Fine Mapping of Milk Production QTL on BTA6 by Combined Linkage and Linkage Disequilibrium Analysis

H. G. Olsen,¹ S. Lien,² M. Svendsen,¹ H. Nilsen,¹ A. Roseth,¹
M. Aasland Opsal,¹ and T. H. E. Meuwissen²

¹Department of Animal and Aquacultural Sciences,
Agricultural University of Norway,
N-1432 Aas, Norway

²Centre for Integrative Genetics and Department of Animal Science,
Agricultural University of Norway

ABSTRACT

Combined linkage and linkage disequilibrium analysis were used to refine the position of a previously detected QTL affecting milk production traits on bovine chromosome 6. Through a series of single- and multitrait and single- and multipoint QTL analyses, the QTL could be positioned to a 7.5-cM interval surrounded by the markers BMS2508 and FBN12. The most significant results were found for fat percentage and protein percentage. This effect seemed to be caused by a QTL allele embedded in one specific marker haplotype that caused a reduction in fat and protein yields and a concomitant increase of milk yield, thus resulting in a marked reduction of fat and protein percentages. (**Key words:** cattle, fine mapping, linkage disequilibrium, milk)

Abbreviation key: BTA = bovine (*Bos taurus*) chromosome, IBD = identical by descent, LA = linkage analysis, LD = linkage disequilibrium, logL = logarithm of the likelihood.

INTRODUCTION

Most important health and production traits in livestock are quantitative by nature—i.e., regulated by a large number of genes and also affected by the environment—that causes a continuous phenotypic distribution. Genes, or closely linked groups of genes, affecting such traits are called QTL. Although the effect of each of these genes on the genotypic variance is generally low, QTL significantly affecting important traits have been identified in a large number of livestock species. Usually, the first step toward mapping a QTL is to perform a genome scan using linkage analysis. With

this approach, a parent and an offspring generation are genotyped for a moderately dense marker map covering all chromosomes, and the inheritance of the marker alleles from parents to offspring is traced. A significant difference in phenotypes among groups of offspring having inherited alternate alleles of a marker from their common parent indicates linkage to the marker of a QTL affecting the trait (for details, see e.g., Georges et al., 1995; Knott et al., 1996). However, the mapping resolution achieved by this approach is low because the distances between markers are relatively large. More importantly, increasing the map density usually will not significantly improve the precision because there will be only very few recombinations between closely linked markers. Typically, confidence intervals for the most likely QTL positions are in the order of 20 to 30 cM. Because most practical implementations of marker information require the QTL to be mapped to an interval of 1 to 2 cM, methods of higher precision are needed to refine the region.

Linkage disequilibrium (LD) mapping allows the utilization of all recombinations that occurred during the generations before marker genotyping started and is thus a powerful method for fine mapping. Meuwissen and Goddard (2000) postulated that a base population in linkage equilibrium undergoes a mutation at the QTL, creating a novel QTL-allele embedded in one specific marker haplotype. Due to recombinations in the following generations, the original haplotype will remain only for markers close to the QTL. Thus, in the current generations, these marker alleles will be in linkage disequilibrium with the QTL alleles. The LD can be detected by estimating the effects of the marker haplotypes on the quantitative trait. Haplotypes with identical marker alleles are expected to have a similar effect on the trait because the identical marker alleles imply that the chromosomal region is inherited in a manner that is identical by descent (IBD) from an ancient common ancestor, and the haplotypes are therefore expected to carry similar QTL alleles. However, as

Received: April 28, 2003.

Accepted: August 11, 2003.

Corresponding author: H. G. Olsen; e-mail: hanne-gro.olsen@iha.nlh.no.

LD approaches are highly affected by factors such as population admixture causing spurious long-distance disequilibria, the extent of LD along the chromosome, and how correct the linkage phases are estimated, pure LD analysis may report a number of false positives (Perez-Enciso, 2003).

