CEACAM18 as candidate for the Holstein calving QTL on BTA18

***Xiaowei Mao*[12], Naveen Kumar Kadri[1], DirkJan de Koning[2], Goutam Sahana[1], Bernt Guldbrandtsen[1]**
[1]Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University
[2]Swedish University of Agricultural Sciences, Uppsala, Sweden

ABSTRACT: A genome-wide association study of Nordic Holstein cattle SNP chip genotypes and imputed next generation sequencing (NGS) variants identified carcinoembryonic antigen-related cell adhesion molecule 18 (*CEACAM18*) on BTA18 as the gene most highly significantly associated with direct calving traits. Phenotypes used were estimated breeding values (EBV) for six direct calving traits and one compound index trait. A SNP by SNP mixed model approach was first applied using HD genotypes. Haplotypes in the significant region were fitted in a mixed model. Finally, NGS variants in the significant region were utilized to precisely locate causative mutations. Results identified 21 QTL regions associated with one or more calving traits on 16 autosomes. These findings contribute to an improved understanding of the genetic architecture of the calving traits. They may help in improving calving performance in dairy breeding programs.
Keywords: Calving**;** Holstein**;** Genome-wide association

Introduction

Reduced calving performance not only causes considerable economic loss due to veterinary treatment costs, calf loss and lower production of cows affected by dystocia, but also results in reduced animal welfare. Due to influences from both calf and maternal genetic components, calving performance is analyzed in terms of direct and maternal components. The direct genetic calving effect is the effect of the calf's genotype, whereas the maternal genetic effect is the effect of the dam's genotype on calving. Several linkage-based QTL mapping studies and genome-wide association study (GWAS) have detected quantitative trait loci (QTL) underlying calving traits (Thomasen et al. 2008; Cole et al. 2011; Sahana et al. 2011). These studies indicate that at approximately 57Mb on BTA18 a QTL with a large effect on direct calving traits is segregating in Holstein cattle. The objective of the present study was to detect QTL affecting six direct calving traits and one compound index trait in the Nordic Holstein population to a higher precision utilizing HD genotypes and imputed NGS variants.

Materials and Methods

Animals and phenotypes. EBVs for calving traits were used as response variables. They were predicted using best linear unbiased prediction (BLUP) procedures with a sire model. EBVs for direct genetic effect of three traits: calving ease (CE), calf size (CS) and stillbirth (SB) were analyzed. Two EBVs for direct effect were predicted for each of these three traits (CE, CS and SB): first (F) and later (L) (combined second up to fifth) parities. For instance, EBV for a direct genetic effect (D) for calve size in first parity was denoted as DCSF. Besides, a birth index (BI) was calculated, which is a compound index describing a sire's total direct additive genetic effect by combining DSBF, DSBL, DCEF, DCEL, DCSF, and DCSL. The sub-traits were combined by relative economic weights (see http://www.nordicebv.info).

Imputation of HD genotypes and NGS variants. The imputation of 50k SNP to HD genotypes and NGS variants was done in two steps. 50k genotypes (46,702 SNPs after quality control) for 12,322 Nordic bulls were imputed to HD genotypes(734,077 SNPs) using IMPUTE2 (Howie et al. 2009). A reference population with HD genotypes was available for 2,036 bulls (902 Holstein, 735 Nordic Red and 399 Danish Jersey). In the second step, whole genome sequences from 529 dairy cattle from a combination of data from Aarhus University and from the 1,000 Bull Genomes Project were used as reference population. The imputed HD genotypes were imputed to the NGS level (25,607,739 variants) using IMPUTE2. In our final dataset, there were 5,571 animals which have both EBV and genotype.

Single marker analysis for HD genotypes. A mixed model analysis (MMA) (Yu et al. 2006) was applied to find SNPs associated with the EBVs. SNP was successively fitted in the MMA as fixed effects along with a polygenic component in the model. The software package DMU (available at http://dmu.agrsci.dk) was used for the analysis. Testing the significance for association between each marker and trait was done by a t test against null model with a marker effect of zero. For the purpose of reducing family-wise error rate (FWER) caused by multiple testing, a Bonferroni correction which was done by $-\log_{10}(0.05/\text{total number of markers})$ was applied. The number of markers used in the Bonferroni correction was 664,204.

Test for multiple QTL on BTA18. In order to examine whether multiple QTL associated with EBV are segregating in the targeted regions, the highly significant SNP or SNPs from results obtained using above model were added in the MMA as a cofactor. The remaining SNPs were successively tested in the model as fixed effects.

