J. Dairy Sci. 98:1–11
http://dx.doi.org/10.3168/jds.2015-9599

Genetic dissection of milk yield traits and mastitis resistance QTL on chromosome 20 in dairy cattle[1]

Naveen K. Kadri, Bernt Guldbrandtsen, Mogens S. Lund, and Goutam Sahana[2]
Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, 8830 Tjele, Denmark

ABSTRACT

Intense selection to increase milk yield has had negative consequences for mastitis incidence in dairy cattle. Due to low heritability of mastitis resistance and an unfavorable genetic correlation with milk yield, a reduction in mastitis through traditional breeding has been difficult to achieve. Here, we examined quantitative trait loci (QTL) that segregate for clinical mastitis and milk yield on *Bos taurus* autosome 20 (BTA20) to determine whether both traits are affected by a single polymorphism (pleiotropy) or by multiple closely linked polymorphisms. In the latter but not the former situation, undesirable genetic correlation could potentially be broken by selecting animals that have favorable variants for both traits. First, we performed a within-breed association study using a haplotype-based method in Danish Holstein cattle (HOL). Next, we analyzed Nordic Red dairy cattle (RDC) and Danish Jersey cattle (JER) with the goal of determining whether these QTL identified in Holsteins were segregating across breeds. Genotypes for 12,566 animals (5,966 HOL, 5,458 RDC, and 1,142 JER) were determined by using the Illumina Bovine SNP50 BeadChip (50K; Illumina, San Diego, CA), which identifies 1,568 single nucleotide polymorphisms on BTA20. Data were combined, phased, and clustered into haplotype states, followed by within- and across-breed haplotype-based association analyses using a linear mixed model. Association signals for both clinical mastitis and milk yield peaked in the 26- to 40-Mb region on BTA20 in HOL. Single-variant association analyses were carried out in the QTL region using whole sequence level variants imputed from references of 2,036 HD genotypes (BovineHD BeadChip; Illumina) and 242 whole-genome sequences. The milk QTL were also segregating in RDC and JER on the BTA20-targeted region; however, an indication of differences in the causal factor(s) was observed across breeds. A previously reported F279Y mutation (rs385640152) within the growth hormone receptor gene showed strong association with milk, fat, and protein yields. In HOL, the highest peaks for milk yield and susceptibility to mastitis were separated by over 3.5 Mb (3.8 Mb by haplotype analysis, 3.6 Mb by single nucleotide polymorphism analysis), suggesting separate genetic variants for the traits. Further analysis yielded 2 candidate mutations for the mastitis QTL, at 33,642,072 bp (rs378947583) in an intronic region of the caspase recruitment domain protein 6 gene and 35,969,994 bp (rs133596506) in an intronic region of the leukemia-inhibitory factor receptor gene. These findings suggest that it may be possible to separate these beneficial and detrimental genetic factors through targeted selective breeding.

Key words: clinical mastitis, association study, milk yield, cattle

INTRODUCTION

Intense selection to increase milk yield and milk quality in dairy cows has had negative consequences on other functional traits, including the incidence of clinical mastitis (**CM**; Oltenacu and Broom, 2010). Attempts to improve CM resistance through traditional breeding have not been very successful due to low heritability estimates for CM and unfavorable genetic correlations between CM resistance and milk yield (Lund et al., 1999; Heringstad et al., 2000; Heringstad et al., 2005). Therefore, CM remains the most frequent and costly disease in dairy production, affecting 18% of all dairy cows (USDA, 2008). Several studies from Nordic countries have reported QTL for CM (Lund et al., 2007; Sahana et al., 2008, 2013; Schulman et al., 2009, 2011; Sodeland et al., 2011). Linkage disequilibrium (**LD**) spreads across large chromosomal regions, especially in Holsteins (de Roos et al., 2008), making it difficult to distinguish the quantitative trait nucleotide from SNP that are in strong LD with the causal variants. Thus, the reported QTL intervals are broad, impeding identification of affected genes and causal polymorphisms. However, the availability of whole-genome sequencing (**WGS**) data for dairy cattle has increased the chance

Received March 20, 2015.
Accepted July 25, 2015.
[1]Part of the PhD study of Naveen K. Kadri.
[2]Corresponding author: goutam.sahana@mbg.au.dk

of identifying causative polymorphisms underlying QTL. Availability of multiple breed data may help in breaking the long stretch of strong LD.

