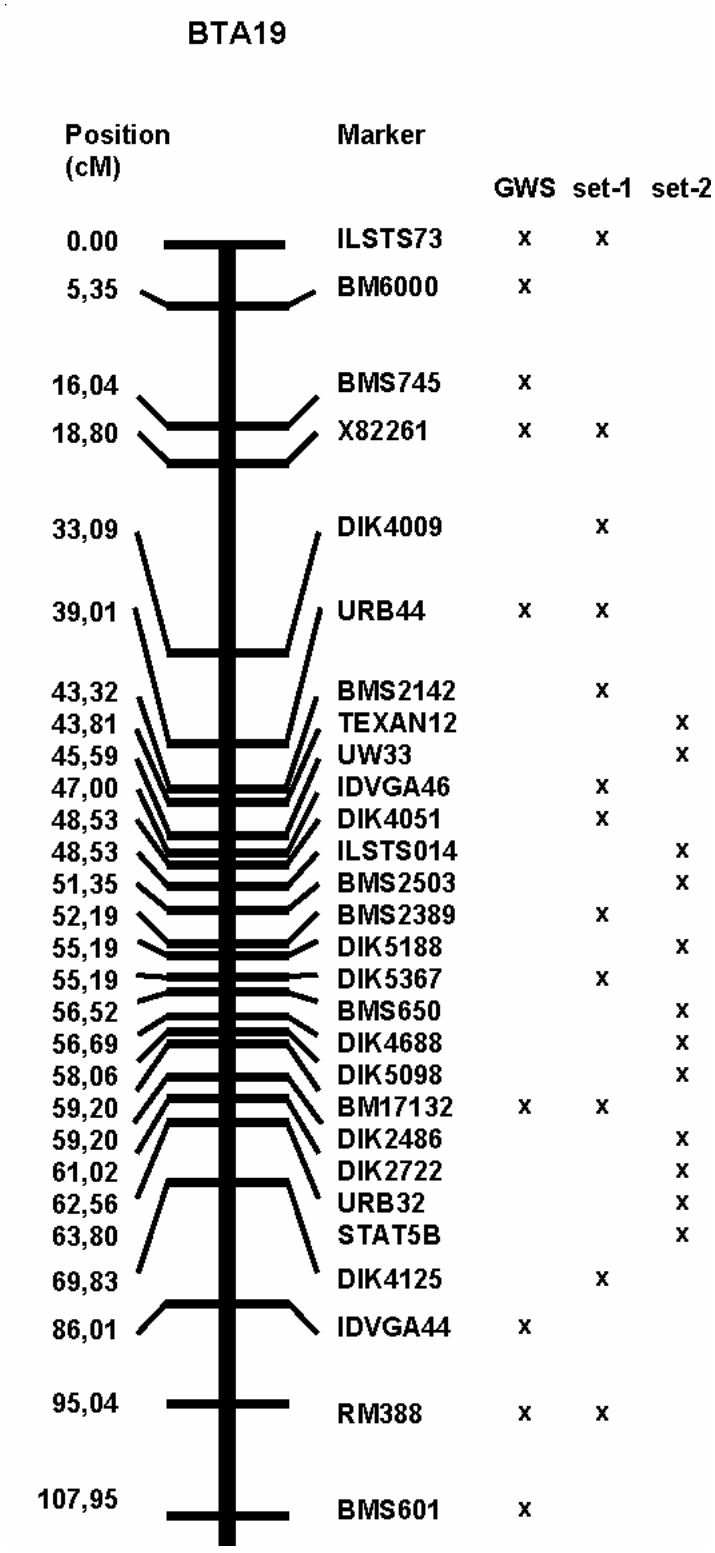

Figure 4: All markers used for the genome-wide scan (GWS) and intensive study (set-1 and set-2) on chromosome 19. Markers' positions (in centiMorgans, cM) were taken from the publicly available linkage map (USDA map).

Table 4: All markers used on BTA19 with marker name, synonym, genebank accession number, position on the USDA linkage map, used primer sequences and remarks.

Marker	Synonym	Accession No.	Position	Forward primer	Reverse primer	Remark
ILSTS73	ILSTS073	L37232	0,000	AGGGCAGGAGTAATCTTTGG	AACAGAGAGTATGGTGGTGG	
BM6000		G18448	5,352	ACAGCAATGCCATGGACC	TGCCATTTGGATGTGTGC	
BMS745		G18744	16,044	TAGGGACTTGTTACCCGTGG	TGCAAGCTGTGAGGAGGAG	
X82261		X82261	18,797	TGCTGCATCAAGTTCTCGAC	TCCCCCACTTCAATTTCAAC	
DIK4009		AB165361	33,091	CAGTGGAGTCACCCTTCTCC	GAGAATTACCCAGCCCCTTG	
URB44	URB044	U21777	39,007	CAATGTGATCTTCGTGTTCTGC	TCCTTCTTCCATCTCAAAGCTG	
BMS2142		G18924	43,814	AAGCAGGTTGATGATCTTACCC	GTCGGCACTGAAAATGATTATG	
TEXAN12		G01757	43,319	ACTGAAGGTGGCAAGAAACG	GTCCATGGGTTGCAAAGAGT	
UW33		L38834	45,590	TTGGGGTAGTCAGAGGTGGA	CACAGAATGGCAGAGACCAA	
IDVGA46	IDVGA-46	X85062	47,000	ACTTTGAAGTGGGCCTGGAT	ACACAAGCCAGGATCATCAC	
ILSTS014		L23488	48,531	GGAGAGACGGGTGCATTC	TTTCTGACCCGACCTCTGTC	
DIK4051		AB165391	49,301	GCAAAAAGAGGTCCGAGATG	AAATTCTGTTTTGGGGACCA	
BMS2503		G18958	51,346	TTGAACAACTACCAGCTTCCC	CATGTGTGAGACGGCACAGT	
BMS2389		G19011	52,192	AATGTTAGGTTTACATGCAGCC	AGGCAATAGGATCTCCACTAGC	
DIK5188		AB166223	55,186	TTTTGTGGCACTCTCCAATG	GAGGCCTTGTAGACCAAGCA	
DIK5367		AB166353	55,266	TTGTGTAGCCAGCATGCAAT	GCCAAGAAGGATGGCATTTA	
BMS650		G18860	56,518	ACAGCAACCACCTCACTCCT	CAGAAGCACACAGAGCCAAG	
DIK4688		AB165852	56,686	ACGTTCACAGACACCATCCA	AGAGGAGGTCCTGGCAGTGT	Allele suppression
DIK5098		AB166150	58,055	ATGGCATCATGTAAGCCTCA	AATCCCCATGAACAGAGGTG	Null allele
BM17132		G18544	59,202	ATCTGCCAGTATCACATCAACA	GTTACTTTTCCAGGCATGAAGC	
DIK2486		AB165026	59,222	TTTGACCATCTCCTAAAAGCAA	TTCACTTGTTTTTCCCCAAA	
DIK2722		AB165158	61,021	GAGAGGGTGCAAACAAGAGC	AGGGTGGGGAGATAGAGCTG	
URB32	URB032	U21770	62,564	GGAAGGGCAACAAAACAAAA	GGGAGGCTCAGGTTATCTCC	
STAT5B		AJ005638	63,802	CGTCTTTCTCCTTGGCAGTC	TGCAACGTGAGTGAACTTCC	
DIK4125		AB165447	69,830	TGTTGCGTTAGCTTCAGGTG	CAGCTATTGAAAACAAACTTACGA	
IDVGA44	IDVGA-44	X85059	86,010	GGGAGAATGGATGGAACCAAAT	TTCGAAGACGGGCAGACAGG	Irregular allele sizes
RM388		G29114	95,035	GGGGACCATCACGTACACTC	GGGACAGCCAGTCTTCTCAG	
BMS601		G18858	107,951	CACTAGGACGATGCTCTCAGG	TCACAAGAGCAATGACGAGG	Null allele

4 Methods

4.1 DNA extraction from semen

The semen straw contents, approximately 200-300 µl, were washed by repeated centrifugation and re-suspension of the pellet in PBS buffer until the supernatant was clear. The pellet was then re-suspended in 900 µl of DNA extraction buffer, and 100 µl of 0.5M DTT (dithiothreitol) and 50 µl of proteinase K were added to the suspension. The solution was incubated over night at 65° C. On the following day, 450 µl of 6M NaCl was added to the solution and it was then centrifuged for 40 min. The supernatant, containing dissolved DNA, was transferred into a new tube and mixed with 3.5 ml of ice cold 100% ethanol, causing the DNA molecules to stick together. The DNA precipitate was "fished out" with a disposable pipet tip, re-suspended in 50 µl of water and again incubated at 37° C over night for complete dissolution. The DNA concentration was measured by absorption at 260 nm with a spectrophotometer.

PBS buffer:	*NaCl*	*16.00 g*
	KCl	*0.40 g*
	Na_2HPO_4	*2.88 g*
	KH_2PO_4	*0.48 g*
	H_2O bidest.	*ad 2.0 l*
	(pH adjusted to 7.4 with 1M HCl)	

DNA Extraction buffer:	*NaCl(2 M)*	*100 mM*
	Tris (pH 8.2)	*10 mM*
	EDTA (pH 8.0)	*2 mM*
	SDS	*1 %*

4.2 Blood lysates from frozen whole blood samples

After thawing and mixing, a 6 µl aliquot of a blood sample was diluted in 500 µl of lysis buffer K. After a one minute centrifugation at 4000 x g, the supernatant was poured off. This procedure was repeated three times. In a final step, 2 µl of proteinase K (10 ng/µl) was added to the pellet, which was then re-suspended in 50 µl of lysis buffer K. The

suspension was incubated over night at 56°C or 2.5 h at 65°C than heated to 94°C for 15 minutes and centrifuged.

Lysis buffer K:	*Tris-HCl (1 M, pH 8.3)*	*8.0 ml*
	$MgCl_2$	*0.3 ml*
	KCl	*10.0 ml*
	Tween 20	*2.0 ml*
	H_2O bidest.	*ad 400 ml*

4.3 DNA extraction from milk and blood samples

DNA extraction from milk pools and whole blood samples was accomplished through the use of QIAamp Blood-Kits (Qiagen), following the manufacturer's protocol. For milk, the cooled sample was centrifuged to remove the upper layer of milk fat. The remaining sample was brought up to a standard volume with physiological NaCl so that all samples would have the volume of the largest sample. This adjustment was applied to all samples extracted in one round. Samples were then treated as follows: 100 µl of Qiagen protease per ml of sample was added and briefly mixed. For blood, 30 µl of the blood sample was treated with 20 µl of Qiagen protease and mixed with 170 µl of PBS buffer.
Subsequent steps were common for both milk and blood. To each ml of sample 1.2 ml of AL buffer was added. This solution was then thoroughly vortexed and incubated for 10 min at 70° C. After the incubation, 1 ml of ethanol was added per ml of sample. The solution was then transferred onto a QIAamp Midi column and placed in a centrifugation tube. The samples were filtrated through the membrane by centrifugation for 3 min at 2000 x g. DNA, bound to the membrane, was washed with AW1 and AW2 buffers in order to remove residual contaminants. In a final step DNA was washed out of the membrane with bidistillate water.

4.4 PCR conditions

4.4.1 Single PCR

Individual genotypes for the animals from FV-ROOT were necessary for the selection of informative markers used later in pool genotyping, for shadow-band correction and for

haplotype analysis. Since the multiplex PCR can yield artifacts and influence allele amplification a single PCR product was needed. The single PCR products were made according to standard protocol (Table 5). PCR reactions were performed in 15 µl final volume. For samples from semen or blood the following conditions were used: the reaction mix was preheated for 15 min at 95°C for activation of hotstart *Taq* polymerase, subjected to 35 cycles (95°C for 60 s, 58°C for 60 s and 72°C for 90 s) and a final extension for 7 min at 72°C and 45 min at 60°C. Single PCR products were combined before electrophoresis into 42 different sets, according to fluorescent label and product length. For milk pools, samples were preheated for 15 min at 95°C, subjected to 42 cycles (95°C for 60 s, 58°C for 60 s and 72°C for 90 s) and a final extension for 7 min at 72°C and 45 min at 60°C. Single PCR products were combined before analysis into sets of a maximum of four markers of non-overlapping fragment size to reduce the formation of artifacts. The possibility of signal penetration between different colours was considered as well, so the PCR products were combined together only when they did not overlap in fragment size across different colours. Thermal cycling was performed on Primus 96^{plus}, MWG-Biotech and PTC-100, MJ Research, Inc. thermocyclers.

Table 5. Standard PCR protocol

Reagents	Concentration	Volume used for the given DNA template (µl)	
		Blood and semen	Milk
H_2O bidest		8.8	7.85
Buffer	10 x	1.5	1.5
dNTP	2 mM	1.5	1.5
MgCl	25 mM	1.2	1.2
Forward primer	10 µM	0.4	0.375
Reverse primer	10 µM	0.4	0.375
Taq polymerase	5 U/µl	0.2	0.2
DNA template	30 ng/µl	1	2

4.4.2 Multiplex PCR

Every multiplex was optimised using test DNA samples. Each marker used in a multiplex was first tested as a single PCR product to establish marker properties and possible presence of artifacts. Optimal multiplex conditions were found by changing the quantity of primers and the annealing temperature. In case of incompatibility of markers, e.g. artefact or allele suppression, the multiplex reaction was split into two or more sets that were again

combined after PCR for joint electrophoresis and analysis. PCR reactions were performed in an 11 µl final volume. The reaction mix was set up by combining 5 µl of so-called "primer mix" and 5 µl of "*Taq* mix". "Primer mix", containing all primers diluted in water, was subject to changes, according to the results of the multiplex optimisation. "*Taq* mix", containing buffer, MgCl, dNTPs and polymerase, was kept standard (Table 6). To the reaction mix 1 µl of DNA template was added and the PCR was started. Amplification was done as follows: the reaction mix was preheated for 15 min at 95°C, subjected to 35 cycles (95°C for 60 s, 58,5°C for 60 s and 72°C for 90 s) and a final extension for 7 min at 72°C and 45 min at 60°C. Note that given annealing temperature (underlined) varied between different multiplexes.

Table 6. Standard "*Taq* mix"

Reagents	Concentration	Volume used for the DNA template from blood and semen (µl)
H_2O bidest		1.94
Buffer	10 x	1.1
dNTP	2 mM	1.1
MgCl	25 mM	0.66
Taq polymerase	5 U/µl	0.2

4.5 Horizontal electrophoresis

In case of optimisation of difficult markers, which needed conditions different from those of the standard one, as well as for the control of PCR successfulness, PCR products were tested on agarose gel stained with Ethidium Bromide (EtBr). Gel concentration was adapted according to size of products. All gels were made with TBE buffer, and the same buffer was used as a running buffer. 7 µl of sample was mixed with 2 µl of blue buffer before loading. The first lane was always loaded with 5 µl of DNA length marker. Electrophoresis was done on Pharmacia, LKB-GPS200/400 equipment. The first standard technique was based on EtBr staining by soaking gels for 10 min. Later, this technique was changed and gels were made by the direct addition of 11.5 µl of EtBr per 50 ml of gel. Gels were then exposed to UV light in a MWG-Biotech UV Transilluminator and a photo was taken for analysis.

TBE buffer:	*Tris*	*162.0 g*
	Boric acid	*27.5 g*
	EDTA	*9.3 g*
	H_2O bidest.	*ad 1.0 l*

Blue buffer:	*Dextran-blue*	*50.0 mg*
	H_2O bidest.	*1.0 ml*

4.6 Vertical electrophoresis

4.6.1 Gel preparation

Polyacrylamide gel was made using 20 ml of gel solution, 16 ml of TEMED and 64 µl of APS. It was poured between two glass plates with a syringe, separated by 0.2 mm thick spacers, and the plates were fixed with three clamps on each side. A square tooth casting comb was installed and fixed with clamps. Polymerization was conducted for 20 min at room temperature followed by 20 min at 61^{o} C.

Gel solution:	*Urea NF*	*33.60 g*
	10 x TBE	*8.00 ml*
	Acrylamide/Bisacrylamide (29:1)	*10.66 ml*
	H_2O bidest.	*ad 80.00 ml*

4.6.2 Preparation of PCR products for analysis

PCR products for gel electrophoresis were diluted with water in a proportion of 1:15 for template DNA extracted from semen and blood and 2:8 for DNA extracted from milk. 1 µl of diluted PCR products was mixed with 1 µl of the standard and 2 µl of formamide blue buffer. Prepared samples were heated for 2 min at 95 ^{o} C and immediately cooled on ice. Subsequently, 1 µl was loaded on the gel.

Formamide blue buffer:	*Dextran-blue*	*10.0 mg*
	Formamide	*1.0 ml*

4.6.3 Electrophoresis in gel

Electrophoretic separation and spectral detection of dye-labelled DNA fragments was made using an ABI Prism 377 DNA Sequencer. The dye-labelled DNA fragments were electrophoresed through the vertical, 0.2 mm thick polyacrylamid gel, and separated according to size. Separation distance was 36 cm. At the end of a gel they passed a region where a laser beam continuously scans across the gel. The laser excites the fluorescent dyes attached to the DNA fragments, causing them to emit light at a specific wavelength for each dye. The light was collected and separated according to wavelength by spectrograph onto a CCD (charge coupled device) camera. The data collection software recorded the light intensities and stored them as electrical signals for processing. Applied values for each run included a gel running speed of approximately 200 bases per hour (bph) or 2400 scans per hour (s/c), and duration of one run averaged 2.5 hours. A total of four dyes can be analysed in one lane. We used Fam, Tet and Hex as fluorescent primer labels and Tamra labelled PCR products as an internal standard. The square tooth casting comb produces 50 slots and usually 48 were used (the rest served as a blind control), so we could analyse half of a 96-well PCR plate in one gel. At first, all uneven lanes were loaded and then the run was started. After 5 minutes the run was stopped and the even lanes were loaded. TBE buffer was used for electrophoresis.

4.7 Capillary electrophoresis

4.7.1 Preparation of PCR products for analysis

PCR products for capillary electrophoresis were diluted with water in a proportion of 1:10. When the PCR products came from two or more sets, 1 µl of each set was diluted in 10 µl of water. From the diluted samples, 1 µl was then pipetted into 10 µl of Formamide/Standard mixture. Tamra labelled PCR products were used as internal standard.

Formamide/Standard mixture:	*Formamide*	*500 µl*
	Standard	*2 µl*

4.7.2 Electrophoresis in capillary

The basic principle of electrophoresis in capillary is the same as in electrophoresis in gel. The whole process takes place in a glass capillary filled with polymer. The current flow moves the injected portion of PCR products through the capillary, and the laser detects the fluorescent day labels. Emitted fluorescence from the dyes is then recorded by a CCD camera, and the intensity of the fluorescence at each point is later analysed. Capillary electrophoresis was performed on an ABI Prism 310/3100 DNA Sequencer.

4.8 Analysis of the data

The fragment length analysis results were processed with ABI *GeneScan*® and ABI *Genotyper*® software programs. With the ABI *GeneScan*® program we assigned the values for the internal standard. The program adjusted the run variations between different lane runs. Assigned standard values were then used by the program to determine the size of the fragments. The data was afterwards processed with an ABI *Genotyper*® program. With this program, genotypes were assigned using the allele size definition for the marker from the category list. All genotypes were exported in formatted table, which could be readily imported into a database.

5 Statistical methods

5.1 Quality control of genotypes

5.1.1 Database applications

We used two systems of control in the process of genotyping. The first system included the double genotyping of some already genotyped animals. After importing the information into the database, a list of conflicts was created containing both genotypes, the old one and the new one. Both genotypes were once more inspected, mistakes traced and a decision made about the right one. When we were not able to find a cause for the repetition conflict, genotyping was repeated again on an independent PCR. For intensive study, all animals were genotyped twice.

The second system included a paternity check. Inconsistencies in inheritance between parents and their offspring were listed by database application. The paternity conflict list was processed in the same manner as the previous system: the genotypes were inspected one more time, and a decision was made concerning the correct genotype or genotyping was repeated to obtain the correct one. Animals with persistent inheritance conflicts were excluded from further analysis.

5.1.2 Mistyping analysis

The mistyping analysis was done as a further quality control of genotypes in the FV-ROOT pedigree (Chapter 3.1). This analysis indicates genotype mistypings that are consistent with Mendelian inheritance and are revealed only by a decrease in pedigree likelihood caused by the excessive recombinations they involve. For this purpose we used the SimWALK2 program (Sobel and Lange, 1996). The mistyping analysis considered the following information: complex relationships within the pedigree, genotypes of all used markers along a chromosome, marker positions and allele frequencies. The necessary input data were made by a database application. Through a multipoint analysis that uses all available data, SimWalk2 reports the overall probability of mistyping at each observed allele. All genotypes marked as implausible with a significant probability were additionally checked and, if needed, corrected, or the genotyping process was repeated.

5.1.3 Analysis with the chrompic option of the CRI-MAP program

In addition to the mistyping analysis by the SimWALK2 program we used the chrompic option of the CRI-MAP program (Lander and Green, 1987) for further checks for improbable double or multiple recombinations. This analysis was performed on complex pedigrees based on GDD-20 or GDD-11. As these pedigrees have substantially more animals, the FV-ROOT SimWALK2 program would not be able to run the mistyping analysis in a reasonable computing time. The CRI-MAP program, just like the mistyping analysis by SimWALK2, considers complex pedigree and marker distances but does not make use of allele frequencies. According to the authors, little information is lost this way, except when the allele is rare. The input file consists of genotypes of animals in a complex pedigree for a set of marker loci. In the case of missing data for an individual at a particular locus, CRI-MAP deduces missing genotypes where possible. If possible genotypes for missing data include a homozygous genotype then all meioses in that individual are treated by CRI-MAP as uninformative for that locus. Again here the CRI-MAP program ignores some available information, but this loss is also reported to be small. The chrompic option is starting by finding the maximum likelihood estimates (MLE) of the recombination fractions for the specified locus order. The program uses MLE of recombination fractions to find the particular phase choice having the highest likelihood for that pedigree. The number of recombinations is listed in the output data, as well as the names of any informative markers, which are out of phase with the closest informative marker on either side. These “out of phase” markers represent candidate data errors, when markers are closely spaced. Genotypes of these possible errors were inspected and, if needed, corrected or independently repeated. For each chromosome interval, a list of chromosomes having a recombination in that interval was made, providing a better overview of incidence of recombinations in certain intervals and families.

5.2 Estimation of allele frequencies

Marker allele frequencies were obtained by simple counting. There were relatively small numbers of unrelated founders, but a comparatively large number of unrelated mothers of genotyped animals, which were not sampled. For unbiased allele frequency estimation we counted only founder alleles and all unrelated and safely deduced maternal alleles. The

known father allele was also used to establish the maternal allele in an animal, which was then counted. For this purpose a database application was used.

5.3 Haplotype analysis

Haplotype analysis was performed with SimWALK2. The haplotype analysis estimates the most likely set of fully typed maternal and paternal haplotypes of the marker loci for each individual in the pedigree. The recombination events within haplotypes are highlighted. Haplotyping by SimWALK2, as described by *Sobel and Lang*e (1996), is done by running a Markov chain over a set of genetic descent graphs for a pedigree. The genetic descent graph specifies the paths of gene flow in a pedigree but omits specification of the actual founder allele travelling down each path. The likelihood of a descent graph is calculated as

$$\Pr(\hat{G} \cap M) = \underset{i=1}{\overset{m}{Trans}}(\hat{G}) \prod \Pr(C_i),$$

where $\Pr(\hat{G} \cap M)$ is the joint likelihood of a descent graph $\hat{G}$ and a marker phenotype vector M, $Trans(\hat{G})$ is the transmission factor and $\Pr(C_i)$ is a probability of connected components of the founder tree graph.

