Genome Wide Association For Casein Index In Milk Of Dairy Cattle

M.H.P.W. Visker[*], G.C.B. Schopen*, J.A.M. van Arendonk* and H. Bovenhuis*

Introduction

The value of bovine milk for human consumption is mainly determined by the content and composition of the milk fat and milk protein. The protein fraction of bovine milk comprises six main proteins: the four caseins (α_{s1}-casein (α_{s1}-CN), α_{s2}-casein (α_{s2}-CN), β-casein (β-CN) and κ-casein (κ-CN)) and the two whey proteins (α-lactalbumin (α-LA) and β-lactoglobulin (β-LG)).

Caseins are important for cheese making; their amounts and relative concentrations determine cheese yield and influence technological properties such as coagulation (Wedholm et al. 2006). Variation in the casein content of milk is highly associated with variation in β-LG content; low concentration of β-LG protein is associated with high concentration of casein (Lundén et al. 1997; Schopen et al. 2009a). A large part of the variation in β-LG content is associated with β-LG protein variants, of which variants A and B are common in most cattle breeds (Farrell et al. 2004). Whey proteins have long been regarded as low-value by-products of cheese production. Although interest for whey is increasing because of its high nutritional value, the dairy industry still prefers milk with high contents of caseins and low contents of whey proteins in order to increase the yield of cheese.

Despite its relevance for the dairy industry, little is known about the genetic background of casein content in milk. It has been shown that casein index is heritable ($h^2 = 0.70$; Schopen et al. 2009a), and a recent linkage analysis (Schopen et al. 2009b) has identified a highly significant QTL for casein index on *Bos taurus* autosome (BTA) 11 and suggestive QTL on BTA 6 and 7. In the present study we have applied a genome wide association analysis for casein index in bovine milk, with the aim to confirm and fine map the QTL that were identified in the linkage analysis, and to possibly identify additional chromosomal regions related to casein index.

Materials and Methods

Population. As part of the Dutch Milk Genomics Initiative, morning milk samples and bloods samples were collected from 1,912 first lactation Dutch Holstein Friesian cows. Cows belonged either to five large or 50 small paternal half-sib families. More details about the population are given by Schopen et al. (2009a). Blood was used to extract genomic DNA. Milk samples were used for determination of, among others, detailed milk protein composition.

[*]Animal Breeding and Genomics Centre, Wageningen University, P.O. Box 338, 6700 AH Wageningen, the Netherlands

Genotypes. A 50K SNP chip was designed by CRV (Arnhem, the Netherlands) and obtained from Illumina (San Diego, CA, USA), and used to genotype all animals with the Infinium assay (Illumina). This approach resulted in 50,856 technically successful SNPs, which were mapped using the bovine genome assembly (BTAU4.0; Liu et al. 2009). The 589 SNPs that mapped to chromosome X, the 231 monomorphic SNPs, and the 393 SNPs with a genotyping rate < 80% were not included in the association analysis. Genotypes of the remaining 49,643 SNPs were available for 1,868 animals.

Phenotypes. Detailed milk protein composition was determined by capillary zone electrophoresis (CZE) as described by Heck et al. (2008). Relative concentrations of the four caseins: α_{s1}-CN, α_{s2}-CN, β-CN and κ-CN, and two whey proteins: α-LA and β-LG were expressed as percentage of total protein weight (w/w). Casein index was calculated as: casein index = (Σcasein / (Σcasein + Σwhey)) x 100, with Σcasein defined as the sum of relative concentrations of α_{s1}-CN, α_{s2}-CN, β-CN and κ-CN, and Σwhey defined as the sum of relative concentrations of α-LA and β-LG (Schopen et al. 2009a).

Not all phenotyped animals had also genotypes, and not all genotyped animals had also phenotypes. As a result, the dataset used in the association analysis consisted of 1,713 animals with both genotypes and phenotypes.