Linkage analysis and linkage disequilibrium analysis can be combined as proposed by Meuwissen et al. (2002). This approach allows the utilization of recombinations occurring both within and outside the pedigreed and genotyped generations (i.e., linkage analysis and linkage disequilibrium analysis, respectively) and also accounts for unknown background genes. Additionally, a combined approach prevents the false positives caused by accidental marker-phenotype associations from LA or the long-distance disequilibria found in cattle (Farnir et al., 2000) that may arise when the approaches are used separately.

In dairy cattle, much effort has been undertaken to detect QTL for milk production traits, and a number of chromosomes have been reported to harbor regions with significant effects. Several studies have reported the presence of one or more QTL on chromosome 6 (**BTA6**) (e.g., Georges et al., 1995; Spelman et al., 1996; Zhang et al., 1998; Velmala et al., 1999; Ron et al., 2001), but the results of the studies differ somewhat with respect to the number of QTL detected, their positions, and to which extent the milk traits are affected by the QTL. Several studies performed in a number of breeds have reported segregation of at least one QTL close to marker BM143 in the middle of BTA6 (e.g., Spelman et al., 1996; Kühn et al., 1999; Velmala et al., 1999; Nadesalingam et al., 2001; Ron et al., 2001). In a previous study in Norwegian Dairy Cattle, linkage analysis was utilized in a genome scan for QTL affecting 5 milk production traits (milk yield, fat percentage, fat yield, protein percentage, and protein yield) in 6 elite sire families with a total of 284 sons (Olsen et al., 2002). A highly significant QTL was detected on chromosome 6 close to marker FBN9, with its most likely position approximately 2 cM downstream of BM143. The 95% confidence interval for the QTL position was found to be 16 cM. Two of the elite sires, which were paternal half sibs, shared a common haplotype in the relevant area that seemed to cause a marked reduction in both fat and protein percentages and a weak increase of milk yield. No effect on fat and protein yield was found. The aim of the present study was to utilize the combined linkage and linkage disequilibrium method of Meuwissen et al. (2002) to refine the position of the detected QTL, and to investigate whether more than one QTL was segregating in this chromosomal segment.

Table 1. Name, relative position (in cM from the leftmost marker), number of alleles and heterozygosity (assuming a population in Hardy-Weinberg equilibrium) for the markers used for fine mapping.

Name	Position	Alleles	Heterozygosity
URB16	0	6	0.66
BM1329	1.01	6	0.66
BMS2508	9.73	7	0.73
FBN12	17.27	9	0.82
BMS1242	18.27	10	0.82
BM143	19.28	13	0.85
BMS690	22.37	10	0.62
DIK82	23.38	9	0.79
FBN13	29.77	9	0.58
BMS470	30.78	8	0.75

MATERIALS AND METHODS

Data

All animals in the study belonged to the dual-purpose Norwegian Dairy Cattle breed. The animals were organized in a granddaughter design consisting of 35 elite sire families that were chosen because of their relatively large number of progeny-tested sons. 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 approximately 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. Six of the families had been used in the previous genome scan (Olsen et al., 2002). Complex family relationships existed among the animals. Several of the sires were paternal halfsibs, and 18 of the sires were also included as sons in older families.

Predicted transmitting abilities of the sons from the 35 families were used as performance information in the analyses. The PTA for the 5 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 AI Association and evaluated using BLUP with a single trait sire-maternal grandsire model.

Marker Map

Initially, 16 publicly available microsatellites in the region around the putative QTL on chromosome 6 were tested for genotyping difficulties, number of alleles, and degree of heterozygosity among the 35 elite sires. Of these, 10 markers (URB16, BM1329, BMS2508, FBN12, BMS1242, BM143, BMS690, DIK82, FBN13, and BMS470) were selected for genotyping in all sires and sons. The relative positions, number of alleles, and heterozygosity of these markers are shown in Table 1. The markers spanned a region of approximately 31 cM

Figure 1. Markers used for previous genome scan (left) and current fine mapping (right) and distances between markers according to Haldane's mapping function. Numbers in boldface to the right indicates the ID number of the marker brackets.

around the putative QTL found in Olsen et al. (2002), thus covering approximately the peak of the test statistic curve of that study. Markers were genotyped using primers and PCR conditions as described by Ma et al. (1996) and Weikard et al. (1997) for URB16 and FBN markers, respectively, and at USMARC Genome Database (2003) for BM and BMS markers. Figure 1 shows the position of these markers as compared to the markers used in the previous genome scan (Olsen et al., 2002).

