CALL FOR PAPERS | Functional Analysis of Sequence Variation

Multisite haplotype on cattle chromosome 3 is associated with quantitative trait locus effects on lactation traits

Miri Cohen-Zinder,1 Ravikiran Donthu,2 Denis M. Larkin,2 Charu Gupta Kumar,2 Sandra L. Rodriguez-Zas,1,2 Kalista E. Andropolis,2 Rosane Oliveira,2 and Harris A. Lewin1,2

1Institute for Genomic Biology, and 2Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois

Submitted 17 December 2010; accepted in final form 1 September 2011

ONE OF THE GRAND CHALLENGES in modern genetics is to identify the molecular genetic basis for complex traits (20). In cattle, the low-to-moderate heritability of quantitative traits under selection and limited sizes of resource populations are largely responsible for mapping most quantitative trait loci (QTL) to large intervals of >10 cM (http://www.genome.iastate.edu/cgi-bin/QTLdb/BT/index). The poor genomic resolution of mapped cattle QTL makes identification of the genes and polymorphisms underlying their effects a difficult and expensive endeavor. Despite these limitations, two QTL with large effects on milk, milk fat, and protein yields were resolved at the gene level with high probability (12, 23), thereby demonstrating the strength of available cattle resource populations and genomic tools for complex traits analysis. These findings have been rapidly implemented by the dairy industry for genetic improvement in milk production traits (24). Moreover, these studies have illuminated our general understanding of the genetic control of complex traits and have provided two new model systems for lactation biology. Therefore, genetic dissection of QTL affecting lactation traits is important both as a biomedical model system and for applications in animal breeding.

Fine-mapping of QTL is a prerequisite step for identifying causative mutations that affect quantitative traits. For fine mapping, genotyping arrays containing more than one million single nucleotide polymorphisms (SNPs) have recently been used for genome-wide association studies (GWAS) in humans to detect markers closely linked to QTL and genes contributing to complex diseases (3, 9, 64). Medium-density SNP arrays are also available for cattle, dogs, and other species and have been used to study the origins and population genetics of domesticated animal breeds (56, 60). High-density SNP arrays have been used to demonstrate that linkage disequilibrium (LD) in humans extends ~12.3 Kbp on average and that this distance may vary between populations (57), whereas in dogs, LD extends over 10 Kbp (36). In cattle, LD is thought to range from 10 to 40 Kbp (31); however, because the genotyping array used for this study contained ~55,000 SNPs (39), corresponding to an average density of 1 SNP every ~50 Kbp, this estimate is uncertain. Khatkar et al. (31) suggested that ~300,000 markers, or an average density of 1 SNP every 10 Kbp, is required to perform a high-power genome scan for monogenic and polygenic traits in Holstein cattle. Such high-density SNP genotyping arrays are only just becoming commercially available as a result of the rapidly increasing information spawned by initiatives to sequence the cattle genome (35). The discovery of millions of cattle SNPs provides an extraordinary opportunity to fine-map QTL previously identified using low-resolution maps of microsatellite markers (21, 50).

Previous studies have identified QTL for milk production traits segregating on BTA3 in different dairy cattle populations (10, 26, 58). Heyen et al. (26) performed a microsatellite-based genome scan that detected QTL associated with milk production, milk composition, and health traits in a North American Holstein-Friesian population consisting of 1,068 progeny tested sons of eight elite sires represented in the U.S. Dairy Bull DNA Repository (DBDR). QTL affecting milk yield (MY) and milk composition traits were localized to a region of
BTA3 between the microsatellite markers BL41 and TGLA263, which are located at 42 and 51 cM, respectively, on the cattle linkage map (28). These results were corroborated by Rodriguez-Zas et al. (49) and Ashwell et al. (4). In two other independent studies of North American Holsteins, Plante et al. (46) and Daetwyler et al. (14) mapped QTL for milk production traits to this same region of BTA3. However, the genes and biological mechanisms underlying these traits have not yet been determined.

In the present study, a combination of mapping and molecular approaches was used to identify candidate genes for QTL effects on MY, fat yield (FY), and protein yield (PY) located within a QTL critical region on BTA3. This strategy allowed us to identify a multisite haplotype that accounts for major effects on MY, FY, and PY within families and at the population level.

METHODS

Selection of a 16.3 Mbp QTL Region on BTA3 for High-density SNP Mapping

To map the region between BL41 and TGLA263 on the cattle genome assembly, cattle whole genome shotgun (WGS) reads GI: 53116256 and GI: 52939457 containing the markers BL41 and TGLA263 were downloaded from the National Center for Biotechnology Information. Shotgun read GI: 53116256 is part of the whole genome scaffold Chr3.35 in cattle genome assembly Btau 3.1, positioned at 27,123,933–27,125,882 bp; shotgun read GI: 52939457 is part of scaffold Chr3.49 positioned at 36,178,440–36,179,190 bp. We extended the region to the end of scaffold chr3.49 and added 4.2 Mbp of assembled sequence upstream of marker BL41 to examine differences in haplotype structure beyond but proximal to the QTL interval mapped by Heyen et al. (26). Microsatellites BL41 and TGLA263 were also found to be located within the mapped and sequenced cattle BAC clones CH240-514H13 and CH240-176P18, respectively (54).

All 27 sequence scaffolds within the targeted region were mapped to the most recent cattle genome assembly (Btau 4.0) using tBLASTn (best hit, E < 1e-50). The order and orientation of scaffolds in Btau 4.0 matched the order we determined for all 27 scaffolds. For further analyses we used the corresponding region of BTA3 in Btau 4.0: 25,391,268 to 41,681,279 bp.

Selection of SNPs on BTA3

A total of 2,500 SNPs from the region flanking and between the markers BL41 and TGLA263 were identified for our studies. Of these SNPs, 2,007 are part of the two million in silico SNPs identified from the alignments of WGS reads of cattle genome assembly Btau 2.0 (55). In addition, 310 SNPs were found by alignment of the nucleotide sequences of Holstein bacterial artificial chromosome (BAC) clones 115K2 and 116A6 from the RPCI-42 BAC library (see below) to cattle genome assembly Btau 3.1; 183 more SNPs were detected by comparing publicly available cattle expressed sequence tags (http://www.ncbi.nlm.nih.gov/dbEST/) to Btau 3.1.

To identify additional SNPs, two Holstein BAC clones from the RPCI-42 library, 115K2 and 116A6, were sequenced using Roche 454 sequencing technology (38). The BAC clones were sequenced to ~20x coverage, and sequence reads were assembled de novo into contigs using Newbler (38).

