DOI: 10.1534/genetics.104.031559

Mapping of a Milk Production Quantitative Trait Locus to a 420-kb Region on Bovine Chromosome 6

Hanne Gro Olsen,* Sigbjørn Lien,*,†,[1] Mathieu Gautier,*,‡ Heidi Nilsen,* Arne Roseth,† Paul R. Berg,† Kristil K. Sundsaasen,† Morten Svendsen§ and Theo H. E. Meuwissen*,†

**Department of Animal and Aquacultural Sciences, Agricultural University of Norway, N-1432 Aas, Norway, †Centre for Integrative Genetics, Agricultural University of Norway, N-1432 Aas, Norway, ‡Laboratoire de Génétique Biochimique et de Cytogénétique, Institut National de Recherche Agronomique, 78352 Jouy en Josas, France and §Geno Breeding and A.I. Organisation, N-1432 Aas, Norway*

Manuscript received May 24, 2004
Accepted for publication October 6, 2004

ABSTRACT

A QTL affecting milk production traits was previously mapped to an interval of 7.5 cM on chromosome 6 in Norwegian dairy cattle. This article aimed to refine this position by increasing the map density in the region by a set of single-nucleotide polymorphisms and analyzing the data with a combined linkage and linkage disequilibrium approach. Through a series of single- and multitrait and single- and multipoint analyses, the QTL was positioned to an interval surrounded by the genes ABCG2 and LAP3. As no recombinations were detected in this interval, physical mapping was required for further refining. By using radiation hybrid mapping as well as BAC clones, the bovine and human comparative maps in the region are resolved, and the QTL is mapped within a distance of 420 kb.

MAPPING and characterization of genes controlling important production, health, and quality traits have become an important field of research in livestock species. The ultimate goal of this work is to identify and characterize the gene(s) affecting the traits and the mutations underlying the genetic variation, thereby yielding important insight into the function and structure of the genome and how the genes interact with each other and the environment. Such genotype information may be an important supplement to the traditional phenotype-based selection systems in livestock, which for some traits are less than optimal due to the quantitative nature and low heritability of these traits. In dairy cattle, most emphasis has been on detecting quantitative trait loci (QTL) affecting milk production, and all autosomal chromosomes have been suggested as harboring QTL affecting one or more of the five milk traits (*i.e.*, milk yield, fat yield and percentage, and protein yield and percentage; Khatkar *et al.* 2004). Several studies have reported the segregation of at least one QTL in the middle of chromosome 6 (BTA6), close to marker BM143 (*e.g.*, Spelman *et al.* 1996; Kühn *et al.* 1999; Velmala *et al.* 1999; Nadesalingam *et al.* 2001; Ron *et al.* 2001). In Norwegian dairy cattle, a highly significant QTL causing a major reduction in fat percentage and protein percentage, as well as a minor increase of milk yield, was previously detected using linkage analysis (Olsen *et al.* 2002). By using the combined linkage disequilibrium and linkage analysis (LDLA) method of Meuwissen *et al.* (2002), the position of this QTL was later refined to a 7.5-cM interval between markers BMS2508 and FBN12 (Olsen *et al.* 2004). In this article, we aim to narrow this position even further by increasing the map density with a set of single-nucleotide polymorphisms (SNPs) and analyzing the data with LDLA methodology.

MATERIALS AND METHODS

Data: All animals in the study belonged to the Norwegian dairy cattle breed. The animals were organized in a granddaughter design consisting of 35 elite sire families. The total number of sons in the study was 1098, ranging from 14 to 71 sons for the smallest and largest families, respectively. The total number of daughters was ~680,000, with an average of 619 daughters per son. The pedigree of each animal in the study was traced back as far as known. Complex family relationships existed between the animals. Several of the sires were paternal half-sibs, and 18 of the sires were also included as sons in older families.

Predicted transmitting abilities (PTAs) of the sons from the 35 families were used as performance information in the analyses. The PTAs for the five milk production traits (milk yield, fat percentage, fat yield, protein percentage, and protein yield) were available from the national genetic evaluation in June 2000 carried out by GENO Breeding and Artificial Insemination Association and evaluated using BLUP with a single-trait sire-maternal grandsire model.

Marker map: To obtain a high marker resolution in the most likely QTL position of BTA6, the microsatellite map of Olsen *et al.* (2004) was increased by a number of SNPs. The

[1]*Corresponding author:* Department of Animal and Aquacultural Sciences, Agricultural University of Norway, Box 5025, N-1432 Aas, Norway. E-mail: sigbjorn.lien@iha.nlh.no

SNPs were generated by amplifying and sequencing introns of genes likely to be positioned to the region on the basis of comparative mapping information (Band *et al.* 2000; Weikard *et al.* 2002). Sequencing was done on eight bull sires of Norwegian dairy cattle using Dye Terminator chemistry and an ABI3730 sequencer (Applied Biosystems, Foster City, CA). An SNP-detection pipeline was constructed on the basis of phred, phrap, and polyphred programs (Nickerson *et al.* 1997). Each chromatogram was processed using phred (Ewing *et al.* 1998) into a fasta sequence file and a list of quality values for each base in the sequence. Phrap (http://www.genome.washington.edu) used this information to assemble the sequences into contigs. The phrap parameter repeat_stringency was set to 0.8, as this maximized the number of sequences allocated to the correct intron. Polyphred was run on the resulting assemblies with the insert deletion detection option. Only sequence variants with a polyphred ranking of 1 or 2 were considered as putative SNPs. Contig assembly and putative SNPs were visually inspected using consed (Gordon *et al.* 1998). If a putative SNP passed visual inspection, a biopython script was used to create an input file for the MassARRAY primer design program, using the polyphred output and contig consensus sequence from phrap. The SNPs were genotyped using matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy (MALDI-TOF MS) assays (Sequenom, San Diego). Table 1 provides GenBank accession numbers and information about position of the SNPs within the sequence, as well as allele frequencies of the markers. Only alleles inherited from dams were used when estimating allele frequencies, to obtain unbiased estimates. Assays for genotyping by MassARRAY are available on request.

