Differential expression of ruminant *ZNF496* variants: Association with quantitative trait locus affecting bovine milk concentration and fertility^{1,2}

M. Golik,*† G. Glick,*† S. Reicher,*† A. Shirak,* E. Ezra,‡ Y. Zeron,§ E. Gootwine,* M. Ron,* J. I. Weller,* and E. Seroussi*³

*Agricultural Research Organization, Institute of Animal Science, Bet-Dagan 50250, Israel

†The Hebrew University of Jerusalem, The Robert H. Smith Faculty of Agriculture, Rehovot, 76100, Israel

‡Israel Cattle Breeders Association, Caesaria Industrial Park 38900, Israel

§Sion, Al Institute, Shikmim 79800, Israel

ABSTRACT

A single nucleotide polymorphism in the intergenic region upstream of the *ZNF496* gene on *Bos taurus* chromosome 7 displayed significant population-wide linkage disequilibrium with milk protein percentage in the Israeli Holstein population. The frequency of the allele associated with increased protein concentration was 10%. This single nucleotide polymorphism was located in the promoter region from which a 10-exon transcript of the bovine and the ovine *ZNF496* genes are transcribed. The gene architecture was similar to the mouse ortholog *Zkscan17*. A 5-exon murine antisense transcript was complementary to the 5' untranslated *Zkscan17* region that included a sequence domain conserved between mouse and ruminants, suggesting a regulatory function. In the bovine *ZNF496* chromosomal region, segregation of a quantitative trait locus (QTL) for milk protein percentage was confirmed in a daughter design sire family. Concordance was not obtained between QTL status of bulls and any of the polymorphisms in the functional elements of *ZNF496*. This excludes these variations as the causative polymorphism under the assumption of no epigenetic effect for this locus. However, *ZNF496* variants were differentially expressed in bovine ovaries, and only the paternal variant was expressed in liver and kidney in a sheep family with polymorphic *ZNF496* sequence. Thus, the search for the mutation underlying the minor QTL allele, which is a top economically favorable allele in Israeli Holstein cattle, may be complicated by the presence of an imprinting center in this QTL confidence interval.

Key words: quantitative trait locus, linkage disequilibrium mapping, *ZNF496*, protein concentration

INTRODUCTION

Weller et al. (2008a,b) previously identified QTL affecting economic traits on *Bos taurus* autosome (BTA) 7 in the Israeli Holstein cattle. Several QTL affecting milk protein and fat yield and percentage were mapped to the centromeric half of this chromosome. Population-wide linkage disequilibrium (LD) was tested for 213 SNP in this chromosomal region by analysis of the genetic evaluations of up to 696 bulls using a linear model. Single nucleotide polymorphisms in the zinc finger protein 496 (*ZNF496*) promoter showed significant LD for milk protein percentage ($P < 10^{-10}$).

The *ZNF496* promoter contains SCAN, Kruppel-associated box (KRAB), and zinc finger domains, and physically interacts with the *NSD1* histone methyltransferase to form a transcriptional co-repressor. The *NSD1* histone methyltransferase is involved in the gigantism disorders and childhood acute myeloid leukemia (Nielsen et al., 2004; Losson and Nielsen, 2010). The *ZNF496* promoter also indirectly functions as a transcriptional activator physically interacting with *Jumonji* (*JMJ*)/*Jarid2* and inhibiting the transcriptional repression of *JMJ*, which plays important roles in embryonic development (Mysliwiec et al., 2007). Other physical associations with the arrestin 3, retinal (X-arrestin, *ARR3*), and *ZNF446* gene products have been documented (<http://thebiogrid.org/124291/summary/homo-sapiens/znf496.html>), suggesting involvement of *ZNF496* in the regulation of additional biological processes.

In search for the polymorphism that underlies the BTA7 QTL, we cloned the *ZNF496* gene of cattle and sheep, and demonstrated that this gene transcription has a complex expression pattern typical of imprinted genes, which complicates the uncovering of potential causative mutations for a milk production QTL within this gene.

Received July 27, 2010.

Accepted December 10, 2010.

¹Nucleotide sequence data reported are available in the DDBJ/EMBL/GenBank databases under the accession numbers BN001496; FN908756; FN908757; FN908185; FN908186.

²This article contains supplementary material, which is available at: http://cowry.agri.huji.ac.il/Golik_et_al2010/Supplementary_Table1.xls.

³Corresponding author: Seroussi@agri.huji.ac.il

Table 1. PCR primer pairs for *ZNF496* amplicons

Target ¹	#	Forward 5'→3'	#	Reverse 5'→3'
E2-E6	1	GGTTTTCTGGATGAGCAGGA	2	CGCTTCCTGAAGAATGGAAG
E5-E10	3	ACCTTGCAAAGAGCCAGGAT	4	GGGGAGTCTGAAACCAGTCA
P	5	AAGACCACCTTGCTCCTCT	6	TTCCCAGGAACTGCAAAAC
P	7	TCAACTCCTATTGCGTGCTG	8	AGTTAGTGAGGGCCATGGTG
P	9	GTACCTTCCCATGGTCTCCA	10	TACCCCTTTCTCAGGAGCAC
P	11	AGTCTGACCTGCCTTGGCTA	12	TCCTGAGCAGAGCTGATTGA
P-E1	13	TCAATCAGCTCTGCTCAGGA	14	GTCTCCGCATCTCAGTGGTT
P-I2	15	CAAGGCGGAGAACAGAGAAG	16	TACCGTATGCCAAGGAAAGG
I2-I3	17	AGCAAGGCTTACTCGCTCTG	18	AGAAGGAACAACCCCATTTCC
I3-I4	19	TCCCAAAGGCTGATCTCTGT	20	AGCCCCACTCTCTCCCTTT
I4-I5	21	AATTGCTGGTCTCATCTGG	22	CAGCTGTCTCCAAACCAACA
I5-I6	23	TGACATGTCTTAAAGCAAAGCAAT	24	TTCCCACATGCATATCAA
I6-I8	25	ACATCCCTCCCCTAGCTTGT	26	CTGGTTGAAAGAGGGCAATC
I8-I9	27	AGTACCACCAGGCCTCCTTT	28	CACACATACACGGATGCACA
I9-E10	29	TTCTTGGGTAGGGTTTGCAG	4	GGGGAGTCTGAAACCAGTCA

¹E = exon, P = promoter region, I = intron.

