Revisiting the QTL for Milk Production Traits on BTA6
R. Weikard*, P. Widmann*, J. Buitkamp§, R. Emmerling§ and C. Kühn*

Introduction
A variety of QTL mapping studies in different dairy cattle populations suggest that there are at least three distinct QTL regions on bovine chromosome 6 (BTA6) affecting milk production traits (http://www.animalgenome.org/cattle/). According to the current version of the bovine genome assembly Btau 4.0 (NCBI), one QTL (37.2-38.2 Mb) with a predominant effect on milk protein and fat content is located proximally of marker BM143, whereas a second QTL particularly affecting milk fat and milk protein yield was identified distally of BM143 (44.1-46.6 Mb). Moreover, a third QTL with effect on milk yield and composition was detected in a region enclosing 88.1-88.8 Mb. The multiple QTL architecture on BTA6 impedes dissecting the distinct QTL on a molecular level. Whereas the third QTL region has a strong focus on the casein gene cluster, fine mapping studies targeted at candidate genes underlying the QTL affecting milk composition in QTL region 1 had indicated conflicting results. Mutations in the ABCG2 (ABCG2 Y581S, Cohen-Zinder et al. 2005; Olsen et al. 2005) and SPP1 (OPN3907, Schnabel et al. 2005) genes have been proposed as putative underlying functional mutation. However, OPN3907 identified in the US Holstein population could be excluded in the Israeli Holstein (Seroussi et al. 2008) and Norwegian dairy (Olsen et al. 2007) cattle populations as causative molecular background in QTL region 1. Because of its chromosomal position and key function in energy metabolism, the PPARGC1A gene was discussed as a positional and functional candidate for the QTL region 2. A noncoding gene variant of the PPARGC1A gene (c.1892+19T>C) was identified to be trait-associated in the German Holstein (GH) population indicating that the PPARGC1A gene could be involved in the genetic variation underlying the QTL for milk fat yield on BTA6 (Weikard et al. 2005). However, it remained to be elucidated whether this gene variant is causal for the QTL or if the trait association is due to linkage disequilibrium (LD) with a yet undetected functional polymorphism in close proximity to the SNP. Therefore, the aim of this study was to re-evaluate the QTL region 2 using a higher density map of SNPs in two independent cattle populations, GH and Fleckvieh (GF).

Material and methods
Animals. Genomic DNA was extracted from blood of 1033 first calving daughters of 19 sires (9 sires with first crop daughters only) from the GH and ear samples of 407 first calving heifers comprising 18 families (all first crop daughters) of the GF populations. Yield deviations for the first lactation milk performance traits were included in the analysis.
Screening for SNPs. Single nucleotide polymorphisms (SNPs) were identified by comparative sequencing of genomic DNA pools from GH daughters differing in milk fat yield and DNA from three individuals originating from a GH x Charolais resource

* Leibniz Institute for Farm Animal Biology (FBN), Research Unit Molecular Biology, 18196 Dummerstorf, Germany
§ Bavarian State Research Center for Agriculture, Institute for Animal Breeding, 85586 Grub, Germany

**Genotyping.** Thirty three selected DNA variants distributed across or close to the PPARGC1A gene (QTL region 2) and five SNPs located in genes of QTL region 1 were genotyped using the MassARRAY® iPLEX Gold System on a MALDI-TOF mass spectrometer (Sequenom). Two iPLEX multiplex assays were designed and genotyped by Eurofins Medigenomix GmbH (Germany). Two gene variants (Indels) and 25 microsatellite markers (only in the GH set) distributed across BTA6 were genotyped by PCR-SSLP on the MegaBACE™ 1000 DNA Analysis System (GE Healthcare). The NCAPG I442M variant was analysed using PCR-RFLP according to Eberlein et al. (2009). Furthermore, two gene variants of the DGAT1 gene (DGAT1 K232A and DGAT1 promoter VNTR) known to exert a high impact on milk yield, milk fat yield and content were genotyped by PCR-RFLP or SSP assay, respectively, as described by Kühn et al. (2007).

**Statistical analyses.** To estimate LD between loci, pairwise \( r^2 \)- and D’-values were determined using the program “Power Marker” (Liu and Muse 2005) allowing to integrate multi-allelic markers. LD was calculated separately for maternally and paternally inherited haplotypes in both breeds that were determined by Monte Carlo Markov Chain algorithm (Simwalk2, Sobel and Lange, 1996). Association analyses were performed with selected tagged SNPs and records for phenotypes „Yield Deviation first lactation” applying algorithms implemented in the program package Qxpak (Perez-Enziso and Misztal, 2004) with the following model:

\[
y_{id} = d \text{gat1}_K232A_d + \sum_k \sum_h \lambda_{ikh} g_k + u_i + e_{dhk}
\]

where \(y_{id}\) is the record of individual \(i\) with DGAT1 K232A genotype \(d\), \(\text{dgat1}_K232A_d\) is the fixed effect of the DGAT1 K232A genotype, \(\lambda_{ikh}\) is an indicator variable which is 1 when the allele at the \(h^{th}\) haplotype (1 or 2) of the \(i^{th}\) individual is \(k\) and otherwise 0, \(g_k\) is the allelic effect of allele \(k\) at the target locus, \(u_i\) is the infinitesimal genetic effect of individual \(i\), and \(e_{dhk}\) is the residual.