The starting descent graph with maximum likelihood is found by simulated annealing. Moving between the descent graphs in a Markov chain is performed using three transition rules, T_0, T_1 and T_2 (Sobel and Lange, 1996). The Metropolis mechanism is used to construct a Markov chain that uses these transition rules and possess the correct equilibrium distribution. The Metropolis mechanism divides a step into a proposal stage and an acceptance stage. The movement from one step to another is accepted with Metropolis probability.

As for mistyping analysis, SimWALK2 uses all available data (complex relationships within pedigree, genotypes of all used markers along a chromosome, recombination fractions and allele frequencies) to estimate the most likely haplotypes compatible with the data set. Needed input data were formatted by a database application. The first analysis was done for all 29 autosomes in the FV-ROOT pedigree (Chapter 3.1). All obtained haplotypes were presented graphically for easier tracing of haplotypes in the pedigree tree. The second analysis was done in complex pedigree based on GDD-11 in order to obtain haplotypes necessary for combined linkage disequilibrium and linkage (LDL) mapping in the same pedigree. As already mentioned (Chapter 3.1.1), this pedigree is comprised of 1460 animals out of which 593 animals were genotyped for 12 to 21 markers. In order to

get the most possible haplotypes in a reasonable computing time, complex pedigree based on GDD-11 was divided into five overlapping sub-pedigrees, four sub-pedigrees, including two families, and one sub-pedigree including three families. Overlapping part of all sub-pedigrees were common ancestors i.e. FV-ROOT.

5.4 QTL mapping by means of "selective DNA pooling"

The determination of linkage by the "selective DNA pooling" method is based on the distribution of parental alleles among pooled DNA samples of the extreme phenotypic groups of offspring (Darvasi and Soller, 1994). The estimation of allele frequencies in pooled DNA samples is based on the linear relationship between the initial number of copies of the allele in a pool and final allele band intensity in an acrylamide sequencing gel, determined by densitometry (Lipkin et al., 1998). This estimation is hampered by presence of "shadow" bands i.e. artifactual PCR products derived from microsatellite's genomic tract by deletion or insertion of one or more repeat motifs. Consequently, allele band intensities should be corrected for the presence of overlapping shadow bands. The corrected allele band intensities are then used to test marker-QTL linkage by "selective DNA pooling". The shadow band correction was done as proposed by *Lipkin et al.* (1998). Briefly: The relative intensity ($RI_{n,i}$) of the i^{th} shadow band derived from genomic tract of allele A_n is calculated as:

$$RI_{n,i} = D_{n,i} / D_n,$$

where n is the number of repeats in the native genomic tract of allele A_n, i is the "order" of the shadow band (i = +1 to -3), D_n is the densitometric intensity of the main band derived from the genomic tract of A_n, and $D_{n,i}$ is the densitometric intensity of the i^{th} shadow band derived from the genomic tract of allele A_n. RI was calculated for the first three leading (i = –3 to –1) and the first trailing shadow band (i = +1). Shadow-corrected intensity (CI) of all bands in pool was calculated by the following model:

$$CI_n = D_n - \sum_{i=-3}^{+1} \left(CI_{n-i} RI_{n-i,i} \right),$$

where n, D_n and i are as above, CI_{n-i} is the shadow corrected intensity of D_{n-I}, $RI_{n-i,i,}$ is the RI of the shadow band of A_{n-i} that overlaps the main band A_n.

A test for marker–QTL linkage for an individual sire-marker combination was based on rejecting the null hypothesis, $D = 0$, where D is the difference in sire allele frequencies

between high- and low-daughter pools, as proposed by *Lipkin et al.* (1998) and *Mosig et al.* (2001). Thus, this test can be applied only to sires heterozygous at the given marker. Using normal approximation, the null hypothesis with type I error, α, is rejected if

$$Z_D = D/\mathrm{SE}(D) > Z_{1-\alpha/2},$$

where SE(D) is the standard error of D and $Z_{1-\alpha/2}$ is the ordinate of standard normal distribution such that area from $-\infty$ to $Z_{1-\alpha/2}$, equals $1 - \alpha/2$.

Significance of a marker M_i across j sires is summation over the subset of sire-by-marker combinations for that marker only:

$$\sum_{j=1}^{Sj} Z_{ij}^2 \sim \chi^2_{(Sj)},$$

where s_j is the number of heterozygous sires tested for the j^{th} marker. For the single marker test across sires all tests are two-tail tests as well, with P_j obtained for an individual marker M_j as twice the area of the chi-square distribution, with degree of freedom (d.f.) = s_j from χ^2_j to $+\infty$.

Since the detection of marker–QTL linkage represents a multiple test situation, involving many sires, markers and chromosomes, the usual criteria for setting experiment wise type I error results in a low experimental power (because of the high proportion of rejected true effects). To avoid it "adjusted false discovery rate" (aFDR) was calculated, as proposed by *Mosig et al*. (2001). It takes into account some critiques of false discovery rate (FDR) approach and calculates the FDR as:

$$q < n_2 P_{(h)} / t,$$

where t is the rank number of the ordered test comparisons, $P_{(h)}$ is the P-value corresponding to the null hypothesis of the h^{th} test and n_2 is the number of true cases of linkage analysis. n_2 is calculated by iteration as described by *Mosig et al.* (2001). Markers significant at an "aFDR" of 5% were taken as identifying a chromosomal region containing a QTL (QTLR). For each QTLR, sires were identified that show a significant sire-marker effect using an "aFDR" cut-off of 10%. The allele substitution effect was calculated as the difference between two genotypic groups of the sire's daughters, corrected for the selection intensity, as described by *Lipkin et al.* (1998). A genotypic group is defined according to the allele that daughter inherited from her father. The "selective DNA pooling" method was applied to all pool results in DD-18 (Chapter 3.1, Table 1).

5.5 Approximate interval mapping for selective DNA pooling

The approximate interval mapping (AIM) analysis is performed as described by *Dolezal et al.* (2005). Briefly: A different subset of tested sires will be heterozygous at each marker along a chromosome, so the test statistic will be more or less affected by the number of those sires and their QTL status. The approximate interval mapping method was developed to predict test statistics for markers for which the sire was homozygous or at any other location on the chromosome (T_l), given the observed test statistic of a series of heterozygous sire-by-marker combinations (T_i). The technical derivation of this approach on results from a single sire was as follows:

The prediction of T_l at location *l* from an observation on T_i at location *i*, is: $E(T_l) \sim (1\text{-}2r)^2\ T_i$, i.e. it is only a function of the proportion of recombination between the two locations. Recalling that under the null hypothesis of no linkage, the variance of $T_i \sim \chi^2_{(1)} = 2$, the covariance between the predicted T_l and observed T_i is:

$$\text{cov}(E(T_l),T_i) \approx (1\text{-}2r)2\ \text{var}(T_i) = 2(1\text{-}2r)2.$$

If the test statistic is observed at both locations *i* and *l*, the covariance is the same:

$$\text{cov}(T_l,T_i) \approx (1\text{-}2r)2\ \text{var}(T_i) = 2(1\text{-}2r)2.$$

Combining all of the above results, a multi-point prediction at any location on the chromosome can be made using a simple selection index analogy:

$E(T_l) = \mathbf{b}'\mathbf{t}$, where $\mathbf{t}$ is a vector of all observed test statistics (T_i) with $\mathbf{V}=\text{var}(\mathbf{t})$ the variance covariance matrix of $\mathbf{t}$ with 2 on the diagonal and $(1\text{-}2r_{il})^2$ on off-diagonals, and $\mathbf{b}$ the solution to $\mathbf{V}^{-1}\mathbf{c}$, with $\mathbf{c}$ a vector of covariances between predicted (T_l) and observed (T_i) test statistics ($=2(1\text{-}2r_{il})^2$). Two different Bonferroni-corrected chromosome-wide thresholds for AIM test statistic across sires were used to determine significant marker-QTL linkage: The first threshold is determined by $\chi^2_{(0.5/nM,\max Si)}$, where *nM* is number of markers investigated for respective chromosome and *maxSi* is the number of informative sires at the most informative marker for chromosome (AIM05). The second threshold is determined by $\chi^2_{(0.5/nM,nS)}$, where *nS* is total number of sires investigated for the respective chromosome (AIM05*). The two thresholds are equivalent if *maxSi*= *nS*. The approximate interval mapping method was applied to all single marker test results.

5.6 Linkage map construction

The build option of the CRI-MAP program was applied in order to confirm the marker order given by the published linkage map, i.e. USDA linkage map (Itoh et al., 2005; Ihara et al., 2004). Analysis started with "ordered loci", which were held fixed, and inserted new loci in each possible interval in the map. The fixed orders were ascertained by comparisons with results from the published linkage map, a high-resolution radiation hybrid map (Itoh et al., 2005; Everts-van der Wind et al., 2005) and the whole genome shotgun sequence results for a corresponding chromosome. For the inserted loci, the order having the highest log10 likelihood was found and inserted into the map. The next locus was tried in the same manner. Alternative orders with comparable likelihoods are also identified. We also attempted to separate the markers that were at the same position on the USDA linkage map for BTA19, when possible, but these loci tended to have low informativity in our data too. The best order, which integrates all available information, was chosen as the correct one, but the alternative analyses were performed as well (Results).

5.7 Linkage analysis

Linkage analysis was done with the QTL Express program, which is based on a linear regression method to map a QTL in the most common mapping designs, e.g. half-sib, F2, sib-pair etc. We used a model with one QTL for the half-sib design. Interval mapping is a two step procedure, which determinates the identity by descent (IBD) probabilities first and then fits a statistical model to the observations and IBD coefficients (Seaton et al., 2002). This model puts a putative QTL on every cM along the chromosome and computes the test statistic at each position. The test statistic is standard F-test, with the degrees of freedom equal to the number of the half-sib families in the numerator and the number of the common parents, informative at a given location, in the denominator. The mean QTL effect result was given for each common parent.

An analysis for one chromosome was first conducted across all families. Analyses were done family-wise for significant or indicative results across families to determine the sire's QTL status. A 95% chromosome-wise significance threshold was calculated on 10,000 permutations, and the bootstrapping option on 10,000 iterations was used to determine the confidence interval. Input data were written in the appropriate format by a database application and consisted of genotypes for given markers, recombination fractions between

markers and the phenotype data. The corrected breeding values (cBV= breeding value – half of mother's breeding value) were used as a phenotype and were weighted by their respective reliabilities. Interval mapping was applied to the GDD.

5.8 Combined linkage disequilibrium and linkage mapping

Combined linkage disequilibrium and linkage (LDL) mapping based on the variance component approach was performed as described by *Lee and van der Werf* (2004; 2005; 2006). Briefly: A mixed linear model for detecting QTL was applied. A vector for phenotypic observations was modelled as:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u} + \sum_{i=1}^{NQ} \mathbf{Z}\mathbf{q}_i + \boldsymbol{e},$$

where $\boldsymbol{y}$ is a vector of phenotypic observations on the trait of interest, $\boldsymbol{\beta}$ is a vector of fixed effects, $\boldsymbol{u}$ is a vector of random polygenic effects for each individual, NQ is the number of QTL, $\boldsymbol{q}_i$ is a vector of random effects due to i^{th} QTL and $\boldsymbol{e}$ are residuals. The associated variance-covariance matrix of all observations for a given pedigree and marker genotype set was then modelled as:

$$\mathbf{V} = \mathbf{ZAZ'}\sigma^2_u + \mathbf{Z}\,\mathbf{GRM}\,\mathbf{Z}\,\sigma^2_{qi} + \mathbf{R},$$

where **A** is the numerator relationship matrix based on additive genetic relationships, **GRM** is the genotype relationship matrix whose elements are IBD probabilities between individuals computed for a putative QTL position and conditional on marker information, and $\mathbf{R} = \boldsymbol{I}\sigma^2_e$ ($\boldsymbol{I}$ as an identity matrix). IBD probabilities were estimated by applying a Markov chain Monte Carlo (MCMC) algorithm. They were estimated for every cycle and averaged over all cycles. Estimation of IBD probabilities was based on joint information from LD and linkage. For each putative QTL position a **GRM** was constructed. The most likely QTL position can be estimated as the position with the highest maximum likelihood value across all positions. The log-likelihood ratio test LRT=-2(log(L_0)-log(L_p)) was computed, where L_0 corresponded to the likelihood of the null hypothesis model which assumes Var($\mathbf{q}_i$) = 0 and (L_p) likelihood under alternative model (L_p) at each position p. For the analyses the mutation age and past effective population size were held at 100. Initial homozygosity on each locus was 0.25. Initial burn-in was 100, followed by 1100 iterations. Parameter estimates were collected every 10^{th} round. The combined LDL mapping was applied on complex pedigree based on GDD-11.

Two programs were used:

- combined linkage disequilibrium (LD) and linkage (L) analysis with the random walk approach (ra) and the meiosis Gibbs sampling (ms) - **LDL_rams**,
- combined linkage disequilibrium (LD) and linkage (L) analysis - **LDL**.

The LDL_rams procedure uses the unordered genotypes and generates a **GRM** file by random walk and Gibbs sampling iterations. It takes a lot of computing time for the analysis of big pedigrees and for a few thousand iterations. On the other hand, LDL analysis uses reconstructed haplotypes, made by some external program, e.g. SimWALK2, to generate the **GRM** file and use it in variance component analysis. Practically, this means that the LDL analysis itself is very fast, and all computational time comes from the haplotype analysis, which can be also time-consuming.

6 Results and discussion

6.1 Sampling

We collected milk, blood and semen samples during the course of the EU-project BovMAS to use in different mapping designs: daughter design, granddaughter design and the complex pedigree. For the daughter design, milk samples were collected in two countries, Germany and Austria, from the two resource populations: purebred Fleckvieh (10 families) and the advanced backcross Fleckvieh x Red Holstein (8 families) population. The resulting design consisted of 18 daughter design families (DD-18; DD-18 fam01 to DD-18 fam18) for which we sampled in total 48,190 daughter milk samples (Chapter 3.1, Table 1). The number of sampled daughters varied form 1470 to 6057, with an average of 2677.

For the confirmation of mapping results we sampled blood and semen samples from the sons in nine DD-18 families, totalling 448 samples (Chapter 3.1, Table 1). Besides these samples we also used blood and semen samples collected in the previous mapping projects for granddaughter design families. The total number of sampled sons in 20 granddaughter design families (GDD-20) was 1392. The family size varied from 39 to 145 (69 on average; Chapter 3.1, Table 2).

The set for the intensive study on BTA19 was composed of 11 granddaughter design families (GDD-11). This set of families was created by combining five families sampled for the confirmation of DD-18 results and six families from GDD-20. The total number of animals was 694, with the family size varying from 22 to 96 (63 on average; Chapter 3.1, Table 2). Since there was not enough DNA for 13 animals, only 681 were actually genotyped. For the final haplotype analysis and linkage disequilibrium and linkage (LDL) analysis these 681 animals were connected along with their sires and mothers through ancestors to FV-ROOT, building a so-called complex pedigree based on GDD-11. After this was filtered for those animals genotyped for 12 to 21 markers, the total number of genotyped animals was reduced to 593. Obtained, filtered pedigree had 1460 animals in total (Chapter 3.1.1).

To ensure haplotype tracing from important founders to sires in DD and/or GDD in the haplotype analysis, we collected additional blood and semen samples from all available male ancestors up to the important founders, building a complex pedigree (FV-ROOT) comprising 69 animals.

6.2 Genome wide scan

6.2.1 Individual genotyping

For the genome wide scan we chose 237 microsatellite markers, distributed over all 29 autosomes, from the public database. Prior to the genome wide scan we genotyped our complex five-generation pedigree (FV-ROOT) for all markers. Individual genotyping results were necessary for the selection of markers for pool analysis in each family, as the "selective DNA pooling" mapping method can be applied only on families for which the sire is heterozygous for a given marker. Also, the individual genotypes of FV-ROOT animals were necessary for the shadow-band correction, as well as for the haplotype analysis.

During the process of genotyping we decided to discard 18 markers. The decision to discard the markers was made because:

- Ten markers gave either strong by-products or inconsistent results under the standard PCR protocol. They were discarded both for individual and pool genotyping.
- Five markers showed irregular allele sizes, meaning that neighbouring dinucleotide repeat alleles were separated by 3 bp, so alleles and shadow bands did not overlap and, as such, were not useful for the shadow correction method. These markers were retained, if appropriate, for the haplotype analysis but were discarded for pool analysis.
- We could establish the presence of null alleles at three markers. According to *Callen et al.* (1993), the occurrence of a null allele is due to the mutation on the complementary sequence of one of the primers, which leads to failure in efficient amplification of an allele. The presence of null alleles was recognised by observing a pattern of allele inheritance in a pedigree. A typical case of a null allele was when a father was homozygous for an allele but his offspring were homozygous for a different allele. After repeated genotyping and exclusion of genotyping errors, the presence of a segregating null allele was determined. The microsatellite allele(s) can be detected by testing new forward and/or reverse primers, as we did for two markers (BM1500 and LMU2402). By making different combinations of old primers with null alleles and newly ordered ones, we were able to ascertain the mutated primer sequence. Markers with null alleles were partly held for the haplotype analysis but discarded for the pool analysis.

6.2.2 Pool genotyping

A total of 582 pools were constructed for all traits in two duplicates from the daughter samples in DD-18. From these, 144 pools were for two main traits, milk yield (MY) and protein percent (PP), also in two duplicates. Pools were genotyped only for markers for which the family sire was heterozygous according to the results of individual genotyping. A total of 219 markers were considered for the genome wide scan for MY and PP in four pools of the first duplicate. Altogether, 17,569 pool genotypes of the first duplicate were produced. The second duplicate was genotyped for the confirmation of the results where the single marker tests showed QTL linkage. In total 4331 confirmation pool genotypes were produced for 83 markers. So, for MY and PP in the first and second duplicate 21,898 pool genotypes were obtained.

During the process of pool genotyping, an additional 10 markers were found to be problematic in certain families. These markers were excluded from pool analysis in affected families but used for analysis in the remaining informative families.

6.3 QTL mapping by "selective DNA pooling"

A total of 21,898 pool genotypes were combined into 4695 single marker tests. The pool genotypes that showed inconsistent patterns between two replicates, i.e. large differences in estimated allele frequencies, were reanalysed and, if necessary, retyped. The retyping was done only if inspection of genotypes indicated a genotyping error. If the cause of the large variance between replicates couldn't be resolved, single marker tests were excluded when they exceeded an arbitrary value of 0.012. On average, 10% of the tests were excluded due to a large variance. In total 3701 tests had a correct variance between the two replicates. Out of 3701 tests, 531 were significant with a probability value of $P<0.05$ and 235 with a probability value of $P>0.01$. The significant results were distributed over all 29 autosomes. By combining adjusted false discovery rate (aFDR) and the approximate interval mapping (AIM) we detected 31 QTL regions distributed across 26 chromosomes.

6.4 Haplotype analysis and identity by descent (IBD) mapping

Haplotype analysis was performed by the SimWALK2 program for all 29 autosomes in the complex pedigree (FV-ROOT). All known pedigree relations were used, in order to establish "strong" haplotypes and to derive the best possible genotypes when genotype information was not available. Haplotype analysis was able to assign the origin of a chromosome region to a Fleckvieh or Red Holstein ancestor. Obtained haplotypes were graphically presented for all chromosomes. The haplotype flow from a founder to an actual generation was much easier to track in the graphics, and recombination sites were easier to observe.

Because of the strong founder influence and the optimal time since the introgression of his alleles into the population, the advanced backcross Fleckvieh x Red Holstein (ABFV) population provides the great opportunity for IBD mapping. If there is an introgression of the QTL alleles from the Red Holstein into the ABFV population they will occur within the specific haplotype context. The inherited haplotype will not be the same in all the offspring, so a simple comparison of the inherited haplotype parts can reveal the common region shared by the offspring. The IBD mapping presumes that the QTL is already mapped in the specific region and the QTL segregating sires are identified.

As mentioned above, all 29 autosomes were scanned for QTL by "selective DNA pooling". Also, all family sires were haplotyped in the context of a complex pedigree for all 29 autosomes. Following the principles of IBD mapping, the results of mapping by "selective DNA pooling" in DD-18 were compared with the results of haplotype analysis in order to detect the QTL regions possibly introgressed by the Red Holstein founder. All markers, found significant by the single marker test and approximate interval mapping (AIM) for QTL affecting PP and/or MY in ABFV families were checked for their origin, determined by haplotype analysis. If the haplotype analysis indicated the RH origin, we investigated these regions closely. All eight ABFV families were mapped by the "selective DNA pooling" method and haplotyped. In the late phase of data evaluation it was noticed that two ABFV families have an unusually high number of significant sire by marker combinations, but they have a high variance between the two pool replicates (>0.0012). Since these two families were sampled in Austria with different sampling logistics, we presumed an error in the sampling process and omitted these families in further analyses. All results presented below were only for six ABFV families sampled in Bavaria.

There were a total of eight QTL regions, which were closely inspected for possible RH introgression: BTA01 proximal, BTA05 distal, BTA09 distal, BTA10 proximal, BTA19 central, BTA23 central to distal, BTA25 central and BTA28 distal. All regions are briefly presented below with significance levels presented by an "adjusted false discovery rate" (aFDR) cut-off of 5% or 10% and AIM thresholds: AIM05 and AIM05* (Chapter 5.4 and 5.5).

BTA01: Approximate interval mapping (AIM) results suggested that there is a single QTL affecting both PP and MY, but with a higher significance for PP. The AIM-statistic maximum was at marker BM1312 (61.85 cM) for both traits but according to the AIM05 threshold, the QTL is not significant (Fig. 5A and B). On the other hand, single marker test results in DD-18 showed significant linkage between marker BM1312 and the QTL affecting both PP (at "aFDR" level of 5%) and MY (at "aFDR" level of 10%). The QTL segregates in one FV and one ABFV family. The haplotype analysis showed that one segregating ABFV family received a RH haplotype for the appropriate QTL region. The haplotype introduced by the RH founder is coupled with a negative effect for PP (- 0.013). There are also three other ABFV family sires, which received the same RH haplotype but didn't show significant or even an indicative result (Fig. 5C). The described distribution of segregating families and haplotypes is consistent with a pure FV origin of the QTL.

BTA05: A highly significant QTL (at "aFDR" level of 5% and AIM05*) affecting PP was detected on chromosome 5, at the marker BM2830 (116.91 cM; Fig.6A). Some families are informative and significant (at "aFDR" level of 5%) for two or three adjacent markers and thus make a large fragment of the chromosome significant. The overall shape of AIM-statistic curves seems to indicate the segregation of two independent QTL: a central one affecting both MY and PP and a distal one affecting only PP. Both QTL segregate both in FV and ABFV families. ABFV families, segregating for the central QTL affecting MY and PP, don't have a common RH haplotype in the possible QTL region, so this QTL segregates independent of the RH haplotype. There are two pure FV families and two ABFV segregating for the distal PP-QTL. Both ABFV families received the RH allele at the most distal marker associated with a negative effect in both families (Fig. 6C). The third ABFV family received the RH haplotype on the distal half of the chromosome, but is not segregating for the PP-QTL. Pure FV families also segregate at the BM2830 bracket. The literature describes the opposite situation, the distal QTL affecting MY and the central

one affecting PP (review by Khatkar et al., 2004). According to the described distribution of segregating families and haplotypes, we conclude that there is no strong support for the introgression of the genetic active QTL variant by the RH founder.

BTA09: The AIM-statistic curve indicated the segregation of two linked QTL for the PP, both significant at AIM05 level (Fig. 7A). The first peak is positioned at approximately 44 cM (marker BMS555) and the second one at approximately 91 cM (marker BM4208). Contrariwise, the family-wise AIM-statistic pointed towards a single QTL. In fact, there are three families (two ABFV and one FV family) with significant effects at both positions and one FV family informative at both positions but significant only at the first position. One of the significant ABFV families received the RH haplotype at marker BM4208. The AIM-statistic curve for the MY indicated the presence of a single QTL with the most likely position at the marker BM7209 (81.56 cM; Fig. 7B). There are two families responsible for the effect on MY, one FV and one ABFV family. The ABFV family was without RH haplotype at the marker. Family-wise AIM-statistics also suggested a distinct QTL affecting MY. Irrespective of the number of segregating QTL, the results of the haplotype analysis on BTA09 supported the FV origin of these QTL (Fig. 7C).