MATERIALS AND METHODS

Bioinformatics

The BLAST family of programs was used for database searches on the NCBI/NIH server (<http://www.ncbi.nlm.nih.gov/BLAST/>). Sheep sequences were also located in the International Sheep Genomics Consortium sequence data repository (<https://isgdata.agresearch.co.nz/>). Sequence of clones and draft contig sequences were downloaded and assembled in a *GAP4* database (Staden et al., 2000), in which all the sequence data relevant to this project were accommodated. This database was used for comparison of PCR amplification of the exon at genomic versus the transcript level. Repetitive sequences were filtered out from the genomic sequence using Repeat Masker server (<http://ftp.genome.washington.edu/>). Predicted amino acid sequences of the 4 *ZNF496* proteins were aligned using the ClustalW (Thompson et al., 1994) and the output analyzed by the Boxshade program. Cellular localization was predicted by PSORT (Nakai and Horton, 1999; <http://psort.nibb.ac.jp/>).

Primer Design, PCR Amplification, and Sequencing

Oligonucleotide primers were designed using Primer Designer 3.0 (Scientific & Educational Software). The DNA templates were PCR amplified in 33 cycles of 92°C, 1 min; 63°C, 1min; 72°C, 2 min. Reverse-transcription polymerase chain reaction (RT-PCR) and sequencing of bovine and ovine mRNA was performed as described previously (Golik et al., 2006; Reicher et al., 2010, respectively), using PCR primers 1 and 2, or 3 and 4 for amplifying the 5' or 3' exons of the *ZNF496* transcript, respectively, as detailed in Table 1. Tissue sections were collected at the abattoir a few minutes

after slaughter and kept in liquid nitrogen. Ovaries were stabilized in RNAlater (Qiagen GmbH, Hilden, Germany). Total RNA extraction was carried out from samples (50 mg) immediately following homogenization in Trizol (Gibco-BRL, Gaithersburg, MD), according to manufacturer's instructions. Reverse transcription of total RNA (1 µg) was performed using 1 U of avian myeloblastosis virus reverse transcriptase (Promega, Israel) as specified in the manufacturer's protocol. The PCR products were size-separated by electrophoresis on 1.2% agarose gels stained with ethidium bromide and were extracted from the gel using QIAquick gel extraction kit (Qiagen, Hilden, Germany) according to manufacturer's instructions, and sequenced from both directions with the ABI BigDye Terminator Sequencing Kit (Applied Biosystems) on an ABI 3730.

Detection and Genotyping of SNP

Large-scale genotyping was performed using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometer (MassARRAY, Sequenom Inc.), and the amplification primers are given in Supplementary Table 1 (available online; see footnote 2 on page 2092). This table also describes the SNP identity, genomic position, the number of bulls genotyped, the minimum allele frequency, and whether the allele distribution corresponded to Hardy-Weinberg equilibrium.

For SNP array analysis, DNA was extracted from the semen of 912 Holstein bulls used for artificial insemination in Israel (http://www.icba-israel.com/cgi-bin/bulls/en/bl_main.htm). Twelve young bulls were excluded from LD analyses, as they have not yet completed their progeny tests. The sample included sires born in Israel, as well as international sires origi-

nating from France (4), Germany (2), the Netherlands (26), and the United States (27). The sire DNA was genotyped using the BovineSNP50 BeadChip (Illumina, Inc., San Diego, CA), which included 54,001 SNP as described previously (Weller et al., 2010). Smaller-scale characterization of SNP within the candidate genes was also performed by DNA sequencing of PCR products as described previously (Golik et al., 2006). It should be noted that positions of 27 SNP overlapped between the mass spectrometry and the BeadChip methods.

Statistical Analysis of SNP Effects on Economic Traits

A total of 696 AI Israeli Holstein bulls were genotyped for SNP of bovine *ZNF496*. Estimated breeding values for milk, fat, and protein were computed by a multitrait animal model analysis of the entire Israeli Holstein population (Weller and Ezra, 2004). These were the standard industry evaluations computed in June 2009. The current EBV are available from http://www.icba-israel.com/cgi-bin/bulls/en/bl_main.htm. Evaluations for fat and protein percent were derived from the evaluations for the production traits, as described previously (Weller and Ezra, 2004). The following fixed linear model was used to estimate the effect associated with this SNP for each of 5 traits analyzed:

$$Y_{ijk} = aS_{ij} + bB_k + c(B_k)^2 + e_{ijk}, \quad [1]$$

where Y_{ijk} is the genetic evaluation of sire i with SNP genotype j and birth year k for each trait; S_{ij} is the number of "A" alleles of the SNP ($j = 0, 1$, or 2) for sire i ; B_k is the sire birth year k ; a , b , and c are regression effects; and e_{ijk} is the random residual for sire i . The linear and quadratic effects of the sire birth year were included to account for genetic trends in the population.

All 225 SNP (Supplementary Table 1; available online; see footnote 2 on page 2092) were also analyzed individually for the 5 milk production traits and fertility by the multiple-trait canonical transformation program (MTC) animal model REML program of I. Misztal (University of Georgia, Athens, GA). In addition to bulls with valid genotypes, all known parents, grandsires, and paternal granddams of these bulls were included in the analysis model. The analysis model was:

$$Y_{ijk} = aS_{ij} + g_{ij} + e_{ijk}, \quad [2]$$

where g_{ij} is the additive genetic effect of animal i , and the other terms are as defined for equation [1]. Two groups were determined for animals with unknown par-

ents. All effects were assumed to be random, and the additive genetic effects were computed with inclusion of the relationship matrix. The substitution effect was computed as $\left\{ \sigma_a^2 / [2p(1-p)] \right\}^{0.5}$, where σ_a^2 = the variance of the SNP effect and p = the minor allele frequency.