Association analysis. Genome wide association was analyzed in two steps. In the first step, a single SNP analysis was performed using the SNPassoc package (González et al. 2007) in R with the linear model:

$$y^*_{ij} = sire_i + SNP_j + e_{ij} \qquad (1),$$

where y* was the phenotype adjusted for the systematic environmental effects: days in lactation, age at first calving, calving season and herd; sire was the fixed effect of sire; SNP was the fixed effect of SNP genotype; and e was the random residual ($e_{ij} \sim N(0,\sigma^2_e)$). The systematic environmental effects that were used to adjust the phenotypes were estimated using an animal model in ASReml (Gilmour et al. 2002) for all 1,912 cows with phenotypes, as described by Schopen et al. (2009a). The sire effect was included in the model to account for family effects.

The genomewise false discovery rate (FDR) was obtained using the qvalue package (Storey and Tibshirani 2003) in R, which used the *P* values obtained from the linear model for all 49,643 SNPs. Associations with a FDR < 0.01 were considered significant.

In the second step of the genome wide association analysis, all regions containing SNPs that were significantly associated with casein index were further analyzed using ASReml (Gilmour et al. 2002), to account for all genetic relationships among animals, with the animal model:

$$y_{ijklmno} = \mu + b_1 * lactst_i + b_2 * e^{-0.05 * lactst_i} + b_3 * ca_j + b_4 * ca_j^2 + season_k + scode_l + SNP_m + herd_n + animal_o + e_{ijklmno} \qquad (2),$$

where y was the (unadjusted) phenotype, μ was the overall mean, lactst was the covariate describing the effect of days in lactation, ca was the covariate describing the effect of age at first calving, season was the fixed effect of the class of calving season (June – August 2004,

September – November 2004, or December 2004 – February 2005), scode was the fixed effect accounting for differences in genetic level between groups of proven bull daughters and young bull daughters, SNP was the fixed effect of SNP genotype, herd was the random effect of herd, animal was the random additive genetic effect of animal, and e was the random residual. The variance-covariance structure of additive genetic effects was Var(animal) = $A\sigma^2a$, where A was the matrix of additive genetic relationships between animals and σ^2a was the additive genetic variance.

Results and discussion

Association analysis for casein index in bovine milk was performed using genotypes of 49,643 SNPs for 1,713 animals. Significant associations (FDR < 0.01) were detected on BTA 2, 6, 11, and 20 (Figure 1) when using model (1). Associations remained significant on BTA 6 and 11 when accounting for all genetic relationships among animals using model (2).

On BTA 6 the SNP most significantly associated with casein index was positioned at 88.5 Mbp and explained 2.8 % of the genetic variance. On BTA 11 the SNP most significantly associated with casein index (-log10(*P* value) = 134.37) was positioned at 107.2 Mbp and explained 60.6 % of the genetic variance.