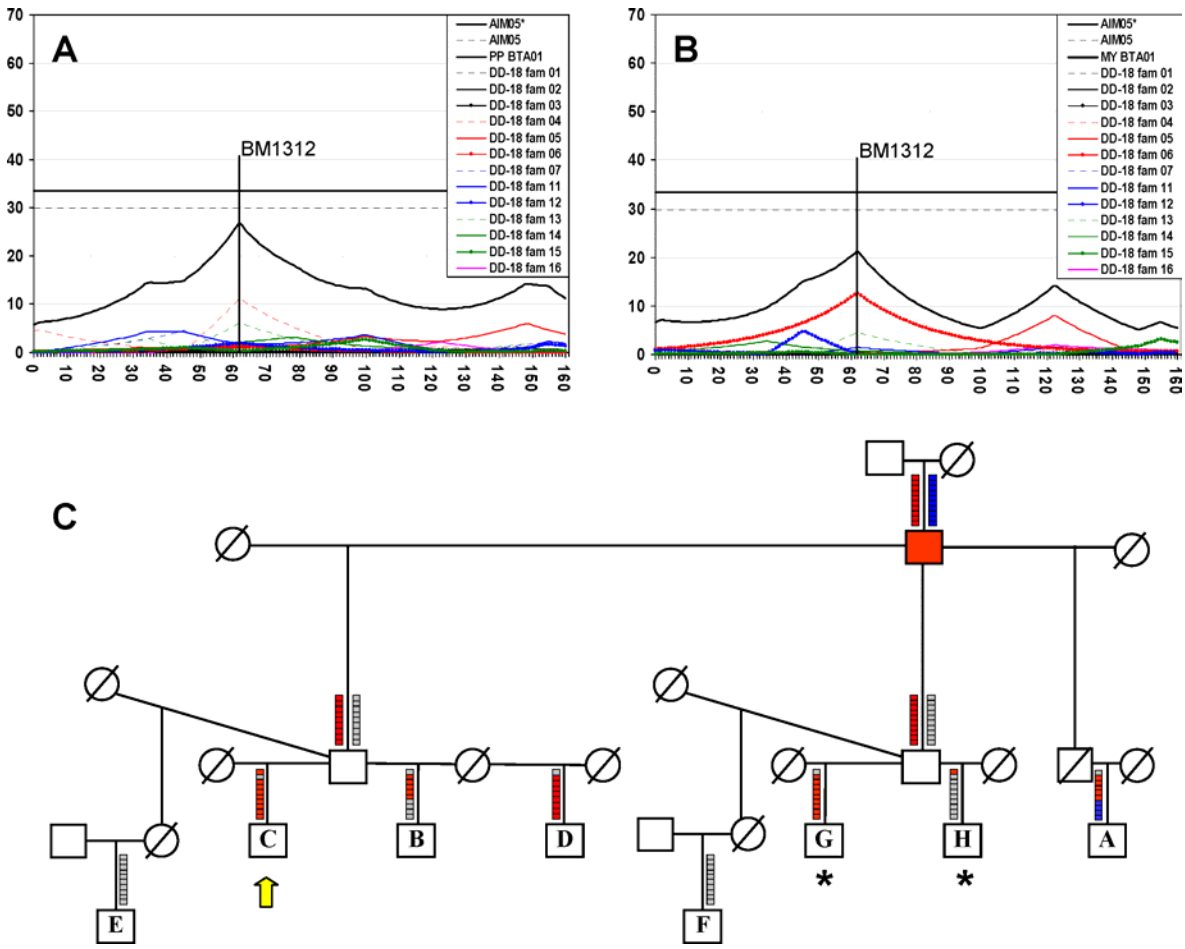

Figure 5: Identity by descent (IBD) mapping for chromosome 1 (**BTA01**). Results from the approximate interval mapping (AIM) for protein percentage (PP; A) and for milk yield (MY; B) are shown. The AIM-statistic maximum for both PP and MY is at marker BM1312 (61,85 cM). The haplotype analysis for BTA01 is shown only for the advanced backcross population (ABFV; C). The important founder "REDAD" is marked red and eight family sires are marked A-H. REDAD's haplotype coming from Red Holstein is marked in red, REDAD's haplotype coming from Fleckvieh is marked in blue and non-REDAD haplotypes are marked in grey. Each square in the haplotype presents one marker used in the analysis. Paternal haplotypes are placed left and maternal right. In the pedigree, squares represent males and circles female animals. Symbols for non-genotyped animals are crossed with a line. One ABFV family, segregating for the quantitative trait locus (QTL) affecting PP and MY, is marked with a yellow arrow. Families excluded from analysis are marked with an asterisk (*).

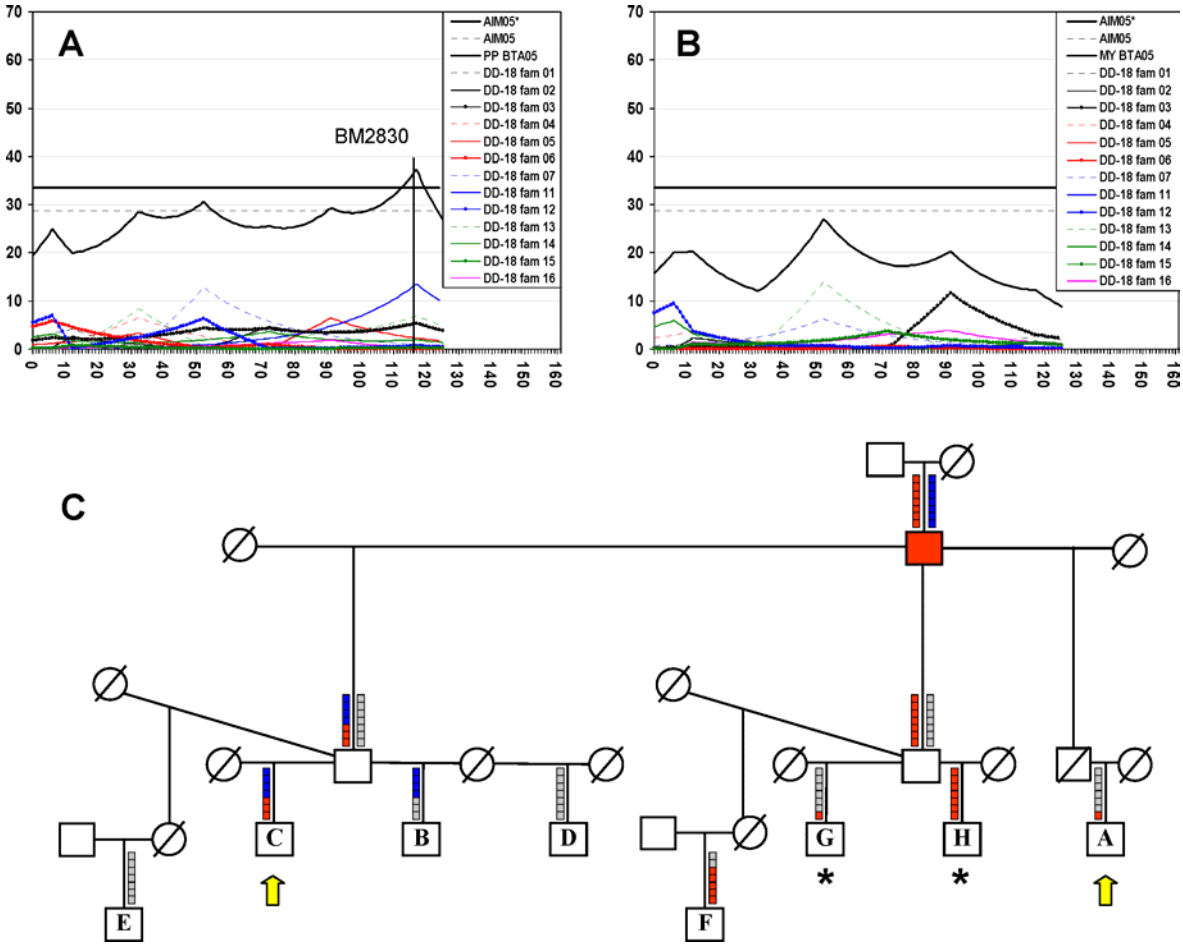

Figure 6: Identity by descent (IBD) mapping for chromosome 5 (**BTA05**). Results from the approximate interval mapping (AIM) for protein percentage (PP; A) and for milk yield (MY; B) are shown. The AIM-statistic maximum for protein percent (PP) is at marker BM2830 (116,91 cM). The haplotype analysis for BTA05 is shown only for the advanced backcross population (ABFV; C). The important founder "REDAD" is marked red and eight family sires are marked A-H. REDAD's haplotype coming from Red Holstein is marked in red, REDAD's haplotype coming from Fleckvieh is marked in blue and non-REDAD haplotypes are marked in grey. Each square in the haplotype presents one marker used in the analysis. Paternal haplotypes are placed left and maternal right. In the pedigree, squares represent males and circles female animals. Symbols for non-genotyped animals are crossed with a line. One ABFV family, segregating for the quantitative trait locus (QTL) affecting PP and MY, is marked with a yellow arrow. Families excluded from analysis are marked with an asterisk (*).

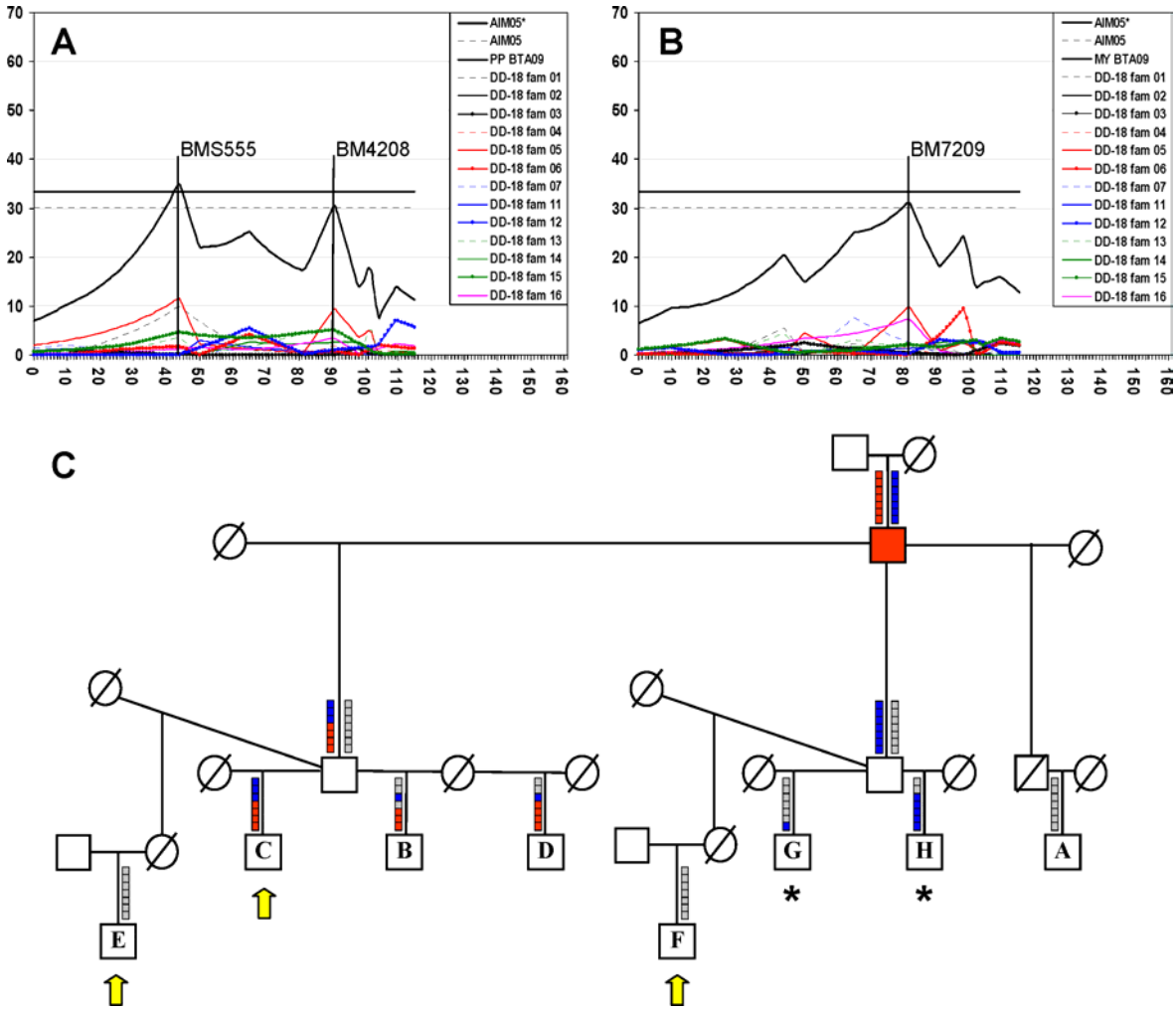

Figure 7: Identity by descent (IBD) mapping for chromosome 09 (**BTA09**). Results from the approximate interval mapping (AIM) for protein percentage (PP; A) and for milk yield (MY; B) are shown. The AIM-statistic curve shows two peaks for protein percent (PP): at app. 44 cM (marker BMS555) and at app. 91cM (marker BM4208). The AIM-statistic curve for MY has its maximum at 82 cM (marker BM7209). The haplotype analysis for BTA09 is shown only for the advanced backcross population (ABFV; C). The important founder "REDAD" is marked red and eight family sires are marked A-H. REDAD's haplotype coming from Red Holstein is marked in red, REDAD's haplotype coming from Fleckvieh is marked in blue and non-REDAD haplotypes are marked in grey. Each square in the haplotype presents one marker used in the analysis. Paternal haplotypes are placed left and maternal right. In the pedigree, squares represent males and circles female animals. Symbols for non-genotyped animals are crossed with a line. One ABFV family, segregating for the quantitative trait locus (QTL) affecting PP and MY, is marked with a yellow arrow. Families excluded from analysis are marked with an asterisk (*).

BTA10: Single marker test results indicated one significant QTL affecting MY and two affecting PP at the proximal half of the chromosome (at "aFDR" level of 5%). More conservative thresholds used by AIM detected no significant QTL on BTA10 (Fig. 8 A and B). The family-wise AIM suggested just one QTL affecting both MY and PP. The positioning of the QTL proximal on the chromosome is primarily due to three heterozygous ABFV families which also share the proximal ~35 cM of haplotype coming from the RH founder (Fig. 8C). The other three not segregating ABFV families received either the distal half of RH haplotype or no RH haplotype at all. Haplotype analysis and mapping by "selective DNA pooling" clearly suggest an introgression of this QTL from the RH population within an interval from 0-35 cM.

BTA19: The AIM analysis on the BTA19 detected a highly significant QTL affecting PP (AIM05*). The most probable position for the QTL lies between 20 and 70 cM with the AIM-statistic maximum on the marker URB44 (39.01 cM; Fig. 9A). Both segregating families are ABFV families and they share the same RH haplotype in this region (Fig. 9C). ABFV families showing no significant or indicative results for the QTL affecting PP are also without RH haplotype at marker URB44. None of the pure FV families shows linkage between PP-QTL and this marker. Besides the two segregating ABFV families, family-wise AIM indicates a possibility for one more segregating family of purely FV origin at the adjacent marker (BM17132). The AIM also detected significant linkage (AIM05) between marker BM17132 (59.20 cM) and the MY (Fig. 9B). Out of two ABFV families segregating for the MY-QTL, one did and the other did not receive the RH haplotype on this marker. According to this mapping result, we selected BTA19 as the most promising candidate for the intensive study of possible introgression of the QTL allele by the RH founder into the FV population.

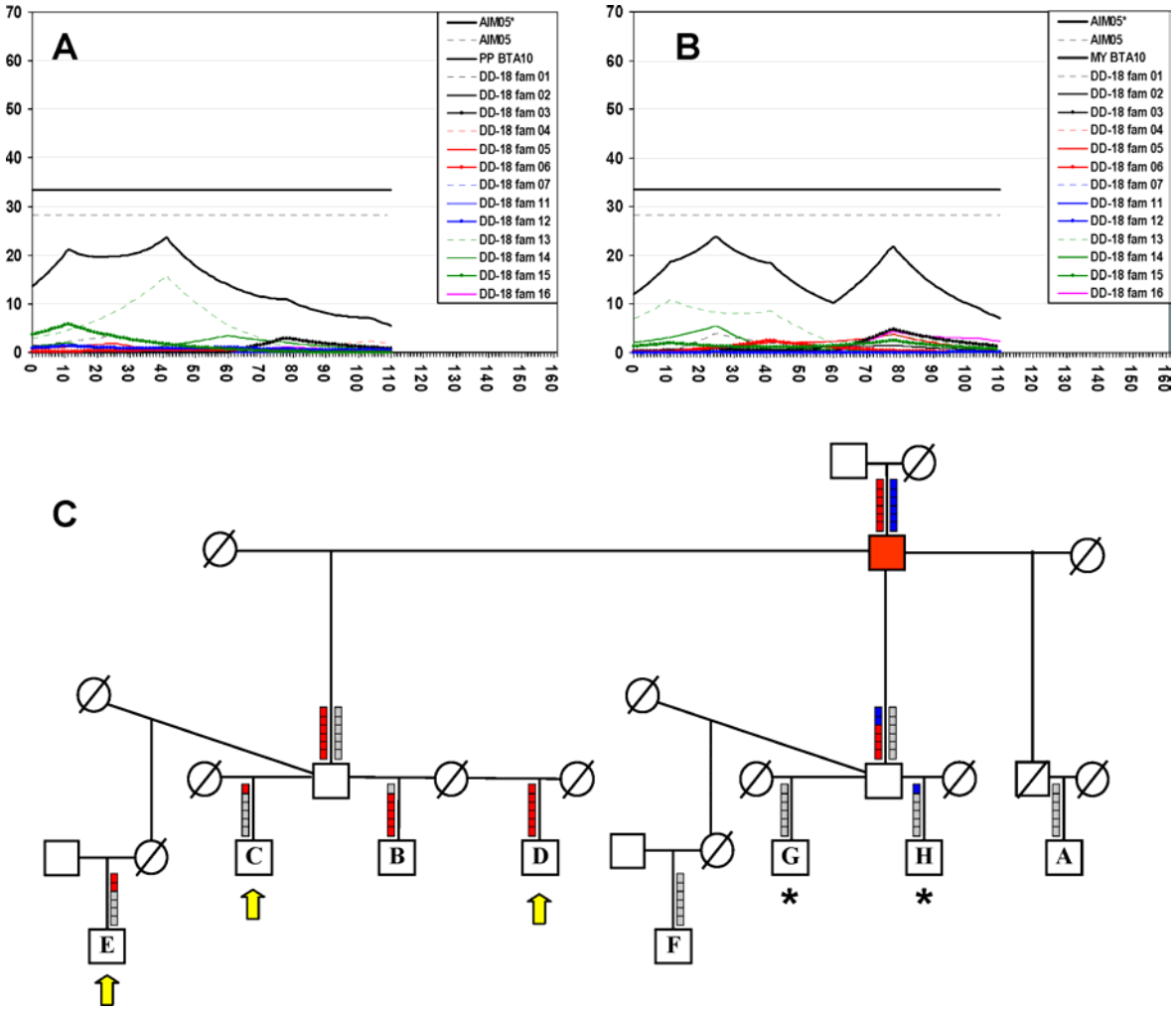

Figure 8: Identity by descent (IBD) mapping for chromosome 10 (**BTA10**). Results from the approximate interval mapping (AIM) for protein percentage (PP; A) and for milk yield (MY; B) are shown. The haplotype analysis for BTA10 is shown only for the advanced backcross population (ABFV; C). The important founder "REDAD" is marked red and eight family sires are marked A-H. REDAD's haplotype coming from Red Holstein is marked in red, REDAD's haplotype coming from Fleckvieh is marked in blue and non-REDAD haplotypes are marked in grey. Each square in the haplotype presents one marker used in the analysis. Paternal haplotypes are placed left and maternal right. In the pedigree, squares represent males and circles female animals. Symbols for non-genotyped animals are crossed with a line. One ABFV family, segregating for the quantitative trait locus (QTL) affecting PP and MY, is marked with a yellow arrow. Families excluded from analysis are marked with an asterisk (*).

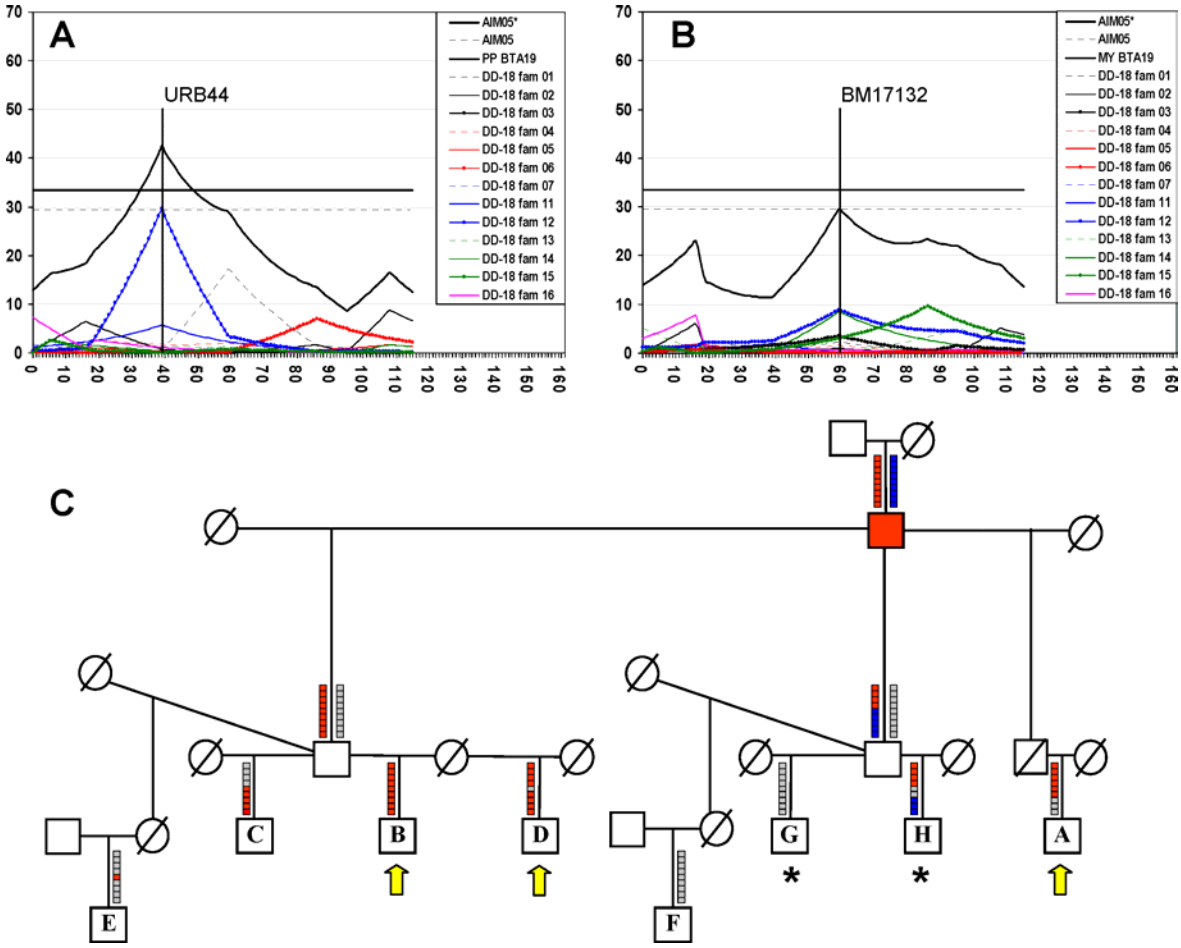

Figure 9: Identity by descent (IBD) mapping for chromosome 19 (**BTA19**). Results from the approximate interval mapping (AIM) for protein percentage (PP; A) and for milk yield (MY; B) are shown. The highest AIM-statistic peak for the PP is on the marker URB44 (39.01 cM). The AIM-statistic curve for MY has its maximum at app. 59 cM (marker BM17132). The haplotype analysis for BTA19 is shown only for the advanced backcross population (ABFV; C). The important founder "REDAD" is marked red and eight family sires are marked A-H. REDAD's haplotype coming from Red Holstein is marked in red, REDAD's haplotype coming from Fleckvieh is marked in blue and non-REDAD haplotypes are marked in grey. Each square in the haplotype presents one marker used in the analysis. Paternal haplotypes are placed left and maternal right. In the pedigree, squares represent males and circles female animals. Symbols for non-genotyped animals are crossed with a line. One ABFV family, segregating for the quantitative trait locus (QTL) affecting PP and MY, is marked with a yellow arrow. Families excluded from analysis are marked with an asterisk (*).

