β-Lactoglobulin content of bovine milk is affected by multiple mutations on BTA11

H. Bovenhuis1 and N. Bedere1.
1Animal Breeding and Genomics Centre, Wageningen University, Wageningen, the Netherlands.

ABSTRACT: About 80% of the differences in β-lactoglobulin content of milk are due to genetic factors. Results from a GWAS showed that a region on BTA11 which contains the β-lactoglobulin gene has a major effect on β-lactoglobulin content. In this study we fine map this region using imputed 777k SNP data and we constructed haplotypes based on QTag-SNPs in order to capture the genetic variation associated with this region. We identified 4 haplotypes (A1, A2, B1 and B2) with significantly different effects on β-lactoglobulin content. This suggests that this region contains multiple mutations that affect milk β-lactoglobulin content. The difference in β-lactoglobulin content between the extreme haplotype groups A1A1 and B2B2 is 3.2%. The haplotypes explain 93% of the genetic variation in β-lactoglobulin content and therefore provide an efficient selection tool to reduce β-lactoglobulin content of milk and to increase cheese production.

Keywords: dairy cattle; β-lactoglobulin; fine mapping; haplotypes

Introduction

Milk is a unique product that contains many essential ingredients such as vitamins, minerals, essential amino acids and fatty acids (German et al. (2002)). Most dairy cattle breeding goals focus on improving the yields of milk volume, fat and protein (Miglior et al. (2005)) and detailed milk composition is not part of the breeding goals. However, it might be of interest to select for cows that produce milk with a specific composition which has added value for the dairy industry. E.g. a low β-lactoglobulin (β-lg) content of milk is associated with higher cheese production (van der Berg et al. (1992), Wedholm et al. (2006)) and reduced fouling rate of heating equipment (Elofsson et al. (1996)). A further reason for reducing β-lg content is that β-lg is known to be a major milk allergen (Wal, (1998)).

Schopen et al. (2009) showed that β-lg content of milk is strongly affected by genetic factors: the heritability is 0.80. A genome wide association study by Schopen et al. (2011) identified a region on BTA11 with a major effect on β-lg content. This region contains the β-lg gene which codes for the β-lg protein. Aschaffenburg and Drewry (1955) were the first to describe two distinct forms of the β-lg protein (A and B). Several studies showed that β-lg protein variants A and B are associated with β-lg content: the β-lg B variant is associated with a lower β-lg content (e.g. Bobe et al. (1999); Lunden et al. (1997), Heck et al., (2009)).

Schopen et al. (2011) found that after adjusting for the β-lg protein variants a significant proportion of the genetic variance remains associated with this region. This suggest that the mutations responsible for the differences between β-lg A/B protein variants either are not the causal mutations or that this region contains multiple mutations with an effect on β-lg content. The recent availability of a high density (777k) SNP array enables to fine map the targeted region on BTA11 and investigate if one or multiple mutations are responsible for the observed effects.

This study aims to fine map a chromosomal region on BTA11, associated with β-lg content, using 777k SNP data and to investigate if one or multiple mutations are responsible for the observed effects. To this end we developed a strategy to build haplotypes that tags all genetic variation associated with a chromosomal region and tested for the number of haplotypes with distinct effects on β-lg content.

Materials and Methods

Animals and Phenotypes. The present study was part of the Dutch Milk Genomics Initiative. Morning milk samples of 1,713 Holstein Friesian heifers on 383 commercial herds were analyzed for detailed milk protein composition. The β-lg content was determined by CZE as described by Heck et al. (2008).

Genotypes. A 50k SNP chip developed by CRV (cooperative cattle improvement organization, Arnhem, the Netherlands) was used to genotype 1,736 cows as well as the sires of the cows using the Infinium assay (Illumina, USA). In addition, 55 of the sires of these cows were genotyped with the BovineHDbeadChip (777k, Illumina, USA). For imputing the 1,736 cows from 50k to 777k a reference population of 1,333 animals was available. The reference population included the 55 sires. Other animals in the reference population were provided by CRV. For imputation and phasing BEAGLE was used (Browning and Browning, 2009).

Additionally, genotypes for the two SNPs responsible for the amino acid changes in the β-lg protein variants A vs B and 8 other SNPs associated with β-lg content, identified by Ganai et al. (2008), were available for 1,611 cows. For 125 cows these SNP genotypes were missing and imputed and phased using BEAGLE (Browning and Browning, 2007). The positions of the SNPs were based on the Btau 4.2 assembly.

