Fine-mapping of a candidate region associated with milk-fat composition on *Bos Taurus Autosome* 17

***S. I. Duchemin*[*,†], M. H. P. W. Visker[*], J. A. M. Van Arendonk[*], and H. Bovenhuis[*]**
[*]Animal Breeding and Genomics Centre, Wageningen University, Wageningen, Netherlands, [†]Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

ABSTRACT. A genomic region associated with milk-fat composition has been detected on BTA17. However, no candidate gene or causal variant has been identified so far. Our goal was to fine-map this region by using imputed 777k SNP genotypes. Phenotypes and genotypes were available on 1,581 animals. Phenotypes consisted of gas chromatography measurements of C8:0, and genotypes consisted of the imputed 777k SNP. Single SNP analyses were run using an animal model to retrieve significant SNP. Based on these SNP, haplotypes were constructed. Haplotypes suggested the presence of one QTL within the fine-mapped region. This QTL explained 9% of the genetic variation in C8:0. Although many genes are present in this QTL region, most of these genes have not been characterized yet. Therefore, no clear candidate gene associated with milk-fat composition could be identified.
Keywords: Cattle, Milk fatty acids, High density genotyping

Introduction

Bovine milk-fat is composed of more than 400 different fatty acids (FA; Jensen, 2002). Variation in the content of several identified FA in milk is affected by genetic factors. Stoop et al. (2008) showed that individual FA have heritability estimates that range from 0.22 to 0.71.

Some genes have been recognized as having large effects on milk-fat composition, such as *acyl-CoA: diacylglycerol acyltransferase1* (***DGAT1***) located on BTA14, and *stearoyl-CoA desaturase1* (***SCD1***) located on BTA26 (e.g., Schennink et al., 2008). In addition, several other regions of the bovine genome have been associated with milk-fat composition but have not been characterized yet (e.g., Bouwman et al., 2012). By fine-mapping these regions, it might be possible to identify candidate genes (Ishii et al., 2013) for milk-fat composition.

A region significantly associated with fatty acids in milk has been detected on BTA17 (Bouwman et al., 2012). However, no candidate gene or causal variant has been identified so far. The aim of our study was to fine-map this candidate region on BTA17 by using imputed 777k SNP genotypes.

Material and Methods

Animals and Phenotypes. Morning milk samples were retrieved from 1,581 first-lactation *Holstein-Friesian* cows from 398 herds throughout the Netherlands between May and June 2005. At least 3 cows per herd were sampled. Milk fatty acids were measured by gas chromatography, as described by Schennink et al. (2007), and were measured as weight proportion of total fat (%wt/wt). Here we focus on C8:0.

Genotypes and Imputation. The DNA of 55 sires and 1,795 daughters belonging to our experimental population was genotyped with a 50k SNP chip (Illumina, San Diego, CA). A reference population of 1,333 animals, belonging to CRV (Arnhem, the Netherlands) and including the 55 sires, was genotyped additionally with a 777k SNP chip (Illumina). Our experimental population was imputed to 777k SNP genotypes with Beagle version 3.2.2 (Browning and Browning, 2009). After the exclusion of animals with pedigree errors, 1,581 animals were imputed from 50k to 777k SNP genotypes. Imputation of BTA17 increased the number of SNP genotypes from 1,584 (50k) to 22,240 (777k). The positions of the imputed SNP were based on the bovine genome assembly UMD 3.1.

Fine-Mapping. The fine-mapping was performed by running single SNP analyses with the following animal model:

$$y_{ijklmno} = \mu + b_1 * dim_{ijklmno} + b_2 * e^{-0.05*dim_{ijklmno}} + b_3 * afc_{ijklmno} + b_4 * afc^2_{ijklmno} + season_k + scode_l + SNP_m + herd_n + a_o + e_{ijklmno} \quad (1)$$

where y is the dependent variable; μ is the overall mean; *dim* is the covariate describing the effect of days in milk, modeled with a Wilmink curve (Wilmink, 1987); *afc* is the covariate describing the effect of age at first calving; *season* is the fixed effect of calving season (June – August 2004, September – November 2004, or December 2004 – February 2005); *scode* is the fixed effect accounting for differences in genetic level between groups of proven bull daughters and young bull daughters; *SNP* is the fixed effect of SNP