BTA23: The AIM-statistic curve indicates a highly significant QTL (AIM05*) affecting both PP and MY, at chromosomal region 20-60 cM. There is the overall "double peak" shape with one peak at position 24 cM and the other at position 52 cM (Fig. 10A and B). One ABFV and one FV family show the effect for PP at 24 cM. The ABFV family has a RH haplotype on this marker (Fig. 10C). Three FV and one ABFV families segregate for the QTL at 52 cM. This QTL segregates with FV haplotypes. The family-wise AIM suggests that there is a single QTL affecting both traits and mapping most probably central to the distal region of BTA23.

BTA25: The QTL with the highest genome-wide significance for PP according to the AIM approach lies on BTA25 (Fig. 11A). Two highly significant FV families show effects at marker BM737 (31.6 cM). The other two families, one FV and one ABFV, show an effect at BMS1353 (46.44 cM), and they are both not informative for marker BM737. The ABFV family received a complete chromosome from the RH founder. Two not significant ABFV families also received the same haplotype from the RH founder (Fig. 11C). This and the absence of similar QTL in other breeds (review by Khatkar et al., 2004) suggest that the QTL variance at BTA25 is of FV origin.

BTA28: The AIM approach was able to detect a highly significant QTL (AIM05*) affecting PP at the distal half of BTA28 (Fig. 12A). Four families show significant effects for one out of the three most distal markers on BTA28 and no significant effect for MY (Fig. 12B) at the same markers. Two out of four segregating families are ABFV families and share the distal half of BTA28 which comes from the RH founder (Fig. 12C) but have an opposite direction of the QTL effect. The distribution of haplotypes and QTL effects over families suggest rather a FV origin of the genetic active QTL variant.

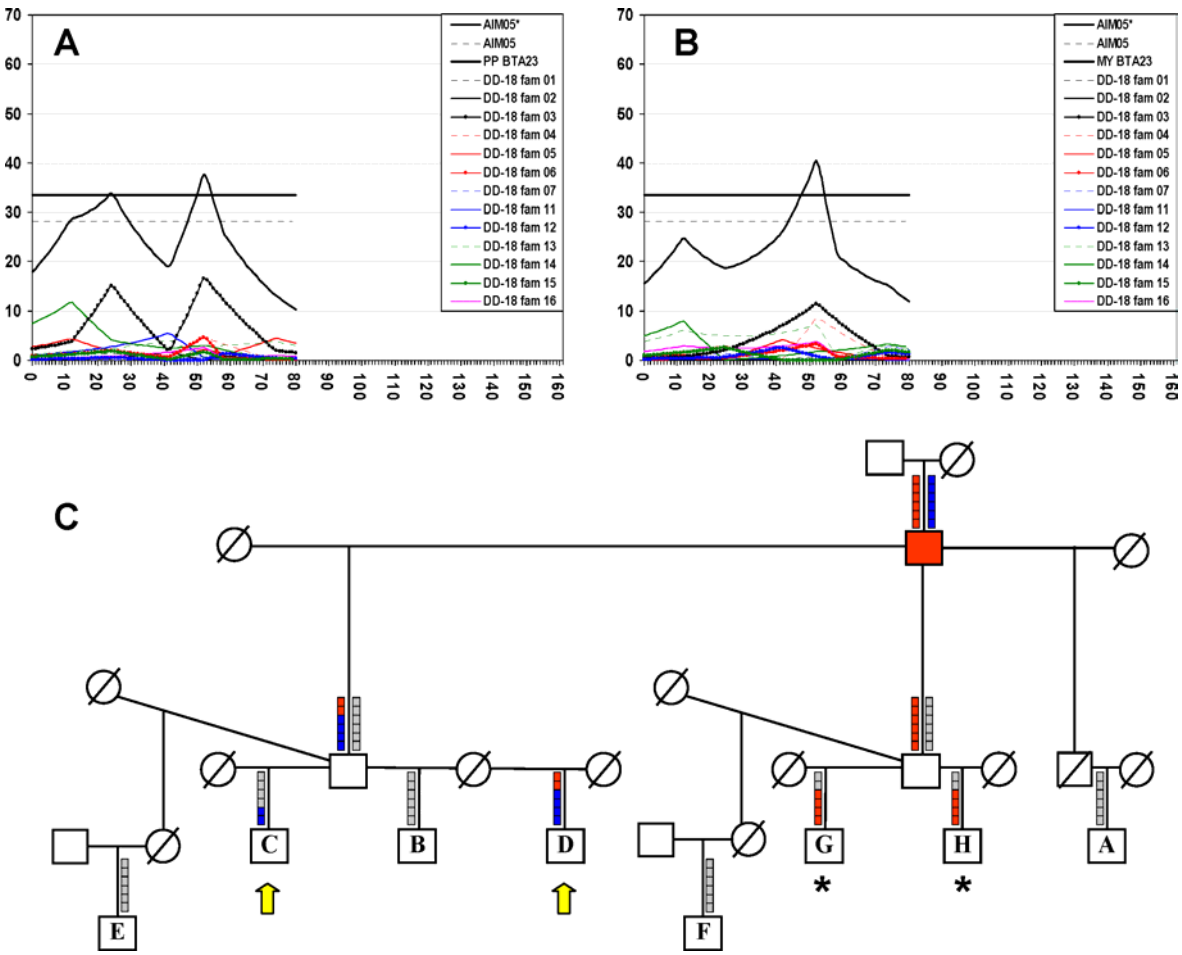

Figure 10: Identity by descent (IBD) mapping for chromosome 23 (**BTA23**). Results from the approximate interval mapping (AIM) for protein percentage (PP; A) and for milk yield (MY; B) are shown. The haplotype analysis for BTA23 is shown only for the advanced backcross population (ABFV; C). The important founder "REDAD" is marked red and eight family sires are marked A-H. REDAD's haplotype coming from Red Holstein is marked in red, REDAD's haplotype coming from Fleckvieh is marked in blue and non-REDAD haplotypes are marked in grey. Each square in the haplotype presents one marker used in the analysis. Paternal haplotypes are placed left and maternal right. In the pedigree, squares represent males and circles female animals. Symbols for non-genotyped animals are crossed with a line. One ABFV family, segregating for the quantitative trait locus (QTL) affecting PP and MY, is marked with a yellow arrow. Families excluded from analysis are marked with an asterisk (*).

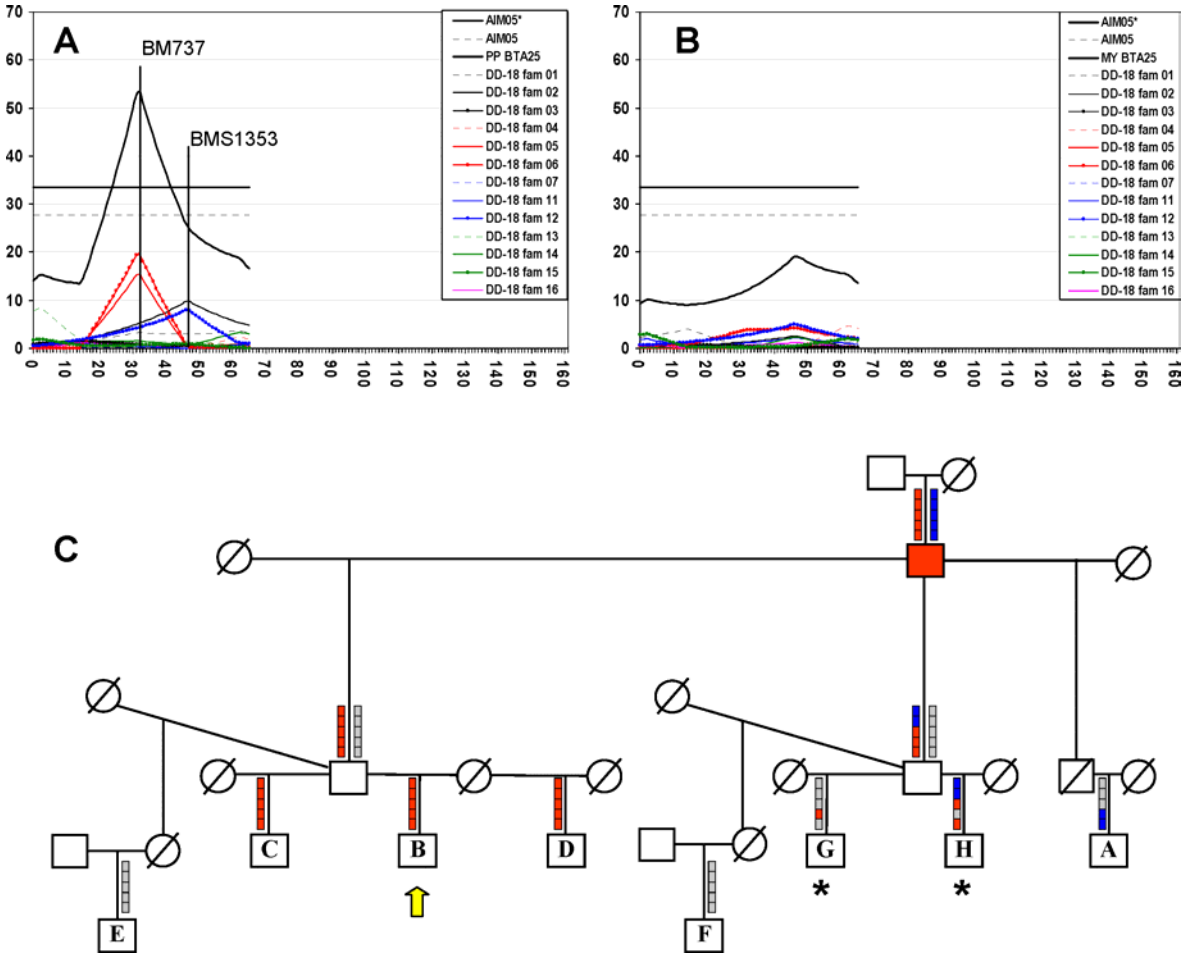

Figure 11: Identity by descent (IBD) mapping for chromosome 25 (**BTA25**). Results from the approximate interval mapping (AIM) for protein percentage (PP; A) and for milk yield (MY; B) are shown. The position of the two most significant markers, BM737 and BMS1353, are also shown (A). The haplotype analysis for BTA25 is shown only for the advanced backcross population (ABFV; C). The important founder "REDAD" is marked red and eight family sires are marked A-H. REDAD's haplotype coming from Red Holstein is marked in red, REDAD's haplotype coming from Fleckvieh is marked in blue and non-REDAD haplotypes are marked in grey. Each square in the haplotype presents one marker used in the analysis. Paternal haplotypes are placed left and maternal right. In the pedigree, squares represent males and circles female animals. Symbols for non-genotyped animals are crossed with a line. One ABFV family, segregating for the quantitative trait locus (QTL) affecting PP and MY, is marked with a yellow arrow. Families excluded from analysis are marked with an asterisk (*).

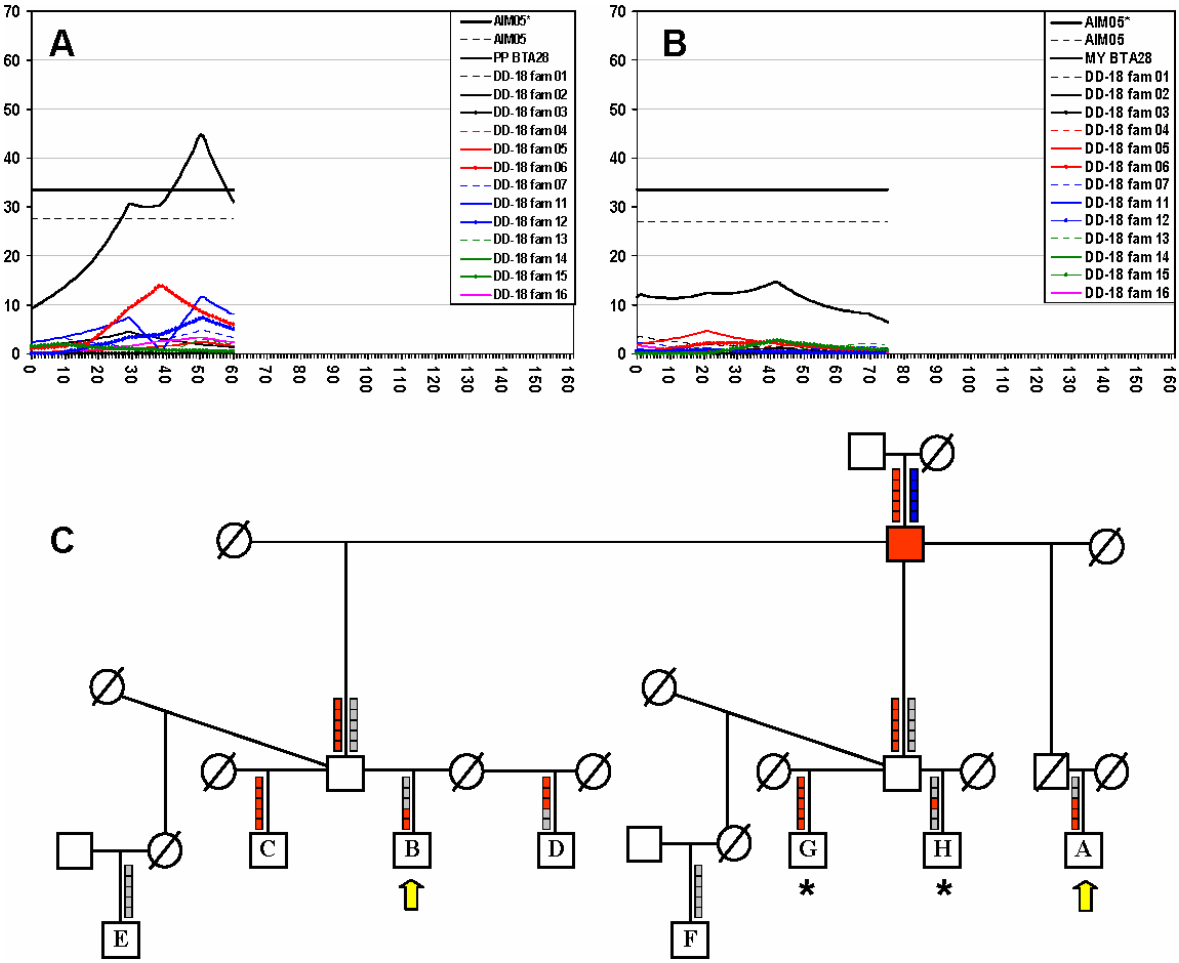

Figure 12: Identity by descent (IBD) mapping for chromosome 28 (**BTA28**). Results from the approximate interval mapping (AIM) for protein percentage (PP; A) and for milk yield (MY; B) are shown. The haplotype analysis for BTA28 is shown only for the advanced backcross population (ABFV; C). The important founder "REDAD" is marked red and eight family sires are marked A-H. REDAD's haplotype coming from Red Holstein is marked in red, REDAD's haplotype coming from Fleckvieh is marked in blue and non-REDAD haplotypes are marked in grey. Each square in the haplotype presents one marker used in the analysis. Paternal haplotypes are placed left and maternal right. In the pedigree, squares represent males and circles female animals. Symbols for non-genotyped animals are crossed with a line. One ABFV family, segregating for the quantitative trait locus (QTL) affecting PP and MY, is marked with a yellow arrow. Families excluded from analysis are marked with an asterisk (*).

As already mentioned, the QTL mapping by "selective DNA pooling" was performed in different stages (Results 6.3). After an initial genome wide scan, BTA19 was chosen for further intensive study as the most promising candidate for the RH introgression. The reason to choose BTA19 for further work included the good concordance between the single marker test and AIM results on the one hand and haplotype analysis for two ABFV families, segregating for the QTL affecting PP, on the other hand. None of the FV families in DD-18 was significant for this marker. The primary goal of this study was to analyse the chromosomal region with possible introgression of a new allele from Red Holstein to Fleckvieh population and not to confirm the QTL already reported in other studies. There are reported QTL affecting PP and/or MY in Holstein populations on BTA01, BTA05, BTA09, BTA10 and BTA23 (review by Khatkar et al., 2004). There is one study on BTA25 in Israeli Holstein (Ron et al., 2004) and two studies on BTA28 in US-Holstein (Rodriguez-Zas et al., 2002; Ashwell et al., 2004), reporting the QTL affecting MY. Further work on one of these chromosomes could mean the confirmation of already published results rather than possibly testing a new QTL region. Only for BTA19 were there no reports on QTL affecting either PP or MY (review by Khatkar et al., 2004). The results of an intensive study on the BTA19 are described below in detail.

6.5 Results on BTA19

6.5.1 Identity by descent mapping

For chromosome 19 we genotyped nine markers in the genome wide scan. The "selective DNA pooling" results in DD-18 suggested a linkage between marker URB44 at 39.01 cM and the QTL affecting the PP (PP-QTL) in only two ABFV families with a negative effect on the PP (-0.017 and -0.00001, respectively). None of the pure Fleckvieh families showed linkage on URB44.

Approximate interval mapping results showed the presence of a highly significant QTL affecting PP (AIM05*). The most probable position of the PP-QTL was estimated to be between 20 and 70 cM, with the highest peak on the marker URB44 at 39.01 cM (Fig. 9A). The evaluations of family-wise AIM-statistic curves showed that two ABFV families (DD-18 fam11 and DD-18 fam12) are contributing to the highest effect on marker URB44. Family-wise AIM also indicates the possibility of one more segregating family of purely FV origin at the adjacent marker (BM17132).

According to the single marker test and AIM results, all the ABFV sires were grouped into two groups. In the first group were two sires showing a significant effect on the PP: DD-18 fam11 (P=0.0168) and DD-18 fam12 (P<0.00001). The second group consisted of four non-significant ABFV sires. Sires' haplotypes were compared for these two groups in order to possibly localise the PP-QTL better (similar to Riquet et al. 1999; Fig. 13). In the first group, one sire received the whole haplotype block from the Red Holstein founder and another sire got the proximal part, including the URB44 and the next distal marker, BM17132. The two significant families shared the same haplotype in the vicinity of URB44. Four non-significant sires, sorted by size of the received RH haplotype, got:

- The proximal and distal part of RH haplotype block but not the central part (URB44 and BM17132)
- Only the distal part, including BM17132
- Short chromosomal fragment in the vicinity of marker URB44, possibly also the next proximal marker
- Didn't receive the RH haplotype at all

If one sire received the Red Holstein haplotype on marker BM17132 and the other one on URB44 but neither is significant for the QTL affecting PP while the significant sires got Red Holstein haplotype on both markers, this could indicate that the QTL lies between these two markers, in a region of approximately 20 cM (Fig. 13). Of course, this inference is correct only under the assumption that all performed analyses are accurate and there are neither false positives nor false negatives. Both PP-QTL segregating families were selected for the intensive study.

6.5.2 Initial interval mapping

Initial interval mapping on BTA19 was conducted by QTL Express in GDD-20 with the data derived from previous mapping studies on the Fleckvieh population. A total of 16 markers were used for the analysis. All families were not genotyped for all the markers (Fig. 14), since all marker genotypes did not come from a single project, but were collected from various projects conducted at our institute. The interval mapping procedure by QTL Express and similar linear regression based programs are able to combine different families genotyped for different marker sets into one across-family analysis. Corrected estimated

breeding values (cBV) were used as phenotypes. Initial interval mapping provided indications that there might be two QTL on BTA19, one QTL affecting PP at approximately 55 cM and a second one affecting MY and PY at approximately 102 cM. Both had an F-statistic score between 2 and 3 (Fig. 14).

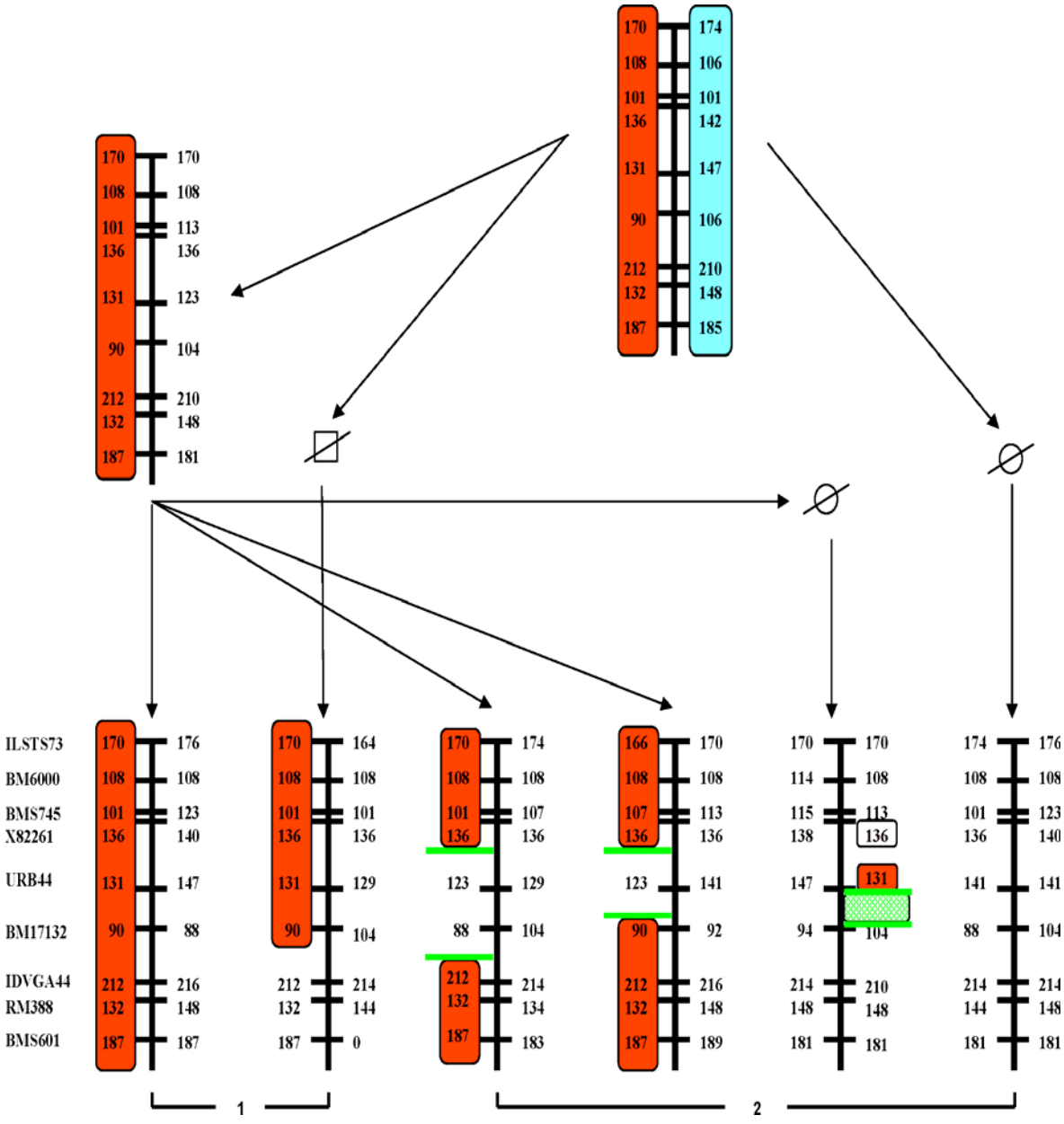

Figure 13: Identity by descent (IBD) analysis in an advanced backcross Fleckvieh x Red Holstein (ABFV) population for chromosome 19 (BTA19) in detail. Two family sires heterozygous for the quantitative trait locus affecting protein percent (PP-QTL; 1) and non significant family sires for PP-QTL (2) are shown. REDAD's haplotype coming from Red Holstein is marked in red and coming from Fleckvieh in blue. A possible location of PP-QTL is marked in green. In the pedigree, squares represent male and circles female animals. Signs for non-genotyped animals are crossed with a line.