Nominal 5 and 1% significance levels were determined empirically for the effect of the SNP at position 39,956,467 bp on BTA7 by permutation analysis. One thousand samples were generated from the data with the relationships and genetic evaluations for protein percent intact, but with genotypes assigned randomly among the genotyped bulls. In addition to individual testing of each SNP, the false discovery rate was computed based on the nominal 1% significance level.

Determination of concordance between a polymorphism and the QTL is based on comparing the marker genotypes to the QTL genotypes for individuals with known QTL genotypes. Quantitative trait loci genotypes for AI bulls can be determined by application of either a daughter or granddaughter design. Eleven Israeli Holstein sires were analyzed previously in the QTL region by the daughter design (Ron et al., 2004). None of these bulls were heterozygous for the QTL affecting protein concentration, and only sire 2283 was heterozygous for the SNP at position 38,243,702. The daughter design was applied to 2 additional sire families, 3652 and 3881, heterozygous for the SNP at position 38,243,702. Regression-based interval mapping was performed by the program of R. Spelman (Spelman et al., 1996). The estimated breeding values of the daughters from the standard June 2009, evaluation were analyzed for the 5 milk production traits and female fertility. All records were weighted equally. Each family was analyzed separately for each trait with the genetic markers on BTA7 as described previously (Weller et al., 2008a). Because these results were used to confirm the results of the LD analysis, significance of the within family contrasts were determined by one-tailed t -tests.

RESULTS

Allele Substitution Effects and Interval Mapping

Population-wide LD for economic traits was tested for 225 SNP located on BTA7 by analysis of the genetic evaluations of up to 696 Holstein bulls by the linear model. Genotyping was performed using Sequenom mass spectrometer. A SNP (A $\rightarrow$ T, build 4.0 position 39,956,467), in the intergenic region upstream the *ZNF496* gene (GeneID: 526480), displayed significant LD with milk production traits, using the linear model for analysis (Table 2). The lowest LD probability ($P <$

Table 2. Substitution effect of genetic markers on sire genetic evaluations for economical traits using a linear model, REML, and the daughter design analyses

Analysis	<i>ZNF496</i> promoter SNP BTA7:39,956,467			Daughter design, sire 3652 (n = 51)	BeadChip SNP BTA7:38,243,702		
	Linear model (n = 696)	REML ¹ (n = 696)			Linear model (n = 900)	REML ² (n = 900)	
	Substitution effect ± SE	Substitution effect ³	% of the total variance		Substitution effect	Substitution effect ± SE ³	Substitution effect ³
kg of milk	−169 ± 60*	355.0	4.5*	−110	−116 ± 51*	ND ⁴	ND
kg of fat	6.4 ± 2.1*	12.1	3.6*	0.8	6.7 ± 1.9*	0.0	0.0
kg of protein	4.0 ± 1.5*	0.0	0.0	1.5	4.5 ± 1.3*	14.0	8.8
% of fat	0.12 ± 0.02**	0.20	10.3**	0.04	0.10 ± 0.02**	0.20	8.2
% of protein	0.09 ± 0.01**	0.12	13.7**	0.05*	0.08 ± 0.01**	0.11	12.6
Fertility	−1.9 ± 0.4**	2.4	5.3*	0.6	1.8 ± 0.4**	0.5	0.2

¹This analysis also included 782 ancestors without genotypes.

²This analysis also included 951 ancestors without genotypes.

³The sign is consistent throughout all the analyses and refers to the favorable allele, except REML, which does not indicate the direction of the effect. For the REML analysis, the substitution effect was computed as $\left\{\sigma_a^2/[2p(1-p)]\right\}^{0.5}$, where σ_a^2 = the variance of the SNP effect and p = the minor allele frequency.

⁴Not done.

* $P < 0.05$; ** $P < 0.0001$.

10^{-10}) was observed for milk protein percentage. The REML analyses for protein percent did not converge for 12 SNP, leaving 213 SNP with valid results. The nominal 1% significance level for the SNP effect based on the permutation analysis was 4% of the variance. Thirteen SNP with effects greater than this were found, for a false discovery rate of 0.16 (2.13/16). Of these, 5 were within 1 Mbp of the SNP at position 39,956,467. The REML analysis, which, unlike the linear model analysis, accounts for the family relationships, estimated a similar substitution effect for protein percentage for this SNP (0.12%). This SNP explained 14% of the variance among the bull genetic evaluations (Table 2), which was more than any of the other SNP analyzed, except one. No effect was found for protein yield in the REML analysis. The frequency of the allele with positive effects on fat and protein concentration was 10%.

Daughters of 3 bulls heterozygous for this SNP (2283, 3652, and 3881) were analyzed by the daughter design (Weller et al., 1990; Ron et al., 2001). The effect associated with this chromosomal region was significant by interval mapping ($P < 0.05$ for the single-tailed t -test) for the protein percentage trait only within the smaller family of sire 3652 (n = 51, Table 2), and neither significant effect nor trend was detected within the 2283 (n = 325) and 3881 (n = 109) families. It is expected that the effects detected by a daughter design should be smaller, because the genetic evaluations of cows are more highly regressed, as compared with bulls.

Population-wide LD for economic traits in this critical region on BTA7 was also analyzed by the linear model using the BovineSNP50 BeadChip (Weller et al.,

2010) and the breeding values of 900 Holstein bulls. Although SNP located near the *ZNF496* locus did not display significant effects on milk production traits, a more centromeric SNP (A → C, build 4.0 position 38,243,702) displayed significance at $P < 10^{-12}$ (Table 2). The frequency of the C allele was 10%, which is typical of the rare haplotype associated with increase of protein percentage in milk (see the haplotype described in the section entitled “Isolation and genotyping of bovine *ZNF496* sequence variations”). These results correspond to the fact that most of the SNP genotyped by the Sequenom mass spectrometer near this gene did not display significant LD for the milk production traits, and did not have allelic frequencies similar to the SNP at position 39,956,467. Segregation status for the SNP at position 38,243,702 was identical to the SNP at position 39,956,467 in 97% of the 655 sires that were analyzed by both the mass spectrometry and the BeadChip methods. Thus, it is not surprising that the magnitude of the effects were also very similar. In the REML analysis, a substitution effect for protein percentage similar to the linear model was detected, and this SNP explained 13% of the variance among the bull EBV (Table 2). Hence, based on the REML and linear model analyses, we were able to confirm segregation of a QTL for protein percentage near the *ZNF496* gene. Estimated effects by the REML analysis for milk, fat, and protein production, and fertility were not consistent for the 2 SNP, and also differed from the linear model analyses. This is apparently due to the fact that more animals were included in the analysis of BeadChip SNP, and confidence intervals for effects from the