Marker order and map distances among markers were estimated using the CRI-MAP 2.4 program (Green et al., 1990), with map distances based on Haldane's mapping function. Average distance between markers was 3.42 cM, ranging from 1 to 8.7 cM. However, no

- 1 recombination was detected between markers FBN12 and BMS1242 (i.e., the markers surrounding bracket 4) in our data set. The order of these markers were set according to Weikard et al. (2002), and the distance between them set to 1 cM, as the mapping methods required some recombination between markers.
- 2

Statistical Analyses

- 3 **Single QTL-single-trait analysis.** Each of the 5 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 based on marker information.
- 4
- 5 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 IBD probabilities of pairs of haplotypes were calculated from marker and/or pedigree information (for details, see Meuwissen and Goddard, 2001; and 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 depended on the effective population size and the number of generations since the base populations, 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 i th position of the QTL. The last step was to calculate the likelihood of the data using restricted maximum likelihood. The model of the performance data was expressed as
- 6
- 7
- 8
- 9

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

where $\mathbf{y}$ is a $(n \times 1)$ vector of records (i.e., PTA for the milk trait in question), μ is the overall mean, $\mathbf{1}$ is a vector of 1's, $\mathbf{h}$ is a vector of random haplotype effects of dimension $q \times 1$, where q is the number of different haplotypes, $\mathbf{Z}$ is a $(n \times q)$ incidence matrix relating observations and haplotype effects, $\mathbf{u}$ is a vector of random polygenic effects, and $\mathbf{e}$ is a vector of residuals. The variances of $\mathbf{h}$, $\mathbf{u}$, and $\mathbf{e}$ 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, $\mathbf{A}$ is the additive genetic relationship matrix, and $\mathbf{R}$ 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 logarithm of the likelihood ($\log L$) 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 based on a model containing only background genes and the test statistic formed as the difference in logL between the model with and without a QTL. The marker bracket with the highest logL difference was taken as the most likely QTL position. The significance level of the test statistic was determined from the chi-square distribution, as the logL ratio times 2 is chi-squared distributed with 1 degree of freedom (i.e., the number of degrees of freedom is equal to the difference in number of parameters fitted in the model with a QTL vs. the model with no QTL). Because the QTL were already found to be highly significant in the study of Olsen et al. (2002), no specific methods for determining the significance threshold were used. Therefore, a nominal *P*-value of 5% was considered to be significant.

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 3 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 PTA of bull *i* for milk, fat, and protein yield, $\mathbf{b}$ is a vector of fixed effects, which includes the mean for each of the 3 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 3 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. Following Goddard (2001), it was assumed that all vectors of haplotype effects $\mathbf{h}_{ij}$ are along the same line, instead of pointing in all directions of their 3-dimensional space. This substantially reduces the number of parameters to be estimated, and the assumption is true when the QTL is bi-allelic (i.e., all haplotype effect vectors point to either the effect of QTL allele 1 or QTL allele 2 and thus lay along the line connecting these 2 points). The assumption is only approximately true when the QTL is multi-allelic, but 2 of the QTL alleles are much more frequent than the others. If the $\mathbf{h}_{ij}$ effects all lay along the same line, the variance-covariance matrix of $\mathbf{h}_{ij}$ can be modeled by:

$$\text{Var}(\mathbf{h}_{ij}) = \mathbf{H} = \mathbf{v}\mathbf{v}',$$

where $\mathbf{v}$ is the (3×1) direction vector of the line along which all $\mathbf{h}_{ij}$ are modeled. Thus, ASREML estimated only the 3 parameters of the direction vector $\mathbf{v}$, instead of the 6 parameters contained by the full variance-covariance matrix $\mathbf{H}$.

