Proceedings, 10th World Congress of Genetics Applied to Livestock Production

The Hunt for a Functional Mutation Affecting Conformation and Calving Traits on Chromosome 18 in Holstein Cattle

J.B. Cole[1], J.L. Hutchison[1], D.J. Null[1], P.M. VanRaden[1], G.E. Liu[2], S.G. Schroeder[2], T.P. Smith[3], T.S. Sonstegard[2], C.P. Van Tassell[2], and D.M. Bickhart[1].
[1]Animal Improvement Programs and [2]Bovine Functional Genomics Laboratories, ARS, USDA, Beltsville, MD and [3]Roman L. Hruska U.S. Meat Animal Research Center, ARS, USDA, Clay Center, NE

ABSTRACT: Sequence data from 11 US Holstein bulls were analyzed to identify putative causal mutations associated with calving and conformation traits. The SNP ARS-BFGL-NGS-109285 at 57,589,121 bp (UMD 3.1 assembly) on BTA18 has large effects on 4 measures of body shape and size, 2 measures of dystocia, longevity, and lifetime economic merit. This region includes a human sialic acid Ig-like lectin 6 (Siglec-6) gene. We sequenced a homozygote for the minor allele, 4 carriers, and 3 non-carrier bulls from the same family. Three unrelated carrier bulls also were sequenced. Tandem duplications, insertions, and deletions were detected using custom analysis software that uses paired-end read alignments and split-read mapping. One duplication CNV and two tandem duplication events were detected within Siglec-6. Predicted tandem duplications present in the carrier animals suggest that portions of two exons and a connecting intron within the Ig-like protein domains of Siglec-6 may have been duplicated.
Keywords: calving traits; dairy cattle; quantitative trait loci; whole genome sequencing

Identification of a novel QTL affecting calving traits

Many studies have reported on QTL affecting calving traits in several populations of Holstein cattle (Kühn et al., 2003; Schnabel et al., 2005; Holmberg and Andersson-Eklund, 2006; Kolbehdari et al., 2008; Thomasen et al., 2008; Purfield et al., 2014; Seidenspinner et al., 2011). Cole et al. (2009) initially identified a quantitative trait locus (QTL) on BTA18 with large effects on body depth, lifetime net merit, productive life, rump width, sire and daughter calving ease, stature, and strength in US Holsteins (e.g., Figure 1). The single nucleotide polymorphism (SNP) marker ARS-BFGL-NGS-109285 (rs109478645) at 57,589,121 bp on the UMD 3.1 assembly consistently tracks the QTL. In addition to large additive effects on these traits, genetic correlations among traits differed when computed at the chromosome and whole-genome levels (Table 1). Additional studies have confirmed the presence of the QTL in various Holstein populations (Brand et al., 2010; Sahana et al., 2011; Purfield et al., 2014), and Qanbari et al. (2011) identified a signature of selection in the same region of the cattle genome as the QTL.

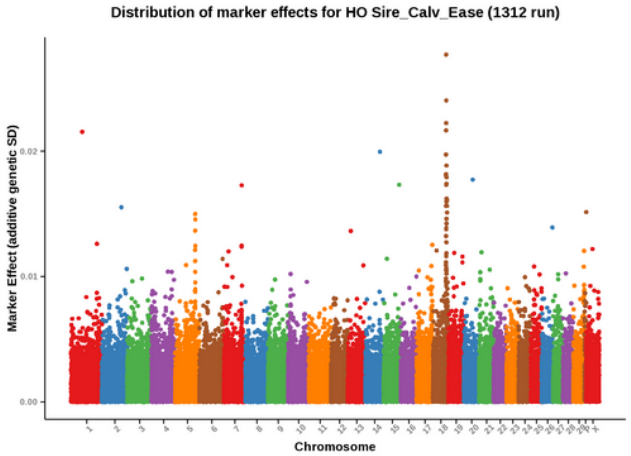

Figure 1. Manhattan plot showing SNP effects in additive genetic standard deviations for sire calving ease in US Holsteins from the December 2013 genetic evaluation run.

Table 1. Correlations among BTA18-specific[1] (above the diagonal) and genome-wide (below the diagonal) sire estimated breeding values for net merit (NM), longevity (PL), sire (SCE) and daughter calving ease (DCE), stature (STA), strength (STR), body depth (BD), and rump width (RW).

	NM	PL	SCE	DCE	STA	STR	BD	RW
NM	0.88	-0.64	-0.67	-0.44	-0.53	-0.6	-0.46	0.88
PL	0.71		-0.72	-0.71	-0.48	-0.62	-0.67	-0.55
SCE	-0.32	-0.31		0.81	0.69	0.79	0.79	0.79
DCE	-0.56	-0.44	0.58		0.57	0.73	0.75	0.69
STA	0.16	0.00	0.25	-0.13		0.82	0.83	0.78
STR	0.02	-0.09	0.28	0.03	0.72		0.95	0.9
BD	-0.01	-0.18	0.30	0.05	0.78	0.91		0.88
RW	0.11	-0.02	0.23	-0.07	0.68	0.71	0.72	0.11

[1]BTA18 = *Bos taurus* autosome 18.