Table 3. Genomic organization of the *ZNF496* gene on BTA7¹

Intron		Exon		Intron	
Sequence	No.	Size	Sequence	Size	
acgacacaaatgtttttcagGCGCCT	1	180	GAGGAGgtgagtgcggccggcgccgc	168	
gtctgtctttgtcttttcagCTCATT	2	226	TGTGAGgtgaggagccggagctggc	1,380	
tgattttctgtttgaaatagGGTCGT	3	117	TCGAAGgtaccacagagtgttaatta	348	
agccctctctgggattgcagCTCAAG	4	427	CAATGGgtgagcagaggggctgggc	411	
tcctttctgatctcttcagTCCTGC	5	184	ACCCAGgtagggtgccatgggagtg	6,161	
tactgtctctgtgttcagAACCGG	6	77	CCCCGgtgagtgtttctggggcag	25,468	
ccttcctctctgtgaatagATGATG	7	133	CTCCAAGtaagacttaactcagctct	500	
tcattttctctctggaaagACTTTC	8	108	ACTTGGgtgaggaatgataccagaac	1,083	
accttggtgctttctgtgtagTGGGAG	9	114	GCCCAGgtaaggaattcttaggagg	5,936	
	10	3,451	AGGTGCAAAATAAAATTTGACTTGC		
			TGAATT+poly (A) tail		

¹Exon and intron sizes are given in base pairs. Intron and exon sequences are written in lowercase and uppercase letters, respectively. The first and last 2 bases of introns (gt and ag for donor and acceptor splice sites, respectively) are in bold type. The polyadenylation site (AATAAA) is in bold and underlined type. The genomic size of the gene was 46,472 bp.

REML analyses are greater, because the analysis model also included the additive polygenic effect.

Cloning of Long Variants of the *ZNF496* Gene

BLASTN search of ruminant *dbEST* using human *ZNF496* mRNA (GenBank Accession No. NM_032752) and its orthologous bovine genomic sequence identified 41 expressed sequence tags (GenBank Accession Nos. EE805798, DY516712, DY148335, DN515039, CB168685, DY465768, BM090220, GO719870, GO715741, EH168881, EE854440, GO682070, GO684854, BF654452, DN514372, DY466869, DN534605, EE222597, DY476627, BG691593, EE377856, DV916651, DV933966, EH126093, CX952359, CK727881, DY166125, CU653278, DY165480, GO744973, BG694264, BP111411, GO351570, CB434683, CB418460, GO767834, GO739593, BE487172, DY126130, BE668425, AW430191), which indicated that the transcript described for this bovine gene (GenBank Accession No. XM_002689056.1) was missing exons in its 5' end. Following these findings, we predicted the existence of a long *ZNF496* transcript and submitted this putative transcript as a third-party annotation (TPA; Table 3, GenBank Accession No. BN001496). Ten exons formed this transcript, and all exon-intron borders displayed typical donor and acceptor sequence motifs (Table 3). This assembled cDNA was of 5,017 bp, and the fifth ATG codon was the putative initiator of translation according the orthologous human protein (Figure 1). This codon was located on the fourth exon. The last exon had a stop codon (TGA) in orthology to the human transcript and had a typical polyadenylation site (AATAAA at base 4994). The bovine transcript could encode a polypeptide of 587 AA with the predicted molecular mass of 66.4 kDa. To corroborate the TPA prediction, we designed PCR primers

(#1–4, Table 1) and were able to RT-PCR amplify a transcript spanning exons 2 to 10 (GenBank Accession Nos. FN908756; FN908757; FN908185), using total mRNA extracted from bovine mammary gland biopsy and ovine ovary as the cDNA templates (Golik et al., 2006). The SNP in position 39,956,467, which is associated with protein percentage (Table 2), is located 1,846 bp upstream of the transcription start site of the long bovine *ZNF496* transcript and, therefore, is within its predicted promoter region.

Alignment of the *ZNF496* orthologs (Figure 1) from different species indicated that the closely related bovine and ovine orthologs (98% identity, 99% similarity) were more similar to the human (83–84% identity, 88–89% similarity) than to murine ortholog (72% identity, 79% similarity). Ruminant orthologs shared sequence motifs that suggest nuclear localization: a bipartite nuclear localization signal (NLS) and 2 shorter Pat4 and Pat7 sequence motifs (Figure 1), and the *PSORT* computer program predicted nuclear localization of these orthologs with scores of 76%. Three types of domains typical to the *ZNF496* architecture (Nielsen et al., 2004; Mysliwiec et al., 2007) were conserved in all orthologs (Figure 1): 1) SCAN, which is a leucine-rich N-terminal motif that plays an important role in controlling the assembly of complexes of transcriptional regulators; 2) Kruppel-associated box (KRAB) that functions to repress transcription by recruiting a transcriptional co-repressor; and 3) C2H2-type zinc fingers of the DNA-binding motif.