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:

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{Z}\mathbf{u} + \sum_i \mathbf{I}_i \mathbf{X}_i \mathbf{q}_i + \mathbf{e},$$

where $\mathbf{y}$, $\mathbf{1}$, μ , $\mathbf{Z}$, $\mathbf{u}$, and $\mathbf{e}$ is as for the single QTL-single-trait analyses, $\sum_i$ denotes summation over all possible QTL positions (bracket midpoints), $\mathbf{q}_i$ is a vector of haplotype effects, $\mathbf{X}_i$ is a known incidence matrix relating allelic effects with records, and $\mathbf{I}_i$ is an indicator variable, where $\mathbf{I}_i = 1$ ($\mathbf{I}_i = 0$) indicates (no) QTL at position *i*. $\text{Var}(\mathbf{q}_i) = \mathbf{G}_i \sigma_q^2$, where $\mathbf{G}_i$ is the IBD matrix between the haplotypes at the *i*th position. The variance components and haplotype effects were estimated using Gibbs sampling with 500,000 cycles, and $\mathbf{I}_i$ was sampled with a Metropolis-Hastings step (Meuwissen et al., 2001). As opposed to the paper of Meuwissen and Goddard (2002), the prior probability of having a QTL in bracket *i* ($\mathbf{I}_i = 1$) depended on the length of the bracket. As a QTL previously had been mapped to the relevant area of BTA6 (Olsen et al., 2002), the total prior probability of detecting a QTL in the genotyped area of 31 cM was 100%. The prior of each bracket was then set equal to the relative length of that bracket. Flat priors were assumed for σ_u^2 and σ_e^2 , whereas the prior distribution of σ_q^2 was inverse-chi-squared with 4.2 degrees of freedom (Meuwissen and Goddard, 2002). After discarding the initial 10,000 cycles from the MCMC chain as burn-in, the fraction of cycles with $\mathbf{I}_i = 1$ gave an estimate of the posterior probability of a QTL at position *i*. A significant QTL was present in a bracket if the posterior probability of that bracket exceeded 0.5.

Multiple QTL-multiple-trait analysis. Multiple QTL-multiple-trait analysis was performed according to Meuwissen and Goddard (accepted). This method is an extension of the multiple QTL-single-trait method. The data were modeled as:

$$\mathbf{y}_i = \mathbf{X}_i\mathbf{b} + \mathbf{u}_i + \sum_j (q_{ij1} + q_{ij2}) \mathbf{v}_j + \mathbf{e}_i,$$

where $\mathbf{y}_i$, $\mathbf{X}_i\mathbf{b}$, $\mathbf{u}_i$, and $\mathbf{e}_i$ are the same as for single QTL-multitrait analysis, $\sum_j$ denotes summation over all possible QTL positions, $\mathbf{v}_j$ is the direction vector for the effects of the QTL alleles at position *j*, q_{ij1} and q_{ij2} are the sizes of the QTL effects for the paternal and maternal allele of animal *i* at position *j* along the direction vector $\mathbf{v}_j$. As in the single QTL multitrait analysis, the number of parameters was reduced by assuming that the effects of the QTL alleles on each of the traits were lying on the same line instead of pointing in all directions of their 3-dimensional space. This reduction

Table 2. Single QTL – single trait analysis: The most likely marker brackets for each of the five milk production traits and the log-likelihood (logL) ratio and nominal P-values for the most likely position.

Trait	Bracket	Markers	LogL	P-value
Fat %	3	BMS2508-FBN12	36.4	1.4×10^{-17}
Protein %	3	BMS2508-FBN12	65.8	1.8×10^{-30}
Fat yield	6	BM143-BMS690	7.52	1.0×10^{-4}
Protein yield	5	BMS1242-BM143	6.66	2.6×10^{-4}
Milk yield	9	FBN13-BMS470	3.65	7.0×10^{-3}

of the number of parameters was more important in this analysis, since the effects of several QTL were estimated simultaneously. For details, see Meuwissen and Goddard (2003).