Marker order and map distances were estimated using the CRI-MAP 2.4 program (Green *et al.* 1990) with map distances based on Haldane's mapping function (Figure 1). The average distance between markers was 2.43 cM, ranging from 0 to 16.1 cM. No recombination was detected between markers RAP1GDS1 and BM1329, as well as between markers ABCG2, LAP3, and HCAP-G in our data set. The likely order of these markers was determined by using information from Weikard *et al.* (2002) and radiation hybrid mapping results described herein. As the LDLA mapping method requires some recombination between markers, these distances were set somewhat arbitrarily to 0.01 cM. In several of the genes, two or more SNPs were genotyped. These SNPs were treated as individual markers, and all analyses were performed in the same manner for such an SNP as for the other, more distant (microsatellite) markers. Although the distance between such SNPs was in the range of a few hundred bases and thus recombination can be regarded as nonexistent, the recombination fractions between SNPs in the same gene were still set to 0.001 cM to fulfill the demand for some distance between the markers.

Linkage disequilibrium map: A linkage disequilibrium (LD) map for the studied region was constructed by first calculating the amount of LD (measured as D'; Hedrick 1987) for all possible marker pairs and then, for each marker, finding the average D' with that marker and all markers residing within 5 cM in each direction from that marker.

Statistical analyses: *Single QTL—single-trait analysis:* Each of the five milk traits was analyzed separately using the combined linkage and linkage disequilibrium method of Meuwissen *et al.* (2002). In short, the method consisted of the following steps: First, the linkage phases of all sires and sons were estimated on the basis of marker information. In the second step, the midpoint of each marker bracket was regarded as a putative position for a QTL. For each putative QTL position (*i.e.*, midpoint of marker bracket), the identical-by-descent (IBD) probabilities of pairs of haplotypes were calculated from marker and/or pedigree information and allele frequencies (for details, see Meuwissen and Goddard 2001; Meuwissen *et al.* 2002). Only the bracket midpoints were considered, since for a dense marker map individual positions within the bracket would have similar probabilities. The IBD probability depends on the effective population size and the number of generations since the base population, which were both assumed to be 100. For each bracket, a complete matrix of IBD probabilities between all haplotypes at the putative QTL position was obtained. This matrix is denoted G_i, where the subscript i reflected the fact that the probabilities depended on the ith position of the QTL. The last step was to calculate the likelihood of the data using restricted maximum likelihood (REML). The model of the performance data was expressed as

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{Zh} + \mathbf{u} + \mathbf{e},$$

where $\mathbf{y}$ is a $(n \times 1)$ vector of records (*i.e.*, PTAs for the milk

TABLE 1

Genes included in the marker map

Gene symbol	GenBank no.	SNP denotation	Base	Frequency
SEC24B	AY744575	SEC24B_231	C	0.24
			T	0.76
		SEC24B_341	A	0.24
			G	0.76
		SEC24B_367	C	0.43
			T	0.57
CENPE	AY744568	CENPE_101	C	0.70
			T	0.30
RAP1GDS1	AY744576	RAP1GDS1_299	A	0.14
			G	0.86
NFKB1	AY744571	NFKB1_462	C	0.90
			T	0.10
		NFKB1_580	C	0.31
			T	0.69
ABCG2	AY744566	ABCG2-F_49	A	0.97
			C	0.03
		ABCG2-R_256	A	0.23
			G	0.77
LAP3	AY744570	LAP3_281	A	0.64
			G	0.36
		LAP3_529	A	0.36
			G	0.64
		LAP3_572	A	0.37
			G	0.63
		LAP3_581	A	0.36
			C	0.64
HCAP-G	AY744569	HCAP-G-R_318	A	0.72
			C	0.28
		HCAP-G-R_339	C	0.11
			T	0.89
PPARGC1	AY744572	PPARGC1-F_67	C	0.70
			G	0.30
	AY744573	PPARGC1-R_186	A	0.11
			C	0.89
APBB2	AY744567	APBB2_482	A	0.25
			C	0.75
PTPN13	AY744574	PTPN13_89	G	0.04
			T	0.96
		PTPN13_280	A	0.60
			G	0.40

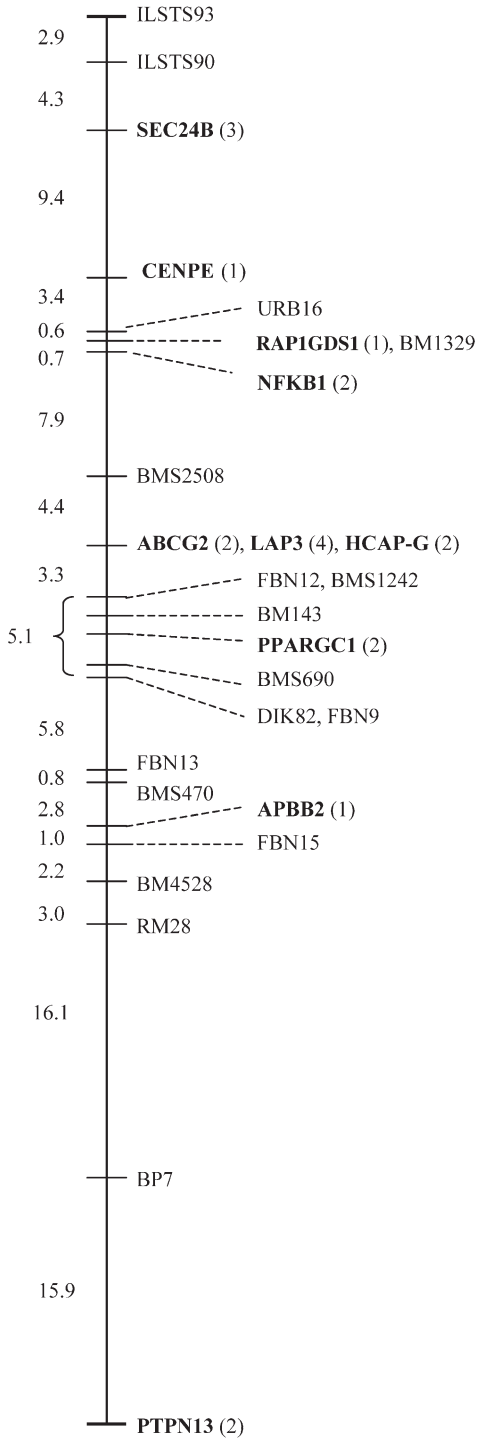

Figure 1.—Marker map. Genes are in boldface type, and numbers of SNPs per gene are in parentheses.