In total, 1,647 cows had both phenotypic and genotypic information and were used for the association study. Based on results by Schopen et al. (2011) we focused in the current study on the region from 75 Mb till 110 Mb on BTA11. In total 9,925 SNP genotypes were available of which 872 SNP were homozygous in our population and therefore not included in the association study.

Association study. The single SNP association study was performed using the following model:
where $y_{klmno}$ was the $\beta$-lg content, $\mu$ is the mean for $\beta$-lg content, $dim_{klmno}$ is the covariate describing the effect of days in milk, $ca_{klmno}$ is the covariate describing the effect of age at first calving, $season_{m}$ is the fixed effect of calving season ($k=1, 2$ or $3$), $scode_{i}$ is the fixed effect of sire group ($l=1, 2$ or $3$), $SNP_{m}$ is the fixed effect of the SNP, $animal_{n}$ is the random additive genetic effect of animal $n$, $herd_{o}$ is the random herd effect and $e_{klmno}$ is the random residual effect. The animal effects were assumed to be distributed as $N(0, \sigma_{A}^2)$, herd effects were assumed to be distributed as $N(0,1\sigma_{herd}^2)$ and the residuals were assumed to be distributed as $N(0, \sigma_{e}^2)$, where $A$ is the additive genetic relationships matrix and $I$ the identity matrix. In the association analysis, the variance components were fixed to estimates obtained from model [1] without the SNP effect. The statistical package ASReml (Gilmour et al., 2006) was used to perform the analysis.

**Haplotype Analysis**

To identify SNPs that capture the genetic variation in $\beta$-lg content associated with the tail part of BTA11 a stepwise approach was adopted. For this purpose we zoomed in on the region from 100 till 110 Mb -lg content associated with the tail of the genetic variation in $\beta$-lg protein. After the first analysis, the phenotype was adjusted for the effect of the most significant SNP, which contained 2,897 SNPs of which 313 were non-polymorphic. Subsequently variance components were re-estimated and the association study was repeated with variance components fixed at their new values. In analogy to “tag SNPs”, i.e. a limited set of SNPs that capture the genetic variation associated with a genomic region (Ballding, 2006), we defined “Q-Tag SNPs” as the set of SNPs identified by the described procedure that capture the genetic variation of a chromosomal region.

Haplotypes were constructed based on Q-Tag SNPs and effects of these haplotypes were estimated. The association of haplotypes with $\beta$-lg content was estimated using the following animal model:

$$y_{klmno} = \mu + \beta_1 dim_{klmno} + \beta_2 e^{0.05 dim_{klmno}} + \beta_3 ca_{klmno} + \beta_4 ca_{klmno} + season_{m} + scode_{i} + SNP_{m} + animal_{n} + herd_{o} + e_{klmno}$$

where $y_{klmno}$ was the $\beta$-lg content, $\mu$ is the mean for $\beta$-lg content, $dim_{klmno}$ is the covariate describing the effect of days in milk, $ca_{klmno}$ is the covariate describing the effect of age at first calving, $season_{m}$ is the fixed effect of calving season ($k=1, 2$ or $3$), $scode_{i}$ is the fixed effect of sire group ($l=1, 2$ or $3$), $SNP_{m}$ is the fixed effect of the SNP, $animal_{n}$ is the random additive genetic effect of animal $n$, $herd_{o}$ is the random herd effect and $e_{klmno}$ is the random residual effect. The animal effects were assumed to be distributed as $N(0, \sigma_{A}^2)$, herd effects were assumed to be distributed as $N(0,1\sigma_{herd}^2)$ and the residuals were assumed to be distributed as $N(0, \sigma_{e}^2)$, where $A$ is the additive genetic relationships matrix and $I$ the identity matrix. In the association analysis, the variance components were fixed to estimates obtained from model [1] without the SNP effect. The statistical package ASReml (Gilmour et al., 2006) was used to perform the analysis.

**Results and Discussion**

The average $\beta$-lg content of the milk samples was 3.50% (w/w%) and 8.34% of the protein consists of $\beta$-lg. The estimated heritability of $\beta$-lg content based on the current data set is 0.78 and the proportion of the variation explained by differences between herds is 0.05. These estimates are very similar to the values reported by Schopen et al. (2009) using largely the same data.