Resource Populations and SNP Genotyping

For QTL fine-mapping, we used North American Holstein sires in the DBDR collection (n = 263) that are unrelated for at least one generation and sires and offspring of DBDR family 1 (F1, n = 190), family 2 (F2, n = 129), and family 5 (F5, n = 54). These families were previously shown to segregate QTL for milk production traits (26). All samples were genotyped using MassARRAY iPLEX Gold technology (30).

Reconstruction of Haplotypes

A custom Perl script was used to infer putative maternal and paternal alleles in offspring from SNP genotyping results. A locus was considered as informative for inheritance of paternal and maternal alleles if the sire was homozygous and its offspring was either homozygous or heterozygous at the same locus, or when a sire was heterozygous and its offspring was homozygous. The combination of paternal and maternal alleles in each offspring formed the putative paternal and maternal haplotypes inherited by the offspring. Extended maternal haplotypes of the progeny tested sons were defined by differentiating maternally inherited alleles from paternally inherited alleles.

The unique set of maternal haplotypes was determined by calculating genetic distances between all the maternal haplotypes using the CLUSTALDIST tool of Biology Workbench (http://workbench.sdsc.edu/). Maternal haplotypes with genetic distance > 0.01 were classified as unique. Haplotype sharing between at least two sires was defined with HapBlock software (67) using default parameters as previously described (45, 66).

Statistical Analyses: Association Studies, Linkage Mapping, and LD Analysis

Linkage maps were developed for F1 and F2 for the 16.3 Mbp QTL region using CRIMAP v. 2.5, which is a revised version of CRIMAP v. 2.4 (22) modified by Jill Maddox and Ian Evans (University of Melbourne, Melbourne, Australia). No linkage map was created for F5 due to the small family size. The linkage maps for F1 and F2 were then used as input to QTL Express (52) using a regression model for half-sib data to verify the QTL effects within the 16.3 Mbp of QTL region on BTA3 detected previously by Heyen et al. (26). We performed 5,000 permutations to calculate 1% and 5% experiment-wise significance thresholds. Calculation of the 95% confidence interval (CI) was performed as described by Lander and Botstein (34).

Haplovie software (6) was used to define LD blocks according to Gabriel et al. (17). Haplotype of the sires and imputed maternal haplotypes were used as input. First, estimates of LD (D’ values) and 95% CI bounds were generated for all possible combinations of SNPs within the 16.3 Mbp region. A SNP pair was considered to be in strong LD if the value fell within the 95% confidence limits 0.70–0.98. An LD block was defined as a region in which 95% of SNP pairs showed strong LD (6, 17).

Breeding values for the lactation traits, expressed as predicted daughter yield deviations (DYDs), were obtained for DBDR sires and offspring of F1, F2, and F5 from the USDA Animal Improvement Program Laboratory (http://www.aipl.arsusda.gov/query/public/tdb.shtml). Student’s t-test was used to test for associations between haplotypes and DYDs within families (JMP7.0 software, SAS Institute). For allele-association studies, only individuals homozygous for alternative alleles at each informative SNP locus were included in the analysis. Informative SNP loci were defined as those with at least 10 homozygous individuals for each allele within F1, F2, and the Holstein sire population, or at least five homozygous individuals in F5. Analysis of informative SNPs for association with DYDs for MY, FY, and PY was performed in F1, F2, F5, and in the Holstein sire population using a t-test. The allele-specific associations were made by comparing DYD means of the two groups of homozygotes for alternative alleles. A locus was considered to be in strong LD if the value fell within the 95% confidence limits 0.70–0.98. An LD block was defined as a region in which 95% of SNP pairs showed strong LD (6, 17).

One-way ANOVA was used to estimate the effect size (r2) and allele substitution effect (α2) in F1 and F2 for SNPs in RAPIA...
(RAP1AP25) and ADORA3 (ADORA3P26) (JM7.0 software, SAS Institute), where (α/2) is half of the difference between the least squares means of the two homozygous genotype classes.

**Sequence Capture and DNA Sequencing**

A custom-designed NimbleGen Sequence Capture 385K array (Roche NimbleGen, Madison, WI) was used to isolate sequences from two targeted chromosome segments within the 16.3 Mbp QTL critical region on BTA3. Primary sequence data were extracted from cattle genome Build 4.0 (region 1: 30,706,774–31,080,623; region 2: 34,065,565–36,461,414). The total length of the targeted region was 2.77 Mbp. In the targeted region, 1.82 Mbp (66%) were covered with probes after removing repetitive sequences. Sequence capture, library preparation, and sequencing on a 454 Genome Sequencer FLX system were performed according to the manufacturers’ protocols (NimbleGen Array User’s Guide, Sequence Capture Array Delivery; 454 Life Sciences, Branford, CT).

Alignment to the reference assembly, mapping, and detection of polymorphisms was performed using GS Reference Mapper (Genome Sequencer Data Analysis software version 2.0.00, Roche). To identify SNPs, a stringent analysis was performed using minimum overlap length of 100 bp, minimum overlap identity of 95%, and a contig threshold of 1,000 bp. Build 4.0 of BTA3 was used as reference sequence for subsequent analysis.

**Integrated Haplotype Score Analysis**

Integrated haplotype score (iHS) was calculated using a program developed by Voight et al. (59). Imputed maternal haplotypes of offspring with ≥20 heterozygous informative loci and haplotypes of sires were used as input. To predict missing genotypes within the maternal haplotypes, we used fastPHASE software (51). For each SNP, the allele that was found in the Hereford reference genome sequence Build 4.0 was arbitrarily assigned as the “ancestral” allele, while the alternate allele was regarded as the “derived” allele.

**Identification of Polymorphisms in Promoter Regions**

SNPs located within −5,000 bp to +500 bp of the transcription start site (TSS) of candidate genes in the QTL region were selected for an analysis of their position relative to transcription factor binding sites (TFBS). Transcription start sites were downloaded from the UCSC Genome Browser (refGene table for the cattle genome). Promoter extraction was performed as previously described (33). Polymorphisms in cattle promoters were analyzed to determine if TFBSs within the promoter region were disrupted by the allelic differences or resulted in the gain of a new TFBS. In addition, promoters in orthologous human and mouse genes were scanned for the presence/absence of these TFBSs.