Initial bioinformatics analysis suggested that sequestration of leptin by a sialic acid-binding immunoglobulin-type lectin (Siglec) may result in longer gestation lengths and, in turn, increased calf birth weights. However, the initial hypothesis could not be directly addressed due to the lack of birth weight observations for calves born in the US.

Maltecca et al. (2011) reported that the same QTL had the largest effect of all assayed SNP on sire (direct) gestation length, which provided additional support to the hypothesis that high calf birth weights were driving the observed phenotypic effects. In a recent study, Cole et al.

(2014) used selection index predictors of sire PTA for calf birth weight as phenotypes for a genome-wide association study of birth weight. Their results show that the QTL on BTA18 has a large effect on birth weight.

Identification of the causal variant with next-generation sequencing

Next-generation sequencing has been used to successfully identify causal variants associated with novel fertility haplotypes in US dairy cattle (Sonstegard et al., 2013; McClure et al., 2014), and a similar approach was used to search for the causal variant associated with the BTA18 QTL.

Genetic resources. A pedigree- and haplotype-based scheme was used to select 5 carrier and 3 non-carrier bulls from the same family for sequencing, as well as 3 additional, unrelated carrier bulls (Figure 2). The 8 related animals were selected based on their predicted transmitting ability (PTA) for sire calving ease (SCE), the reliability of the PTA, and the range of PTA in their proven sons. The 3 additional bulls were selected based on their PTA for SCE and their haplotype status for the QTL. Small values are preferred for SCE because they correspond to difficulty-free births. The breed average for US Holsteins is 8%, which means that 8% of Holstein calves are born with some degree of difficulty (Van Tassell et al., 2003).

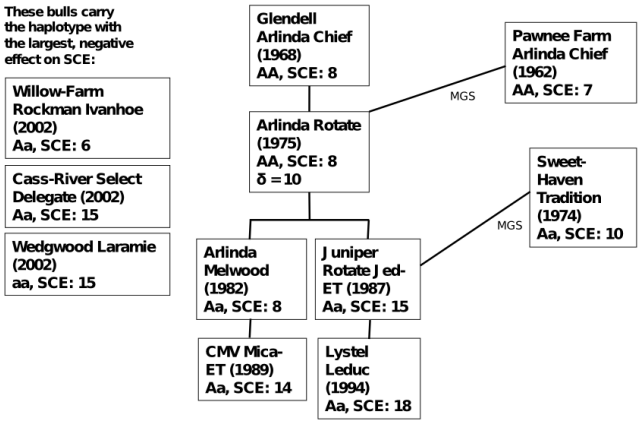

Figure 2. Animals selected for sequencing based on their haplotypes (left column) or pedigree (right column).[1]

[1]A(a) = major (minor) allele for the SNP with the largest effect on sire calving ease (SCE); MGS = maternal grandsire; δ = difference between the PTA SCE of a bulls best and poorest proven sons.

DNA extraction and sequencing. Genomic DNA was extracted from frozen semen purchased commercially or obtained from the Cooperative Dairy DNA Repository (Beltsville, MD). Paired-end libraries were created using the recommended protocol for the Illumina Tru-seq DNA sample prep kit. Samples were sequenced to 20X target coverage in 100 x 100 SBS reactions on the HiSeq 2000 (Illumina, Inc., San Diego, CA).

Analysis of sequence data. Sequence reads were aligned to the UMD 3.1 reference genome using MrsFAST (Hach et al., 2010), and discordantly mapping reads were identified with custom Java programs. VariationHunter (Hormozdiari et al., 2010) was used to identify variants based on read orientation and alignment distance within clusters of read pairs that aligned in the same location.

Results of sequencing. Sequence coverage on the Illumina HiSeq 2000 (Illumina, Inc., San Diego, CA) for all bulls ranged from approximately 7X to 35X coverage (Table 2). Tandem duplications, insertions and deletions were detected using paired-end read alignments and split-read mapping.

Table 2. Sequence metrics for 11 Holstein bulls sequenced on the Illumina HiSeq 2000.

Bull name	Total reads	Coverage
Chief	333,628,731	23.03
Arlinda Chief	981,726,824	35.41
Tradition	390,387,538	14.01
Rotate	~476,000,000	17.00
Melwood	~448,000,000	16.00
Jed	656,190,604	23.66
Mica	433,353,161	15.63
Leduc	767,440,677	27.68
Rockman Ivanhoe	195,769,690	7.06
Delegate	377,380,110	13.61
Laramie	371,477,172	13.39

One duplication CNV and two different tandem duplication events were detected within the gene. Seroussi et al. (2010) and Hou et al. (2011) also have reported the presence of structural variants in this region of the genome. Predicted tandem duplications present in the carrier animals suggest that the portions of two exons and a connecting intron within the Ig-like protein domains of the Siglec-6 gene may have been duplicated. Some heterozygotes with desirable SCE also have deletions near the N-terminal end of the protein.