Mouse Noncoding RNA From the Opposite Strand is Overlapping a Conserved Sequence Element in the First Exon Within the Untranslated Region of the Bovine *ZNF496* Transcript

To investigate the function of the 5' noncoding region of the long *ZNF496* transcript, we *BLASTN*-searched

Figure 1. Alignment of predicted amino acid sequences of the *ZNF496* proteins. The human and mouse reference sequences were aligned with the ovine and bovine orthologs (GenBank Accession Nos. NP_116141; NP_766529; CBM42648; CBM43246, respectively). Black and gray boxes indicate identity and similarity between the amino acid sequences, respectively. White boxes indicate nonconservative amino acid changes between the proteins. Dots indicate gaps introduced by the alignment program. SCAN, KRAB, and zinc fingers domains are delineated. Nuclear localization signals (pat4 pat7 NLS and a bipartite NLS) are annotated in light gray.

the *fRNAdb* database of noncoding RNA (ncRNA) sequences using exons 1 to 3 of the bovine *ZNF496* mRNA as a query. We found 4 murine sequences producing significant alignments (88–92% identity, GenBank Accession Nos. AK011602, AK017115, AK030920, AK163429) with a 113-bp element in the first exon of the 5' untranslated region (UTR) of the bovine *ZNF496* transcript. Examination of these murine transcripts indicated that their 5-exon genomic structure has typical donor and acceptor sequence motifs (Table 4), which indicated that they are transcribed from the strand opposite to that of *Zkscan17* with an overlap over the first 2 exons of the 5' UTR of the reference *Zkscan17* transcript (GenBank Accession No. NM_001130529). Mouse and cattle *ZNF496* orthologs share a similar 10-exon genomic structure with nontranslated exons at their 5' ends. This genomic structure, including its representative *cis*-antisense transcripts, is viewable in the FANTOM3 interface (<http://fantom3.gsc.riken.jp/db/annotate/main.cgi?masterid=F630207M06>).

Isolation and Genotyping of the Bovine *ZNF496* Sequence Variations

The promoter region and the exons of bovine *ZNF496* of 5 heterozygotes (sires 2283, 3259, 3652, 3881, and

5190) and 2 homozygotes (sires 3883, 7185) for the rare allele of the SNP (BTA7:39,956,467) upstream of bovine *ZNF496* were sequenced. The PCR primers are given in lines 5 to 29 of Table 1. All of these sires also carried an 18-bp indel in exon 2. Twenty-three variations were detected in these sires. In total, 15 polymorphisms were segregating in all of the heterozygotes, of which 11 were located in the promoter region (Table 5). The rare haplotype that consisted of these variations was evident from the 2 homozygotes and was possibly present in all heterozygotes. Haplotype analysis indicated that the insertion of 18 bp in the second exon (+ allele) was in gametic association with the QTL allele increasing the protein percentage. The rare haplotypes of sires 2283, 3652, and 3881 were identical, even though only sire 3652 was segregating for the QTL as determined by the daughter design analysis. Thus, concordance with the QTL could not be established for any of the polymorphisms detected so far.

Differential Expression of *ZNF496* Variations

The observed lack of concordance between the QTL and the 18-bp indel in exon 2 of bovine *ZNF496* should have excluded the insertion allele as the causative QTL mutation, assuming Mendelian inheritance without

Table 4. Genomic organization of the *ZNF496* cis-antisense transcript on MMU11¹

Intron		Exon		Intron	
Sequence	No.	Size	Sequence	Size	
acgacacaaatgtttttc ag GCGCCT	1	1,225	GTCCTG gt gagtacggagagagaagg	304	
tctgtgtttcttctgcat ag AACCTG	2	122	TAGTAG gt atggagttaagggtcgag	518	
tgattcttccacattt gag GGATCG	3	108	CTTCAG gt acctccctcatacaggag	1,715	
gtctactgagtgttcc ag GACAGC	4	97	TGGCAG gt acgggatactcaagtat	535	
	5	1,100	GATGCCTG <u>AATAAA</u> ACTCCATTTTGT		

¹Exon and intron sizes are given in base pairs and correspond to GenBank Accession No. AK163429. Intron and exon sequences are written in lowercase and uppercase letters, respectively. The first and last 2 bases of introns (gt and ag for donor and acceptor splice sites, respectively) are in bold type. A putative polyadenylation site (AATAAA) is in bold and underlined type. The genomic size of the gene was 5,724 bp.

epigenetic effects. However, sequence analysis of transcripts derived from ovaries of 5 cows with nucleotide polymorphisms in the last exon indicated that the variations observed in the genomic DNA were not equally expressed in the 3 cows that were heterozygous for the genomic sequence of *ZNF496* exon 10. In the case of c.2234C > T, although the homozygous cow expressed the common “C” variant (Figure 2A), the heterozygous cow expressed only the rare (allele frequency 5%, SNP genomic position 39,910,492) “T” variant (Figure 2B). This differential expression may indicate the existence of a complex mechanism, such as classical imprinting or copy number variation (CNV), and is typical of

situations in which the missing variation is either not expressed or quickly degraded. Conservation of the differential expression in other ruminates may point to the general importance of this mechanism in controlling *ZNF496* transcription. To determine if such pattern of expression is conserved in other ruminants, we investigated the expression of genomic variations in the last exon of ovine *ZNF496* in a sheep family (Figure 2C). A chromatogram of the genomic DNA sequences of heterozygote revealed an ambiguous trace at position corresponding to the c.1719G > A mutation and at the positions that followed an indel. The short variant of this indel had a 3-bp deletion