**Fine mapping.** Figure 1 shows the results of the association study based on 9053 SNPs located in the target region from 75 Mb till 110 Mb on BTA11. The lead SNP (rs110066229) i.e. Q-Tag SNP1 is located in the third exon of $\beta$-lg gene and is one of the 2 mutations responsible for the difference between $\beta$-lg protein variants A and B (Ganai et al., 2008). The effect of this SNP on $\beta$-lg content was highly significant as illustrated by the -log(P-value) (Figure 1). After adjusting $\beta$-lg content for the effect of Q-Tag SNP1 the estimated additive genetic variance dropped from 1.126 to 0.113. The association study based on $\beta$-lg content adjusted for Q-Tag SNP1 resulted in a highly significant effect for SNP rs110144148 with a -log(P-value) of 12.9. This SNP will be termed Q-Tag SNP2. The highly significant effect of Q-Tag SNP2 shows that not all variation associated with this chromosomal region is captured by the difference between the $\beta$-lg protein variants A and B. Q-Tag SNP2 is located at 107.3Mb, i.e. distal from the lactoglobulin gene.

![Figure 1 Fine mapping the tail part of the BTA 11 for $\beta$-lg content using imputed 777k SNP genotypes.](image)

**Haplotypes.** Q-Tag SNP1 and Q-Tag SNP2 were used to construct 4 possible haplotypes: GG, GA, AG and AA where the first letter refers to Q-Tag SNP1 and the second to Q-Tag SNP2 (Figure 2). The estimated allelic effects of Q-Tag SNP1 ($\beta$-lg protein variants A and B) and the haplotypes are shown in Figure 2. The statistical analysis implicitly assumes additive allelic effects which seems justified for the effect of this chromosomal region (e.g. Heck et al. (2009)). It is estimated that having one copy of the $\beta$-lg protein variant A results in a 1.43% higher $\beta$-lg content as compared to having $\beta$-lg protein variant B. This corresponds to a difference between AA and BB (i.e. having two copies) of 2.86%. This agrees with the difference of 2.84% reported by Heck et al. (2009) using...
largely the same data as the current study but based on β-lg genotype effects.

![Diagram of haplotypes and allelic effects](image)

**Figure 2** Estimated allelic effects for β-lg protein variants and for haplotypes on β-lg content. Percentages indicate the frequencies of the alleles/haplotypes.

Pairwise comparison of the predicted haplotype effects show that all four haplotypes have significantly different effects on β-lg content. The GG and the GA haplotypes differentiate β-lg protein variant A and are therefore referred to as A₁ (GG) and A₂ (GA) (Figure 2). The two other haplotypes AG and AA differentiate β-lg protein variant B and are referred to as B₂ (AG) and B₁ (AA). The estimated difference between haplotype GG and AA is 1.6% and therefore the expected difference between cows with haplotype combination A₁A₁ and B₁B₁ is 3.2%. When modelling the haplotypes as random effects we estimated that they explain 93% of the genetic variation in β-lactoglobulin content.

The effect of the haplotypes on protein content was not significant (p=0.06) but there was a highly significant effect of the haplotypes on casein content and the casein index. The casein index can be seen as a proxy for the efficiency of cheese production. The difference in casein index between protein variants AA and BB is 3.15%. Van Den Berg et al. (1992) showed that β-lg BB milk results in approximately 3% more cheese as compared to β-lg AA milk. The difference in casein index between A₁A₁ and B₁B₁ haplotype groups is 3.57%. These results indicate that milk of B₁B₁ cows has the same protein content than milk of A₁A₁ cows but contains less β-lg and more casein which makes milk of B₁B₁ cows more suitable for cheese production.

**Conclusion**

In this study we fine mapped a chromosomal region on BTA11 associated with β-lg content using 777k SNP data. The lead SNP is one of the 2 mutations responsible for the difference between β-lg protein variants A and B. However, not all variation associated with this chromosomal region is captured by the difference between the β-lg protein variants A and B. Results show that 4 haplotypes with different effects on β-lg content can be distinguished: A₁, A₂, B₁, and B₂. The presence of 4 haplotypes suggest that there at least 2 mutations in this regions responsible for the observed effects. In this study we did not identify the causal mutations but we described a procedure to identify haplotypes with distinct effects on the target trait. These haplotypes can be used to efficiently select cows producing milk with lower β-lg content and higher cheese yield, even if knowledge on causal mutations is lacking.

**Literature Cited**