Candidate Haplotype analysis for HD genotypes. Using marker haplotypes adds power and precision to QTL mapping compared to just using single markers (Akey et al. 2001). The analysis was done in three steps to find the causative haplotype. First, the linkage phase of the markers was inferred using BEAGLE (Browning and Browning (2007)). Next, Haploview (Barrett et al. 2005)

was used to construct and visualize haplotype blocks. Then, to avoid problems due to rare haplotypes, haplotype blocks in the significant regions were fitted in the linear mixed model as random effects to identify the most associated region. Finally, each haplotype within the haplotype block was compared to the other haplotypes clustered together.

Single marker analysis for NGS variants on BTA18. To further narrow down the QTL region and to identify the causative mutation, NGS variants within a significant region identified by previous analysis were analyzed by successively fitting each variant in the MMA as a fixed effect. Bonferroni correction was applied to calculate the threshold defining significance.

Results and Discussion

Single marker analysis for HD genotypes. Single marker association analysis revealed 21 QTL regions associated with one or more traits on 16 autosomes (Figure 1). The QTL region on BTA 18 had the most significant association with all seven traits. In that region, SNP rs136283363 was the most significant across all seven traits. For BI, SNP rs136283363 had the lowest P-value (2.04×10^{-59}). SNP rs136283363 had an allele substitution effect (G to A substitution) of half an additive genetic standard deviation for all traits. For each extra A allele, the breeding values for DCEF, DCEL, DSBF, DSBL and BI rose, while decreasing values for DCSF and DCSL. Thus bigger calf size was associated with more difficult calvings and increased risk of still birth. When fitting SNP rs136283363 as a cofactor in the model, no other SNP on BTA18 remained significant. This suggests that only a single QTL was segregating on BTA18 in the Nordic Holstein population.

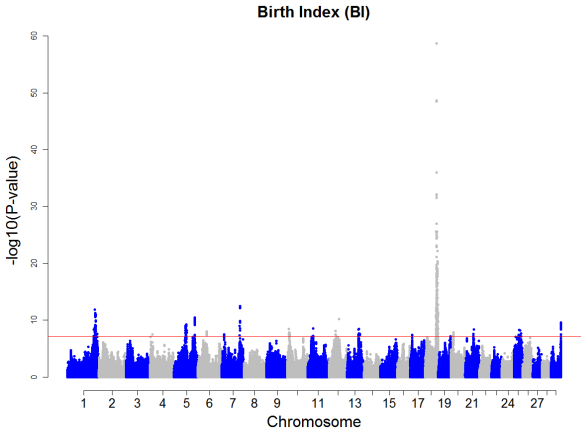

Figure 1. Genome-wide scan for Birth Index with -$\log_{10}$ of the P-value analysis for association with SNP using HD genotypes. Chromosomes are shown in alternating shades for clarity. The red line demonstrated the genome-wide threshold for the present study.

Haplotype analysis for HD genotypes. The corresponding associated haplotype was GGGAAGA ($P = 3.23 \times 10^{-59}$) which were located from left to right at 57,520,290, 57,523,817, 57,527,294, 57,531,340, 57,535,046, 57,548,213 and 57,565,406 bp on BTA18. The genomic region for haplotype GGGAAGA covers the carcinoembryonic antigen-related cell adhesion molecule 18 gene (*CEACAM18*) and part of cytoplasmic tRNA 2-thiolation protein 1 gene (*CTU1*). *CEACAM18* gene is a member of the carcinoembryonic antigen (*CEA*) gene family, which belongs to the immunoglobulin gene super family. It contains two major subgroups based on sequence comparisons. One subgroup of genes in this family encode the pregnancy-specific glycoproteins (PSG). Increasing amounts of PSG are expressed by the placenta during pregnancy, from which they are secreted into the maternal bloodstream (Thompson et al. 1991). Besides, during embryonic development cell adhesion molecules are considered important. They function in intercellular interactions which happen between migrating and fixed cells in the determination of animal form during organogenesis (Van Nagell et al. 1982). Studies in mouse (Finkenzeller et al. 2003) demonstrated that *CEACAM1* and PSG are related to innate and adaptive immunity, and take part in the control of the maternal immune system to avoid rejection of the semi-allotypic embryo (Wessells et al. 2000).