trait in question); μ is the overall mean; **1** is a vector of 1's; **h** is a vector of random haplotype effects of dimension $q \times 1$, where q is the number of different haplotypes; **Z** is a $(n \times q)$ incidence matrix relating observations and haplotype effects; **u** is a vector of random polygenic effects; and **e** is a vector of residuals. The variances of **h**, **u**, and **e** are $\mathbf{G}_i\sigma_h^2$, $\mathbf{A}\sigma_u^2$, and $\mathbf{R}\sigma_e^2$, respectively, where $\mathbf{G}_i$ is the matrix of IBD probabilities among haplotypes; **A** is the additive genetic relationship matrix; and **R** is a diagonal matrix with n_j^{-1} on the diagonals (n_j is the number of daughters of bull j). For each marker bracket, the log-likelihood (logL) of a model containing a QTL as well as background genes was calculated by maximizing the likelihood with respect to the variance components using the ASREML package (Gilmour *et al.* 2000). The likelihood of the alternative hypothesis of no QTL was calculated on the basis of a model containing only background genes, and the logL ratio was formed as the difference in logL between the models with and without a QTL. Then, the test statistic LOD score was obtained by dividing the logL ratio by ln 10. The marker bracket with the highest LOD score was taken as the most likely QTL position. The significance level of LOD score was determined from the chi-square distribution, as the LOD score $\times$ 2 $\times$ ln 10 is chi-square distributed with 1 d.f. (since one less variance component was fitted in the model without the QTL). Because the QTL were already found to be highly significant in the study of Olsen *et al.* (2002), a nominal *P*-value of 1% was considered to be a significant result.

Single-QTL multiple-trait analysis: The ASREML program was also used to perform the multiple-trait analysis, using the same IBD matrix ($\mathbf{G}_i$) as in the single-trait analysis. The multitrait analysis included the three yield traits only. The data were modeled as

$$\mathbf{y}_i = \mathbf{X}_i\mathbf{b} + \mathbf{u}_i + \mathbf{h}_{i1} + \mathbf{h}_{i2} + \mathbf{e}_i,$$

where $\mathbf{y}_i$ is a (3×1) vector of PTAs of bull i for milk, fat, and protein yield; **b** is a vector of fixed effects, which includes the mean for each of the three traits; $\mathbf{h}_{i1}$ and $\mathbf{h}_{i2}$ are (3×1) vectors of effects of the paternally or maternally inherited haplotype of bull i on each of the three traits, respectively; and $\mathbf{u}_i$ and $\mathbf{e}_i$ are (3×1) vectors of random polygenic effects and residuals of bull i on each trait, respectively. For further details, see Olsen *et al.* (2004).

Multiple-QTL single-trait analysis: All traits were analyzed with the multiple-QTL single-trait method proposed by Meuwissen and Goddard (2002). The data were modeled as

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{Zu} + \Sigma_i I_i \mathbf{X}_i\mathbf{h}_i + \mathbf{e},$$

where **y**, **1**, μ, **Z**, **u**, and **e** are as for the single-QTL single-trait analyses; Σ_i denotes summation over all possible QTL positions (bracket midpoints); $\mathbf{h}_i$ is a vector of haplotype effects; $\mathbf{X}_i$ is a known incidence matrix relating haplotype effects with records; and I_i is an indicator variable, where $I_i = 1$ ($I_i = 0$) indicates (no) QTL at position i. $\mathrm{Var}(\mathbf{h}_i) = \mathbf{G}_i\sigma_h^2$, where $\mathbf{G}_i$ is the IBD matrix between the haplotypes at the ith position. The variance components and haplotype effects were estimated using Gibbs sampling with 500,000 cycles; and I_i was sampled with a Metropolis-Hastings step (Meuwissen *et al.* 2001). As opposed to the article of Meuwissen and Goddard (2002), the prior probability of having a QTL in bracket i ($I_i = 1$) depended on the length of the bracket. As a QTL previously had been mapped to the relevant area of BTA6 (Olsen *et al.* 2002, 2004), the total prior probability of a QTL in the genotyped area was assumed to be 100%. The prior of each bracket was proportional to the length of that bracket. Flat priors were assumed for σ_u^2 and σ_e^2, while the prior distribution of σ_h^2 was inverse chi square with 4.2 d.f. (Meuwissen and Goddard 2002). After discarding the initial 10,000 cycles from the Monte Carlo Markov chain as burn-in, the fraction of cycles with $I_i = 1$ gave an estimate of the posterior probability of a QTL at position i. A significant QTL was considered present in a bracket if the posterior probability of that bracket exceeded 0.5; *i.e.*, the posterior probability of having a QTL in that bracket is larger than the probability of no QTL.

Radiation hybrid mapping: To confirm the order of the closely linked markers and refine the comparative map between bovine chromosome 6 (BTA6) and human chromosome 4 (HSA4) in the QTL region, 10 genes were mapped

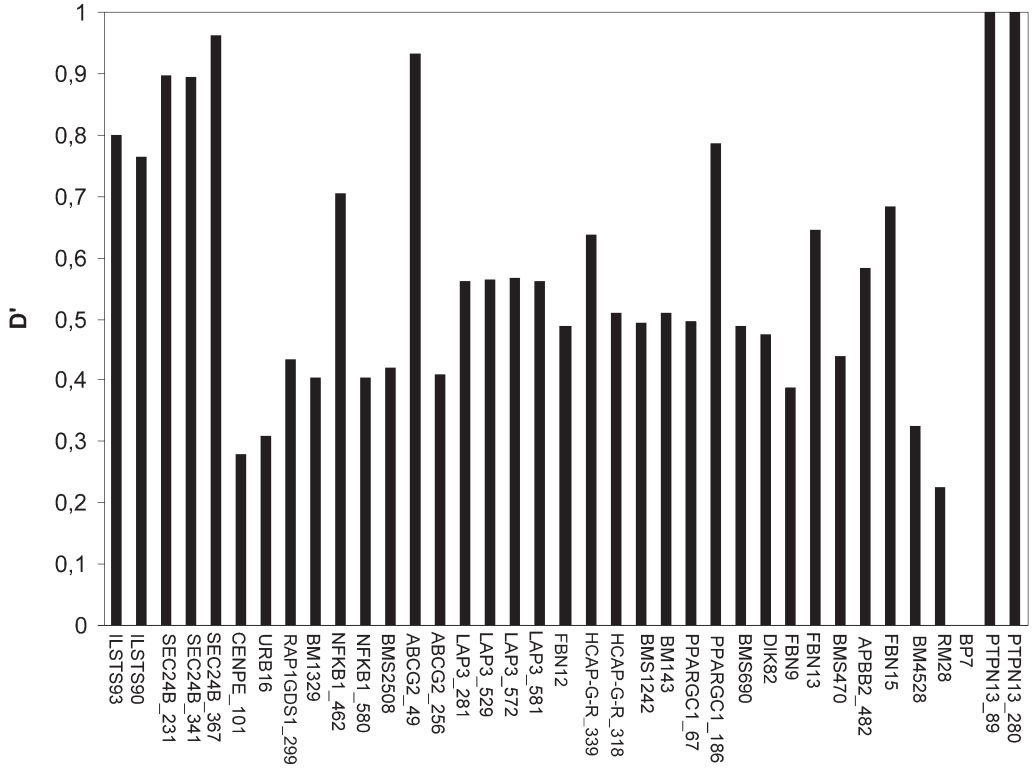

FIGURE 2.—Linkage disequilibrium map of the genotyped BTA6 region. The *y*-axis indicates the average *D*′ between the named marker and markers within a distance of 5 cM.