RESULTS

Initially, each of the milk production traits was analyzed separately using a single QTL analysis as described by Meuwissen et al. (2002). The log-likelihood ratio was formed for each marker bracket and trait by calculating the difference in logL between the model with a QTL in the midpoint of that bracket and the model without a QTL fitted. The results revealed a highly significant QTL in marker bracket 3 (BMS2508 – FBN12) for fat percentage and protein percentage (Table 2 and Figure 2). The logL ratio for fat percentage was 36.4 (nominal P -value = 1.4×10^{-17}), whereas the test statistic for protein percentage was 65.8 (nominal $P = 2 \times 10^{-30}$). However, as shown in Figure 2, the peaks for the percentage traits were rather broad, with brackets 2, 4, and 5 almost as likely as bracket 3. For fat yield, bracket 6 (BM143 – BMS690) showed a slightly significant logL ratio of 7.52 (nominal $P = 10^{-4}$). Protein

yield had the highest test statistic of 6.66 (nominal $P = 2.6 \times 10^{-4}$) in bracket 5 (BMS1242 – BM143). These traits also showed rather broad peaks with significant test statistics for several brackets. For milk yield, the most likely position was found in bracket 9 (FBN13 – BMS470), with a logL ratio of 3.65 (nominal $P = 7 \times 10^{-3}$).

The 3 yield traits (i.e., milk, fat, and protein yield) were analyzed simultaneously by a single QTL–multitrait analysis. As shown in Figure 3, the multitrait analysis yielded a logL ratio curve with similar shape as for the single trait analyses of the percentage traits shown in Figure 2. However, the peak of the likelihood curve was shifted slightly to the right, with bracket 4 (FBN12 – BMS1242) as the most likely position (logL ratio 30.28, nominal $P = 7.1 \times 10^{-15}$). As can be seen from Figure 3, the marker brackets adjacent to this peak had very similar test statistics. The curve had a somewhat bimodal shape, with a broad peak extending from brackets 2 to 6, and a smaller peak at bracket 8.

To test for the presence of several QTL on BTA6, all 5 traits were analyzed individually with a multiple QTL–single-trait model (Meuwissen and Goddard, 2002). With this approach, a sharp peak was found for both percentage traits in bracket 3, with a posterior probability of 0.98 and 1.0 for fat% and protein%, re-

Figure 2. Single QTL–single-trait analysis. Protein percentage (■), fat percentage (▲), protein yield (○), fat yield (◆), milk yield (---). Arrows indicate approximate marker positions.

Figure 3. Single QTL–multitrait analysis (milk yield, fat yield and protein yield).

Figure 4. Multiple QTL-single-trait analysis. Protein percentage (■), fat percentage (▲), protein yield (○), fat yield (—), milk yield (---).

spectively (Figure 4). Minor peaks were also found for protein yield in bracket 3, and for several traits in brackets 6 and 8. However, the posterior probabilities were smaller than 0.5 for these marker brackets, i.e., the Bayesian analysis indicated that the presence of a QTL was less likely than no QTL.

As a final step, multiple QTL-multitrait analysis was performed for the 3 yield traits according to Meuwissen and Goddard (accepted). Again, a highly significant QTL was found in bracket 3 (Figure 5), with a posterior probability of 0.98. Bracket 6 had a probability of 0.36, whereas the posterior probabilities for all other brackets were close to zero. Although these results gave some indications of a second QTL toward the rightmost end of

Figure 5. Multiple QTL-multitrait analysis (milk yield, fat yield, and protein yield).