Family-wise analyses were conducted in order to determine the QTL status for each family (similar to Schnabel et al., 2005). Altogether, six families were heterozygous for QTL affecting PP (PP-QTL). There were also six families heterozygous for the QTL affecting MY and PY but there was just one family heterozygous for both QTL. Out of the six PP-QTL segregating families three families were ABFV families and three sires were pure FV sires. Surprisingly, two of three ABFV sires didn't get the Red Holstein haplotype at all. Results from initial interval mapping indicate, together with the family-wise AIM results, the possibility this PP-QTL is already present in FV. To test this possibility further, all six families segregating for PP-QTL were included in the set for the intensive study.

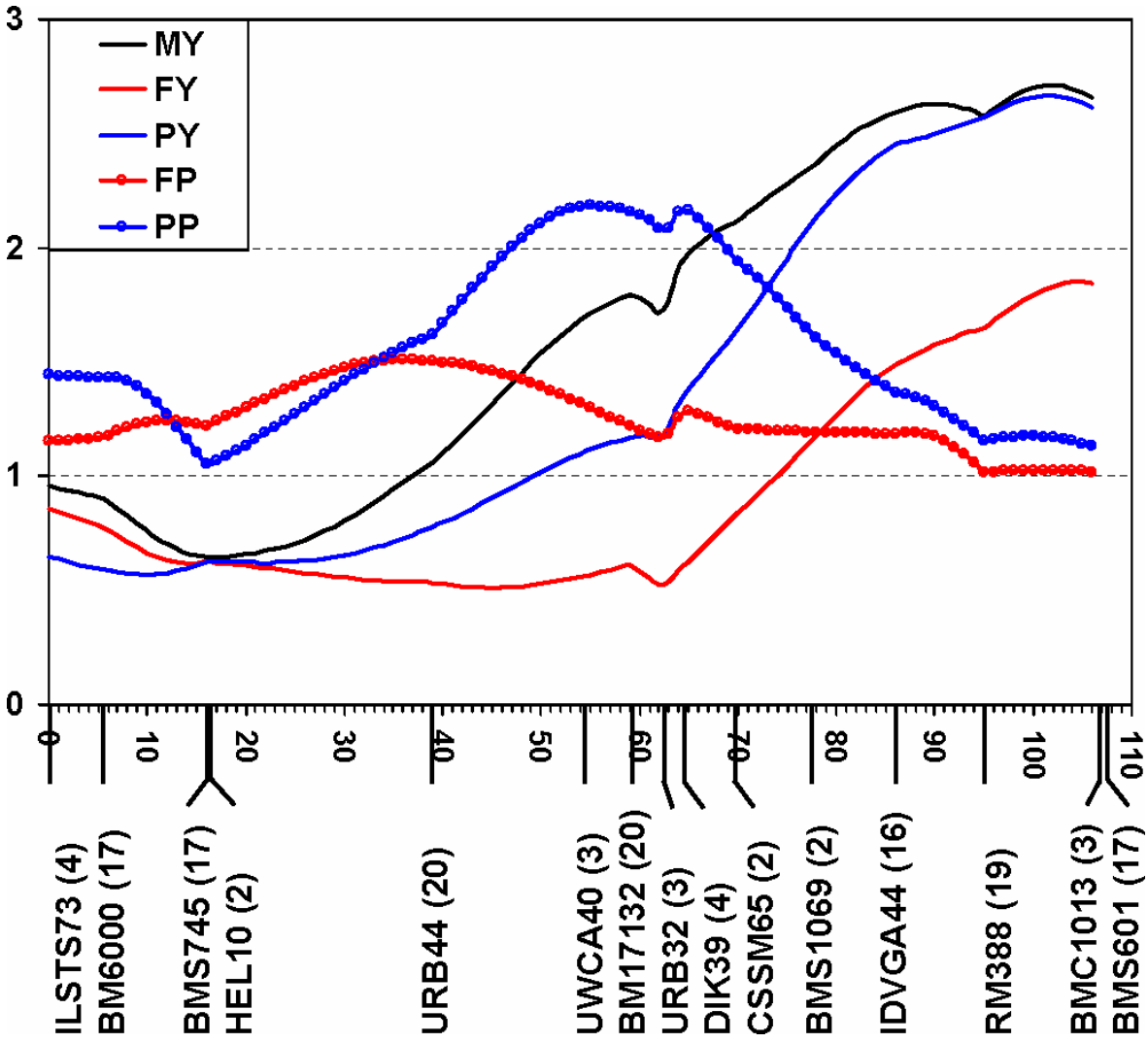

Figure 14: Initial interval mapping in granddaughter design (GDD-20). Interval mapping results for milk yield (MY), fat yield (FY), protein yield (PY), fat percent (FP) and protein percent (PP) are shown. Positions of markers used in the analysis are denoted on the X-axis, and the numbers of families, genotyped for each marker, is denoted. F-statistic values are presented on the Y-axis, and chromosome length is in centiMorgans (cM) on the X-axis.

6.6 Intensive study on BTA19

6.6.1 Set of animals for intensive study

Two families, according to the results of "selective DNA pooling", and six families, according to results from initial interval mapping, were chosen as the set of families for the intensive study. An additional three families, two closely related and one unrelated were chosen for the set to contribute to fine mapping by maternal haplotypes. These families were non-significant in the "selective DNA pooling" but are very important for the genetic active Fleckvieh population. In total, 11 families were chosen for the intensive study. The chosen design was a granddaughter design (GDD-11; Chapter 3.1, Table 2). GDD-11 consisted of a total of 681 animals, genotyped for 21 markers. These animals, together with their sires and mothers, were connected through ancestors to the FV-ROOT, thus building a complex pedigree based on GDD-11. All animals which have been genotyped for 12 to 21 markers were filtered out, yielding a total of 593 animals. This created a complex pedigree with a total of 1460 animals (Chapter 3.1.1).

6.6.2 The set of markers used for the intensive study (set-1)

The first set (set-1) consisted of 12 markers (Chapter 3.3, Fig. 4; Chapter 6.6.6, Fig. 17). Seven markers were chosen on BTA19 in the region of a possible location of the QTL affecting the PP. The QTL location was delimited by IBD mapping to a region of approximately 20 cM, between the markers URB44 and BM17132. Five of seven markers were located between URB44 and BM17132 and additional two are flanking this bracket. Three markers already used in the genome wide scan (X82261, URB44 and BM17132) were included in the set, adding up to a total of 10 markers. This set was used for genotyping the GDD-11. All markers were chosen with an intention to be used in multiplex PCR, so they were combined to fit with each other according to the fragment length and fluorescent dye.

Seven of the GDD-11 families were already genotyped for marker RM388 and one family for marker ILSTS73. The statistical programs used were able to combine different families genotyped for different marker sets into one across-family analysis (QTL Express) and reconstruct genotypes where possible or treat them simply as missing genotypes (SimWALK2, CRI-MAP, LDL_rams). Hence, set-1 for the analysis of the data included 12 markers (Chapter 3.3, Fig. 4; Chapter 6.6.6, Fig. 17).

6.6.3 First interval mapping in GDD-11

Interval mapping in GDD-11 was performed with the QTL Express program, with corrected estimated breeding values (cBV) as phenotypes. The set one with a total of 12 markers was used for the analysis (Chapter 3.3, Fig. 4; Chapter 6.6.6, Fig. 17). Interval mapping results across families confirmed the presence of the QTL affecting PP at the approximate position of 55 cM and the F-statistic test from 2.67 (Fig. 15). A chromosome-wise significance threshold was calculated on 10,000 permutations, and the QTL effecting PP was highly significant (P=0.0082). The bootstrapping option on 10,000 iterations placed the 95% confidence interval of the QTL position between 0 and 63 cM (Fig. 15). We observed that the bootstrapping procedure implemented in the QTL Express program often produces an extra peak on the beginning and/or the end of the chromosome. These peaks are most likely method artifacts, and they were ignored in this analysis. Therefore, the interval between 18 an 60 cM was considered to be the 95% confidence interval and the interval between 46 and 60 cM to be the interval with the best results. Again, the family-wise analyses were conducted in order to determine the QTL status for each family. Out of the eight families previously selected for GDD-11 because they were heterozygous for the PP-QTL, seven families confirmed the results. Three had a highly significant QTL effect ($P<0.01$) and another four had indicative results. Only one ABFV family, previously mapped for PP-QTL, didn't confirm the QTL effect. This ABFV family was marked as segregating for the QTL according to the results of the interval mapping in GDD-20, but it didn't share the Red Holstein haplotype. The results of the interval mapping in GDD-11 point towards possible genotyping errors in this family, leading to the false positive QTL detection in previous projects.

6.6.4 Determining QTL-marker phase and marker haplotypes

For determining QTL-marker phase and marker haplotypes the SimWALK2 program was used. The analysis was performed in FV-ROOT for 12 markers. Once more, haplotypes of the PP-QTL segregating ABFV families could be traced to the common Red Holstein founder. ABFV families that didn't demonstrate a PP-QTL effect through interval mapping in GDD-11 have received a completely different haplotype or the haplotyped differs in the central part of the chromosome. The region for a possible QTL location could be narrowed to approximately 16 cM (Fig. 16) through further comparison of sires' haplotypes between

the group with and group without a PP-QTL effect. For the segregating families with a BTA19 of FV origin, haplotypes could neither be traced to the common founder nor alleles confirmed as identical by state.

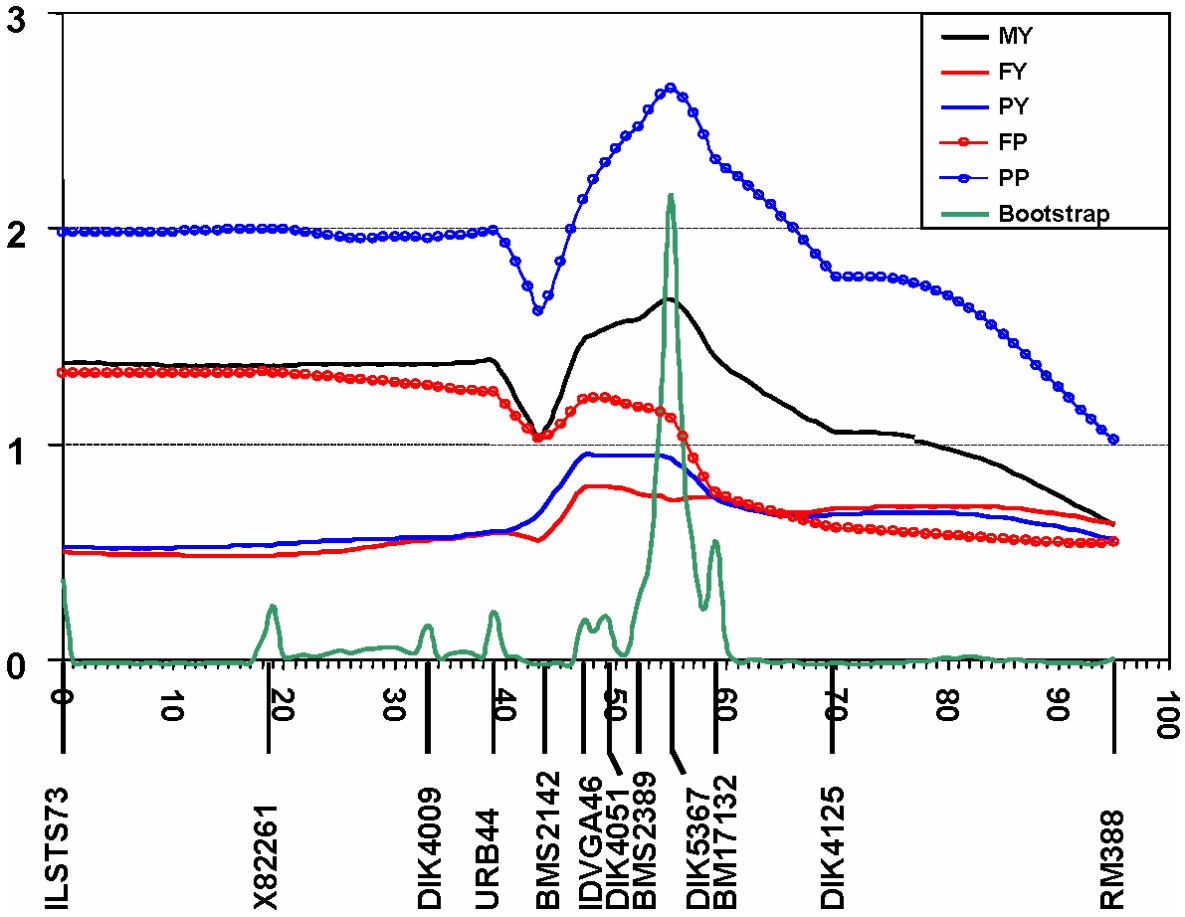

Figure 15: First interval mapping in granddaughter design (GDD-11). Interval mapping results for milk yield (MY), fat yield (FY), protein yield (PY), fat percent (FP), protein percent (PP) and the results of bootstrapping procedure for all families are shown. The number of bootstrapping samples has been rescaled. Positions of markers used in the analysis are denoted on the X-axis. F-statistic values are presented on the Y-axis and chromosome length in centiMorgans (cM) on the X-axis.

6.6.5 Second marker set (set-2)

In order to refine the PP-QTL position, a second marker set, composed of 12 markers, was chosen to better cover the candidate region (Chapter 3.3, Fig. 4; Chapter 6.6.6, Fig.17). All markers were chosen with the intention to be used in multiplex PCR. During the genotyping process two markers were discarded: DIK4688 was rejected because of inconsistent results under the standard multiplex PCR protocol and DIK5098 was rejected because of inconsistent results and the possible presence of a null allele. During the quality

control of the genotypes one further marker (UW33) was eliminated from the analysis due to the excessive number of recombinations it produced. Finally, the set-2 used in analysis had nine markers.

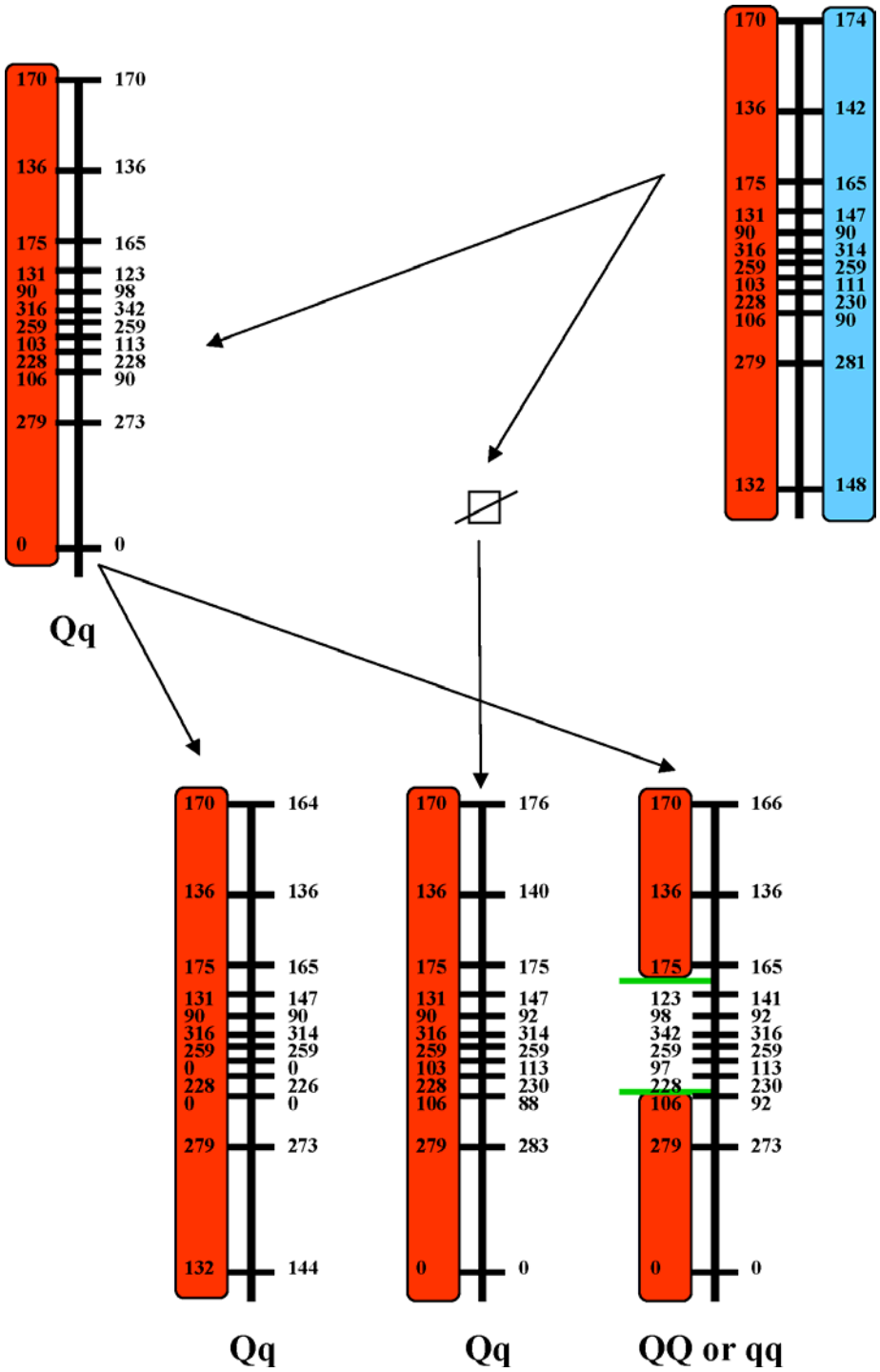

Figure 16: Identity by descent analysis in advanced backcross Fleckvieh x Red Holstein (ABFV) families of granddaughter design (GDD-11) for chromosome 19 (BTA19). Comparison of haplotypes only from sires who received the Red Holstein haplotype (marked in red) narrowed down the QTL region to 16 cM. The haplotype marked in blue comes from Fleckvieh. The QTL status of animals is shown: three sires heterozygous (Qq) for the quantitative trait locus affecting protein percent (PP-QTL) and sire with unknown PP-QTL status (QQ or qq).

6.6.6 The linkage map construction

As already mentioned, the marker positions on the linkage map were taken from the publicly available USDA linkage map (http://www.marc.usda.gov/genome/genome). Two different strategies were considered to separate the markers that were at the same position on the USDA linkage map and to confirm the orders of the used markers on BTA19. The first strategy included a comparison of results from the published linkage map, high-resolution radiation hybrid map (Itoh et al., 2005; Everts-van der Wind et al., 2005) and the whole genome shotgun sequence (WGS) results for a corresponding chromosome. The second strategy included the application of the build option of the CRI-MAP program. The following conclusions have been made:

- The marker TEXAN12 (43.81 cM) is located behind the marker BMS2142 (43.31 cM) on the USDA linkage map. But according to WGS results, the position of the marker TEXAN12 should be in front of BMS2142. This result was confirmed by applying the build option of the CRI-MAP program to our own data. Therefore, for the future analysis we exchanged the positions of TEXAN12 and BMS2142 and used the map with TEXAN12 at 43.31 cM and BMS2142 at 43.81 cM.
- The marker ILSTS014 and the marker DIK4051 are both at the same position (48.53 cM) on the USDA linkage map. They were parted according to the WGS results, which placed the marker ILSTS014 in front of DIK4051. Finally, ILSTS014 was at 48.53 cM and DIK4051 at 49.30 cM.
- The marker DIK5188 and the marker DIK5367 are both at 55.19 cM on the USDA linkage map. This pair of markers was parted according to the results of high-resolution radiation hybrid mapping. On our map, marker DIK5188 was at the position of 55.19 cM and DIK5367 at 55.27 cM.
- The marker BMS17132 and the marker DIK2486 are both located at 59.20 cM on the USDA linkage map. They were parted according to WGS results, with marker BMS17132 at 59.20 cM and DIK2722 at 59.22 cM.

The marker pairs located at the same position on the linkage map were also less informative in our data and their positions could not be ascertained by the build option of the CRI-MAP program. During the decision-making process, the results of high-resolution radiation hybrid mapping were favoured over the WGS sequence results. The WGS results were used when there was no information on the radiation hybrid map or there was

information for only one marker for the marker pair in question. The distance in cM between two markers was calculated by counting the base pairs from all sequence blocks which we could safely localise between two markers in the WGS sequence and dividing the resulting number by the number of base pairs in one cM. The number of base pairs in one cM was calculated from the total number of base pairs (in WGS sequence) divided by the total length of chromosome 19 in cM. On the BTA19 one cM corresponds to approximately 0.5 million base pairs. The marker map used in the subsequent analysis is shown in figure 17.

6.6.7 Final interval mapping

The final interval mapping in GDD-11 was performed with a total of 21 markers (set-1 and set-2) with the QTL Express program, using the corrected estimated breeding values (cBV) as phenotypes. The interval mapping results across the families confirmed the presence of the QTL affecting PP at the approximate position of 55 cM and the F-statistic test from 2.37 (Fig. 18). A chromosome-wise significance threshold was calculated on 10,000 permutations, and the QTL with effect on PP was significant (P=0.025). The bootstrapping option on 10,000 iterations was used to determine confidence interval. The 95% confidence interval of the QTL position was placed in a broad range from 0 to 95 cM. As discussed in Chapter 6.3.3, we observed that the bootstrapping procedure implemented in the QTL Express program often produces an extra peak on the beginning and/or the end of the chromosome, most likely the result of method artifacts. This is due to the relatively high proportion of bootstrap samples right at the beginning and on the very end of the investigated region. Ignoring these peaks, the 95% confidence interval was placed in the interval between 20 and 65 cM, and the interval between 54 and 62 cM was the interval with the best results (Fig. 18). Family-wise analysis showed similar results to those from previous interval mapping, with five families significant or indicative for the PP-QTL. Two previously indicative families showed only very weak effect. The control families were without effect. One ABFV family, chosen according to the initial interval mapping results but without RH haplotype, was without effect, confirming, once more, that it was genotyping error that led to a false positive result and the selection of this family for intensive study. Even though we used a closely spaced marker map, the PP-QTL position couldn’t be refined by interval mapping. The reason for this lies in the fact that there are

only a few informative recombinations between closely spaced markers (Olsen et al., 2004).