Table 5. Genotypes and QTL status of the SNP 39,956,467 rare allele carriers

Genetic marker	Position ¹	Bull number and QTL status ²									
		5190 ND		3881 NS		3259 ND		2283 NS		3652 S	
BeadChip SNP BTA7	38,243,702	C	A	C	A	C	A	C	A	C	C
SNP, <i>ZNF496</i> intron8	39,917,453	C	C	C	C	C	T	C	C	C	C
SNP, <i>ZNF496</i> intron8	39,917,570	C	C	C	C	C	C	C	T	C	C
SNP, <i>ZNF496</i> intron8	39,918,152	C	C	C	A	C	C	C	C	C	A
SNP, <i>ZNF496</i> intron7	39,918,469	C	C	C	C	C	C	C	A	A	A
SNP, <i>ZNF496</i> intron6	39,944,866	G	A	G	A	G	A	G	A	G	G
SNP, <i>ZNF496</i> intron5	39,951,114	C	C	C	G	C	C	C	G	C	C
SNP, <i>ZNF496</i> intron3	39,952,525	G	C	G	C	G	C	G	C	G	G
SNP, <i>ZNF496</i> intron2	39,952,728	C	G	C	C	C	C	C	C	C	C
18 bp indel, <i>ZNF496</i> exon2	39,954,067-85	+	–	+	–	+	–	+	–	+	+
9 bp indel, <i>ZNF496</i> intron1	39,954,414-23	–	+	–	+	–	+	–	+	–	–
SNP, <i>ZNF496</i> promoter	39,955,084	A	G	A	G	A	G	A	G	A	A
SNP, <i>ZNF496</i> promoter	39,955,129	A	G	A	G	A	G	A	G	A	A
SNP, <i>ZNF496</i> rs43517615	39,955,354	C	T	C	T	C	C	C	T	C	C
SNP, <i>ZNF496</i> promoter	39,955,427	T	C	T	C	T	C	T	C	T	T
SNP, <i>ZNF496</i> rs29013512	39,955,552	A	G	A	G	A	G	A	G	A	A
SNP, <i>ZNF496</i> promoter ³	39,956,467	T	A	T	A	T	A	T	A	T	T
SNP, <i>ZNF496</i> promoter	39,956,729	C	A	C	A	C	A	C	A	C	C
SNP, <i>ZNF496</i> promoter	39,957,657	C	T	C	T	C	T	C	T	C	C
SNP, <i>ZNF496</i> promoter	39,957,713	C	C	C	A	C	C	C	A	C	C
SNP, <i>ZNF496</i> promoter	39,958,123	T	A	T	A	T	A	T	A	T	T
SNP, <i>ZNF496</i> promoter	39,958,187	C	A	C	A	C	A	C	A	C	C
SNP, <i>ZNF496</i> promoter	39,958,372	ND	ND	T	C	T	C	T	C	T	T
SNP, <i>ZNF496</i> rs29013511	39,958,639	G	C	G	C	G	C	G	C	G	G

¹The positions (bp on build 4) in bold type were analyzed both by sequencing and MS.
²Alleles in bold type correspond to the rare haplotype. ND = not determined, NS = non-segregating, S = segregating for the QTL.
³The SNP analyzed in Table 2.

(GenBank Accession No. FN908186), which resulted in the deletion of the serine residue at position 407 similarly to the human and mouse orthologs. Analysis of the mother genomic DNA (Dam DNA) revealed only the short variant. However, sequence analysis of transcripts from kidney and liver of the daughter revealed only the sequence of the paternal long variant (Figure 2C). In another unrelated individual, only the short variant was detected, indicating that this variant transcript could be expressed in liver (data not shown). An ambiguous base in the cDNA (Figure 2C, dashed arrow) showed that this position does not match the sequence of the genomic DNA obtained from lymphocytes, which may be explained either by somatic mutation or RNA editing (Farajollahi and Maas, 2010). Hence, a complex

mechanism of regulation of *ZNF496* variants was also observed in the sheep mRNA expression.

DISCUSSION

Milk production traits are determined by the contributions of a large number of genes and environmental factors. Nevertheless, the size of the effect of the QTL we describe on BTA7 places this locus in the top 10 QTL detected in the Holstein cattle genome. Three independent analyses validate this QTL. Two of them were population-wide LD analyses of SNP, including mass spectrometry for 225 SNP in a sample of 696 sires and BovineSNP50 BeadChip analysis for 900 sires. The third line of evidence was based on the daughter de-

Figure 2. Differential expression of *ZNF496* variations in cattle and sheep. Chromatograms of genomic and cDNA sequences of bovine (A, B) and ovine (C) *ZNF496*. The sequences around SNP 39,910,998 for DNA expressed in the ovaries (cDNA) and genomic (DNA) sequences in cow were compared for a cow homozygous for the common variation (A) and a cow heterozygous for the rare variation (B). The arrow below the chromatogram of the polymorphic DNA points to this SNP position. Only the rare allele was detected in the cDNA of the heterozygous cow. Sequence chromatograms derived from the last exon of ovine *ZNF496* (C) were compared between the genomic DNA extracted from daughter (DNA) and mother (Dam DNA) blood lymphocytes and transcripts expressed in the daughter tissues (kidney and liver cDNA). Solid arrow and line annotate the c.1719G > A mutation and the positions that followed an indel variation, respectively. The dashed arrow indicates the position where the cDNA sequence differs from the genomic DNA sequence. Color version available in the online PDF.

sign (Ron et al., 2001). In both REML analyses, based on both the mass spectrometry and the BeadChip SNP, a similar substitution effect for protein percentage was detected, and the 2 SNP analyzed (Table 2) explained about 13% of the variance among the bull EBV. Discrepancies between LD and REML analyses are expected, because the LD analysis does not include the polygenic effect. Related animals will tend to have similar genotypes for a specific marker, but will also have a common polygenic effect. The reason that the magnitude of the LD effect was lower in the daughter design analyses was due to the fact that EBV of cows are much more regressed than the genetic evaluations of bulls (Ron et al., 2001). Of all genomic effects we observed in BovineSNP50 BeadChip analysis, when sorted according to their fraction of total variance of protein percentage as calculated by REML analysis (data not shown), this BTA7 QTL was the eighth in importance following BTA6 (Cohen-Zinder et al., 2005), BTA14 (Weller et al., 2003), BTA3 (36 Mb), BTA19 (14 Mb), BTA9 (25 Mb), BTA18 (56–63 Mb), and BTA1 (1 Mb). However, unlike most of these other effects, the minor QTL allele was the economically favorable allele. With 10% frequency of the minor allele, selection for this allele should have a significant economical effect.

Various studies have suggested that daughter yield deviations (**DYD**) or deregressed evaluations should be analyzed for estimation of QTL effects (e.g., Bolormaa et al., 2010), because QTL effects estimated by analysis of EBV will be biased. Israel and Weller (1998) demonstrated by simulation that QTL estimates derived by analysis of DYD are also biased, and that probability values of the effects obtained were equal for EBV and DYD. Therefore, following Ron et al. (2004) we decided to analyze EBV throughout.