by radiation hybrid (RH) mapping using the 3000-rad panel described by WILLIAMS *et al.* (2002). The genes (and their position in megabases) from the start of HSA4 (according to the July 2003 version of the UCSC Genome Browser; http://genome.cse.ucsc.edu/) were as follows: LAP3 (17.563), HCAP-G (17.6), SLIT2 (20.01), PPARGC1 (23.54), PTPN13 (87.97), MLLT2 (88.39), NUDT9 (88.80), IBSP (89.18), SPP1 (89.36), and ABCG2 (89.47). RH data were analyzed using the CarthaGène software described by SCHIEX and GASPIN (1997).

Physical map construction: Genes were also PCR screened on the BAC INRA library as described previously by EGGEN *et al.* (2001) to build a physical map in the QTL region. The screening allowed the identification of contigs mapping to the region of interest in the INRA BAC-based first-generation bovine physical map published recently by SCHIBLER *et al.* (2004) (http://locus.jouy.inra.fr/fpc/cattle/). This map is further anchored on the international physical map that serves as a basis for the bovine genome-sequencing consortium (http://www.bcgsc.ca/lab/mapping/bovine). Additionally, to confirm and refine the bovine-human comparative map, bovine BAC ends from international (LARKIN *et al.* 2003) and INRA BAC clones (M. GAUTIER, personal communication) in the actual region were mapped *in silico* using BLAST (ALTSCHUL *et al.* 1997).

RESULTS

To refine the position of the milk production QTL reported in OLSEN *et al.* (2004), the BTA6 map was extended by 20 SNPs situated in a total of 10 genes. This extended map covering ~90 cM is shown in Figure 1.

Figure 2 shows a linkage disequilibrium map of the studied region. Average *D*′ between a marker and the surrounding markers varied from ~0.2 to 1.0, with an average of 0.57 for the entire interval. *D*′ for one marker was set to 0 as the distance to the nearest marker exceeded 5 cM.

A series of analyses were performed to refine the QTL position. Initially, all five milk traits (*i.e.*, milk yield, fat yield, fat percentage, protein yield, and protein percentage) were analyzed separately using the single-QTL analysis. For each marker bracket and trait, a LOD score was obtained by first calculating the difference in logL (logL ratio) between the model with a QTL in the midpoint of that bracket and the model without a QTL fitted and then dividing the logL ratio by ln 10. The marker bracket with the highest LOD score was expected to contain the QTL. Similar to the findings of the two previous studies (*i.e.*, OLSEN *et al.* 2002, 2004), protein percentage and fat percentage yielded highly significant results (Figure 3), whereas fat and protein yield were nonsignificant (results not shown). While the former studies also detected some segregation for milk yield, no significant results were found in this study (not shown). As shown in Figure 3, the shape of the test statistic curve is similar for the two percentage traits, but with an even higher LOD score for protein percentage than for fat percentage. A marked peak can clearly be seen for both traits at position ~33.6 cM. This corresponds to the interval LAP3_281–LAP3_529, *i.e.*, the interval between the two first SNPs in the LAP3 gene. For protein percentage, the LOD score at this point is 38.9, whereas that of fat percentage is 23.3. A dramatic drop in LOD score was seen for the remaining brackets within the LAP3 and HCAP-G genes. Still the test statistic remains fairly high for several of the surrounding brack-

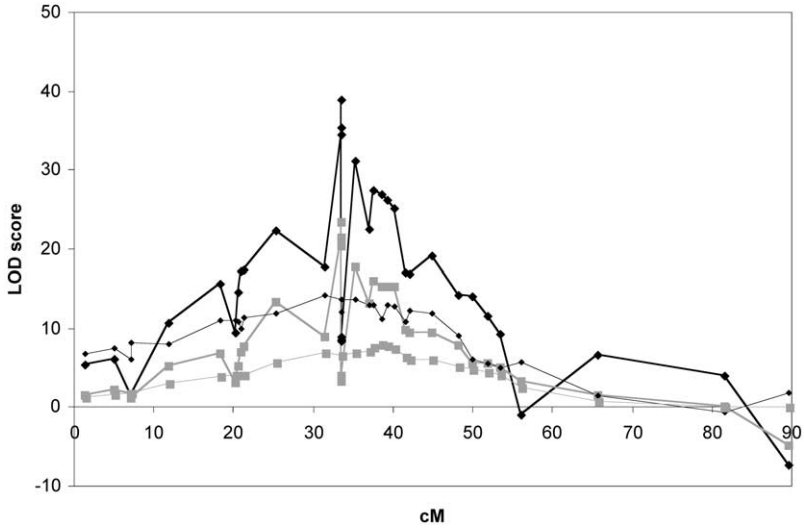

FIGURE 3.—Single-QTL single-trait analysis of protein percentage (◆) and fat percentage (■) using LDLA (thick lines) and linkage analysis (thin lines). Points illustrate bracket midpoints.

ets, as the LOD score exceeds 30 for protein percentage and 20 for fat percentage for most brackets between ABCG2-F_49 and FBN12. This is a very narrow interval with no recombination detected between the genes ABCG2, LAP3, and HCAP-G and with ~3.3 cM between HCAP-G and FBN12 in our map. To test the advantage of the combined approach over pure linkage analysis, the percentage traits were also analyzed with a model using linkage information only. As shown in Figure 3, these analyses yielded much flatter curves with broader peaks and less well-defined maxima than those obtained by LDLA and clearly visualize the benefit of the combined approach.

Second, a single-QTL multitrait analysis was performed for milk yield, protein percentage, and fat percentage. The milk traits are known to be highly correlated, with a positive correlation between the three yield traits and a negative correlation between yield and percentage traits. Thus, by using a composite hypothesis combining information from all traits simultaneously, the power to detect QTL and the accuracy of estimating QTL positions might be improved, as the QTL may have pleiotropic effects on all the milk traits. This three-trait analysis yielded a LOD score curve similar in shape to that of the two percentage traits run individually, with a peak for the bracket LAP3_281–LAP3_529, a relatively high score for the surrounding brackets, but with the same drop in LOD score for the remaining LAP3 and HCAP-G intervals (results not shown). For the bracket HCAP-G-R_318–HCAP-G-R_339 the logL did not converge and the LOD score was not obtained. Thus the multitrait approach confirmed the results from the single-trait approach, but was not able to refine the QTL position further.