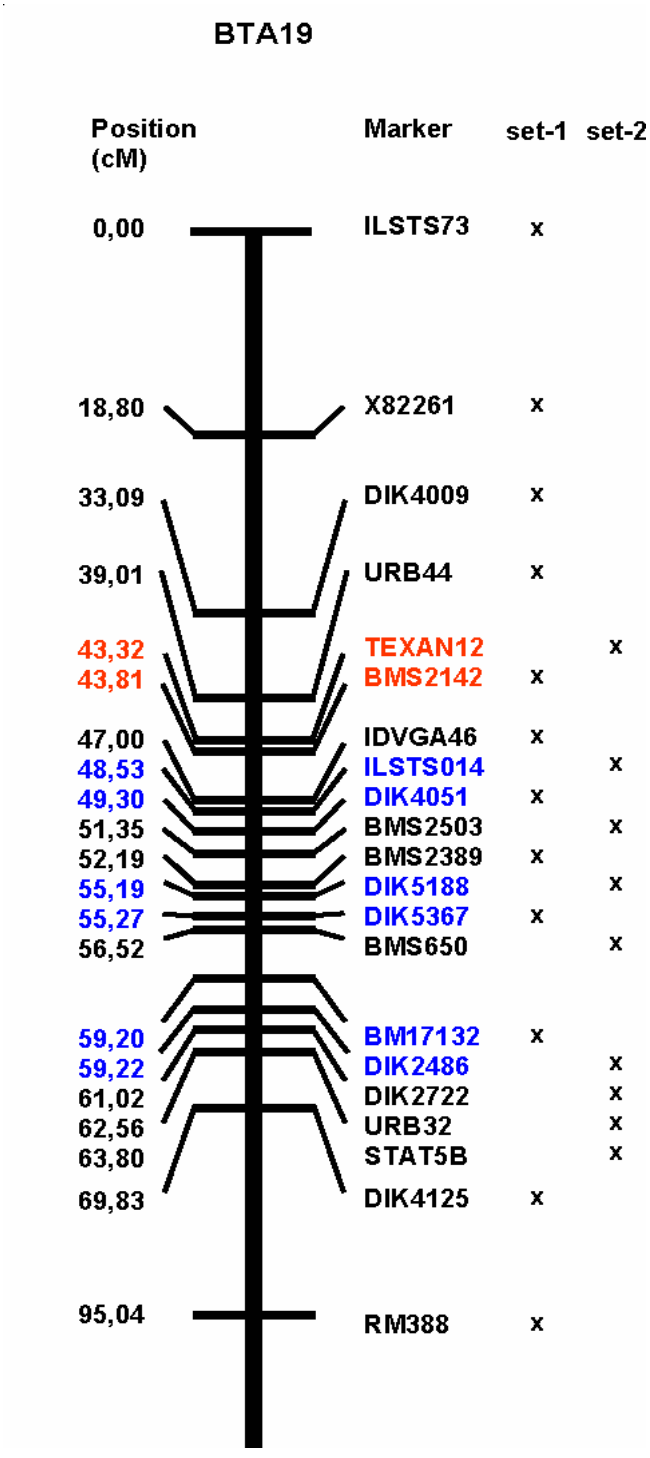

Figure 17: Actual marker map used in all subsequent analysis in GDD-11. All markers are shown here with positions assessed by the linkage map construction strategies. The marker pair highlighted in red swapped their position on our map as compared to the position on the USDA linkage map. Marker pairs highlighted in blue share the same position on the USDA linkage map.

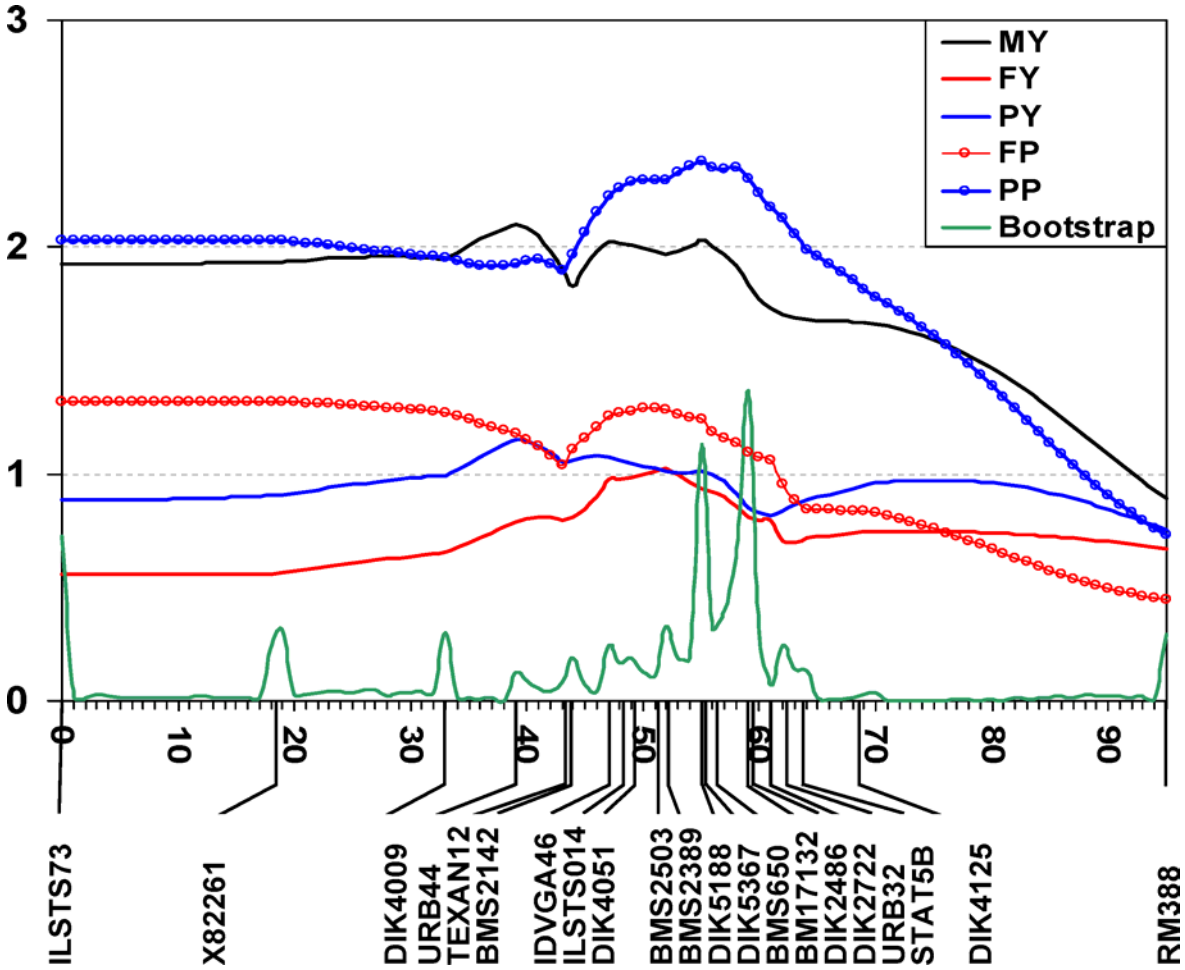

Figure 18: Final interval mapping in granddaughter design (GDD-11). Interval mapping results for milk yield (MY), fat yield (FY), protein yield (PY), fat percent (FP), protein percent (PP) and the results of the bootstrapping procedure for all families are shown. The number of bootstrap samples has been rescaled. Positions of markers used in analysis are denoted on the X-axis. F-statistic values are presented on the Y-axis and chromosome length is in centiMorgans (cM) on the X-axis.

6.6.8 Final haplotype analysis

Final haplotype analysis was done in a complex pedigree based on GDD-11, using 21 markers in order to obtain haplotypes for the combined LDL mapping. Since this pedigree includes 1460 animals from which 593 were genotyped for 12 to 21 markers, the SimWALK2 was not able to perform the haplotype analysis in a reasonable computing time. Therefore, it was necessary to divide the pedigree into five overlapping sub-pedigrees: four sub-pedigrees with two families and one with three families (Chapter 5.3). Overlapping part of all sub-pedigrees was FV-ROOT, which served, at the same time, as the control for the haplotype analysis. All haplotypes were summarised into one table by the database application. They were used as reconstructed haplotypes for LDL analysis.

6.6.9 Combined linkage disequilibrium and linkage analysis

Two programs were used for the combined linkage disequilibrium and linkage (LDL) analysis: the LDL_rams1.76 program and the LDL1.42 program. Both were kindly provided by Mr. Lee and Mr. van der Werf. (School of rural science and agriculture, University of New England, Australia). The LDL_rams program uses unordered genotypes for the analysis, and the LDL program requires reconstructed haplotypes. For the analyses the mutation age and past effective population size were held at 100. Initial homozygosity on each locus was 0.25 (Chapter 5.8). Both analyses were performed using corrected breeding values as the phenotypes.

6.6.9.1 Combined linkage disequilibrium and linkage analysis with the random walk approach and meiosis Gibbs sampling – LDL_rams

The analysis was performed in complex pedigree based on GDD-11 with 21 markers. Because the pedigree is composed of 1460 animals, out of which 593 animals are genotyped for 12 to 21 markers, it was not possible to perform the analysis with a high number of iterations. LDL_rams was started with 1100 iterations and an initial burn-in of 100. Parameter estimates were collected every 10^{th} round. The LDL_rams analysis locates the PP-QTL at position 53.69 cM and a log-likelihood ratio test ($LRT=-2(\log(L_0)-\log(L_p))$ value of 8.46 (Fig. 19). According to *Olsen et al.* (2004) the significance level of the LRT value is chi-square distributed with one degree of freedom. Assuming this probability distribution, PP-QTL is highly significant (P=0.0036).

To calculate the confidence interval of the QTL position, the LOD score was calculated as LRT divided by ln 10 (Olsen et al., 2005). The LOD-score was 1.8 at the LRT maximum. Using 1-LOD drop-off criteria (Lander and Botstein, 1989), we calculated the confidence interval for the QTL position (Fig. 19). The confidence interval calculated by 1-LOD drop-off lies between approximately 51 and 55 cM. The LDL_rams analysis was not only able to confirm the PP-QTL at 53.69 cM, it was also able to considerably refine the PP-QTL position with a confidence interval in the region of 4 cM.

The second peak at approximately 67 cM is most probably the consequence of insufficient iteration steps and an unstable **GRM** matrix. In the simulation studies (Lee and van der Werf, 2005), it is shown that the MCMC sampling incorporated in the LDL_rams program

locates the QTL at the correct position and gives very similar results to the results obtained with "true" haplotypes. The simulation was performed in a complex pedigree, with basis population of 100, 200 and 800 animals genotyped for 10 markers. The Gibbs sampler carried out 5000 samplings (discarding the first 1000), and **GRM** elements were sampled every 10^{th} round (Lee and van der Werf, 2005). The MCMC algorithm linearly scales the number of possible haplotypes with an increase in the number of markers and size of the mapping population, as well as the underlying effective size of the mapping population (Sobel and Lange, 1996). All mentioned parameters – marker number (21), mapping population (1460) and effective population size (250-320; Pirchner, 2002) are higher than the parameters used in the simulation study by *Lee and van der Werf* (2005). Therefore, our design needed a large number of samplings to obtain good convergence of the IBD probabilities (e.g. 10,000), which could not be attained with our computing capacity (Linux operating system, 3.6 GHz, 4096 MB RAM, 73 GB hard drive) in a reasonable amount of time. This problem should be overcome with the use of "true" or rather "most likely" haplotypes. Externally produced haplotypes can be used by the LDL program, which means we can run the analysis with the best possible haplotypes. In addition, the demanding computational time of LDL_rams could be avoided and alternative analysis could be performed.

Also, according to *Olsen et al.* (2005), a large test statistic in the surrounding area, as on the second peak at 67 cM, could be due to the presence of an additional QTL, or alternatively to carryover effects. To test the possibility of a second QTL we would need a program for two QTL mapping that is based on the LDL method. The presence of the carryover effect, i.e. effect of one large QTL on neighbouring intervals, could be the consequence of a lack of information in this area i.e. a sparse marker map (Fig. 19).

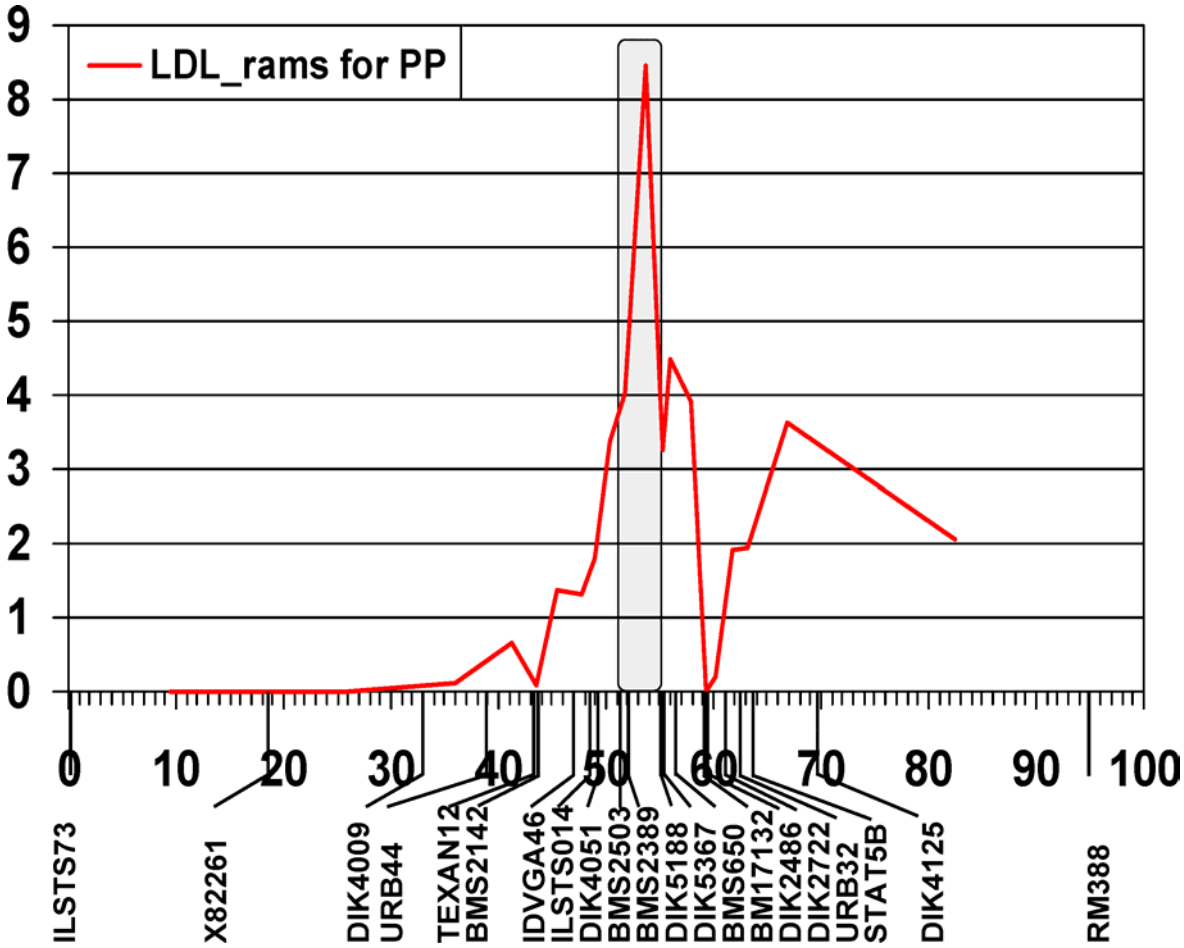

Figure 19: Results of a combined linkage disequilibrium and linkage analysis (LDL) by the LDL_rams program for a quantitative trait locus affecting the protein percent (PP-QTL). Positions of markers used in analysis are denoted on the X-axis. The 1-LOD drop-off confidence interval for PP-QTL is marked in grey. The log-likelihood ratio test values (LRT) are presented on the Y-axis, and the chromosome length is in cM on the X-axis.

6.6.9.2 Combined linkage disequilibrium and linkage analysis - LDL

Haplotypes prepared by SimWALK2 for five sub-pedigrees were controlled and summarised in one table by the database application. They were used as reconstructed haplotypes for the LDL analysis. The same pedigree like in the analysis by LDL_rams program, consisting of 1460 animals, was used. The LDL analysis located the PP-QTL at the same position as the LDL_rams analysis (53.69 cM). The LRT value is 11.63, and the PP-QTL is highly significant at this position (P=0.0006). The second peak at 55.89 cM, which is also derived by LDL_rams analysis, is now more prominent and highly significant as well. The LRT value for this peak is 10.38, and the corresponding P value is 0.0013. The calculated LOD-score is 2.52 at position 53.69 cM and 2.25 at position 55.89 cM.

Because of the second peak, the 1-LOD drop-off confidence interval now lies between approximately 50 and 59 cM.

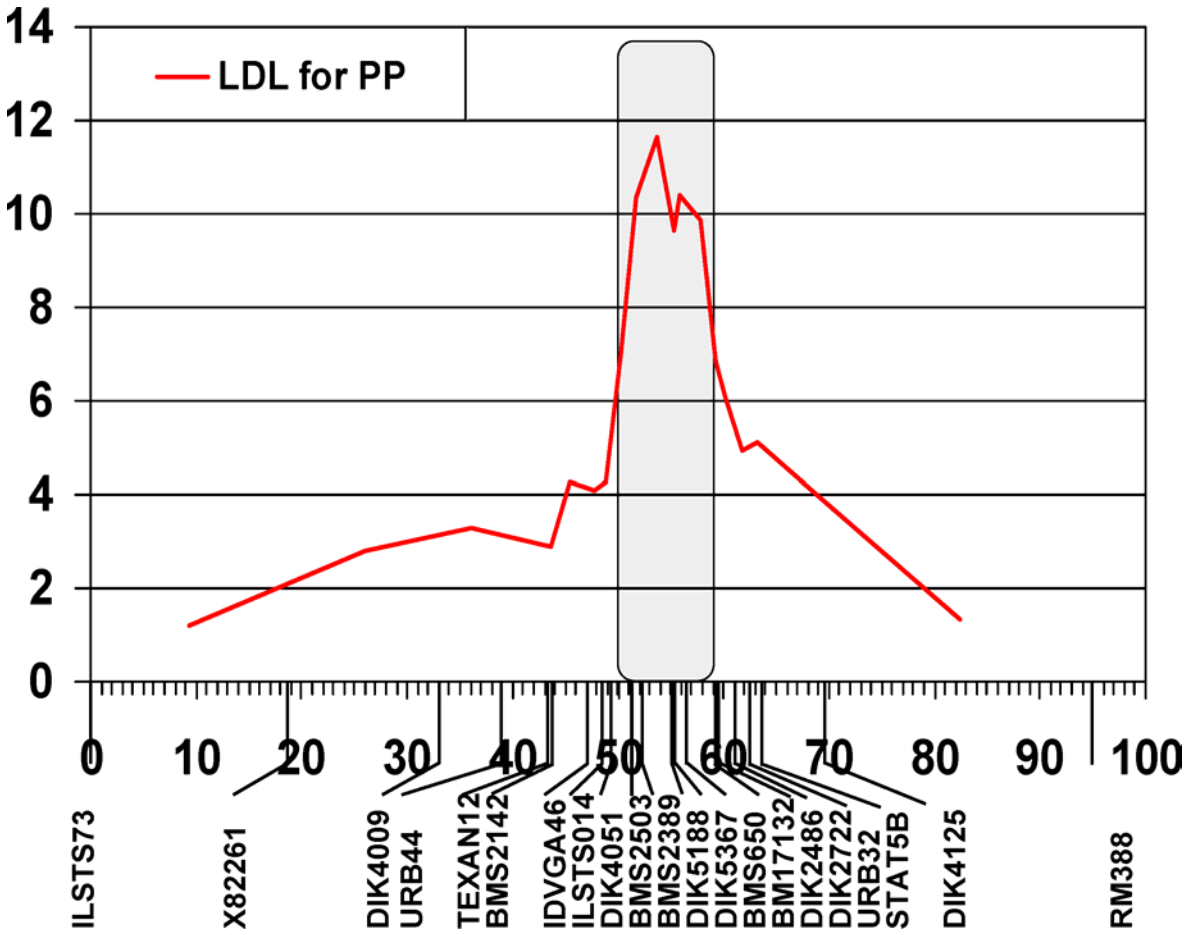

Figure 20: Results of a combined linkage disequilibrium and linkage analysis (LDL) by the LDL program for a quantitative trait locus affecting protein percent (PP-QTL). Positions of markers used in the analysis are denoted on the X-axis. The 1-LOD drop-off confidence interval for PP-QTL is marked in grey. The log-likelihood ratio test values (LRT) are presented on the Y-axis, and the chromosome length is in cM on the X-axis.

This neighbouring peak at 55.89 cM also appears in the analysis with the LDL_rams program, but it is barely significant when using very small sampling numbers (P=0.034) and it does not fit into the 95% confidence interval. *Meuwissen et al.* (2002) have shown that when using only linkage disequilibrium analysis or only linkage analysis, spurious peaks can occur. These peaks are filtered away by combined LDL analysis. As the most acceptable solution they determined that data must conform to both linkage disequilibrium analysis and linkage analysis concepts at the putative QTL position. The combined LDL method can only use the LD value existing in the mapping population. If there is no substantial LD between the markers used, the mapping result will be based primarily on

linkage information. To test this possibility we made a linkage disequilibrium (LD) map for the studied region. The LD map was constructed by first calculating the amount of LD (measured as D'; Hedrick 1987) for all possible marker pairs and then, for each marker, finding the average D' with that marker and all markers residing within 5 cM in either direction from that marker (Olsen et al., 2005; Fig. 21). The D' for the markers with a distance to the nearest marker exceeding 5 cM was set to 0.

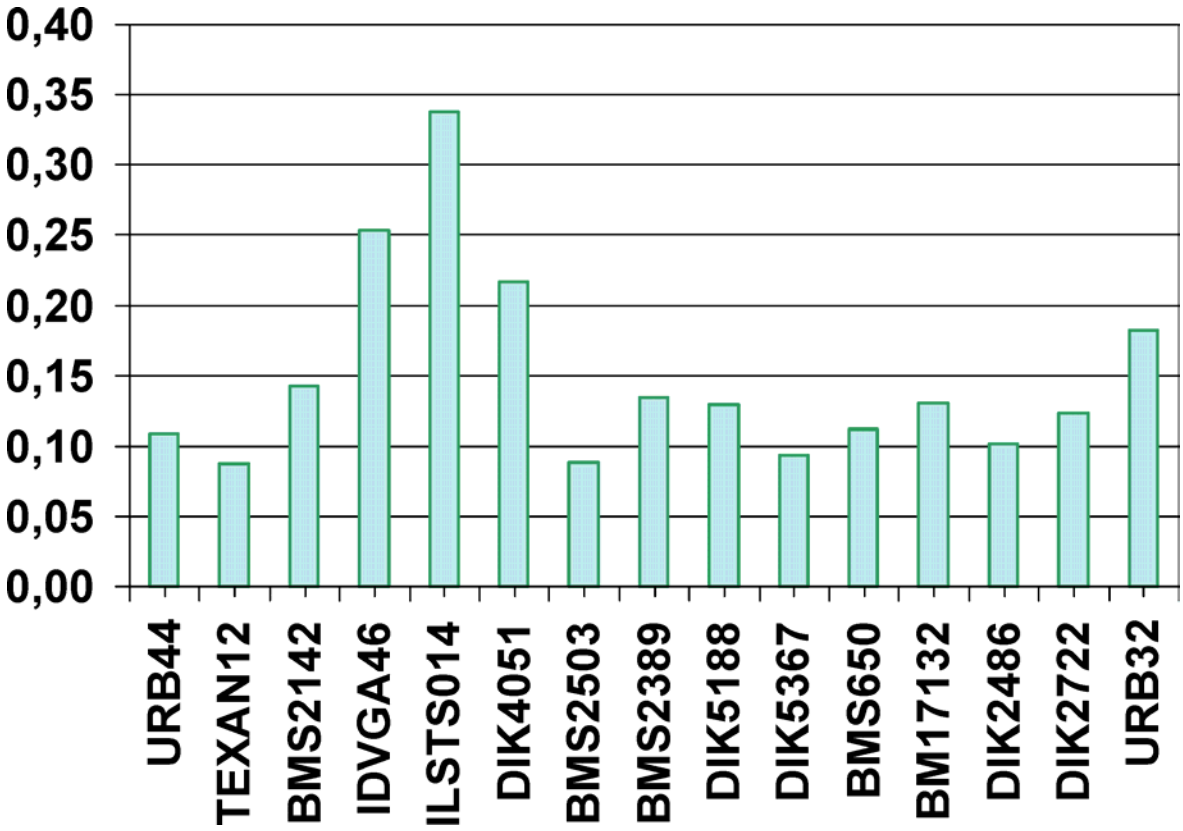

Figure 21: Linkage disequilibrium map for the 15 markers on chromosome 19 (BTA19). The average D' values (Y-axis) between the named markers (X-axis) and markers within a distance of 5 cM.