Our initial LD observation pointed to the significant effect ($P < 10^{-10}$) for SNP in the *ZNF496* promoter, which prompted us to examine this candidate gene, taking into account that such LD was expected to decrease the QTL confidence interval to less than one map unit (Ron and Weller, 2007). Our recent results from BovineSNP50 BeadChip analysis indicated that the LD peak extends 2 map units further toward the centromere, which places a large number of candidate genes within the confidence interval. Nevertheless, the *ZNF496* remains an attractive candidate, because the differential expression of its variants, which is typical of imprinted genes, may explain the lack of concordance between the suggested QTL haplotype and the QTL status observed in the daughter design. The *ZNF496* gene product represses transcription by recruiting the transcriptional co-repressor *NSD1* that was also implicated in genomic imprinting (Baujat et al., 2004). Another important developmental gene interacting with *ZNF496*

is *JMJ*, which has essential roles in the development of multiple tissues and in cell proliferation during development (Takeuchi et al., 2006). Thus, the interactions of *ZNF496* with chromatin regulators and developmental genes make this gene a likely candidate to affect morphogenesis of organs (http://func.med.harvard.edu/human/genes/list_functional_scores/162714). Indeed, *Zkscan17* expression changes during mammary gland development. The pattern of *Zkscan17* expression was analyzed by Ron et al. (2003) during different physiological stages of the mouse mammary gland at virgin (d 42), pregnancy (d 14), lactation (d 10), and involution (d 4) using a DNA array (Tanaka et al., 2000). The UTR of 5'-end of *Zkscan17* transcript was represented by clone H3126E03 in this array. The expression of this gene changed significantly ($P < 0.01$) during this gland development and differentiation. The maximal 2.2-fold change of the normalized intensities of this clone was detected between the lactation and involution stages.

The genomic structure we reported for the bovine and ovine *ZNF496* genes uncovered a complex 5' UTR that stretched over 4 exons and contained a 113-bp element conserved in other mammals. Hence, the murine antisense transcript that overlapped this element suggests 5' UTR involvement in antisense regulation of *ZNF496* in other mammals as well. Imprinted genes are associated with antisense transcripts (Reik and Walter, 2001), and this antisense transcription, therefore, may correspond to the differential expression of variants typical of imprinted genes that we observed in the bovine and ovine *ZNF496* loci. The 18-bp insertion in exon 2 that was part of the rare haplotype associated with the QTL allele of increased protein concentration is a variation that may play a role in 5' UTR regulation.

Concordance between QTL status of segregation and heterozygosity of the candidate variation is considered to be a critical test for identifying the polymorphism that underlies a QTL (Ron and Weller, 2007; Seroussi, 2009). For the 18-bp insertion, as well as for all other variations on the haplotype associated with increased milk protein concentration, concordance was not obtained, and lack of concordance is expected to extend to the other candidate genes along the rare haplotype that is associated with the favorable QTL allele. This would have excluded them from underlying the QTL under the model of simple Mendelian inheritance. However, the differential expression pattern we observed for the *ZNF496* gene complicated this analysis. Imprinted genes are often part of imprinted clusters (Reik and Walter, 2001). The potential existence of an imprinting center in the confidence interval of the QTL we described on BTA7 should, therefore, be taken into consideration in the search of the underlying mutation, and may explain

why carriers of the QTL haplotypes that were expected to be inherited from the same familial background did not segregate using the daughter design analysis. When considering imprinting models for the *ZNF496* chromosomal region, maternal imprinting was the model that was concordant with the data. In case of paternal imprinting, no segregating QTL effect would be detected in heterozygous sires. This was also in concordance with the observation in sheep that on the mRNA level only the paternal allele was expressed (Figure 2B). However, this simple model cannot explain the observation that 2 bulls were heterozygous for the rare haplotype and did not display QTL segregation by the daughter design analysis. This lack of segregation may be explained either by existence of an identical haplotype that is not associated with the favorable QTL allele, or by a complex epigenetic model that confers partial imprinting to this allele by interacting with unknown factors.

CONCLUSIONS

In the bovine *ZNF496* chromosomal region on BTA7, the presence of a QTL affecting economic traits of dairy cattle was demonstrated by population-wide LD analysis and by analysis of a sire family by the daughter design. Whole genome analysis using BovineSNP50 BeadChip indicated that the less frequent QTL allele is an important economically favorable allele in Israeli Holstein cattle. The 5' UTR of the *ZNF496* gene transcript consisted of 3 exons, which included a conserved element involved in anti-sense transcription. A 18-bp insertion in this UTR was observed in the *ZNF496* gene variant that was associated with the favorable allele. However, concordance was not obtained between QTL status of bulls and any of the polymorphisms we detected in this gene. This excludes these variations as the causative polymorphism under the assumption of no epigenetic effect for this locus. Nevertheless, *ZNF496* variants were differentially expressed in bovine ovaries, and only the paternal variant was expressed in liver and kidney in a sheep family with polymorphic *ZNF496* sequence. Thus, the search for the mutation underlying may be complicated by the presence of an imprinting center in the QTL confidence interval.

ACKNOWLEDGMENTS

Contribution from the Agricultural Research Organization, Institute of Animal Science, Bet Dagan, Israel, No. 569/10 is acknowledged. This research was supported by a grant from the Israel Milk Marketing Board and the European Sixth Research and Technological Development Framework Programme, Proposal No. 016250-2 SABRE. We thank I. Misztal and R.

Spelman for use of their computed programs. MassArray genotyping of SNP was performed by T. Koch and E. Ben-Asher (Genome Knowledge Center, the Weizmann Institute of Science, and the Crown Human Genome Center, Rehovot, Israel). BeadChip genotyping of SNP was performed by A. Schein and N. Avidan (Pharmacogenetics and Translation Medicine Center, The Rappaport Institute for Research in the Medical Sciences, Technion, Haifa, Israel).