Although both the single- and multitrait analyses revealed LAP3_281–LAP3_529 as the most significant bracket, the rather large test statistic of the surrounding brackets could be due to the presence of additional QTL or alternatively be due to carryover effects of one large QTL on neighboring intervals. Therefore, a multiple-QTL analysis was performed. Only protein percentage was analyzed in this and further analyses, as fat percentage had showed a curve almost identical in shape and three yield traits were close to nonsignificant in previous analyses. As shown in Figure 4, this analysis strongly points at LAP3_281–LAP3_529 as the most likely position, with no evidence for a second QTL in any of the other intervals.

For each bracket, the vector of haplotype effects (**h**) was estimated from the matrix of IBD probabilities between haplotype pairs and the trait records using ASREML for the single-QTL analyses. Haplotypes with IBD probabilities >0.95 were considered as being the same; thus the actual number of haplotypes was less than twice the number of genotyped individuals. The number of haplotypes per bracket varied from 104 for the interval FBN12–BMS1242 to 1912 for the bracket ILSTS93–ILSTS90. For the LAP3_281–LAP3_529 interval (*i.e.*, the bracket showing the highest LOD score), the total number of haplotypes was 119.

The effects of each of the haplotypes of the LAP3_281–LAP3_529 interval were investigated to identify haplotypes carrying QTL alleles with major impact on milk production. Most estimates of **h** were around the mean, while haplotypes 4, 51, 52, 54, 73, and 78 showed a distinct reduction of protein percentage. The most common of these was haplotype 4, which was found in 123 individuals. The others were found in one individual each. By comparing the mean estimated effect of these six haplotypes to the mean of all others, the extreme haplotypes were found to cause a reduction of the percentage traits corresponding to 0.06 percentage points protein and 0.09 percentage points fat. All haplotypes carried the base G at LAP3_281 and A at LAP3_529. Upon closer examination, individuals carrying this extreme-effect haplotype also shared the same alleles in a larger area

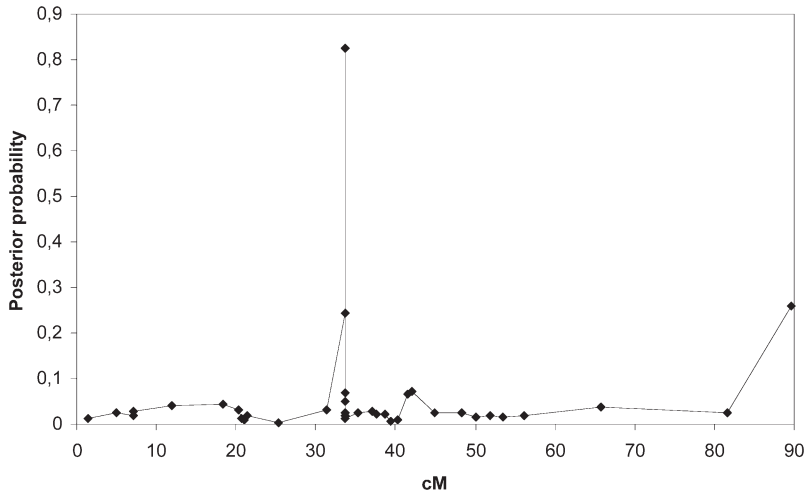

FIGURE 4.—Multi-QTL analysis for protein percentage.

spanning from marker ABCG2-F_256 to HCAP-G-R_339 consisting of the bases G, G, A, A, A, A, and T for these markers, respectively. Notably, marker ABCG2-F_49 makes one boundary of the haplotype because both A and C were present in individuals carrying the extreme haplotype. The relationship of each haplotype with the one showing the most negative effect (no. 4) was also investigated. In Figure 5, these relationships are plotted along the *x*-axis, and the haplotype effects for protein percentage (expressed as deviations from mean PTA) are plotted on the *y*-axis. Figure 5 shows that haplotypes with average effect on the trait are also less related to haplotype 4 than those carrying the highly negative allele.

Comparative studies have shown that the relevant QTL region on BTA6 corresponds to two segments of conserved synteny on HSA4 (BAND *et al.* 2000; WEIKARD *et al.* 2002). One large block from HSA4q22 (containing ABCG2, IBSP, and SPP1) is inverted in bovine and represents a chromosomal boundary to which a block from HSA4p15 (containing LAP3 and PPARGC1) is attached (WEIKARD *et al.* 2002). Our radiation hybrid mapping results confirm these results by mapping LAP3, HCAP-G, SLIT2, and PPARGC1 (from HSA4p15) together with IBSP, SPP1, and ABCG2 (from HSA4q22) to the relevant QTL region on BTA6. Additionally, our RH mapping positions MLLT2 and NUDT9 together with PTPN13 on the telomeric part of BTA6.

As an attempt to refine the QTL position on BTA6, we constructed a physical map based on BAC clones. Initial screening of the INRA BAC pool DNA identified at least one positive BAC clone for each of the marker genes: SPP1 (clones bI0151A04, bI0156B02, bI0777H03, and bI0792H07 belonging to INRA contig ctg3063), IBSP (bI 0946E12 belonging to ctg2199), LAP3 (bI0455H01, bI0459C10, bI0540H05, bI0603B01, and bI0965H10 belonging to ctg509), and HCAP-G (bI0134C03 and bI0778A10 belonging to ctg2554). ABCG2 failed to amplify INRA BAC pool DNA but the gene could be mapped *in silico* (see below). Three of these four INRA contigs (*i.e.*, ctg3063, ctg509, and ctg2554) were anchored to the same international physical map contig ctg503, which