**Genotyping of Promoter Polymorphisms**

Genotyping of promoter SNPs RAP1AP25 and ADORA3P26 was performed on DNA samples collected from 35 randomly selected Holstein cows in the University of Illinois Dairy Cattle Research Unit. All procedures were conducted under protocols approved by the University of Illinois Institutional Animal Care and Use Committee. DNA was extracted from whole blood as described by Miller et al. (40). For allele discrimination, TaqMan MGB primers and probes (available upon request) were designed for the two SNPs using Custom TaqMan SNP Genotyping Assays (Applied Biosystems). Allele discrimination assays were performed according to a standard protocol (Real-Time PCR Systems, Chemistry Guide, Applied Biosystems).

**Gene Expression Analysis**

Quantitative RT-PCR was performed to compare expression of RAP1A, ADORA3, OVGP1, and C3Horf88 in RNA samples prepared from mammary gland biopsies (day 7 and 30 relative to parturition) and liver tissue (day −14, 7, and 30 relative to parturition). Tissue biopsies were collected from the 35 cows described above, RNA was extracted, and cDNAs were prepared as described previously (8, 37). Oligonucleotides for RNA expression studies (available upon request) were designed using Primer Express software (Applied Biosystems, Foster City, CA).

Comparisons of transcript levels in liver and mammary gland tissues at each time point for genotyped animals were performed using the MIXED procedure of SAS (SAS Institute) as described by Bionaz and Loor (8). The final model included fixed effects of time point relative to parturition, genotype for SNPs RAP1A P25 and ADORA3 P26, tissue, and random effect of cow:

\[
GE = \mu + \text{TOME}_i + \text{GENOTYPE}_j + CO\text{W}_{k} + \text{TISSUE}_{l} + \varepsilon_{ijk}
\]

where \(\mu\) is intercept (overall mean); \(i = \text{day 7 or day 30 relative to parturition; j} = \text{genotype A/A or B/B for SNPs RAP1A P25 or ADORA3 P26; k} = \text{cow, and l} = \text{mammary gland or liver tissue. The interaction of time*genotype was tested and found to be not significant and was not included the final model.}

**RESULTS**

**Distribution of SNPs in F1, F2, F5**

Three paternal half-sib families consisting of 373 offspring were genotyped for 2,500 SNPs mapped within the 16.3 Mbp QTL region of BTA3 (Table 1). Among them, 2,462 (98.5%) had ascertained genotypes, of which 1,520 (60.8%) were heterozygous in at least one of the DBDR sires or their progeny-tested sons (Table 1). Of 1,520 heterozygous SNP loci, 998 (65.6%) were polymorphic in at least one of the three sires. The average distance between the 1,520 heterozygous SNP loci was 10.8 Kbp, with a median distance of 0.8 Kbp. The largest gap (630.7 Kbp) at position 26,381,530 (Build 4.0; equivalent to position 1,030,113 in the 16.3 Mbp region) was due to five adjacent SNPs that were not polymorphic in any genotyped sire or offspring.

**Reconstruction of Sire and Dam Haplotypes and QTL Effects**

Extended haplotypes across the 16.3 Mbp region of each sire were reconstructed by determining paternal alleles that cosegregated among his offspring using the order of SNPs in Btau 4.0 as a template (Table 1). Imputed maternal haplotypes were reconstructed using a subset of 1,011 SNPs that were heterozygous in at least one individual, sire, or offspring and had no missing genotypes in the three sires or inconsistent segregation in offspring. Among the 322 imputed maternal haplotypes, 182 had ascertained genotypes, of which 1,520 (60.8%) were heterozygous in at least one individual, sire, or offspring and were classified as unique (genetic distance ≥ 0.01). The unique haplotypes were used for estimation of allele frequencies and for LD analysis (see below).

**Table 1. Heterozygosity of 1,520 SNPs in 376 sires and offspring**

<table>
<thead>
<tr>
<th>Sire, Offspring</th>
<th>Polymorphic SNP Loci</th>
<th>Samples, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>630 (41.4%)</td>
<td>1</td>
</tr>
<tr>
<td>S2</td>
<td>593 (39.0%)</td>
<td>1</td>
</tr>
<tr>
<td>S3</td>
<td>354 (23.2%)</td>
<td>1</td>
</tr>
<tr>
<td>Offspring of S1</td>
<td>1,452 (95.5%)</td>
<td>190</td>
</tr>
<tr>
<td>Offspring of S2</td>
<td>1,438 (94.6%)</td>
<td>544</td>
</tr>
<tr>
<td>Offspring of S3</td>
<td>1,402 (92.2%)</td>
<td>54</td>
</tr>
</tbody>
</table>

SNP, single nucleotide polymorphism; S, sire.
The set of 1,011 SNPs used to impute haplotypes was then used to identify haplotype blocks shared by two or more sires. Over the entire 16.3 Mbp region all six sire haplotypes were found to be unique (Fig. 1). However, a total of six haplotype blocks common to at least two sires were identified. Sire 5 has two shared haplotype blocks with S1 that cover only a small fraction of the region and no shared haplotype blocks with S2 (Fig. 1). Sires 1 and 2 have the largest shared haplotype block (9.67 Mbp), which contains 930 SNPs with identical alleles (Fig. 1).

**Linkage Analysis and QTL Mapping**

To reconfirm the presence of QTL in the BL41 to TGLA263 interval we first constructed linkage maps for F1 and F2 using the genotypes of all informative SNPs in these families. The genetic distance calculated for the entire 16.3 Mbp region using CRIMAP v. 2.5 was estimated to be 15.3 and 11.5 cM in F1 and F2, respectively. The map distance between BL41 and TGLA263 was calculated to be 8.6 and 7.9 cM in F1 and F2, respectively (Table 2).

Multimarker regression QTL analysis revealed broad peaks exceeding the experiment-wise significance threshold over the entire 15.3 cM interval in F1 for MY. For FY and for PY the peaks were at or slightly above the significance threshold over the interval between BL41 and TGLA263. A similar pattern was observed in F2 for MY and PY. The detected effects were similar to those found previously (26).

**Mapping the QTL Critical Region Using Association Studies**

Association studies were performed in F1, F2, F5, and 263 DBDR Holstein sires to assist in fine-mapping QTL.
effects (Fig. 2, A–D). Two subregions located within the 16.3 Mbp QTL region between BL41 and TGLA263 contained SNPs with significant effects on MY and composition traits (Figs. 1 and 2, A–D). These subregions were also within the map intervals for the QTL defined using multilocus regression analysis.