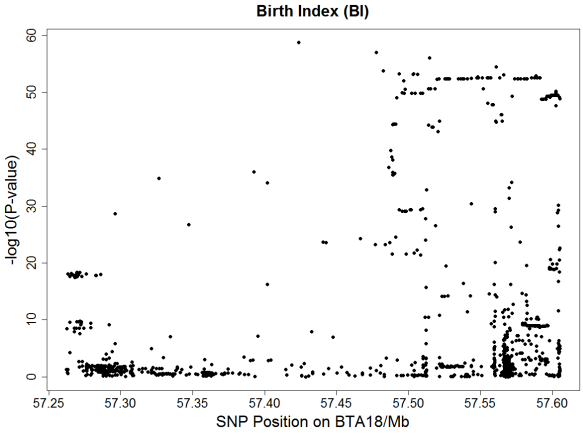

Figure 2. Single SNP analysis for significant region on BTA18 for Birth Index using NGS variants. Y axis is –$\log_{10}$ of the P-value, X axis is the SNP position.

Single marker analysis for NGS variants on BTA18. NGS variants were used to locate the causative mutation BTA18 QTL, In general, the QTL regions resulting from NGS variants analysis matched those from HD genotypes, but were more narrowly defined. For example, the NGS variants located in the *CEACAM18* gene still remained significantly (P-value 4.48×10^{-58}) (Figure 2). The P-values were exactly the same for variants at position 57,537,865, 57,538,528, 57,540,779 and 57,540,847 within *CEACAM18* gene, which was due to high LD among these variants. The analysis of NGS variants in this region confirmed the results from HD genotypes and indicated that NGS data provided more information about the causative mutation compared to HD. However, the ranking of variants or regions according to P-value was not consistent between NGS variants and HD genotypes. The variant at 57,423,824 bp on BTA18 had the lowest P-value (1.84×10^{-59}) in NGS variants analysis, which was different from the single SNP and haplotype analysis using HD genotypes. The most significant NGS variant was located in the intron of *Bos taurus* kallikrein-related peptidase 7 (*KLK7*) gene, its function related to calving is

not known. This inconsistency might be due to some imputation errors due to limited number of full genome sequences available as reference.

Conclusion

This GWAS with imputed HD genotypes and NGS variants confirmed several QTL that have been reported in previous studies. A candidate haplotype approach using HD genotypes, the gene *CEACAM18* at 57Mb on BTA18 had the most significant association with direct calving traits in the Nordic Holstein population. This was afterwards confirmed by analyzing NGS variants. However, we did not succeed in identifying the causative mutations even with full sequence information. It was difficult to separate the true causative mutation from the variants having high LD with this dense genotype. Targeted deep re-sequencing and functional studies could be helpful in finding the causal factor. This finding provides an understanding of the genetic architecture underlying dairy calving performance and could be useful for improving the genomic prediction models.

Literature Cited

Akey J., Jin L. and Xiong M. (2001). *Eur. J. Hum. Genet.* **9**, 291–300.

Barrett J. C., Fry B., Maller J., et al. (2005). *Bioinformatics* **21**, 263–265.

Boelling D., Nielsen U. S., Pösö J., Eriksson J.-å, et al. (1985). Genetic Evaluation of Calving Traits in Denmark , Finland and Sweden. , 179–184.

Browning S. R. and Browning B. L. (2007). *Am. J. Hum. Genet.* **81**, 1084–1097.

Cole, J.B., G.R. Wiggans, L. Ma, et al. (2011). BMC Genomics. 12:408.

Finkenzeller D., Fischer B., Lutz S., et al. (2003).

Howie B. N., Donnelly P. and Marchini J. (2009). *PLoS Genet.* **5**, 15.

Sahana G., Guldbrandtsen B. and Lund M. S. (2011). J. Dairy Sci. 94, 479–86.

Thomasen, J.R., B. Guldbrandtsen, P. Sørensen, et al. (2008). J. Dairy Sci. 91:2098–2105.

Thompson J. A., Grunert F. and Zimmermann W. (1991). *J. Clin. Lab. Anal.* **5**, 344–366.

Van Nagell J. R., Donaldson E. S., Gay E. C., et al. (1982). *Cancer* **49**, 2077–2090.

VanRaden P. M. (2008). *J. Dairy Sci.* **91**, 4414–23.

Wessells J., Wessner D., Parsells R., et al. (2000). *Eur. J. Immunol.* **30**, 1830–1840.

Yu J., Pressoir G., Briggs W., Bi I., et al. (2006). *Nat. Genet.*