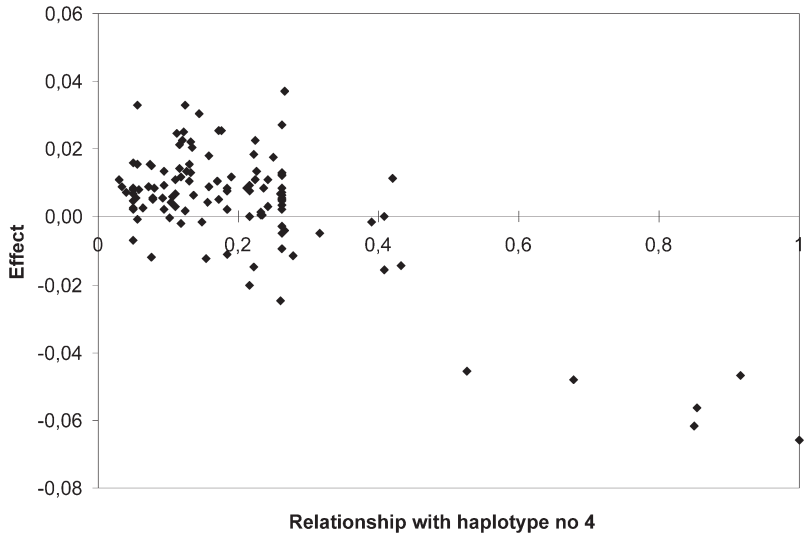

FIGURE 5.—Haplotype relationships and effects for protein percentage (single-QTL analysis). The relationship of the haplotype with the most extreme effect (no. 4) with all others is plotted on the *x*-axis, while the effects of each haplotype for protein percentage (as measured in deviations from mean PTA) are shown on the *y*-axis.

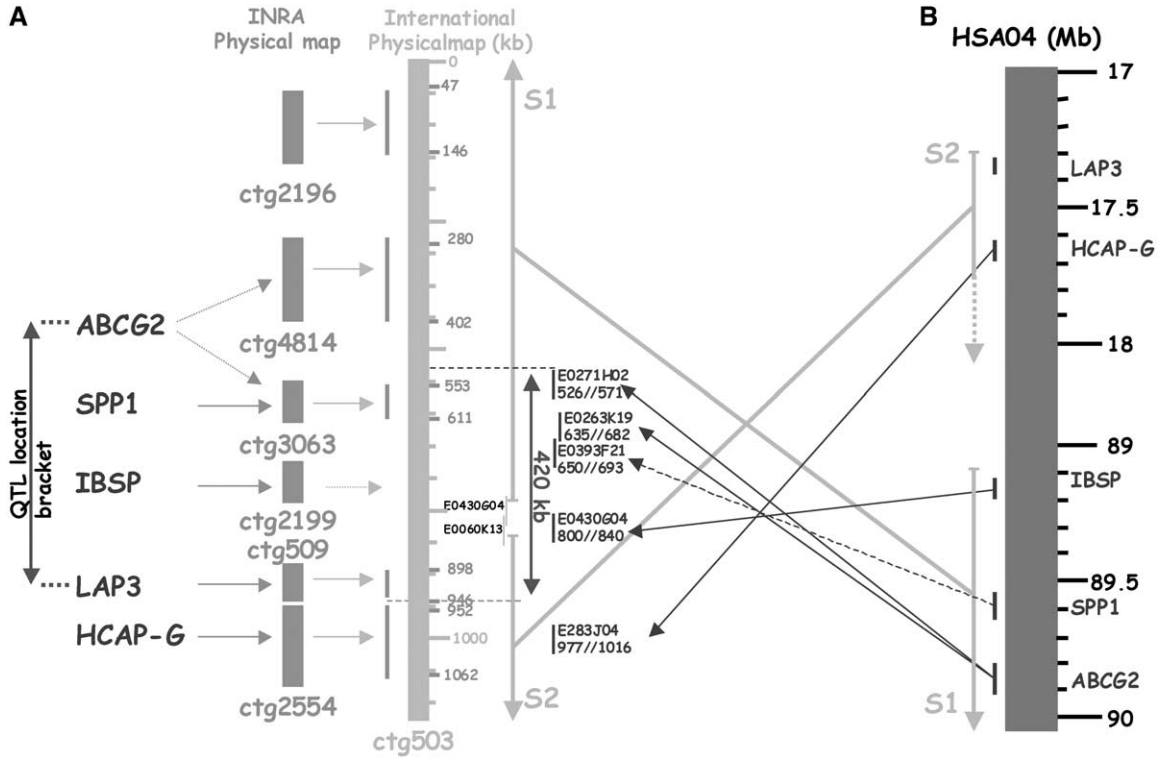

FIGURE 6.—Physical map of the QTL bracket. The left side (A) shows assignment of genes surrounding the QTL bracket to INRA BAC contigs and their anchorage to the international physical map. Locations of the five genes on HSA4 are indicated on the right (B). *In silico* identification of BACs from the international contig ctg503 is indicated by an arrow connecting the gene location on HSA4 to the BAC location on the international physical map. Gene information content and other BAC end sequence similarities with anonymous human sequences allow the identification of two blocks of conserved synteny named S1 and S2 whose boundaries could be restricted to the two BAC clones E0476I20 and E060K13.

contains 433 BAC clones and spans ~1300 kb (data from the international physical map version released in January 2004, available at http://www.bcgsc.ca/lab/mapping/data). Two additional INRA contigs (ctg2196 and ctg4814) have clones common to international BAC contigs and were thus precisely positioned on contig ctg503 (see Figure 6). BLAST of human mRNA sequences against BAC end sequences allowed *in silico* mapping of the human genes ABCG2 (NM_004827) to bovine BAC clones E0271H02 (BZ863331) and E0263K19 (BZ873148), IBSP (NM_004967) to BAC clone E0430G4 (CC547012), SPP1 (NM_000582) to BAC clone E0393F21 (CC592687), and HCAP-G (NM_022346) to BAC clone E0283J04 (BZ 884618). A precise comparative map of the bovine region of interest could thus be drawn with the identification of two blocks of conserved synteny named S1 and S2 in Figure 6. These two blocks were supported by information from genetic mapping, RH mapping and sequence similarities among bovine BAC ends, and genomic sequences on human HSA4 (45 hits for S1 and 22 hits for S2, respectively). These results confirm and greatly refine the boundaries of these two blocks of conserved synteny to a region of <100 kb from position 800 to 878 kb on ctg503 close to the IBSP gene. BAC end sequences from the two overlapping BAC clones E0430G04 (CC547012) and E0060K13 (BZ916340) match significantly against HSA4 positions 89.2 and 17.4 Mb, respectively. Moreover, the order of bovine BAC clones is concordant with respect to significant sequence hits inside each block of conserved synteny, and distances seem to be rather well conserved among the two species. Summarizing the results from BAC mapping, the borders of the QTL position are determined by the mapping of ABCG2 (position 526–682) and INRA contig ctg509 containing LAP3 (position 898–946) in the international BAC contig ctg503. The QTL in interval ABCG2–LAP3 is thus estimated to span a fragment of at maximum 420 kb on BTA 6 (Figure 6).