Figure 21 shows that the D' values for the genotyped marker on BTA19 usually do not exceed 0.15. Only four markers have D' values higher than 0.15 (IDVGA46, ILSTS014, DIK4051 and URB32). Markers IDVGA46 and ILSTS014 show very high D' values compared to other markers. An explanation for this is that both markers are poorly informative. Both have three alleles, out of which one allele has a frequency of over 95%.
LD results shown here (Fig. 21) are inconsistent with results from *Farnir et al.* (2000). They found a substantial LD in the Dutch black-and-white population, extended over several tens of centiMorgans, with high D' values (>0.21) on chromosome 19. A lower LD can arise either from a lower marker density or a higher effective population size. The

marker map used here is relatively dense, but it is still not adequately dense to allow the use of the LD in the Fleckvieh population. The reason probably lies in the fact that the effective population size of the Fleckvieh is estimated high, 250-300 (Pirchner, 2002), in contrast to the estimated effective population size of 50 in the Dutch black-and-white population (Biochard, 1996).

The peak at approximately 67 cM, which appears with LDL_rams analysis, disappears when using the most probable haplotypes. This makes us conclude that the peak arose most probably due to the small number of samplings used (1100 samplings).

With the LDL program we were also able to run also a family-wise analysis, but, as occurred in the earlier case, the PP-QTL segregated in the ABFV as well as in the FV families. The PP-QTL was most probably introduced into the ABFV through their RH founder, but it also exists in the FV population. We couldn't establish a common founder or common haplotype for the significant FV families i.e. the density of the marker map used is powerless to detect a Fleckvieh-wide LD at the QTL.

6.6.9.3 Checking for possible associated effects

The information concerning possible associated effects of the QTL is valuable for a functional candidate gene analysis and a possible implementation in a marker assisted selection (MAS) program. To estimate the possible associated effects of this mapped QTL we analysed our data for all available traits. Analysis was performed for the following traits: milk yield (MY), fat yield (FY), protein (PY), fat percent (FP), milk somatic cell count (SCC), milkability (MA), persistency (PE), productive life (PL), maternal non-return rate (mNR), paternal non-return rate (pNR), maternal calving ease (mCE), paternal calving ease (pCE), maternal stillbirth (mSB) and paternal stillbirth (pSB). Out of all analysed traits we achieved significant results for the QTL affecting three traits (Fig. 22):

- QTL for milkability at 59.21 cM with a LTR value of 14.51 (P=0.0001)
- QTL for productive life with two distinct peaks at 53.69 cM and a LRT value of 9.72 and at 57.86 cM with a LTR value of 8.38 (P=0.002 and P=0.004, respectively)
- QTL for milk somatic cell count at 25.94 cM with a LTR value of 6.73 (P=0.009).

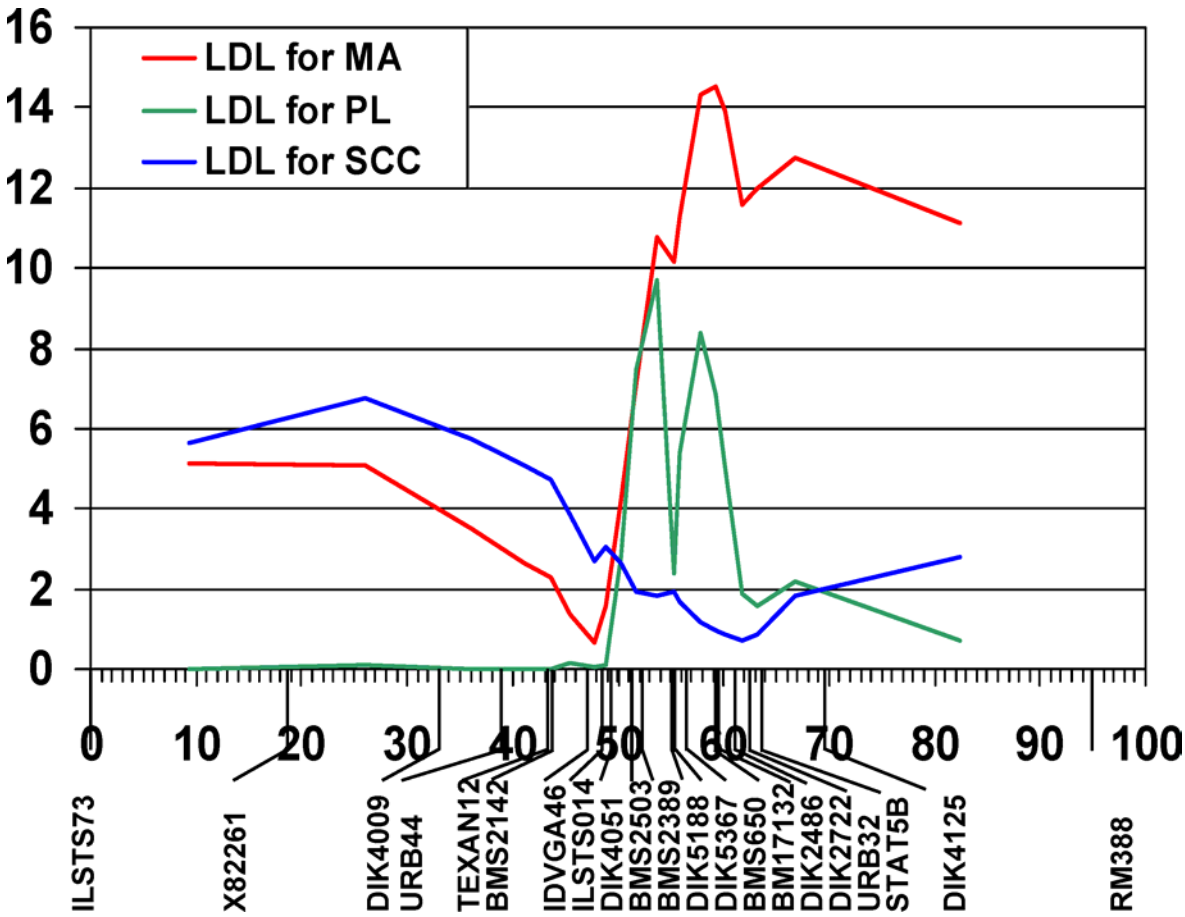

Figure 22: Results of a combined linkage disequilibrium and linkage analysis (LDL) by the LDL program for a quantitative trait locus affecting milkability (MA), productive life (PL) and somatic cell count (SCC). The positions of the markers used in the analysis are denoted on the X-axis. The log-likelihood ratio test values (LRT) are represented on the Y-axis and the chromosome length in cM is represented on the X-axis.

The QTL with an effect on milkability (MA-QTL) has a very high LTR value, and it is highly significant. We analysed the data with QTL Express program in order to test whether the PP-QTL and MA-QTL are one QTL with an effect on both traits or two separate QTL. The interval mapping results across families for both traits indicated the estimated sire effects go in the same direction. The F-test statistical curve in the family-wide analysis showed:

- Three families are significant for both QTL
- One family has significant results for MA-QTL and indicative results for PP-QTL
- One family is indicative for both QTL
- One family has significant results for MA-QTL, but it is without any effect for PP-QTL.

In accordance with these results, we concluded that most likely one QTL has an effect on both PP and MA with a stronger effect on MA. However, the question of whether one QTL

has an effect on two traits or two distinct QTL exist can only be resolved by the application of a two QTL model (Gautier et al., 2005) or a multi-trait multi-QTL model (Meuwissen and Goddard, 2004).

Also, there are reports of a QTL affecting udder conformation trait at approximately 67 cM (Schrooten et al., 2000), which could be associated with QTL for the milkability detected here.

Productive life is difficult to estimate because it is a complex trait dependent on many productivity, fertility and conformation traits. Since it is based on the productivity traits it could be mapped together with some production traits as well. A previously mapped QTL for ovulation rate at 65 cM (Kirkpatrick et al., 2000) could also be connected to the productive life QTL. Ovulation rate and fertility, in general, have influence on the productive life, i.e. cows with good reproductive performance and high production have a longer productive life.

There was also an earlier indication of a somatic cell count QTL located proximally on BTA19 (Bennewitz et al., 2003), which is in good concordance with the QTL for the SCS detected here. The marker map used for the analyses here was not dense enough in the proximal part of the chromosome for the QTL position to be better resolved.

6.6.10 Candidate gene identification

Bovine chromosome 19 contains one of the most conserved syntenies among mammalian chromosomes with most of its homologous genes found on human autosome 17 (HSA17) and small segments on HSA5 and HSA10 (Larkin et al., 2006). This situation presents great opportunity for candidate gene identification. On the other hand, it is necessary for the QTL to be finely mapped as one centiMorgan of the chromosome length can include numerous genes. In the search for candidate gene we limited us only to the region of 3 cM in size, between two markers: BMS2389 and DIK5188, where the likelihood ratio test (LRT) value was at its maximum. Sequences of both markers were compared with HSA17 and were found in region 17p12 to 17p11.2, both located in just one contig. The size of localised region on HSA17 is 2580 kb (from 14,610 to 17,190 kb). There are totally 46 genes in this region on the HSA17. Genes in this region (with descriptions taken out of Online Mendelian Inheritance in Man (OMIM) database) include:

- PMP22 (*peripheral myelin protein 22*) - encodes an integral membrane protein that is a major component of myelin in the peripheral nervous system,

- TRIM16 (*tripartite motif-containing 16*) - this gene was identified as an oestrogen and anti-oestrogen regulated gene in epithelial cells stably expressing oestrogen receptor. Expression of this gene was detected in most tissues but his function has not yet been determined,
- ADORA2B (*adenosine A2b receptor*) - encodes an adenosine receptor, which is an integral membrane protein that stimulates adenylate cyclase activity in the presence of adenosine,
- NCOR1 (*nuclear receptor co-repressor 1*) - encodes a protein that mediates ligand-independent transcription repression of thyroid-hormone and retinoic-acid receptors by promoting chromatin condensation and preventing access of the transcription machinery,
- PIGL (*phosphatidylinositol glycan anchor biosynthesis*) - encodes an enzyme that catalyzes the second step of glycosylphosphatidylinositol biosynthesis,
- UBB (*ubiquitin B*) – encodes ubiquitin, one of the most conserved proteins known, which is required for ATP-dependent, nonlysosomal intracellular protein degradation of abnormal proteins and normal proteins with a rapid turnover,
- TRPV2 (*transient receptor potential cation channel*) - encodes an ion channel that is activated by high temperatures above 52 degrees Celsius and may be involved in transduction of high-temperature heat responses in sensory ganglia,
- TNFRSF13B (*tumor necrosis factor receptor superfamily*) – encodes a protein that is a lymphocyte-specific member of the tumor necrosis factor (TNF) receptor superfamily,
- FLCN (*foliculin*) - mutations in this gene are associated with the Birt-Hogg-Dube syndrome, which is characterized by fibrofolliculomas, renal tumors, lung cysts, and pneumothorax,
- COPS3 (*COP9 constitutive photomorphogenic homolog subunit 3*) – the protein encoded by this gene possesses a kinase activity which phosphorylates regulators involved in signal transduction.

The possible candidate region is very large and includes about 46 genes, which is a limiting factor for the discussion of positional and functional candidate genes. Beside none of the mentioned genes could be associated with protein percent and milkability in dairy cattle. There are also series of genes encoding for hypothetical proteins and zinc finger proteins in this region. As shown by *Björnström and Sjöberg* (2002) the zinc finger

structures are important for intact functional interaction between genes, e.g. oestrogen receptors and prolactin activated STAT5B (STAT5B is located at app. 64 cM on BTA19). They also show that point mutations within the DNA-binding domain of zinc finger proteins can alter the gene interaction and so should be considered as possible cause for altered phenotypes. As there are two genes encoding for zinc finger proteins, seven genes encoding for hypothetical proteins and three genes encoding for proteins with unknown function in this region, the QTL region should be fine mapped before we consider one of them to be our candidate genes.

6.6.11 Final overview and future actions

The IBD QTL mapping method was successfully applied in humans (de Vries et al., 1996; Fallin et al., 2001), cattle (Riquet et al., 1999; Li et al., 2004) and pigs (Nezer et al., 2003). The approach is based on assumption that the observed QTL effect is due to the mutation or migration that happened in the past, g generations ago, and was spread in/over population by reproduction. As a consequence, b individuals in the present generation carrying the desirable allele should also share the haplotype of size $2/gb$ in Morgans, on average, in the vicinity of mutation (Dunner et al., 1997). The IBD method was not always applied with the same success in cattle. The fine mapping of QTL with effect on milk production on BTA14 (Riquet et al., 1999) was hampered by the selection of the mapping population consisting of the Dutch Holstein-Friesian population and the New Zealand Holstein-Friesian population. As later proved, the haplotype of one of the New Zealand sires was coincidentally identical by state with the haplotypes of Dutch sires, leading to the erroneous QTL localisation (Farnir et al., 2002). On the other hand, *Li et al.* (2004) reported a successful application of the IBD method in mapping the QTL for backfat on chromosome 2, 5, 6, 19, 21, and 23 in a commercial cattle population. This population was developed from an Angus base and is expected to derive from one or a limited number of founders. It was also under selection for over 30 years, which should be an extra factor contributing positively to the IBD mapping (Li et al., 2004). The IBD mapping presented here differs from the one proposed by *Riquet et al.* (1999) in the fact that we compared the haplotype of highly related sires so we were able to include the haplotypes of the non-segregating sires into comparison in order to refine the QTLR as much as possible. Also the chosen mapping population – an advanced backcross population Fleckvieh x Red Holstein (ABFV) – is meant to represent the unique opportunity for IBD mapping, as the

influence of the founder in such a population can be substantial. Also, the ABFV is under selection for here examined milk production traits.

The feasibility to use the IBD mapping method depends on the extent of the linkage disequilibrium (Li et al., 2004). Analysis in the Dutch Holstein-Friesian population revealed surprisingly high levels of LD extended over several tens of centiMorgan (Farnir et al., 2000). They also found significant associations between nonsyntenic loci. These findings were confirmed in the North American Holstein population (Vallejo et al., 2003). Similar results were found in two Japanese beef breeds, Japanese Black and Japanese brown (Odani et al., 2006). Analyses in the U.K. Holstein population found the LD only for the distances smaller than approximately 10 cM and never between nonsyntenic loci (Tenesa et al., 2003). Three studies (Tenesa et al., 2003; Olsen et al., 2005; Khatkar et al., 2006) reported a high level of LD on BTA06. As it is known that BTA06 harbours many QTL affecting milk production traits it is suggested that selection for these traits could have generated LD in this region even with substantial amount of the background LD in the population (Tenesa et al., 2003). As already discussed by *Odani et al.* (2006) the usefulness of LD depends on the degree of LD, the distribution and heterogeneity of LD across the genome and its relationship with genetic maps. Our results on BTA19 do not agree with the prediction from *Farnir et al.* (2000) who expects that a situation similar to the one described in the Dutch Holstein-Friesian population will be encountered in most other dairy cattle populations. Even though they found substantial LD on BTA19 the results in our population contribute to the conclusion that there is a difference in the degree of LD between different populations. There are only two reports about QTL affecting milk production traits on BTA19: fat yield and fat percent in German and French Holstein (Bennewitz et al., 2003) and fat percent in French Holstein (Biochard et al., 2003). This also points towards the fact that there is no high selection pressure on BTA19, unlike on BTA06. Due to the use of artificial insemination and the intense selection for increased milk production most of the cattle breeds are structured as very large half-sib families but with low effective population size (N_e) that produces extensive LD (Farnir et al., 2000). Bavarian and Austrian Fleckvieh breeders use large numbers of tested parents to produce and select along successive cattle generations. Around 400 bulls in Bavaria and 140 bulls in Austria, coming from a broad population of dames, are tested every year. The large number of used parents leads to a high effective size (N_e>250; Pirchner, 2002) and consequently to a low LD degree in the population. In order to use LD we should have a far denser marker map in Fleckvieh. All these observations are leading to the conclusion

that it would be important to check the LD degree in the mapping population previously in order to decide about mapping methods and the required marker density.

The method that combines linkage disequilibrium with linkage analysis was chosen for refining the QTL position. This method in comparison with the method using only linkage was able to refine the QTL position sustainable. As it was already mentioned, the LD degree in our mapping population is low so the most information was, once more, extracted from the data by linkage analysis. The included pedigree information, as discussed by *Lee and van der Werf* (2004) has a big impact on the final mapping result when only linkage is used but is not critical when the LD information is used. Our results suggest that pedigree information should be used whenever available, especially in the case when the LD quantity and LD distribution over the genome in the mapping population is not known.

In order to get closer to identifying the gene(s) affecting the traits or that could itself be used as *LD-marker* or *LE-marker* in marker assisted program (MAS), as described by *Dekkers* (2004), two important questions should be resolved in future. The first is the refinement of the QTL position and the second is resolving the question of the existence of one or two QTL affecting PP and MA. Form the above mentioned conclusions it is clear that the LD information is crucial for the fine mapping of the QTL. The use of LD information in our case implies a denser marker map. Resolving the question of the existence of one QTL effecting both PP and MA or two distinct QTL will also contribute to the refinement of the QTL position. There are some methods designed for this purpose, already used for QTL mapping: the two QTL model (Gautier et al., 2005) and the multi-trait multi-QTL model (Meuwissen and Goddard, 2004; Olsen et al., 2004). The refinement of the QTL position and the precise information of QTL effects would open the way for candidate gene identification.

The Fleckvieh population, as dual-purpose breed, shows quite good performances in milk production with more then 6300 kg milk per lactation (305 days; LKV Bayern). Application of genetic information in a good designed MAS program can lead to improvement in milk production. This can subsequently make Fleckvieh much more competitive on the market and that way have a role in preserving of biodiversity of European cattle breeds. The explored introgression of the favourable alleles from one breed to another, without losing the main breed characteristic, can also contribute to similar studies of minor and endangered breeds.

7 Summary

The aim of this study was to identify quantitative trait loci (QTL) affecting milk production traits in one advanced backcross Fleckvieh x Red Holstein (ABFV) population, that are identical by descent (IBD), according to both origin and effect.

The IBD mapping approach was applied to the ABFV. This population provided us a unique opportunity for IBD mapping, because of its strong founder influence and the optimal time since the introgression of the founder's alleles into the population.

The IBD mapping approach presumes an already mapped QTL. "Selective DNA pooling" and approximate interval mapping (AIM) in daughter design (DD) were used to localise the QTL regions for two main traits, milk yield (MY) and protein percent (PP).

Haplotypes were produced by the SimWALK2 program. The results of the haplotype analysis were compared to the mapped QTL regions and are presented here. Chromosome 19 (BTA19) was chosen for further intensive study as the best candidate for possible introgression of QTL alleles by the Red Holstein (RH) founder.

In accordance with the results of the IBD mapping and the initial interval mapping on BTA19, a total of 11 families were chosen in the granddaughter design (GDD-11) for the intensive study as well as 21 markers in two sets.

Interval mapping confirmed the presence of the PP-QTL, but it was unable to refine the QTL position obtained by IBD mapping (region of app. 20 cM).

Combined linkage disequilibrium and linkage (LDL) mapping was applied to refine the PP-QTL position. Through this method we were able to locate a highly significant PP-QTL in the region of approximately 4 cM with a program which use sampled haplotypes and 9 cM with a program which makes use of externally produced haplotypes.

Analyses were performed on all available traits in order to test for possible side effects of the mapped QTL. The highest significance was achieved for milkability. Through analysis of the estimated sires' effects, it appears that the PP-QTL and the QTL affecting milkability are one QTL which has an effect on both traits.

8 Zusammenfassung

Mehrstufige QTL-Kartierungsstrategie in einer fortgeschrittenen Rinder-Rückkreuzungspopulation

Das Ziel dieser Studie war die Identifizierung von „Quantitative Trait Loci“ (QTL) mit einem Effekt auf die Milchleistungsmerkmale in einer fortgeschrittenen Fleckvieh x Red Holstein (ABFV) Rückkreuzungspopulation, welche „Identical by Descent“ (IBD) sowohl in Bezug auf die Herkunft als auch auf den Effekt sind.

Die IBD-Kartierung wurde in einer ABFV durchgeführt. Diese Population gibt uns eine einmalige Chance für die IBD-Kartierung wegen der starken Stammvatereinflüsse und des optimalen Zeitpunkts seit der Introgression der Allele des Stammvaters in die Population.

Die IBD-Kartierung setzt schon kartierte QTL voraus. Das „selective DNA pooling“ und das „Approximate Interval Mapping“ (AIM) im „Daughter Design“ (DD) wurden für die Lokalisation der QTL-Regionen von zwei Hauptmerkmalen, Milch Leistung (MY) und Eiweiß Prozent (PP), benutzt. Die Haplotypen wurden mit Hilfe des SimWALK2-Programms erstellt. Die Ergebnisse der Haplotypenanalyse wurden mit Ergebnissen der QTL-Kartierung verglichen und werden hier präsentiert. Das Chromosom 19 (BTA19) wurde als bester Kandidat für eine mögliche Red Holstein Introgression für eine intensive Untersuchung ausgewählt.

Insgesamt wurden 11 Familien in einem „Granddaughter Design“ (GDD) und 21 Marker in zwei Sets anhand der Ergebnisse der IBD-Kartierung und der anfänglichen Intervallkartierung ausgewählt.

Die Intervallkartierung hat die Anwesenheit eines PP-QTL bestätigt, war aber nicht in der Lage die QTL Position, die wir mit der IBD-Kartierung (ungefähr 20 cM) erzielten, zu verfeinern.

Um die PP-QTL Position zu verfeinern haben wir zusätzlich eine kombinierte „Linkage Disequilibrium und Linkage“ (LDL) Analyse durchgeführt. Mit dieser Methode waren wir in der Lage einen hochsignifikanten PP-QTL in einer Region von ungefähr 4 cM (mit durchschnittliche Haplotypen) bzw. 9 cM (mit extern produzierten Haplotypen) zu kartieren.

Diese Analysen wurden für alle Merkmale, die uns zu Verfügung stehen, durchgeführt, um mögliche Nebeneffekte des QTLs zu testen. Die höchste Signifikanz wurde für das Merkmal der Melkbarkeit erreicht. Die familienweise Analyse hat gezeigt, dass es sich wahrscheinlich um einen QTL mit Effekt auf beide Merkmale handelt.

9 Acknowledgment

I would like to thank Prof. Dr. Dr. Martin Förster for giving me the opportunity to carry out my work on Institut für Tierzucht of the Ludwig-Maximilians-University, Munich, for the critical revision of the manuscript and the shown patience.

I would particularly like to thank Dr. Ivica Međugorac for the numerous and patient explanations, helpful suggestions and the constant support. His inspiration and knowledge led me and mentored me through my whole work. Especially, I want to thank him for the detailed corrections of the manuscript.

I would like to thank Dr. Stefan Krebs for the all the help and explanations, especially with laboratory work and molecular-genetic database inquiries, as well as manuscript corrections.

Special thanks to Angelika Gabler, Verena Gomeringer, Doris Seichter, Claudia Veit and Marijo Vrebac for all help in the work but also for being wonderful colleges and supportive friends.

Thanks a lot to Ms Andorfer, Ms Damian, Mr Dinkel, Ms Niedenführ, Ms Nanou, Ms Schröter, and Ms Zierahn for good cooperation and technical work.