LD Across the 16.3 Mbp QTL Region

The 188 reconstructed dam and sire haplotypes were used to calculate LD across the entire 16.3 Mbp region. We identified 80 LD blocks containing two or more SNPs (Fig. 1, Supplemental Table S1). The blocks contained 767 SNPs covering 1.96 Mbp (11.6%) of the 16.3 Mbp region (mean length = 24.5 Kbp, median length = 4.2 Kbp). The vast majority of LD occurs within the two subregions defined by the association analysis (Fig. 1). The largest LD block (186.5 Kbp, block 38) falls within subregion 2, which is also located within the 9.67 Mbp shared haplotype block of S1 and S2 (Fig. 1). Much lower levels of LD were found in the remainder of the 16.3 Mbp region, which may be due in part to the varying SNP densities across the region. Relationships between the LD blocks, SNPs within them, and DYDs are presented below.

Resequencing of Targeted Regions

For detection of additional genetic polymorphisms in the QTL region we resequenced S1 DNA from the two subregions containing the vast majority of SNPs having significant effects on the lactation traits (Figs. 1 and 2; Build 4.0 coordinates: 30,706,774–31,080,623; 34,065,565–36,461,414). A total of 663,641 reads having an average read length of 211 bp after linker trimming were generated, covering 1.8 Mbp (64%) of the targeted region at 21-fold redundancy. Classification of the SNPs detected using sequence capture technology is presented in Table 3. We selected 35 putative SNPs located in exons or promoters of genes in subregions 1 (8 SNPs) and 2 (23 SNPs), i.e., those most likely to be functional, from this set and genotyped them on the 263 DBDR Holstein sires.

Association Studies Within Families and in the 263 DBDR Sires

The relationships between the SNPs and allele effects associated with yield traits in families and the DBDR sire population were then examined. Among the 35 SNPs identified by resequencing, five had significant effects on one or more of the yield traits (Fig. 3). In the 263 DBDR sires, 22 SNPs were associated with effects on MY and composition traits (FDR $P \leq 0.2$, Table 4). None of the SNPs were present in subregion 1. Nineteen SNPs were located in eight distinct LD blocks in subregion 2 with the remaining three SNPs located between them (Table 4). These results indicate that effects associated with alleles in subregion 1 in families were due to linkage with alleles in subregion 2, thus excluding

---

Table 2. Genetic mapping of the QTL critical region

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offspring, $n$</td>
<td>190</td>
<td>129</td>
<td>54</td>
<td>373</td>
</tr>
<tr>
<td>Informative offspring, $n$</td>
<td>182</td>
<td>112</td>
<td>52</td>
<td>346</td>
</tr>
<tr>
<td>Observed recombinants in 16.3 Mbp region, $n$*</td>
<td>24 (15.2 cM)</td>
<td>11 (11.5 cM)</td>
<td>2 (NA)†</td>
<td>37</td>
</tr>
<tr>
<td>Recombinants between BL41 and TGLA263, $n$†</td>
<td>12 (8.6 cM)</td>
<td>9 (7.9 cM)</td>
<td>1 (NA)†</td>
<td>22</td>
</tr>
</tbody>
</table>

*Genetic distances are indicated in parentheses. †Linkage map was not created for F5 due to low number of offspring and recombinants. NA, not applicable.

---

Fig. 2. Association of SNPs with breeding values for milk yield, fat yield, and protein yield in the 16.3 Mbp QTL critical region. Daughter yield deviation (DYD) values for milk yield (MY, blue diamonds), fat yield (FY, red squares), and protein yield (PY, green triangles) were compared for the 2 classes of homozygotes. A: Dairy Bull DNA Repository (DBDR) sires ($n = 263$). B: family 1 ($n = 190$). C: family 2 ($n = 129$). D: family 5 ($n = 54$). The significance threshold corresponding to $P \leq 0.05$ is shown as a horizontal blue line. Red and blue open bars indicate candidate subregions 1 and 2, respectively.
**Table 3. Classification of new SNPs detected using sequence capture technology**

<table>
<thead>
<tr>
<th>Classification of new SNPs detected using sequence capture technology</th>
<th>SNP Detection Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNPs detected by alignment to reference chromosome (BTA3), n</td>
<td>3,120</td>
</tr>
<tr>
<td>SNPs detected in targeted region (QTL region), n</td>
<td>2,769 (88.0%)</td>
</tr>
<tr>
<td>SNPs detected compared to reference genome, n</td>
<td>687 (25.0%)*</td>
</tr>
<tr>
<td>SNPs detected in S1, n</td>
<td>2,082 (75.0%)*</td>
</tr>
<tr>
<td>New SNPs detected in S1, n</td>
<td>1,453 (70.0%)†</td>
</tr>
<tr>
<td>SNPs in promoters</td>
<td>22</td>
</tr>
<tr>
<td>SNPs in exons</td>
<td>25</td>
</tr>
<tr>
<td>SNPs in introns</td>
<td>177</td>
</tr>
</tbody>
</table>

*Fraction of SNPs in targeted region. †Fraction of SNPs detected in S1.

subregion 1 as a candidate region. Among the significant SNPs in subregion 2, eight were located in six different genes (Table 4; considered further below).

Within families, allele effects were detected by comparing DYDs in homozygotes of each class (i.e., A/A cf. B/B). This permitted assignment of positive or negative effects associated with each allele to their respective paternal haplotypes. The direction of effects on MY for all significant allele associations was consistently positive for the haplotype shared by S1 and S2, although only four SNPs in common had significant effects among the offspring of both sires. In contrast, all 29 SNPs on the shared haplotypes of S1 and S2 that were associated with effects on FY were in the opposite direction to the effects on MY. Thus, alleles for high MY and low FY segregate in half-sib families and the DBDR sire population were then considered further below).

Overlaps between the SNP associations detected in the three half-sib families and the DBDR sire population were then determined (Fig. 4, Table 4). For all the traits, no SNP was consistently associated with any trait in the three families and the population. However, SNP P25 (rs#42538923) was significantly associated with MY in F1, F2, and the DBDR sires. There were only 12 individuals homozygous for P25 in F5 so the sample size was too small to obtain meaningful results. Three additional SNPs with significant effects on MY in both F1 and DBDR sires and three SNPs with significant effects on MY in both F1 and F2 but not in DBDR sires were identified (Fig. 4). All SNPs with significant effects on FY were private to F1 or the DBDR sires.

**Candidate Genes and Polymorphisms for the QTL on BTA3**

Five SNPs, 1160, P25, HQ4, HQ1, and P28, had consistent effects on lactation traits in F1 and/or F2 and/or the DBDR sire population (Table 4). These SNPs are located in a 307 Kbp region between LD blocks 38 and 46, which includes four annotated genes: RAP1A, ADORA3, OVGP1, and C3H1orf88 (Fig. 3). Because of its position within this region and borderline significance on FY in both the DBDR sire population and F1, **ADORA3** SNP P26 was also considered for candidate gene analysis (Table 4).