DISCUSSION

Former studies in Norwegian dairy cattle have positioned a QTL affecting milk production traits to a 7.5-cM interval surrounded by the microsatellites BMS2508 and FBN12 in the middle part of BTA6. The QTL seems to cause a major decrease of fat and protein percentages and a minor increase of milk yield (OLSEN *et al.* 2002, 2004). The aim of this article was to refine the QTL position further to pinpoint candidate genes for molecular characterization. To achieve this, the BMS2508–FBN12 interval was subdivided by including SNPs in the

genes ABCG2, LAP3, and HCAP-G. Of these, LAP3 was genotyped for four SNPs, whereas two SNPs were genotyped in both ABCG2 and HCAP-G. Thus, the original bracket was divided into nine smaller brackets.

Linkage disequilibrium-based approaches are commonly used for fine mapping of QTL. LD mapping is expected to be especially useful in livestock, where high levels of LD are found to extend over tens of centimorgans (*e.g.*, Farnir *et al.* 2000). The LD map presented in Figure 2 shows a substantial amount of LD over the entire chromosomal region, indicating that LD mapping may be useful in our data. None of the markers are in complete disequilibrium with each other; thus it might be possible to map the QTL quite accurately.

All single-QTL analyses yielded similar patterns, *i.e.*, a highly significant test statistic from ABCG2-F_49 to LAP3_529 and with a marked peak for bracket LAP3_281–LAP3_529, a dramatic drop in LOD score for the interval from LAP3_529 to HCAP-G-R_339, and then a new peak for HCAP-G-R_339–FBN12 and onward. We believe that for the brackets with low LOD scores (*i.e.*, those between LAP3_529 and HCAP-G-R_339), there is enough information in the data to exclude these brackets as QTL positions. Thus, the QTL must be localized somewhere between the first SNP of ABCG2 (*i.e.*, ABCG2-F_49) and the second SNP of LAP3 (*i.e.*, LAP3_529). The second peak could be due to the presence of additional QTL in this area or alternatively be due to a carryover effect on neighboring intervals. The first alternative is excluded by the multi-QTL analysis, which clearly points to LAP3_281–LAP3_529 as the only possible position. Thus the data probably do not contain enough information here to prevent the carryover effect. However, although all analyses clearly point toward one specific bracket, the test statistic is fairly high also for the two preceding brackets (*i.e.*, the brackets ABCG2-F_49–ABCG2-F_256 and ABCG2-F_256–LAP3_281). As the linkage analysis program detected no current recombination in this area, there may not be enough information in our data to position the QTL exactly, and the QTL could be situated anywhere in the interval bracketed by the ABCG2 and LAP3 genes.

In our first study, linkage analysis suggested a QTL position close to marker FBN9, with a confidence interval ranging approximately from BMS2508 to FBN13 (Olsen *et al.* 2002). In the subsequent fine-mapping study, LDLA mapping performed on an extended animal material and denser marker map shifted the position leftward to an interval between BMS2508 and FBN12, but, however, with some smaller, nonsignificant peaks in the vicinity of FBN9 (Olsen *et al.* 2004). Finally, the present study localizes the QTL to a narrow interval between ABCG2 and LAP3 and excludes the possibility of a second QTL in the FBN9 region. This leftward shift in position is probably due to two factors: first, that linkage analyses usually yield rather imprecise position estimates due to the lack of recombinations; and second, that at least for this QTL the large effect causes a considerable carryover effect to the neighboring brackets. These problems seem to be overcome in the present study, especially when the multi-QTL mapping method was used.

Several groups have reported the presence of one or more QTL for milk production on BTA6 using linkage analysis. QTL close to the casein gene cluster are detected in several studies (*e.g.*, Kühn *et al.* 1996; Velmala *et al.* 1999). A previous study in Norwegian dairy cattle also reported an association between protein yield and a specific casein haplotype (Lien *et al.* 1995). This association was not confirmed in our studies, however. The discrepancy could be due to the low marker density in this region or a lack of segregating sires. Several studies have detected a second QTL close to marker BM143 (*e.g.*, Spelman *et al.* 1996; Kühn *et al.* 1999; Velmala *et al.* 1999; Nadesalingam *et al.* 2001; Ron *et al.* 2001). Ron *et al.* (2001) detected a confidence interval for protein percentage as narrow as 4 cM surrounding BM143 in two families using a daughter design. As BM143 is ~4 cM rightward of our QTL, these results might reflect the same QTL segregating in different breeds. Comparing results from different studies is important to confirm the presence of a QTL on a chromosomal area, but, as discussed above, comparing positions obtained by linkage analyses may be difficult due to the poor resolution. Also the mode of action of this QTL is debated. According to the literature, it is commonly thought of as affecting the percentage traits (Nadesalingam *et al.* 2001), but the results of several studies suggest that the primary effect is on milk yield, with little effect on fat and protein yield, such that fat and protein percentages are affected indirectly (*i.e.*, Georges *et al.* 1995; Zhang *et al.* 1998; Velmala *et al.* 1999; Nadesalingam *et al.* 2001). Only by identifying the gene(s) behind the QTL and investigating their function, the true nature of the QTL can be discovered. The narrow positioning of the QTL in the present study will clearly simplify this work.

In the absence of a complete bovine genome sequence, one feasible way of characterizing the QTL would be to extract information from the already sequenced human or mouse genome. Unfortunately, comparative mapping in this region is complicated by both rearrangements and inversions, and the most likely QTL position seems to be in the boundary of two blocks from HSA4q22 and HSA4p15. Our physical mapping results (Figure 6) map the QTL within a distance of 420 kb and resolve the bovine and human comparative map in the region. The comparative regions in human contain a very limited number of known genes consisting of LAP3 in the block from HSA4p15 and IBSP, MEPE, SPP1, PKD2, and ABCG2 in the block from HSA4q22. Among these, PKD2 (polycystin 2) seems to have a function that fits best with the QTL effect. The gene encodes an integral membrane protein involved in cell-cell/ma-

trix interactions, which interacts with polycystin 1 (PKD1) to produce cation-permeable currents that may modulate intracellular calcium homoeostasis and other signal transduction pathways (Hanaoka *et al.* 2000; Nauli *et al.* 2003). Studies are underway to reveal the molecular mechanisms responsible for the QTL effect, involving both gene expression and characterization of bovine PKD2 in individuals carrying differerent QTL alleles.

A QTL with large impact on milk production is mapped to an interval of 420 kb between the genes ABCG2 and LAP3 on bovine chromosome six. This is a very narrow region harboring only six known genes on the comparative human map.