genotype; *herd* is the random effect of herd and is assumed to be distributed as ~N(0, $\mathbf{I}\sigma^2_{herd}$); a is the random additive genetic effect of animal and is assumed to be distributed as ~N(0, $\mathbf{A}\sigma^2_a$); **A** is the relationship matrix which is based on pedigree information of 12,548 animals (4 generations) provided by CRV; and *e* is the random residual effect and is assumed to be distributed as ~N(0, $\mathbf{I}\sigma^2_e$). Additive genetic and herd variances were estimated with ASReml 3.0 (Gilmour et al., 2009) without the inclusion of SNP information, and the resulting estimates were fixed in model (1).

Haplotype Analysis. The SNP with the highest significance was defined as "QTagSNP1", i.e., a SNP that tagged the variation associated with the chromosomal region. We corrected for QTagSNP1 in model (1) by including it as a fixed effect, and ran a second round of single SNP analyses. After this second round, if other SNPs became significant, the SNP with the highest significance was defined as "QTagSNP2". We corrected for QTagSNP2 in the already extended model with QTagSNP1, by further including QTagSNP2 as a fixed effect. This procedure was repeated until no additional significant SNP were retrieved. A SNP was considered significant at –Log_{10}(P-value) ≥3.0. Finally, we constructed haplotypes based on the identified QTagSNP.

We estimated the effects of haplotypes by running the following animal model:

$$y_{ijklmnop} = \mu + b_1 * dim_{ijklmno} + b_2 * e^{-0.05*dim_{ijklmno}} + b_3 * afc_{ijklmno} + b_4 * afc^2_{ijklmno} + season_k + scode_l + haplo1_m + haplo2_n + herd_o + a_p + e_{ijklmnop} \quad (2)$$

where all variables are as previously described for model (1), except for: *haplo1* is the random effect of the first haplotype and is assumed to be distributed as ~N(0, $\mathbf{I}\sigma^2_{haplo}$); and *haplo2* is the random effect of the second haplotype and is assumed to be distributed as ~N(0, $\mathbf{I}\sigma^2_{haplo}$). The first and second haplotypes were jointly used to estimate the haplotype variance σ^2_{haplo}. Additionally, haplotypes were tested to check whether they differed from each other. Differences were assessed using Student's t-tests at a significance level of α=0.05.

Candidate genes. Candidate genes for BTA17 were retrieved from the National Center for Biotechnology Information (**NCBI**) website.

Results and Discussion

Fine-mapping. The fine-mapping of BTA17 was possible because we combined high-density SNP genotyping with imputation. Imputation increased the number of SNP from the 50k to the imputed 777k by at least 10 times. Bouwman et al. (2012) analyzed the same data by using 50k SNP genotypes, and identified a QTL associated with C8:0. With the imputed 777k SNP genotypes, we found new SNP that were in stronger association with C8:0 than the SNP found by Bouwman et al. (2012). With these new significant SNP, the location of the QTL was refined to the region located between 29.0 and 34.0 Mbp on BTA17 (Figure 1).

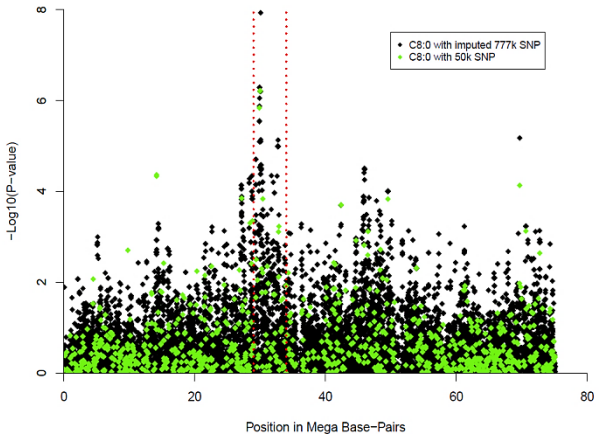

Figure 1: Fine-mapping of BTA17 associated with C8:0. Associations with C8:0 are overlaid between the imputed 777k SNP and the 50k SNP genotypes. The marked region between red dotted lines (29.0 to 34.0 Mbp) is the region we focused on to refine the location of the QTL.