**RAP1A.** **RAP1A** encodes a member of the Ras family of small GTPases (11) and is known to act as a central regulator of mammary gland development (29). SNP P25 (rs#42538923) is a C>G polymorphism located in the **RAP1A** promoter region at position −632 upstream to its TSS (−632C>G) in a binding site of the MYB transcription factor. The **RAP1A** −632G polymorphism was associated with increased DYD for MY in F1, F2, and the DBDR sire population (Table 4). We then calculated the additive effect of **RAP1A** −632C>G on milk production traits in F1 and F2 using DYDs of the sons. The proportion of the DYD variance explained by **RAP1A** −632C>G in F1 was 1.5% for MY, and the size of allele substitution effect (a2) = 118.94 Kg (P = 0.04). Similarly, the size of the effect for MY in F2 was 1%, with an allele substitution effect = 123.3 Kg (P = 0.05). **RAP1A** mRNA was found to be highly expressed in the liver tissue of 35 multiparous Holstein cows compared with mammary gland at days 7 (P < 0.005) and 30 (P < 0.005) relative to parturition, but there was no effect of genotype on expression levels (**RAP1A** −632C>G genotypes were G/G = 10; G/C = 17; CC = 8). The only SNP significantly associated with FY in the DBDR sire population was SNP1160 (rs#42453232) (Table 4). This SNP is an A>C substitution located in intron 4 of the **RAP1A** gene (IVSA4+2037A>C) and was in perfect LD with **RAP1A** −632C>G (D’ = 1.00). The IVSA4+2037C polymorphism was associated with higher MY in the DBDR sire population.

**ADORA3.** The **ADORA3** gene encodes a member of adenosine receptors that are involved in intracellular signaling pathways and physiological functions including stimulation of gluconeogenesis and inhibition of lipolysis (48). SNP P26 is a G>T substitution, located at position −39 upstream to the TSS of cattle **ADORA3** (−39G>T). **ADORA3** −39G>T was found to be in strong LD with **RAP1A** −632C>G (D’ = 0.80). The

![Fig. 3. Candidate genes harboring SNPs with significant effects on lactation traits. Arrows indicate the position and orientation of genes. Asterisks indicate the location of SNPs associated with 1 or more lactation traits in DBDR sires or QTL heterozygous sires (FDR P ≤ 0.2).](http://physiolgenomics.physiology.org/ by 10.220.33.4 on June 20, 2017)
ADORA3 −39G allele, associated with a negative effect on FY, is located on the haplotype shared between S1 and S2. The ADORA3 −39T allele is associated with increased FY in the DBDR Holstein sires and in F1 (Table 4), but slightly below the significance thresholds used for the association analysis (FDR P = 0.13 in F1 and FDR P = 0.31 in the DBDR sire population). The proportion of the DYD variance in FY explained by ADORA3 −39G>T in F1 was 1% and had an allele substitution effect of 4.9 Kg (P < 0.05). The ADORA3 −39T allele creates a PAX2 TFBS (gGGTTTgat) at positions −44 to −36 from the TSS. ADORA3 mRNA was found to be highly expressed in the mammary gland of 35 multiparous Holstein cows relative to liver tissue at day 7 after parturition (P < 0.05), but no effect of genotype was found on expression levels (ADORA3 −39G>T were G/G = 11; G/T = 17; T/T = 7).

OVGP1. This gene encodes a 120 kDa glycoprotein that is synthesized by oviductal epithelial cells and may be regulated by estrogen (2). The SNP HQ4 (rs#43332986), an A>G polymorphism at position 1208 of exon 9 of cattle OVGP1, encodes an H→R substitution at position 403 (H403R) of the OVGP1 polypeptide. OVGP1 H403R was associated with MY in the DBDR sire population and in F1 (Table 4). This SNP shows strong LD with RAP1A −632C>G (D’ = 0.79) but not with ADORA3 −39G>T. Expression of OVGP1 RNA was detected at trace levels in mammary and liver tissue (data not shown).

C3Hlorf88. The function of C3Hlorf88 (an ortholog of human Cof88) is unknown. SNP HQ1 is a G>C polymorphism at position 255 in exon 3 of C3Hlorf88 that encodes a K→N substitution at amino acid position 85 (K85N). The K85 allele was associated with increased DYD MY in the DBDR sire population and F1 (Table 4). The SNP P28 (rs#43334230) is an A>G polymorphism located in position 41 of the C3Hlorf88 promoter (C3Hlorf88 −41A>G). C3Hlorf88

---

**Table 4. SNPs with significant effects on MY and composition traits (FDR P ≤ 0.2) in 263 DBDR Holstein sires and six haplotypes of the three sires segregating for the QTL on BTA3**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Position in Build 4.0</th>
<th>LD Block</th>
<th>Trait</th>
<th>Allele 1</th>
<th>Allele 2</th>
<th>Gene/SNP Location</th>
<th>Effect in Families</th>
<th>S1 H1</th>
<th>S1 H2</th>
<th>S2 H1</th>
<th>S2 H2</th>
<th>S5 H1</th>
<th>S5 H2</th>
<th>P Value (Raw)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1160</td>
<td>34,173,405</td>
<td>38 PY</td>
<td>A</td>
<td>C</td>
<td></td>
<td>RAP1A intron 4</td>
<td>F1 (MY, FY) F2 (MY)</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>0.001</td>
</tr>
<tr>
<td>P25</td>
<td>34,236,445</td>
<td>38 MY</td>
<td>C</td>
<td>G</td>
<td>+d</td>
<td>RAP1A promoter</td>
<td>F1 (MY, FY) F2 (MY)</td>
<td>C</td>
<td>G+</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>0.005</td>
</tr>
<tr>
<td>P26*</td>
<td>34,320,079</td>
<td>FY</td>
<td>G</td>
<td></td>
<td></td>
<td>ADORA3 promoter</td>
<td>F1 (FY)</td>
<td>T+</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>0.028</td>
</tr>
<tr>
<td>HQ4</td>
<td>34,477,970</td>
<td>43 MY</td>
<td>A</td>
<td>+</td>
<td>G</td>
<td>OVGP1 exon 9</td>
<td>F1 (MY)</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>0.014</td>
</tr>
<tr>
<td>HQ1</td>
<td>34,541,775</td>
<td>46 MY</td>
<td>C</td>
<td>G</td>
<td></td>
<td>C3Hlorf88 exon 3</td>
<td>F1 (MY)</td>
<td>C</td>
<td>G+</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>0.006</td>
</tr>
<tr>
<td>P28</td>
<td>34,595,266</td>
<td>46 MY</td>
<td>A</td>
<td>G</td>
<td></td>
<td>ADORA3 promoter</td>
<td>F1 (FY)</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>0.009</td>
</tr>
<tr>
<td>1043</td>
<td>34,645,934</td>
<td>50 MY</td>
<td>A</td>
<td>C</td>
<td>+</td>
<td>intergenic</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>0.001</td>
</tr>
<tr>
<td>620</td>
<td>34,712,240</td>
<td>50 MY</td>
<td>A</td>
<td>G</td>
<td>+</td>
<td>intergenic</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>0.006</td>
</tr>
<tr>
<td>1153</td>
<td>34,721,511</td>
<td>50 MY</td>
<td>A</td>
<td>+</td>
<td>A</td>
<td>intergenic</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>0.007</td>
</tr>
<tr>
<td>1139</td>
<td>34,725,209</td>
<td>50 MY</td>
<td>A</td>
<td>G</td>
<td>+</td>
<td>intergenic</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>0.005</td>
</tr>
<tr>
<td>1138</td>
<td>34,725,331</td>
<td>50 MY</td>
<td>G</td>
<td>A</td>
<td>+</td>
<td>intergenic</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>0.006</td>
</tr>
</tbody>
</table>