We thank GENO Breeding and Artificial Insemination (A.I.) Association for providing relationship information and PTAs for sons. This project received funding from the Research Council of Norway, GENO Breeding and A.I. Association, and BoviBank.

LITERATURE CITED

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Band, M. R., J. H. Larson, M. Rebeiz, C. A. Green, D. W. Heyen *et al.*, 2000 An ordered comparative map of the cattle and human genomes. Genome Res. **10:** 1359–1368.

Eggen, A., M. Gautier, A. Billaut, E. Petit, H. Hayes *et al.*, 2001 Construction and characterization of a bovine BAC library with four genome-equivalent coverage. Genet. Sel. Evol. **33:** 543–548.

Ewing, B., L. Hillier, M. C. Wendl and P. Green, 1998 Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:** 175–185.

Farnir, F., W. Coppieters, J.-J. Arranz, P. Berzi, N. Cambisano *et al.*, 2000 Extensive genome-wide linkage disequilibrium in cattle. Genome Res. **10:** 220–227.

Georges, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto *et al.*, 1995 Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics **139:** 907–920.

Gilmour, A. R., B. R. Cullis, S. J. Welham and R. Thompson, 2000 ASREML reference manual. ftp.res.bbsrc.ac.uk/pub/aar.

Gordon, D., C. Abajian and P. Green, 1998 Consed: a graphical tool for sequence finishing. Genome Res. **8:** 195–202.

Green, P., K. Falls and S. Crooks, 1990 Documentation for CRIMAP, Version 2.4. Washington University School of Medicine, St. Louis.

Hanaoka, K., F. Qian, A. Boletta, A. K. Bhunia, K. Piontek *et al.*, 2000 Co-assembly of polycystin-1 and -2 produces unique cation-permeable currents. Nature **408:** 990–994.

Hedrick, P. W., 1987 Gametic disequilibrium measures: proceed with caution. Genetics **117:** 331–341.

Khatkar, M. S., P. C. Thomson, I. Tammen and H. W. Raadsma, 2004 Quantitative trait loci mapping in dairy cattle: review and meta-analysis. Genet. Sel. Evol. **36:** 163–190.

Kühn, C., R. Weikard, T. Goldammer, S. Grupe, I. Olsaker *et al.*, 1996 Isolation and application of chromosome 6 specific microsatellite markers for detection of QTL for milk-production traits. J. Anim. Breed. Genet. **113:** 355–362.

Kühn, C., G. Freyer, R. Weikard, T. Goldammer and M. Schwerin, 1999 Detection of QTL for milk production traits in cattle by application of a specifically developed marker map of BTA6. Anim. Genet. **30:** 333–340.

Larkin, D. M., A. Everts-Van Der Wind, M. Rebeiz, P. A. Schweitzer, S. Bachman *et al.*, 2003 A cattle-human comparative map built with cattle BAC-ends and human genome sequence. Genome Res. **13:** 1966–1972.

Lien, S., L. Gomez-Raya, T. Steine, E. Fimland and S. Rogne, 1995 Association between casein haplotypes and milk yield traits. J. Dairy Sci. **78:** 2047–2056.

Meuwissen, T. H. E., and M. E. Goddard, 2001 Prediction of identity by descent probabilities from marker-haplotypes. Genet. Sel. Evol. **33:** 605–634.

Meuwissen, T. H. E., and M. E. Goddard, 2002 Mapping multiple QTL by combined linkage disequilibrium/linkage analysis in outbred populations. Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, Communication 21-18.

Meuwissen, T. H. E., B. J. Hayes and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics **157:** 1819–1829.

Meuwissen, T. H. E., A. Karlsen, S. Lien, I. Olsaker and M. E. Goddard, 2002 Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. Genetics **161:** 373–379.

Nadesalingam, J., Y. Plante and J. P. Gibson, 2001 Detection of QTL for milk production on chromosomes 1 and 6 of Holstein cattle. Mamm. Genome **12:** 27–31.

Nauli, S. M., F. J. Alenghat, Y. Luo, E. Williams, P. Vassilev *et al.*, 2003 Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells. Nat. Genet. **33:** 129–137.

Nickerson, D., V. Tobe and S. Taylor, 1997 PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res. **25:** 2745–2751.

Olsen, H. G., L. Gomez-Raya, D. I. Våge, I. Olsaker, H. Klungland *et al.*, 2002 A genome scan for quantitative trait loci affecting milk production traits in Norwegian Dairy Cattle. J. Dairy Sci. **85:** 3124–3130.

Olsen, H. G., S. Lien, M. Svendsen, H. Nilsen, A. Roseth *et al.*, 2004 Fine mapping of milk production QTL on BTA6 by combined linkage and linkage disequilibrium analysis. J. Dairy Sci. **87:** 690–698.

Ron, M., D. Kliger, E. Feldmesser, E. Seroussi, E. Ezra *et al.*, 2001 Multiple quantitative trait locus analysis of bovine chromosome 6 in the Israeli Holstein population by daughter design. Genetics **159:** 727–735.

Schibler, L., A. Roig, M. F. Mahe, J. C. Save, M. Gautier *et al.*, 2004 A first generation bovine BAC-based physical map. Genet. Sel. Evol. **36:** 105–122.

Schiex, T., and C. Gaspin, 1997 Carthagene: constructing and joining maximum likelihood genetic maps. Proc. Int. Conf. Intell. Syst. Mol. Biol. **5:** 258–267.

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 6 in the Dutch Holstein-Friesian population. Genetics **144:** 1799–1808.

Velmala, R. J., H. J. Vilkki, K. T. Elo, D. J. De Koning and A. V. Mäki-Tanila, 1999 A search for quantitative trait loci for milk production traits on chromosome 6 in Finnish Ayrshire cattle. Anim. Genet. **30:** 136–143.

Weikard, R., C. Kühn, T. Goldammer, P. Laurent, J. E. Womack *et al.*, 2002 Targeted construction of a high-resolution, integrated, comprehensive and comparative map for a region specific to bovine chromosome 6 based on radiation hybrid mapping. Genomics **79:** 768–776.

Williams, J. L., A. Eggen, L. Ferretti, C. H. Farr, M. Gautier *et al.*, 2002 A bovine whole-genome radiation hybrid panel and outline map. Mamm. Genome **13:** 469–474.

Zhang, Q., D. Boichard, I. Hoeschele, C. Ernst, A. Eggen *et al.*, 1998 Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. Genetics **149:** 1959–1973.

Communicating editor: J. B. Walsh