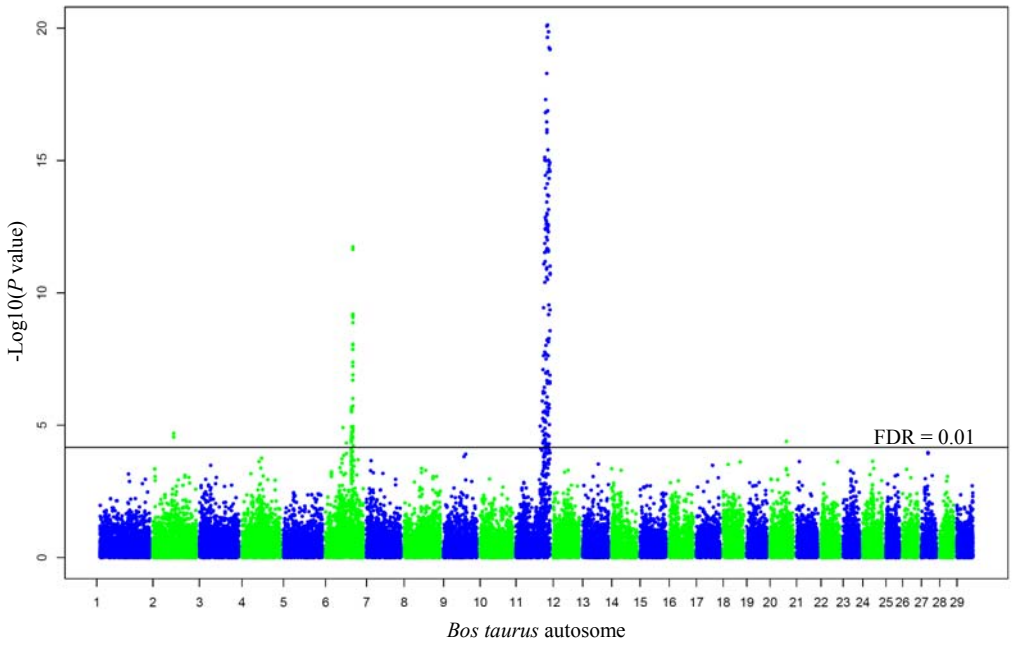

Figure 1: The –log10(P values) for the association of 49,643 SNPs with casein index in bovine milk, on all 29 Bos taurus autosomes. The horizontal grey line corresponds with the threshold level of FDR = 0.01. All -log10(P values) > 20 are not shown.

The association with casein index on BTA 6 was near the location of the casein genes, while the association on BTA 11 was near the location of the *β-LG* gene. Additional analyses showed that the known protein variants of β-CN (A1, A2, B, I) and κ-CN (A, B, E) could almost completely explain the association with casein index on BTA 6, while the protein variants of β-LG (A, B) could almost completely explain the association on BTA 11.

The results of this association analysis confirm the relationship between the contents of caseins in milk and the content of β-LG protein: the most significant association with casein index was on BTA 11 near the location of the *β-LG* gene. In addition, the relationships between the protein variants of β-LG and β-LG content and, consequently, casein content were confirmed: β-LG protein variants A and B accounted almost completely for the association for casein index on BTA 11. Importantly, no significant associations with protein content or yield were detected on BTA 11 (results not shown).

The association with casein index on BTA 11 is consistent with the QTL that was found in the linkage analysis of Schopen et al. (2009b). In addition, the suggestive QTL that was detected on BTA 6 in the linkage analysis was confirmed in the current association analysis, whereas the suggestive QTL on BTA 7 was not. The association analysis did not reveal any new chromosomal regions related to casein index.

Casein index is a measure for the proportion of milk protein present as casein. A higher casein index indicates that the same amount of milk protein will result in a higher cheese yield. Our results show that casein index is predominantly influenced by only one of its underlying components, i.e. β-LG; the role of the caseins is limited, while a role of α-LA seems absent. Thus, increasing cheese yield by means of increasing casein index without affecting protein content can be achieved most efficiently by influencing the content of β-LG, specifically through the use of protein variants A and B.

References

Farrell, H.M., Jimenez-Flores, R., Bleck, G.T. *et al.* (2004) *J. Dairy Sci.*, 87:1641-1674.

Gilmour, A.R., Gogel, B.J., Cullis, B.R. *et al.* (2002) ASReml User Guide Release 1.0. VSN International Ltd, Hemel Hempstead, UK.

González, J.R., Armengol, L., Solé, X. *et al.* (2007) *Bioinformatics*, 23:644-645.

Heck, J.M.L., Olieman, C., Schennink, A. *et al.* (2008) *Int. Dairy J.*, 18:548-555.

Liu, Y., Qin, X., Song, X-Z.H. *et al.* (2009) *BMC Genomics*, 10:180.

Lundén, A., Nilsson, M., and Janson, L. (1997) *J. Dairy Sci.*, 80:2996-3005.

Schopen, G.C.B., Heck, J.M.L., Bovenhuis, H. *et al.* (2009a) *J. Dairy Sci.*, 92:1182-1191.

Schopen, G.C.B., Koks, P.D., Van Arendonk, J.A.M. *et al.* (2009b) *Anim. Genet.*, 40:524-537.

Storey, J.D., and Tibshirani, R. (2003) *Proc. Natl. Acad. Sci. USA*, 100:9440–9445.

Wedholm, A., Larsen, L.B., Lindmark-Månsson, H. et al. (2006) *J. Dairy Sci.*, 89:3296-3305.