Many economically important traits, including milk yield and CM, are polygenic. As a result, variance of these traits may result from the interaction of small changes in the activity of multiple, potentially neighboring, genes that interact and affect the trait in question, making it very challenging to identify individual genes. Single QTL can act through multiple pathways and can affect not only coding regions, but also regulatory regions flanking these genes (Bickel et al., 2011). In recent years, genomic prediction approaches that simultaneously incorporate information from thousands of genetic markers to predict the genetic merit of an animal have come into routine use in dairy cattle breeding (Lund et al., 2011). However, current breeding techniques have not been effective in disrupting unfavorable genetic correlations between CM and milk yield. One factor that has contributed to this problem is the lack of sufficiently detailed genetic information in the models to clarify whether both traits are affected by a single polymorphism (pleiotropy) or by multiple closely linked polymorphisms. Besides, the recognition of correlations between traits depends on the identification of genetic covariance: that is, the gene frequencies of loci contributing positively and negatively to covariance (Bohren et al., 1966; van Binsbergen et al., 2012). Current prediction methods could be improved by incorporating information on important identified causative genetic variants (Brøndum et al., 2015; Zhang et al., 2015).

Sahana et al. (2014) recently reported a QTL on BTA20 affecting CM and SCS in Danish Holstein cattle (**HOL**), with a broad association signal at 32 to 40 Mb and a peak around 35 Mb. In a meta-analysis of 3 breeds [HOL, Nordic Red dairy cattle (**RDC**), and Danish Jersey cattle (**JER**)], the most highly associated SNP with CM was located at 33,398,781 bp (Sahana et al., 2014). Those authors also reported strong evidence for a QTL located near 33 Mb on BTA20 for milk yield, milk fat, and milk protein. Finally, an F279Y mutation (rs385640152) in the growth hormone receptor (**GHR**) gene at 31,909,478 bp has been reported to affect milk, fat, protein, and SCS in HOL (Blott et al., 2003; Rahmatalla et al., 2011).

The large number of reports describing QTL that affect both milk yield and CM susceptibility traits led us to target this region of BTA20 for further study. We hoped to determine whether mutations in a single gene affected both CM and milk traits, or whether multiple causal factors segregated in Nordic cattle breeds that may be separable by more detailed analysis. Our objectives were (1) to fine-map QTL for CM and milk traits (milk yield, milk fat, and milk protein) on BTA20 using imputed WGS variants incorporating data from 3 dairy breeds, and (2) to determine whether QTL for CM and milk traits on BTA20 were due to different but closely linked polymorphisms or to multiple effects of the same polymorphism.

MATERIALS AND METHODS

Animal Populations and Phenotypes

Data of 12,566 genotyped bulls from 3 Nordic dairy cattle breeds (5,966 HOL, 5,458 RDC, and 1,142 JER) and their EBV for 4 indices—CM and milk, fat, and protein yields—were used in the study. The EBV for milk, fat, and protein yields (in kilograms) were estimated by Nordic Cattle Genetic Evaluation (http://www.nordicebv.info) based on production figures derived from routine test-day records and converted to an index for each trait by using a random regression test-day model. Data included are test-day records from d 8 to 365 for first to third lactations. Estimation of EBV for CM was described previously by Sahana et al. (2014). Mean reliabilities for EBV were 0.79 for CM and 0.92 for milk yield traits, respectively. Seventy-five percent of individuals had reliability for CM above 0.76 and above 0.92 for the 3 milk yield indices.

SNP Chip and Genotyping

All animals were individually genotyped by using the Bovine SNP50 BeadChip (Illumina Inc., San Diego, CA), which assays for approximately 54,000 SNP markers, 1,568 of which are on BTA20. Genotyping was done at Aarhus University (Foulum, Denmark) and GenoSkan A/S (Tjele, Denmark). Quality parameters used for selection of SNP in the study were minimum call rates of 90% for animals and 95% for SNP. The minimum accepted GenCall score (Illumina) was 0.60 for individual typings. Individuals with average GenCall scores below 0.65 were excluded. The SNP positions within a chromosome were defined according to the *Bos taurus* genome UMD3.1 assembly (Zimin et al., 2009). After quality control analysis, 46,702 SNP were identified on 29 autosomes, with 1,301 on BTA20.

Haplotype Phasing

Genotype phasing and haplotype clustering were done in a multibreed combined data set, which consisted of genotypes for 12,566 animals for 1,301 SNP remaining after quality control on BTA20. The SNP genotypes were first phased by utilizing information from the pedigree with LINKPHASE and then by using LD with DAGPHASE. Phased genotypes were clustered into 50