Construction of Haplotypes. Haplotypes were constructed based on QTagSNP in this region. QTagSNP1 (*rs109290136*) was significantly associated with C8:0 (-Log_{10}(P-value)= 7.86), and had a minor allele frequency of 0.44. Additionally, QTagSNP2 (*rs135465158*) was significantly associated with C8:0 (-Log_{10}(P-value)=3.19), and had a minor allele frequency of 0.14. After adjusting for both QTagSNP, no other significant SNP was found. A total of four haplotypes were constructed based on the two QTagSNP. Frequencies of haplotypes were 0.12 for A-A, 0.44 for A-G, 0.01 for C-A, and 0.42 for C-G. These haplotypes explained 9.0% of the genetic variance in C8:0. After testing for differences between haplotypes, we concluded that estimated effects for 3 out of 4 haplotypes did not differ from each other. Thus, the haplotypes could be divided in 2 groups with

distinct effects on C8:0: A-G versus the remaining haplotypes. The existence of two groups of haplotypes with distinct effects can be explained by one causal variant. However, we cannot exclude the possibility of multiple causal variants in strong LD.

Candidate genes. We aimed at finding candidate genes associated with C8:0 on BTA17. C8:0 is a FA that is synthesized de novo within the mammary gland of a cow. De novo synthesis is a process by which a FA is elongated from precursors (e.g., C2:0 or C3:0), by adding C2:0. This elongation ends at C16:0 or at C17:0, depending on which were the precursors.

Not all genes on BTA17 have been characterized yet. Among the characterized genes, two candidate genes might be: *AACS* (*acetoacetyl-CoA synthetase*) and *ACADS* (*acyl-CoA dehydrogenase, C2:0 to C3:0 short chain*). These genes seem to be related to fatty acids metabolism but are located outside our QTL region (29.0 to 34.0 Mbp).

From the 29 genes found within our QTL region on BTA17, 18 genes have not been characterized yet. Among the 11 characterized genes, a gene of interest might be *PGRMC2* (Progesterone Receptor membrane component 2) gene. *PGRMC2* is located between 29.87 and 29.89 Mbp on BTA17, and is the closest gene nearby the most significant association with C8:0 retrieved in this study. This gene belongs to the Superfamily *Cytochrome b5-like heme/steroid binding domain*. This superfamily is involved in the fatty acid metabolic process, and oxido-reductase activity. However, *PGRMC2* gene has not been associated to milk-fat composition in dairy cattle.

Conclusion

A QTL that explains 9% of genetic variance in C8:0 had its location refined in this fine-mapping. Because the region of this QTL on BTA17 is still under characterization, no clear candidate gene could be identified. However, the PGRMC2 gene might be an interesting candidate gene.

Acknowledgements

This study is part of the Dutch Milk Genomics Initiative, funded by Wageningen University, the Dutch Dairy Association (NZO), CRV and the Dutch Technology Foundation STW. The first author currently benefits from a joint grant from the European Commission and Cooperative Cattle Improvement Organization CRV (Arnhem, the Netherlands), within the framework of the Erasmus-Mundus joint doctorate “EGS-ABG”.

Literature Cited

Bouwman, A. C., Visker, M. H. P. W., Van Arendonk, J. A. M. et al. (2012). BMC Genetics, 13:93.

Browning, B. L. and Browning, S. R. (2009). Am. J. Hum. Genet., 84:210-223.

Gilmour, A. R., Gogel, B. J., Cullis, B. R. et al. (2009). ASReml Release 3.0.VSN International Ltd, Hempstead, UK.

Ishii, A.,Yamaji, K. , Uemoto, Y. et al. (2013). Anim. Sci. J., 84:675-682.

Jensen, R. G. (2002). J. Dairy Sci., 85: 295–350.

National Center for Biotechnology Information. (2014). http://www.ncbi.nlm.nih.gov/. Accessed on Jan. 2014.

Schennink, A., Heck, J. M. L., Bovenhuis, H. et al. (2008). J. Dairy Sci. 91:2135–2143.

Schennink, A., Stoop, W. M., Visker, M. H. P. W. et al. (2007). Anim. Genet., 38:467-473.

Stoop, W. M., Van Arendonk, J. A. M., Heck, J. M. L. et al. (2008). J. Dairy Sci., 91:385-394.

Wilmink, J. B. M. (1987). Livest. Prod. Sci.16:335–348.