**Results and discussion**

**Allele distribution and LD of SNPs.** Genotyping of 38 targeted SNPs using iPLEX assays was successful for a total of 33 SNPs (calling rate of at least 93%). The average minor allele frequency (MAF) of maternally inherited alleles of genotyped SNPs was 0.240 in the GH and 0.202 in the GF data set. For paternally inherited alleles, average MAF was found to deviate slightly, 0.202 in the GH and 0.155 in the GF data set. LD analysis showed that in both data sets a higher LD was observed between SNPs within the PPARGC1A gene compared to SNPs in gene flanking regions and the QTL region 1. No fixed LD was found between PPARGC1A gene variants and SNPs of genes in 1 Mb vicinity (GBA3, DHX15, LGI, SEPSECS) and between SNPs of ABCG2, SPP1 and NCAK located in QTL region 1, which is contrary to results obtained in the Norwegian dairy population (Olsen et al. 2007). Haplotype blocks tagging SNPs \(r^2 \geq 0.8\) between pairwise SNPs did not differ between maternal and paternal inherited haplotypes but rather between breeds.

**Association analysis.** From each haplotype block one representative SNP was selected for the association analysis, which finally included a total of 24 tagged SNPs. In the GH
population, a significant association of the gene variant ABCG2 Y581S was detected with milk fat (FDR= 0.001) and milk protein (FDR= 0.003) content but not with any of the milk yield traits supporting previous results obtained for the QTL region 1 (Cohen-Zinder et al. 2005, Olsen et al. 2005). However, in the GF data set, the allele ABCG2 581Y associated with a higher milk fat and protein content in GH was fixed. For the SPP1 gene variant OPN3907, we did not find indication for association with milk composition traits supporting the hypothesis that most likely, this variant can be excluded as causative basis for QTL region 1. For QTL region 3, the association of a microsatellite marker located in the CSN3 gene with milk protein content (as reported in numerous previous genome scans) could be confirmed in our study. Regarding QTL region 2, in the GH data set experimentwise significant associations with milk fat and milk protein yield were found for marker BM143 (FDR= 0.015) and SNP RW070 located 5’ upstream of the PPARGC1A gene (FDR 0.08, Fig. 1). These associations are supported by nominally significant effects of five additional SNPs in the interval RW011-DIK4482 on milk yield traits suggesting a possible regulatory function of RW070 for milk fat synthesis.

Figure 1: Additive allelic effects of the SNP RW070 on milk production traits in the GH population

Furthermore, a nominally significant association was found between milk fat yield and the non-synonymous SNP RW023 located in exon 9 of the PPARGC1A gene leading to an amino acid exchange in the protein sequence (P616L). In a previous study a significant increase in the frequency of the allele favourable regarding milk fat yield had also been observed in maternally inherited alleles from dams highly selected for milk production traits compared to a control population (Weikard et al. 2005). In our present study this association was detected in both experimental cattle populations, GH and GF. However, the frequency of the trait-associated allele of RW023 in GF was not sufficiently high to provide evidence for a significant association. Nevertheless, the identical direction of the effects observed in both breeds and the modification of the protein structure due to the amino acid exchange suggest that this mutation presumably may have a trait-affecting impact, which should be reconsidered in further studies. No associations with milk composition traits were observed with markers associated with yield traits, which again underlines that the two QTL in the middle region of BTA6 are unambiguously different regarding the affected traits.

Conclusions

The results of the study reconfirm that in the GH population two distinct QTL are located in relatively close vicinity in the middle region of BTA6, which are not in LD and differ unambiguously regarding their effects on discrete milk production traits. The highly
significant association of the gene variant \textit{ABCG2 Y581S} with milk composition traits provides evidence for the causative molecular relevance of the \textit{ABCG2} gene in QTL region 1 for this breed. In addition, significant and tentative, respectively, associations of \textit{PPARGC1A} gene variants \textit{RW070} and \textit{RW023} with milk yield traits support functional candidacy of \textit{PPARGC1A} as molecular background underlying the QTL in region 2, although the results of our previous study reporting the association of the intronic gene variant c.1892+19T>C with milk fat yield could not be verified. For the GF population, the analysis of the multiple QTL architecture on the molecular level did not provide identical results, except for \textit{RW023}, due to fixed alleles of the respective gene variants and/or limited number of individuals in this study.

The across-breed results of our study indicate exemplarily as shown by the \textit{ABCG2 Y581S} variant that a SNP that had been identified as the most likely causal mutation for milk composition traits in several tested dairy cattle breed populations could possibly have a low MAF or could be fixed in another selected target breed. These results highlight issues requiring consideration when optimizing prediction equations for genomic selection schemes: 1) Rare mutations currently neglected might exert major relevant effects. 2) QTL architecture may vary between populations, which will have an impact on the efficiency of across-breed prediction equations. 3) For pinpointing rare causative variants presumably high density SNP panels will be required to detect SNPs in LD with the causal mutation. Furthermore, our results support multi-breed reference populations as a valuable resource for mapping QTL as suggested by Goddard and Hayes (2009).

**References**