REFERENCES

- Baujat, G., M. Rio, S. Rossignol, D. Sanlaville, S. Lyonnet, M. Le Merer, A. Munnich, C. Gicquel, V. Cormier-Daire, and L. Colleaux. 2004. Paradoxical NSD1 mutations in Beckwith-Wiedemann syndrome and 11p15 anomalies in Sotos syndrome. *Am. J. Hum. Genet.* 74:715–720.
- Bolormaa, S., J. E. Pryce, B. J. Hayes, and M. E. Goddard. 2010. Multivariate analysis of a genome-wide association study in dairy cattle. *J. Dairy Sci.* 93:3818–3833.
- Cohen-Zinder, M., E. Seroussi, D. M. Larkin, J. J. Loo, A. Everts-vander Wind, J. H. Lee, J. K. Drackley, M. R. Band, A. G. Hernandez, M. Shani, H. A. Lewin, J. I. Weller, and M. Ron. 2005. Identification of a missense mutation in the bovine *ABCG2* gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Res.* 15:936–944.
- Farajollahi, S., and S. Maas. 2010. Molecular diversity through RNA editing: A balancing act. *Trends Genet.* 26:221–230.
- Golik, M., M. Cohen-Zinder, J. J. Loo, J. K. Drackley, M. R. Band, H. A. Lewin, J. I. Weller, M. Ron, and E. Seroussi. 2006. Accelerated expansion of group IID-like phospholipase A2 genes in *Bos taurus*. *Genomics* 87:527–533.
- Israel, C., and J. I. Weller. 1998. Estimation of candidate gene effects in dairy cattle populations. *J. Dairy Sci.* 81:1653–1662.
- Losson, R., and A. L. Nielsen. 2010. The NIZP1 KRAB and C2HR domains cross-talk for transcriptional regulation. *Biochim. Biophys. Acta* 1799:463–468.
- Mysliwiec, M. R., T. Kim, and Y. Lee. 2007. Characterization of zinc finger protein 496 that interacts with *Jumonji/Jarid2*. *FEBS Lett.* 581:2633–2640.
- Nakai, K., and P. Horton. 1999. PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* 24:34–36.
- Nielsen, A. L., P. Jørgensen, T. Lerouge, M. Cerviño, P. Chambon, and R. Losson. 2004. *Nizp1*, a novel multitype zinc finger protein that interacts with the NSD1 histone lysine methyltransferase through a unique C2HR motif. *Mol. Cell. Biol.* 24:5184–5196.
- Reicher, S., E. Seroussi, and E. Gootwine. 2010. A mutation in gene *CNGA3* is associated with day blindness in sheep. *Genomics* 95:101–104.
- Reik, W., and J. Walter. 2001. Genomic imprinting: Parental influence on the genome. *Nat. Rev. Genet.* 2:21–32.
- Ron, M., M. Band, M. Shani, J. P. Gregg, H. A. Lewin, and J. F. Medrano. 2003. Gene expression patterns in murine mammary development and involution. Poster presentation in Plant and Animal Genomes XI Conference. San Diego, CA.
- Ron, M., E. Feldmesser, M. Golik, I. Tager-Cohen, D. Kliger, V. Reiss, R. Domochofsky, O. Alus, E. Seroussi, E. Ezra, and J. I. Weller. 2004. A complete genome scan of the Israeli Holstein population for quantitative trait loci by a daughter design. *J. Dairy Sci.* 87:476–490.
- Ron, M., D. Kliger, E. Feldmesser, E. Seroussi, E. Ezra, and J. I. Weller. 2001. Multiple quantitative trait locus analysis of bovine chromosome 6 in the Israeli Holstein population by a daughter design. *Genetics* 159:727–735.
- Ron, M., and J. I. Weller. 2007. From QTL to QTN identification in livestock—Winning by points rather than knock-out: A review. *Anim. Genet.* 38:429–439.

- Seroussi, E. 2009. The concordance test emerges as a powerful tool for identifying quantitative trait nucleotides: Lessons from BTA6 milk yield QTL. *Anim. Genet.* 40:230–234.
- Spelman, R. J., W. Coppieters, L. Karim, J. A. M. van Arendonk, and H. Bovenhuis. 1996. Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. *Genetics* 144:1799–1808.
- Staden, R., K. F. Beal, and J. K. Bonfield. 2000. The Staden package, 1998. *Methods Mol. Biol.* 132:115–130.
- Takeuchi, T., Y. Watanabe, T. Takano-Shimizu, and S. Kondo. 2006. Roles of *jumonji* and *jumonji* family genes in chromatin regulation and development. *Dev. Dyn.* 235:2449–2459.
- Tanaka, T. S., S. A. Jaradat, M. K. Lim, G. J. Kargul, X. Wang, M. J. Grahovac, S. Pantano, Y. Sano, Y. Piao, R. Nagaraja, H. Doi, W. H. Wood III, K. G. Becker, and M. S. H. Ko. 2000. Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. *Proc. Natl. Acad. Sci. USA* 97:9127–9132.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. Improved sensitivity of profile searches through the use of sequence weights and gap excision. *Comput. Appl. Biosci.* 10:19–29.
- Weller, J. I., and E. Ezra. 2004. Genetic analysis of the Israeli Holstein dairy cattle population for production and nonproduction traits with a multitrait animal model. *J. Dairy Sci.* 87:1519–1527.
- Weller, J. I., G. Glick, E. Ezra, Y. Zeron, E. Seroussi, and M. Ron. 2010. Paternity validation and estimation of genotyping error rate for the BovineSNP50 BeadChip. *Anim. Genet.* 41:551–553.
- Weller, J. I., M. Golik, S. Reikhav, R. Domochofsky, E. Seroussi, and M. Ron. 2008a. Detection and analysis of quantitative trait loci affecting production and secondary traits on chromosome 7 in Israeli Holsteins. *J. Dairy Sci.* 91:802–813.
- Weller, J. I., M. Golik, E. Seroussi, E. Ezra, and M. Ron. 2003. Population-wide analysis of a QTL affecting milk-fat production in the Israeli Holstein population. *J. Dairy Sci.* 86:2219–2227.
- Weller, J. I., M. Golik, E. Seroussi, M. Ron, and E. Ezra. 2008b. Detection of quantitative trait loci affecting twinning rate in Israeli Holsteins by the daughter design. *J. Dairy Sci.* 91:2469–2474.
- Weller, J. I., Y. Kashi, and M. Soller. 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J. Dairy Sci.* 73:2525–2537.