Table 3. Frequency and marker alleles for the most common haplotypes (i.e. frequency ≥ 0.03) of marker bracket 3 (BMS2508 – FBN12)

Haplotype ID	Frequency	Alleles
22	0.0614	5 – 1
25	0.0605	3 – 3
5	0.0507	5 – 1
21	0.0457	5 – 1
2	0.0435	3 – 5
3	0.0381	4 – 4
6	0.0318	7 – 2
29	0.0318	6 – 3

the genotyped area, no significant evidence for a second QTL at this position was found.

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 and by Gibbs sampling for the multiple QTL analyses. Haplotypes with IBD probabilities exceeding 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 190 for bracket 5 to 759 for bracket 2. For bracket 3 (i.e., BMS2508 – FBN12), the total number of haplotypes was 481, with a frequency of 0.06 for the most common one. A summary of the most common haplotypes of bracket 3 with their frequencies and marker alleles is given in Table 3.

In an attempt to identify haplotypes carrying QTL alleles with major impact on milk production, haplotype effects, **h** (as estimated by the single QTL analyses) for bracket 3 were plotted for each of the traits. An example is shown for protein percentage in Figure 6, where the haplotype ID are plotted along the X-axis and their effects (expressed as deviations from mean PTA) are plotted on the Y-axis. As shown in Figure 6, most estimates of **h** were around the mean, whereas 8 of them showed a distinct reduction of protein percentage. The ID numbers of these haplotype effects were 5, 338, 342, 343, 344, 352, 387, and 395, respectively. Of these 8 haplotypes, one (denoted haplotype number 5) was found in 113 animals, whereas the other 7 were found in one individual each. By comparing the mean estimated effect of these 8 haplotypes to the mean of all other haplotypes, the extreme haplotypes were found to cause a rather large reduction in PTA of fat yield and protein yield of 3.3 and 2.2 kg, respectively, and a minor increase in milk yield of 21.4 kg. This reduction in kilograms of fat and protein, combined with a small increase of kilograms of milk, lead to major reductions in the percentage traits, with a decrease in fat of 0.07 percentage points and protein of 0.05 percentage points. Studying the marker haplotypes that were associated

Figure 6. Haplotype effects for the combined analysis of milk yield, fat yield and protein yield (single QTL). Haplotype ID numbers are plotted along the X-axis and the effects for protein percentage as measured in deviations from mean PTA on the Y-axis.

with these 8 *h*-values with extreme effects revealed that 7 of the 8 extreme effects were associated with marker allele 5 for BMS2508, allele 1 for FBN12, and allele 6 of BMS1242. Thus, a QTL allele with negative effect on fat and protein and a slightly positive effect on milk yield seems to be embedded in the left part of a common 5-1-6 haplotype spanning brackets 3 and 4. The eighth extreme haplotype (number 352) had alleles 3-0-6 at the same markers, where 0 means that marker FBN12 failed the genotyping in this particular individual. The sire of this animal was heterozygous for FBN12 alleles 1 and 5. Based on genotype information of surrounding markers and recombination fractions, allele 5 has a higher probability of being inherited from sire to son. Thus, a second QTL allele could be embedded in the 3-5-6 haplotype. Alternatively, the effect of this haplotype could be similar to that of haplotype number 5 (alleles 5-1-6) due to the fact that these haplotypes carry the same alleles for all markers downstream of BMS1242.

DISCUSSION

A series of steps have been taken in order to fine map a QTL affecting milk production traits on bovine chromosome 6. The initial single-QTL-single-trait analysis revealed a highly significant QTL affecting fat percentage and protein percentage in marker bracket 3, i.e., an interval of 7.5 cM spanning from marker BMS2508 to FBN12 (Figure 2). No significant effect for any of the yield traits was found at this position. However, the peak for the log-likelihood curve was rather broad, with several of the surrounding brackets

almost as likely as bracket 3. In addition, smaller peaks were found for several of the traits toward the end of the genotyped area.