Strategy for targeted reassembly of chromosome 18

Results from the initial sequencing effort suggested that there are assembly errors in the QTL region on BTA18, possibly due to the presence of repetitive. Several genes involved in leptin signaling may be recovered if the contigs can be properly ordered in a new assembly. The strategy currently being applied is to resequence bovine BAC clones with long-read technology. New contigs encompassing the QTL region of BTA18 will be assembled from the long-read data, and the existing Illumina short-read data will be aligned against the new assembly. A similar strategy was recently applied to a difficult-to-assemble region of the chimpanzee genome (Huddleston et al., 2014).

A set of BAC libraries constructed from the Hereford bull L1 Domino 99375 (registration #: 41170496) and a Holstein calf that map to several regions known to contain misassemblies or marker LD problems in the UMD3.1 ref-

erence assembly will be sequenced at the Roman L. Hruska US Meat Animal Research Center (Clay Center, NE) using a PacBio instrument with P5 chemistry. The assembled PacBio reads will be used to generate larger contigs across the region. These new contigs will be mapped back to the existing reference assembly and confirmed using LD maps from existing SNP genotyping. Sequence data at USDA will be aligned to the new patch regions to identify new markers for use in new genotyping assays. New markers will be validated used validated using Sequenom assays to confirm SNP viability.

Opportunities and challenges

As the cost of sequencing continues to fall it will be tempting to replace genotyping with sequencing, particularly in research settings where there is ready access to computing and laboratory resources. Some authors have proposed the use of genotyping-by-sequencing (e.g., De Donato et al., 2013) in place of SNP genotyping for use in genomic selection and genome-wide association studies, but that may be unwise when rare variants are of interest. The challenge of managing the volume of data produced by modern sequencers also is a concern, particularly when network bandwidth is limited. While the raw reads (e.g., VCF files) can be reduced to binary alignment files, which are much smaller, substantial redundant storage pools are needed in order to protect against data corruption or loss.

Substantial work remains to be done on the annotation of the bovine genome. The *in vivo* function of most cattle genes is inferred from functional studies of homologous proteins in other species. This is valuable in cases where the association of the phenotype and causal variant is clearly characterized, such as the *APAF1* mutation associated with early embryonic loss in Holsteins (Adams et al., 2012), but it is not necessarily that helpful in complex cases. It is not clear how a tandem duplication in a Siglec-6 protein could produce the phenotypes observed in this study, and without a clear understanding of that mechanism it is difficult to state with confidence that a proposed causal variant is the correct one.

Conclusion

Structural variants in and around the Siglec-6 gene are associated with differences in SCE and several other traits. These results demonstrate that sequence data can be used to generate novel hypotheses from quantitative studies. However, higher-quality assemblies and improved annotations are necessary in order to better understand the functional basis for observed QTL effects, particularly those associated with complex phenotypes.

Acknowledgments

Bickhart, Cole, Hutchison, Null, and VanRaden were supported by appropriated project 1265-31000-096-00. Liu, Schroeder, Sonstegard, and Van Tassell were supported by appropriated project 1245-31000-104-00. Smith was supported by appropriated project 5438-31320-012-00.

Literature Cited

Adams et al. (2012). Proc. Plant Anim. Genome XX Conf., abstr. P0555.
Cole et al. (2009). J. Dairy Sci. 92:2931–2946.
Cole et al. (2014). J. Dairy Sci. doi:10.3168/jds.2013-7409.
De Donato et al. (2013). PLoS ONE. doi: 10.1371/journal.pone.0062137.
Hach et al. (2010). Nature Meth. 7:576–577.
Holmberg and Andersson-Eklund. (2006). J. Dairy Sci. 89:3664–3671.
Hormozdiari et al. (2010). Bioinformatics 26:i350–i357.
Hou et al. (2011). BMC Genomics 12:127.
Huddleston et al. (2014). Genome Biol. doi: 10.1101/gr.168450.113.
Kolbehdari et al. (2008). J. Dairy Sci. 91:2844–2856.
Kühn et al. (2003). J. Dairy Sci. 86:360–368.
Maltecca et al. (2011). Anim. Genet. 42:585–591.
McClure et al. (2014). PLoS ONE (Accepted.)
Purfield et al. (2014). Animal 8:224-235.
Qanbari et al. (2011). BMC Genomics 12:318.
Sahana et al. (2011). J. Dairy Sci. 94: 479–486.
Schnabel et al. (2005). Anim. Genet. 36:408–416.
Seidenspinner et al. (2011). J. Anim. Breed. Genet. 126:455–462.
Seroussi et al. (2010). BMC Genomics. 11:673.
Sonstegard et al. (2013). PLoS ONE. 8(1):e5487.
Thomasen et al. (2008). J. Dairy Sci. 91:2098–2105.
Van Tassell et al. (2003). J. Dairy Sci. 86:3366–3373.