*The 6 sire haplotypes are indicated by sire identification number and arbitrarily assigned haplotype number, e.g., S1H1, S1H2. bShared haplotype between S1 and S2. cSNP has integrated haplotype score (iHS) value ≥1.5. d, the allele is significantly associated with the trait. eFalse discovery rate (FDR) P = 0.31. fSegregating allele not determined. gSNP is located outside of haplotypes of the three sires segregating for the QTL on BTA3.
−41A>G was associated with greater DYD MY (Table 4). C3H1orf88 −41A>G and C3H1orf88 K85N are located 191 bp apart and are in perfect LD (D′ = 1.0). However, LD is stronger between C3H1orf88 K85N and RAP1A −632C>G (D′ = 0.77) than between C3H1orf88 −41A>G and RAP1A −632C>G (D′ = 0.62). LD was much stronger between C3H1orf88 K85N and RAP1A IVSA4+2037A>C (D′ = 0.83) than between C3H1orf88 −41A>G and RAP1A IVSA4+2037A>C (D′ = 0.25). Both C3H1orf88 K85N and C3H1orf88 −41A>G were associated with DYD for MY in the DBDR sire population and F1 (Table 4). Expression of C3H1orf88 RNA was detected at only trace levels in mammary and liver tissue (data not shown).

**iHS**

There were 149 SNPs for which iHS was greater or less than zero (Fig. 5). High and low iHS (≥1.5 and ≤−1.5) are generally accepted as strong evidence for a selective sweep, but in the context of our experiment, the sign of the iHS value has no biological meaning (i.e., the ancestral allele is arbitrarily assigned). Ten SNPs had a high positive (≥1.5) or low negative (<−1.5) iHS, of which one SNP is present in subregion 1 and four SNPs are present in subregion 2. Of all 10 SNPs with high or low iHS, only SNP1160 (RAP1A IVSA4+2037A>C) was significantly associated with a trait (PY) in the DBDR Holstein sire population (Table 4). These results indicate that the region in and around the RAP1A gene has undergone a recent selective sweep.

**DISCUSSION**

A series of technical approaches was used to exploit advantages of the dairy cattle breeding population for QTL fine-mapping and the identification of a multisite haplotype affecting lactation traits on BTA3. Our study was facilitated by the availability of millions of in silico SNPs identified by the bovine genome sequencing consortium (55). In the initial screen for polymorphisms in families segregating the QTL we identified 1,520/2,500 (60.8%) SNPs that were heterozygous in at least one of the DBDR sires or progeny tested sons. The median distance of 800 bp between SNPs in the 16.3 Mbp region permitted the most detailed genetic dissection of any cattle QTL region to date. This would roughly be the equivalent of a SNP screen with more than one million informative SNPs, which is greater than the SNP density currently used for human GWAS (3). When we added to that 1,453 new on-target informative SNPs detected by resequencing, we likely achieved close to complete SNP coverage in the QTL critical region of QTL-informative S1. This strategy made detection of the causative quantitative trait nucleotide(s) QTNS, or SNPs very close to the causative mutation(s), highly probable. However, our study does not preclude the possibility that indels, copy number variants (CNVs), or other structural variants in tight LD with the candidate QTNS contribute to the QTL effect(s) (62). For example, CNVs were detected on BTA3 within or close to candidate regions 1 and 2 (5, 15).

To our knowledge, our study represents the first use of targeted resequencing of a large known QTL interval for the detection of candidate QTNS. Resequencing in mouse and human has focused largely on exomes (1, 27) or in some cases a more targeted region such as ~200 Kbp of the KIT locus (13). We found the method to be robust, with 88.5% of SNPs on target located in either candidate region 1 or 2. Although the abundance of repetitive elements allowed for resequencing of only 1.8/2.8 Mbp of the targeted region, the unique sequence was efficiently resequenced at ~21×, yielding 1,453 new high-confidence in silico SNP predictions in S1. These predictions were validated by genotyping a set of 35 SNPs selected from promoter regions and exons. Targeted resequencing of Mbp-scale chromosomal intervals was thus proven to be feasible and should substantially enhance efforts to more precisely define the underlying genetics of complex traits in cattle and other species.

Reconstruction of paternal and maternal haplotypes for the three QTL heterozygous sires and 373 of their sons revealed a high level of diversity within the QTL critical region. The 182 imputed paternal haplotypes, differing by at least 1% of bases, illustrates that haplotype heterozygosity within the 16.3 Mbp QTL critical region remains high despite many generations of selection for lactation traits in the dairy cattle population. By contrast, S1, S2, and S5 showed much more limited diversity within the QTL critical region, although each of the six haplotypes of these three sires is unique across the entire region. A striking finding was the 9.7 Mbp shared haplotype block between S1 and S2 defined by 930 SNPs. The sire of S1 is the great-grand sire of S2 so there is a high probability that this segment is identical-by-descent. We found that segregation of the shared haplotype among the sons of these sires was significantly associated with increased MY, confirming and extending the study of Heyen et al. (26), who used the same animals. These data, combined with the multimarker regression analysis, provide strong evidence that the QTL for MY in these two sires lies within the 9.7 Mbp shared haplotype block.