The milk traits are known to be highly correlated, with a positive correlation among the 3 yield traits as well as between the 2 percentage traits, and a negative correlation between yield and percentage traits. Thus, by using a composite hypothesis combining information from all traits simultaneously in multitrait analyses, the power to detect QTL and the accuracy of estimating QTL positions might be improved, as the QTL may have pleiotropic effects on all traits. Only the 3 yield traits were used for the multitrait analysis, as the percentage traits are functions of the yield traits and would contribute no extra information. The multitrait approach confirmed the QTL found in the single-trait approach but did not refine the QTL position as the shape of the logL ratio curves was similar for the 2 approaches (Figures 3 and 5).

The rather broad curve extending from brackets 2 to 5 could be due to the presence of 2 or more closely linked QTL affecting the trait in the same direction, which could cause a significant test statistic over a large chromosome segment. This was investigated by fitting 2 or more QTL simultaneously. The multipoint QTL analyses were performed for each of the 5 traits individually, as well as for a multitrait analysis including only the 3 yield traits (Figures 4 and 5, respectively). The multi-QTL analysis did not clearly show the presence of more than one QTL in this area, but the fitting of multiple QTL gave a much sharper indication of bracket 3 as the QTL position than the single QTL models. This is probably because the multi-QTL mapping reduced the carryover effect of the QTL to adjacent putative QTL positions (Meuwissen and Goddard, accepted).

All analyses revealed small peaks around brackets 6 to 8. Although no significant test statistics were detected, the fact that the peak remains through all analyses could indicate a second QTL with minor effects in this area.

These results strongly suggested the interval between markers BMS2508 and FBN12 as the most likely position of the QTL. In a study of Israeli Holstein, Ron et al. (2001) detected a highly significant QTL for fat and protein percentages close to BM143, with a 95% confidence interval for the position of a QTL for protein percentage of only 4 cM in two families. The confidence interval for fat percentage was spanning approximately 10 cM in their study. The most likely QTL position in our study, i.e., the midpoint of the interval between markers BMS2508 and FBN12, was situated approximately 4 cM to the left of BM143 in our map. A QTL in the vicinity of BM143 has also been reported in a number of other Holstein populations (e.g., Spelman et

al., 1996; Kühn et al., 1999; Nadesalingam et al., 2001) as well as Finnish Ayrshire (Velmalä et al., 1999). Both in the study of Velmalä et al. (1999) and of Ron et al. (2001), two families were segregating for protein percentage close to BM143. In both cases, however, the two sires had different alleles for BM143, so if they have had an IBD segment containing a common QTL allele, the IBD region is no longer including this marker. Thus, the QTL found in Norwegian Dairy Cattle seems to be the same as that found in these studies.

In a previous genome scan performed on a subset of the families utilized in the current study, the most likely QTL position was found close to marker FBN9 (Olsen et al., 2002). FBN9 is situated between markers DIK82 and FBN13, which was bracket 8 in the current study (Figure 1). Although the QTL position found in the former study was somewhat to the right compared with marker bracket 3 of the present study, a 95% confidence interval for the QTL position found in the previous study spanned approximately brackets 2 to 9 here, and thus the results of the 2 studies are in agreement with each other. In the current study, some smaller, nonsignificant peaks were found around brackets 6 to 8, i.e., close to the most likely position from the genome scan. If these additional peaks are in fact real QTL with minor effects, their effects could have contributed to the shift to the right of the log likelihood peak in the former study.

The combined LA and LD approach narrowed the interval of the QTL from a 16 cM confidence interval in the genome scan (Olsen et al., 2002) to a marker bracket of 7.5 cM in this study. However, in order to identify the gene(s) affecting the traits or utilize QTL information efficiently in marker assisted selection, the QTL position needs to be narrowed down further to 1 to 2 cM. One of several approaches for refining this interval could be to repeat the analyses for individual cM within bracket 3, as only the bracket midpoints were considered with the methods outlined here. However, with a dense marker map, individual positions within a bracket would normally have very similar probabilities of harboring a QTL and would thus yield little new information.