I am grateful to Mr Scott Cornwell and Mr Alex Mclachlan for the language corrections.

I am very grateful for the good cooperation and the willingness to place their samples, data, work and time at our disposal with the following institutions:

- Arbeitsgemeinschaft Süddeutscher Rinderzuchtverbände e.V.,
- Landeskuratorium der Erzeugerringe für tierische Veredelung in Bayern,
- Institut für Tierzucht der Landesanstalt für Landwirtschaft Bayern,
- Milchprüfring Bayern e.V.,
- Tierzuchtforschung e.V. München.

This work was made possible through the financial support from the European Union research project BovMAS (N° QLK5-CT-2001-02379).

Thank to Dr. Ana Međugorac for her support and to Marija, Marko and Ilija Međugorac for being so adorable.

Very special thanks to my parents, Viktor and Branka, and my sister Andrea for boundless love and constant support. Special thanks to my cousin Davor Tomić for his concern and financial support. Support of my family and friends is very much appreciated.

10 References

1. **Almasy L, Blangero J (1998)** Multipoint quantitative-trait linkage analysis in general pedigrees, Am J Hum Genet., 62(5):1198-211.
2. **Andersson L (2001)** Genetic dissection of phenotypic diversity in farm animals, Nat Rev Genet., 2(2):130-8.
3. **Ashwell MS, Heyen DW, Sonstegard TS, Van Tassell CP, Da Y, VanRaden PM, Ron M, Weller JI, Lewin HA (2004)** Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle, J Dairy Sci., 87(2):468-75.
4. **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 (2003)** Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle, Genet Sel Evol., 35(3):319-38.
5. **Biochard D (1996)** Analyse généalogique des races bovines laitières francaises, INRA Prod. Anim., 9: 323-335.
6. **Biochard D, Grohs C, Bourgeois F, Cerqueira F, Faugeras R, Neau A, Rupp R, Amigues Y, Boscher MY, Leveziel H (2003)** Detection of genes influencing economic traits in three French dairy cattle breeds, Genet Sel Evol., 35(1):77-101.
7. **Björnström L, Sjöberg M (2002)** Mutations in the estrogen receptor DNA-binding domain discriminate between the classical mechanism of action and cross-talk with Stat5b and activating protein 1 (AP-1), J Biol Chem., 277(50):48479-83.
8. **Blair MW, Iriarte G, Beebe S (2006)** QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean x wild common bean (Phaseolus vulgaris L.) cross, Theor Appl Genet., 112(6):1149-63.
9. **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 (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(1):253-66.
10. **Bruford MW, Wayne RK (1993)** Microsatellites and their application to population genetic studies, Curr Opin Genet Dev., 3(6):939-43. Review.
11. **Callen DF**, **Thompson AD**, **Shen Y**, **Phillips HA**, **Richards RI**, **Mulley JC**, **Sutherland GR (1993)** Incidence and origin of "null" alleles in the (AC)n microsatellite markers, Am J Hum Genet., 52(5):922-7.
12. **Coppieters W, Riquet J, Arranz JJ, 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, Mamm Genome., 9(7):540-4.
13. **Cowan CM, Dentine MR, Ax RL, Schuler LA (1990)** Structural variation around the prolactin gene linked to quantitative traits in an elite Holstein sire family, Theor. Appl. Genet., 79:577–582.

14. **Darvasi A, Soller M (1994)** Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus, Genetics, 138(4):1365-73.
15. **de Koning DJ, Dekkers JCM, Haley CS (2003)** Designs for QTL detection in livestock and their implications for MAS, Proceedings of an international workshop organised by the Fondazione per le Biotecnologie, the University of Turin and FAO on 17-18 October 2003 in Turin, Italy.
16. **de la Chapelle A, Wright FA (1998)** Linkage disequilibrium mapping in isolated populations: the example of Finland revisited, Proc Natl Acad Sci USA, 95(21):12416-23.
17. **Dekkers JCM (2000)** Quantitative trait locus mapping based on selective DNA pooling, Journal of Animal Breeding and Genetics, 117:1-16.
18. **Dekkers JCM (2004)** Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons, J Anim Sci., 82. Review.
19. **de Vries HG, van der Meulen MA, Rozen R, Halley DJ, Scheffer H, ten Kate LP, Buys CH, te Meerman GJ (1996)** Haplotype identity between individuals who share a CFTR mutation allele "identical by descent": demonstration of the usefulness of the haplotype-sharing concept for gene mapping in real populations, Hum Genet., 98(3):304-9.
20. **Dolezal M, Schwarzenbacher H, Sölkner J, Visscher PM (2005)** An approximate interval mapping procedure for selective DNA pooling 56th Annual Meeting of the European Association for Animal Production 5.6.-8.6.2005 Uppsala, Sweden EAAP-Book of Abstracts No.11 p118.
21. **Dunner S, Charlier C, Farnir F, Brouwers B, Canon J, Georges M (1997)** Towards interbreed IBD fine mapping of the mh locus: double-muscling in the Asturiana de los Valles breed involves the same locus as in the Belgian Blue cattle breed, Mamm Genome, 8(6):430-5.
22. **Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992)** Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups, Genomics, 12(2):241-53.
23. **Ellegren H (2004)** Microsatellites: simple sequences with complex evolution, Nat Rev Genet., 5(6):435-45. Review.
24. **Everts-van der Wind A, Larkin DM, Green CA, Elliott JS, Olmstead CA, Chiu R, Schein JE, Marra MA, Womack JE, Lewin HA (2005)** A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution, Proc Natl Acad Sci USA, 102(51):18526-31.
25. **Falconer DS, Mackay TFC (1996)** Introduction to quantitative genetics, Fourth edition, Pearson education limited.
26. **Fallin D, Cohen A, Essioux L, Chumakov I, Blumenfeld M, Cohen D, Schork NJ (2001)** Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE locus variation and Alzheimer's disease, Genome Res., 11(1):143-51.
27. **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 (2000)** Extensive genome-wide linkage disequilibrium in cattle, Genome Res., 10(2):220-7.
28. **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(1):275-87.

29. **Freyer G, Sorensen P, Kuhn C, Weikard R, Hoeschele I (2003)** Search for pleiotropic QTL on chromosome BTA6 affecting yield traits of milk production, J Dairy Sci., 86(3):999-1008.
30. **Gautier M, Barcelona RR, Fritz S, Grohs C, Druet T, Boichard D, Eggen A, Meuwissen TH (2005)** Fine mapping and physical characterization of two linked quantitative trait loci affecting milk fat yield in dairy cattle on BTA26, Genetics, 172(1):425-36.
31. **George AW, Visscher PM, Haley CS (2000)** Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach, Genetics, 156(4):2081-92.
32. **Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X (1995)** Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing, Genetics, 139(2):907-20.
33. **Garner C**, **McInnes LA**, **Service SK**, **Spesny M**, **Fournier E**, **Leon P**, **Freimer NB (2001)** Linkage analysis of a complex pedigree with severe bipolar disorder, using a Markov Chain Monte Carlo method, Am J Hum Genet., 68(4):1061-4.
34. **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 Res., 12(2):222-31.
35. **Haldane JBS (1919)** The combination of linkage values and the calculation of distance between the loci of linkage factors, Journal of Genetics, 8:299-309.
36. **Hamada H, Petrino MG, Kakunaga T (1982)** A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes, Proc Natl Acad Sci USA, 79(21):6465-9.
37. **Hedrick PW (1987)** Gametic disequilibrium measures: proceed with caution, Genetics, 117(2):331-41.
38. **Huang XQ, Coster H, Ganal MW, Roder MS (2004)** Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (Triticum aestivum L.), Theor Appl Genet., 106(8):1379-89.
39. **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 (2004)** A comprehensive genetic map of the cattle genome based on 3802 microsatellites, Genome Res., 14(10A):1987-98.
40. **Itoh T, Watanabe T, Ihara N, Mariani P, Beattie CW, Sugimoto Y, Takasuga A (2005)** A comprehensive radiation hybrid map of the bovine genome comprising 5593 loci, Genomics, 85(4):413-24.
41. **Jansen RC, Stam P (1994)** High resolution of quantitative traits into multiple loci via interval mapping, Genetics, 136(4):1447-55.
42. **Khatkar MS, Thomson PC, Tammen I, Raadsma HW (2004)** Quantitative trait loci mapping in dairy cattle: review and meta-analysis, Genet Sel Evol., 36(2):163-90. Review.
43. **Khatkar MS, Collins A, Cavanagh JA, Hawken RJ, Hobbs M, Zenger KR, Barris W, McClintock AE, Thomson PC, Nicholas FW, Raadsma HW**

(2006) A first-generation metric linkage disequilibrium map of bovine chromosome 6, Genetics, 174(1):79-85.

44. **Kirkpatrick BW, Byla BM, Gregory KE (2000)** Mapping quantitative trait loci for bovine ovulation rate, Mamm Genome., 11(2):136-9.
45. **Kleppe R, Khorana HG (1971)** Studies on polynucleotides. CXIX. Transcription of short double-stranded deoxyribonucleic acids of defined nucleotide sequences, J Biol Chem, 247(19):6149-56.
46. **Kosambi DD (1944)** The estimation of map distances from recombination values, Annals of Eugenics, 12:172-5.
47. **Lander ES, Botstein D (1989)** Mapping mendelian factors underlying quantitative traits using RFLP linkage maps, Genetics, 121(1):185-99.
48. **Lander ES, Green P (1987)** Construction of multilocus genetic linkage maps in humans, Proc Natl Acad Sci USA, 84(8):2363-7.
49. **Larkin DM**, **Astakhova NM**, **Prokhorovich MA**, **Lewin HA**, **Zhdanova NS (2006)** Comparative mapping of cattle chromosome 19: citogenetic localization of 19 BAC clones, Cytogenet Genome Res., 112(3-4):235-40.
50. **Lee SH, van der Werf JH (2004**) The efficiency of designs for fine-mapping of quantitative trait loci using combined linkage disequilibrium and linkage, Genet Sel Evol., 36(2):145-61.
51. **Lee SH, van der Werf JH (2005)** The role of pedigree information in combined linkage disequilibrium and linkage mapping of quantitative trait loci in a general complex pedigree, Genetics, 169(1):455-66.
52. **Lee SH, van der Werf JH (2006)** An efficient variance component approach implementing an average information REML suitable for combined LD and linkage mapping with a general complex pedigree, Genet Sel Evol., 38(1):25-43.
53. **Levinson G, Gutman GA (1987)** Slipped-strand mispairing: a major mechanism for DNA sequence evolution, Mol Biol Evol., 4(3):203-21. Review.
54. **Li C, Basarab J, Snelling WM, Benkel B, Kneeland J, Murdoch B, Hansen C, Moore SS (2004)** Identification and fine mapping of quantitative trait loci for backfat on bovine chromosomes 2, 5, 6, 19, 21, and 23 in a commercial line of Bos Taurus, J Anim Sci., 82(4):967-72.
55. **Lipkin E, Shalom A, Khatib H, Soller M, Friedmann A (1993)** Milk as a source of deoxyribonucleic acid and as a substrate for the polymerase chain reaction, J Dairy Sci., 76(7):2025-32.
56. **Lipkin E, Mosig MO, 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(3):1557-67.
57. **Litt M, Luty JA (1989)** A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene, Am J Hum Genet., 44(3):397-401.
58. **Liu Z, Reinhardt J, Szyda H, Thomsen H, Reents R (2004)** A marker assisted genetic evaluation system for dairy cattle using a random QTL model, Proceedings of the 2004 Interbull meeting, Bulletin No. 32 p170-174.
59. **Lynch M, Walsh B (1998)** Genetics and analysis of quantitative traits, Sinauer Associates, Inc.
60. **MacCluer JW, van der Berg JL, Read B, Ryder OA (1986)** Pedigree analysis by computer simulation, Zoo Biol., 5:147-160.
61. **Mackay TFC (2001)** Quantitative trait loci in Drosophila, Nat Rev Genet., 2(1):11-20. Review.

62. **Makalowski W (2000)** Genomic scrap yard: how genomes utilize all that junk, Gene, 259(1-2):61-7.
63. **Marri PR, Sarla N, Reddy LV, Siddiq EA (2005)** Identification and mapping of yield and yield related QTLs from an Indian accession of Oryza rufipogon, BMC Genet., 6(1):33.
64. **Medugorac I, Kemter J, Russ I, Pietrowski D, Nuske S, Reichenbach HD, Schmahl W, Forster M (2003)** Mapping of the bovine spinal muscular atrophy locus to Chromosome 24, Mamm Genome., 14(6):383-91.
65. **Meuwissen TH, Goddard ME (2001)** Prediction of identity by descent probabilities from marker-haplotypes, Genet Sel Evol., 33(6):605-34.
66. **Meuwissen TH, Goddard ME (2004)** Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data, Genet Sel Evol., 36(3):261-79.
67. **Meuwissen TH, Karlsen A, Lien S, Olsaker I, Goddard ME (2002)** Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping, Genetics, 161(1):373-9.
68. **Mosig MO, 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(4):1683-98.
69. **Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1986)** Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction, Cold Spring Harb Symp Quant Biol., 51 Pt 1:263-73.
70. **Nezer C, Collette C, Moreau L, Brouwers B, Kim JJ, Giuffra E, Buys N, Andersson L, Georges M (2003)** Haplotype sharing refines the location of an imprinted quantitative trait locus with major effect on muscle mass to a 250-kb chromosome segment containing the porcine IGF2 gene, Genetics, 165(1):277-85.
71. **Norio R (2003a)** Finnish Disease Heritage I: characteristics, causes, background, Hum Genet., 112(5-6):441-56.
72. **Norio R (2003b)** Finnish Disease Heritage II: population prehistory and genetic roots of Finns, Hum Genet., 112(5-6):457-69.
73. **Norio R (2003c)** The Finnish Disease Heritage III: the individual diseases. Hum Genet., 112(5-6):470-526. Review.
74. **Nowak R (1994)** Mining treasures from 'junk DNA', Science, 263(5147):608-10.
75. **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, Anim Genet., 37(2):139-44.
76. **Olsen HG, Lien S, Svendsen M, Nilsen H, Roseth A, Aasland Opsal M, Meuwissen TH (2004)** Fine mapping of milk production QTL on BTA6 by combined linkage and linkage disequilibrium analysis, J Dairy Sci., 87(3):690-8.
77. **Olsen HG, Lien S, Gautier M, Nilsen H, Roseth A, Berg PR, Sundsaasen KK, Svendsen M, Meuwissen TH (2005)** Mapping of a milk production quantitative trait locus to a 420-kb region on bovine chromosome 6, Genetics, 169(1):275-83.
78. **Pareek CS, Pareek RS, Walawski K (2002)** Novel linkage mapping approach using DNA pooling in human and animal genetics. II. Detection of quantitative traits loci in dairy cattle, J Appl Genet., 43(3):309-18. Review.

79. **Paterson AH, Lander ES, Lincoln SE, Hewitt JD, Peterson S, Tanksley SD (1988)** Resolution of quantitative traits into Mendelian factors using a complete RFLP linkage map, Nature, 335: 721-726.
80. **Peltonen L, Palotie A, Lange K (2000)** Use of population isolates for mapping complex traits, Nat Rev Genet., 1(3):182-90. Review.
81. **Pirchner F (2002)** Schätzung inzuchtwirksamer (effektiver) Populationsgrößen aus Genfrequenzschwankungen bei Bayrischem Fleckvieh und Tiroler Grauvieh, Arch. Tierz., Dummerstorf 45, 4:331-339.
82. **Riquet J, Coppieters W, Cambisano N, Arranz JJ, Berzi P, Davis SK, Grisart B, Farnir F, Karim L, Mni M, Simon P, Taylor JF, Vanmanshoven P, Wagenaar D, Womack JE, 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):9252-7.
83. **Rodriguez-Zas SL, Southey BR, Heyen DW, Lewin HA (2002)** Interval and composite interval mapping of somatic cell score, yield, and components of milk in dairy cattle, J Dairy Sci., 85(11):3081-91.
84. **Ron M, Feldmesser E, Golik M, Tager-Kohen I, Kliger D (2004)** A complete genome scan of the Israeli Holstein population for quantitative trait loci by a daughter design, J Dairy Sci., 87: 476-490.
85. **Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988)** Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, Science, 239(4839):487-91.
86. **Schaid DJ (2004)** Linkage disequilibrium testing when linkage phase is unknown, Genetics, 166(1):505-12.
87. **Schlötterer C, Tautz D (1992)** Slippage synthesis of simple sequence DNA, Nucleic Acids Res., 20(2):211-5.
88. **Schnabel RD, Kim JJ, Ashwell MS, Sonstegard TS, Van Tassell CP, Connor EE, Taylor JF (2005)** Fine-mapping milk production quantitative trait loci on BTA6: analysis of the bovine osteopontin gene, Proc Natl Acad Sci U S A, 102(19):6896-901.
89. **Schork NJ, Cardon LR, Xu X (1998)** The future of genetic epidemiology, Trends Genet., 14(7):266-72. Review.
90. **Schrooten C**, **Bovenhuis H**, **Coppieters W**, **Van Arendonk JA (2000)** Whole genome scan to detect quantitative trait loci for conformation and functional traits in dairy cattle, J Dairy Sci., 83(4):795-806.
91. **Schug MD**, **Hutter CM**, **Wetterstrand KA**, **Gaudette MS**, **Mackay TF**, **Aquadro CF (1998)** The mutation rates of di-, tri-, and tetranucleotide repeats in *Drofphila melanogaster*, Mol Biol Evol, 15(12):1751-60.
92. **Seaton G, Haley CS, Knott SA, Kearsey M, Visscher PM (2002)** QTL Express: mapping quantitative trait loci in simple and complex pedigrees, Bioinformatics, 18 339-340.
93. **Sobel E, Lange K (1996)** Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics, Am J Hum Genet., 58(6):1323-37.
94. **Suzuki DT, Griffiths AJF, Miller JH, Lewontin RC (1989)** An introduction to genetic analysis, Fourth edition, W.H. Freeman and company, New York
95. **Szyda J, Liu Z, Reinhardt F, Reents R (2005)** Estimation of quantitative trait loci parameters for milk production traits in German Holstein dairy cattle population, J Dairy Sci., 88(1):356-67.

96. **Tanksley SD, Nelson JC (1996)** Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines, Theor Appl Genet., 92:191–203.
97. **Tautz D, Renz M (1984)** Simple sequences are ubiquitous repetitive components of eukaryotic genomes, Nucleic Acids Res., 12(10):4127-38.
98. **Tautz D, Trick M, Dover GA (1986)** Cryptic simplicity in DNA is a major source of genetic variation, Nature, 322(6080):652-6.
99. **Tautz D (1989)** Hypervariability of simple sequences as a general source for polymorphic DNA markers, Nucleic Acids Researc, 16: 6463-6471.
100. **Tenesa A, Knott SA, Ward D, Smith D, Williams JL, Visscher PM (2003)** Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes, J Anim Sci., 81(3):617-23.
101. **Vallejo RL, Li YL, Rogers GW, Ashwell MS (2003)** Genetic diversity and background linkage disequilibrium in the North American Holstein cattle population, J Dairy Sci., 86(12):4137-47.
102. **Wang, J, Soller M, Dekkers JCM (2002)** Mapping QTL with selective DNA pooling by least squares, 7th World Congress on Genetics Applied to Livestock Production, August 19-23, 2002, Montpellier, France, N° 21-08.
103. **Weber JL, May PE (1989)** Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction, Am J Hum Genet., 44(3):388-96.
104. **Weller JI, Kashi Y, Soller M (1990)** Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle, J Dairy Sci., 73(9):2525-37.
105. **Winter A, Kramer W, Werner FA, Kollers S, Kata S, Durstewitz G, Buitkamp J, Womack JE, 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(14):9300-5.
106. **Yamamoto T, Lin H, Sasaki T, Yano M (2000)** Identification of heading date quantitative trait locus Hd6 and characterization of its epistatic interactions with Hd2 in rice using advanced backcross progeny, Genetics, 154(2):885-91.
107. **Zeng ZB (1994)** Precision mapping of quantitative trait loci, Genetics, 136:1457-1468.
108. **Zhang DX, Hewitt GM (2003)** Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects, Mol Ecol., 12(3):563-84. Review.

11 Appendix

Table 7: Chemicals used in the study and their manufacturer.

Acetone	Roth
Agarose (Seakem LE)	Biozym
Alconox Powdered Precision Cleaner	Alconox Inc.
APS	GATC
boric acid (>= 99,8%, pulv.)	Roth
Desoxyribonukleotide (dNTPs)	peqLab
Dextran-blue	Sigma
Dithiothreitol (DTT)	Sigma
EDTA (>= 99% p.a.)	Roth
Ethyl alcohol (99%)	Roth
Ethidium Bromid (EtBr)	Sigma
Formamide	Sigma
Hot Start DNA Polymerase I (HOT FIREPol)	Solis Biodyne
Isopropanol	Roth
Potassium chloride (KCl)	Merck
Potassium-dihydrogenphosphate (KH_2PO_4)	Merck
Magnesium chloride	Solis Biodyne
Sodium chloride (NaCl)	Roth
Di-Sodiumhydrogenphosphate (Na_2HPO_4)	Merck
NF Acrylamide (30%)	Roth
NF Urea	Roth
Primer	MWG Biotech
Proteinase K (lyophil., >= 30U/mg)	Roth
Buffer for Hot Start DNA Polymerase I	Solis Biodyne
Hydrochloric acid (HCl, 37 % p.a.)	Roth
Sodiumdodecylsulfat (SDS, ~99%)	Sigma
TEMED (>= 99%)	Invitrogen
Tris (>= 99,9% p.a.)	Roth
Tween 20	ICN Biomedials Inc.
Water (double processed tissue culture water)	Sigma

Table 8: **Frequently used abbreviations.**

AB	Advanced backcross
ABFV	Advanced backcross population Fleckvieh x Red Holstein
AB-QTL	Advanced backcross quantitative trait loci analysis
aFDR	Adjusted false discovery rate
A.I.	Artificial insemination
AIM	Approximate interval mapping
ASR	Arbeitgemeinschaft Süddeutcher Rinderzuchtverbände e.V.
BC	Backcross
BTA	*Bos Taurus* autosome
cBV	Corrected breeding values
CCD	Charge coupled device
DD	Daughter design
d.f.	Degrees of freedom
DGAT1	Diacylglycerol acyltransferase
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
FDR	False discovery rate
FP	Fat percent
FV	Fleckvieh
FY	Fat yield
GAS	Gene assisted selection
GDD	Granddaughter design
GHR	Growth hormone receptor
GRM	Genotype relationship matrix
GWS	Genome wide scan
HSA	*Homo Sapiens* autosome
IBD	Identical by descent
LD	Linkage disequilibrium
LDL	Linkage disequilibrium and linkage
LE	Linkage equilibrium
LKV Bayern	Landeskuratorium der Erzeugerringe für tierische Veredelung in Bayern e.V.
LOD	Logarithm of the odds
LRT	Log-likelihood ratio test
MA	Milkability
MAS	Marker assisted selection
mCE	Maternal calving ease
MCMC	Markov chain Monte Carlo

MLE	Maximum likelihood estimates
mNR	Maternal non-return rate
MPR	Milchprüfring Bayern e.V.
mSB	Maternal stillbirth
MY	Milk yield
PBS buffer	Phosphate-buffered saline buffer
pCE	Paternal calving ease
PCR	Polymerase chain reaction
PE	Persistency
PL	Productive life
pNR	Paternal non-return rate
PP	Protein percent
PRL	Prolactin
pSB	Paternal stillbirth
PY	Protein yield
QTL	Quantitative trait locus (loci)
QTLR	Quantitative trait locus (loci) region
REML	Restricted maximum likelihood
RH	Red Holstein
SCC	Somatic cell count
SMA	Spinal muscular atrophy
SSTRs	Simple sequence tandem repeats
TBE buffer	Tris-Borate-Edta Buffer
VC	Variance component
WGS	Whole genome shotgun sequence