The reconstruction of maternal haplotypes from genotyping data of large half-sib families permitted the characterization of LD across the QTL critical region, a refined candidate gene analysis, and detection of a selective sweep. Our LD analysis is of higher resolution than any other study performed to date for the cattle genome, revealing short and longer range patterns of LD. We detected 80 LD blocks with a mean length of 24.5 Kbp, ~10-fold smaller than the 300–500 Kbp reported for European and African cattle breeds (18) but within the range of...
10–40 Kbp reported by Khatkar et al. (31). As expected, there was substantial variation in the size of the LD blocks, with the largest (block 38) being 186.5 Kbp in subregion 2 defined by 56 SNPs. This block is within the 9.7 Mbp haplotype block shared by S1 and S2, and within the 2.4 Mbp fine-mapped region (subregion 2) defined by association studies. Within the fine-mapped QTL critical region LD is extended as compared with flanking regions, likely reflecting a selective sweep in the Holstein breed. This conclusion was firmly supported by the iHS analysis. The relatively short LD median block size of 4.2 Kbp reflects the very high density of SNPs used for the analysis and, if the results are representative of the genome, indicates that the extent of LD may be less than previously thought. If the pattern of LD in Holstein is lower, a substantially higher extent of LD may be less than previously thought. If the pattern of LD in Holstein is lower, a substantially higher number of SNPs may be necessary to capture all QTL by LD in dairy cattle breeding schemes than the 300,000 suggested by Khatkar and coworkers (31).

As a first step to fine-map the QTL critical region we chose to create family-specific linkage maps of SNP markers and a QTL map of the region. This permitted the identification of historical recombinations and a proof of the accuracy of the assembly in the 16.3 Mbp QTL critical region. The genetic distance between BLA1 and TGLA263 was similar to male recombination rates reported for this interval for these and other sires (16, 26). The QTL analysis confirmed the presence of a QTL for MY within the QTL critical region (4, 26), with a broad peak of significance extending beyond BLA1 and TGLA263, but maximum F-values for F1 and F2 between these two markers. In earlier studies, similar results for FY but not MY DYD in offspring of S1 and S2 were obtained for this interval (26, 49). These minor discrepancies in QTL detection and positioning in F1 and F2 could be due to the differences in the statistical methods employed, the statistical power associated with the smaller family size of F2, the partial overlap of informative markers in the two families, or true underlying variability caused by different maternal alleles on opposing haplotypes. The robust F-values and high marker density strongly support the presence of one or more QTL affecting MY, FY, and PY in the QTL critical region. However, we cannot exclude the possibility that additional QTL affecting lactation traits lie within or closely linked to the QTL critical region.

The next step was to more precisely map the position of the QTL by association analysis in the families and a population of 263 sires that do not have S1, S2, or S5 as a parent. An FDR P ≤ 0.2, which corresponds to raw P ≤ 0.014 in our dataset, was used to adjust for multiple testing, as was recently implemented in human complex disease studies (25, 47, 63). Results revealed two subregions within the 16.3 Mbp interval having SNPs with highly significant association to MY and composition traits that overlapped in S1 and S2 (Fig. 2). The two subregions were located between the markers BLA1 and TGLA263, overlapping with the 9.7 Mbp haplotype block and having a cumulative size of ~2.8 Mbp. This region was selected for targeted resequencing to create a comprehensive set of candidate QTNs, of which 35 were added to the family and population analyses. In all, 22 SNPs were found to be associated with MY, FY, and PY in the DBDR sire population, of which 18 were in subregion 2 and none in subregion 1. The most plausible explanation for these results is that the associations with subregion 1 SNPs observed in F1 and F2 are due to linkage effects, but maternal-origin-specific effects (32), paternal imprinting (61), and other genetic and epigenetic factors causing haplotype heterogeneity (41) cannot be excluded. These mechanisms may also explain the restricted overlap of significant SNPs among the two sires and the DBDR population. Thus, the SNPs we have associated to traits are part of a large multisite haplotype on which many SNPs, CNVs, indels, and retrotranspositions may contribute to the phenotype. The contributions of these other genetic and epigenetic sources of variation to the lactation QTL in this region remain to be determined.

We then identified putative candidate genes and putative QTNs by determining overlap in the lists of SNPs associated with effects on lactation traits in the families and the DBDR population. The reasoning behind this approach is that it eliminates effects due to linkage in families and highlights only those SNPs with consistent effects at both the family and population levels. However, because of high LD in the region, associations with tightly linked polymorphisms cannot be ruled out by this criterion alone. To provide additional support for the candidate QTNs we used functional annotation of known genes in the region and gene expression analysis and looked for evidence of a selective sweep of alleles using the iHS approach. By finding common SNPs significantly associated with lactation traits in all three sires and the DBDR population we were able to identify four candidate QTNs and narrow the QTL interval to 307 Kbp. We went further to show that there are six genes in this region and that the shared set of SNPs is contained in two LD blocks (Fig. 3). Furthermore, the SNPs within and proximal to RAP1A showed unambiguous evidence of a selective sweep (Fig. 5).

Examination of the functional annotation of these six genes in the 307 Kbp interval revealed RAP1A as a leading candidate for the MY and PY QTL (Fig. 3). RAP1A is essential for epithelial acinar structure and lumen formation in the human breast (29). Using the gene model for RAP1A we determined that −632C>G is located in a predicted binding site for the MYB transcription factor, which is known to be involved in regulating cell proliferation and differentiation (42). The polymorphism occurs at a pyrimidine residue in the MYB binding consensus sequence PyAACG/TG (7). Although site-directed mutagenesis studies showed reduced affinity of MYB for the binding site when the pyrimidine was changed to a purine (adenine) residue at this position, the difference was not significant (43). We found significant increases in RAP1A expression in liver or mammary tissue after parturition, but host genotype did not affect mRNA expression levels. Allele-specific expression was not measured, leaving the possibility open that transcription levels are affected by the −632C>G mutation, particularly at time points in development and/or the lactation cycle not tested in this study. We speculate that the RAP1A −632C>G promoter mutation acts to increase ductal branching and possibly lumen diameter thus allowing for greater milk production and yield of milk proteins.

In addition to RAP1A −632C>G other SNPs having significant associations with MY and PY, such as the coding SNPs in OVGP1 and C3H1orf88, are likely to be due to the strong observed LD; however, cis-functional effects cannot be ruled out. RAP1A IVSA4+2037A>C is in perfect LD with RAP1A −632C>G, indicating that the associations are due to LD. However, the association of RAP1A IVSA4+2037A>C with
PY in the DBDR sire population but not in F1 or F2, despite the finding that both S1 and S2 are heterozygous, presents a paradox. These results could indicate different QTL for PY and MY, parent-of-origin effects, or other epigenetic phenomena. Also, this SNP was the only one with a high iHS (+1.7) and associated with a lactation trait, indicating that the selective sweep might be associated with PY, MY, or both traits at the population level. Our finding that the \textit{RAP1A} intronic allele with the positive association to DYD PY in the population (the only association with PY in either the population or in families) is in perfect LD with the \textit{RAP1A} promoter allele associated with higher DYD MY in DBDR sires, F1 and F2, does not allow us to distinguish possible pleiotropy of \textit{RAP1A} alleles from a multilocus QTL model. A selective sweep for the polymorphism(s) affecting PY would unquestionably result in a hitchhiking of the polymorphisms affecting MY and FY on the same haplotype.

A compelling case could be made for the \textit{ADORA3} −39G>T mutation as a candidate QTN for FY because SNPs in \textit{RAP1A} were not significant for this trait in families or the DBDR sires. \textit{ADORA3} belongs to a family of G protein-coupled receptors that are involved in intracellular signaling pathways including inhibition of lipolysis (44). The −39G>T mutation in the promoter region of \textit{ADORA3} creates a TFBS for PAX2 at positions −44 to −36 of the TSS. Mammary parenchyma with a targeted deletion of PAX2 developed normal ductal systems when grafted into wild-type host mammary fat pads but failed to undergo higher-order side-branching and lobular development in response to progesterone (53). The candidacy of \textit{ADORA3} is supported by the finding that this gene was highly expressed in mammary gland 7 days after parturition. Although we did not find an effect of genotype on \textit{ADORA3} transcript levels in mammary gland at the time points sampled (allele specificity was not tested) we cannot rule out differences in expression at other time points during the lactation cycle or during development. As for \textit{RAP1A}, a tightly linked indel or structural variant within or proximal to \textit{ADORA3} cannot be excluded as contributing to FY on the basis of our data. Nevertheless, the relationship of \textit{ADORA3} to mammary development and signaling in lipid pathways strongly supports \textit{ADORA3} as a candidate gene for FY.

The genes \textit{ATPFS1} and \textit{WDR77} had no SNPs with significant associations on lactation traits in the DBDR sire population. These genes are thus unlikely to be responsible for the QTL effects in this region. Four SNPs had significant associations with FY in DBDR sires but not in F1 or F2 (all were homozygous in S1 and S2 except SNP896), of which three are intergenic and one is located in intron 1 of \textit{TMEM167B}, a predicted transmembrane protein with no known function. None of these SNPs are in strong LD with \textit{ADORA3} −39G>T or each other. Thus, our data present a complex picture of the QTL affecting FY in this region of BTA3, with \textit{ADORA3} being the most likely candidate and other SNPs possibly serving as markers for long-range enhancers or other unannotated genes.

This \textit{RAP1A} −632C>G mutation accounts for 1.5 and 1.0% of the phenotypic variation in DYD MY in F1 and F2, respectively, and corresponds to approximately +120 Kg milk in a single lactation for animals carrying the G allele. The fraction of the phenotypic variation in DYD FY explained by \textit{ADORA3} −39G>T in F1 was similar to the effect found for \textit{RAP1A} −632C>G. These effects compare favorably to the largest lactation-related QTN in \textit{DGAT1} that accounts for between 18 and 51% of the phenotypic variation in MY and FP, respectively (23), and the \textit{ABCG2} QTN that has an allele substitution effect of ~1.8 Kg FP (12). The relatively large effects of the \textit{RAP1A} −632C>G and \textit{ADORA3} −39G>T SNPs suggest that selection based on these SNPs will be useful in dairy cattle breeding. It is also noteworthy that the “high” \textit{RAP1A} allele for MY is in perfect positive LD with the “low” \textit{ADORA3} allele for FY, which explains the large effect on fat percentage in milk noted previously for this interval (26). Collectively, because of the LD found in the region, the multisite haplotypes defined by the 23 SNPs significantly associated with lactation traits can be deployed to capture all the genetic variation for MY, FY, and PY in this region of BTA3 using marker assisted breeding schemes or genomic selection (21).

Because it is difficult to prove causation of a DNA polymorphism on a quantitative trait in mammals (50), combined approaches, such as the one described here, help to illustrate the reality of complex trait genetics in segregating populations. Recently, Nadeau (41) has termed the irreducibility of QTL “fractal genetics,” which implies that multiple mutations and/or epigenetic factors on a haplotype acting in cis or trans can affect a quantitative phenotype. What emerges from our study is a post-GWAS view of the genetics of complex traits that allows for a more holistic interpretation of how complex traits are regulated at the genomic level. This reassessment is necessary for the field because it prompts careful scrutiny of the purely reductionist approach that attempts to explain all QTL in terms of QTNs or individual structural variants. Our results are completely consistent with this view and offer support for the hypothesis that for traits with low to moderate heritability, such as the lactation traits, relatively few QTL explain more than 1% of the phenotypic variation (19, 21, 65). For such cases, subtle regulatory mutations are likely, such as the one identified here in the promoter region of \textit{RAP1A}. In addition, other closely linked mutations may contribute to QTL effects on the same or different haplotypes, adding another level of complexity to understanding the mechanisms of QTL action. With more complete information on the genomic architecture of QTL regions, a better understanding of the “missing heritability” of complex traits may be forthcoming.

ACKNOWLEDGMENTS

We gratefully acknowledge Dr. Juan Loor for providing blood samples and liver and mammary gland tissues for expression studies (University of Illinois at Urbana-Champaign). We also thank Dr. Jill Maddox and Ian Evans for providing CRIMAP version 2.5 for linkage analysis. Sequence capture was performed by Roche NimbleGen (Roche NimbleGen, Madison, Wisconsin).

SNPs were submitted to dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP) under batch identification no. 1050967, 1050969, 1050970 and 1050980.

GRANTS

This work was supported by the United States Department of Agriculture Cooperative Research Education and Extension Service, Livestock Genome Sequencing Initiative (AG 2009-54480-19875). M. Cohen-Zinder was supported by Vaadia-BARD postdoctoral fellowship award: (F1-393-2007).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

REFERENCES


Identification of novel DNA-copy number variants (CNVs) within the dog genome using exome capture and deep sequencing. *Genome Biol* 10: R165, 2009.