In the present study, with some rather broad intervals and the QTL situated in the second largest bracket, one could argue that the utilization of the combined analysis did not improve the mapping resolution much as compared to a pure linkage analysis. To test the advantage of the combined approach over LA, the same data were analyzed for protein percentage with a model similar to that of the single-trait-single-QTL model but including linkage information only. As shown in Figure 7, the shape of the LogL curve was relatively similar to that of the combined analysis (uppermost curve of

Figure 7. Linkage analysis for protein percentage.

Figure 2), although flatter and with a decreased difference in likelihood between the hypotheses of one versus no QTL. Confidence intervals of the QTL position were calculated for both models using the 1 LOD drop-off criteria (Lander and Botstein, 1989). The interval for the combined analysis contained brackets 3 and 4 only, whereas for the linkage analysis, all brackets, except 1 and 9, were included. Since the combined linkage-linkage disequilibrium mapping approach is especially suited to fine-mapping with a dense marker map, the best approach to refine the QTL position is to increase the marker density between BMS2508 and FBN12 and repeat the analysis on the improved marker map. As the numbers of microsatellites in the region are very limited, effort is now undertaken to generate and genotype a large number of single nucleotide polymorphisms in order to refine the QTL position.

The effect of the QTL found in bracket 3 seems to be due to one specific QTL allele embedded in a haplotype consisting of alleles 5-1-6 at markers BM2508, FBN12, and BMS1242, respectively, i.e., the markers surrounding bracket 3 and 4. This QTL allele seems to cause a decrease in fat and protein yields, as well as a small increase in milk yield. Such an opposite effect on milk yield, and composition would result in a considerable reduction of fat and protein percentages and thus explains the highly significant results obtained for these traits. According to the literature, the QTL is commonly thought of as affecting the percentage traits (Nadesalingam et al., 2001), but the results of several studies suggests that the primary effect is on milk yield, with little effect on fat and protein yield, such that fat and protein percentages are affected (i.e., Georges et al., 1995; Zhang et al., 1998; Velmalä et al., 1999; and Nadesalingam et al., 2001). In Norwegian Dairy Cattle,

however, the percentages seem to be affected mainly through the reduction of fat and protein yield. The QTL allele occurs with a relatively low frequency in the Norwegian population, possibly due to the selection for less but richer milk.

CONCLUSIONS

The QTL for milk production that has previously been mapped to BTA6 in several studies was here fine mapped to a marker bracket surrounded by the markers BMS2508 and FBN12 using combined linkage and linkage disequilibrium analysis. As the relevant marker bracket spans a 7.5-cM region, more markers are needed in this area to further refine the position of the QTL. One specific marker haplotype containing a relatively rare QTL allele has been detected. The QTL allele reduces fat and protein yield and increases milk yield somewhat. This combination of effects implies that the QTL has major effects on the fat and protein percentage of the milk.

ACKNOWLEDGMENTS

We thank GENO Breeding and A. I. Association for providing relationship information and PTAs for sons. The project has received funding from the Research Council of Norway and GENO Breeding and A. I. Association.

REFERENCES

- Farnir, F., W. Coppieters, J.-J. Arranz, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L. Moreau, M. Mni, C. Nezer, P. Simon, P. Vanmanshoven, D. Wagenaar, and M. Georges. 2000. Extensive genome-wide linkage disequilibrium in cattle. *Genome Res.* 10:220–227.
- Georges, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto, A. T. Pasquino, L. S. Sargeant, A. Sorensen, M. R. Steele, X. Zhao, J. E. Womack, and I. Hoeschele. 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. Available: <ftp.res.bbsrc.ac.uk/pub/aar>.
- Goddard, M. E. 2001. The validity of genetic models underlying quantitative traits. *Livest. Prod. Sci.* 72:117–127.
- Green, P., K. Falls, and S. Crooks. 1990. Documentation for CRIMAP, version 2.4. Washington University School of Medicine, St. Louis, MO.
- Knott, S. A., J. M. Elsen, and C. S. Haley. 1996. Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theor. Appl. Genet.* 93:71–80.
- 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.
- Lander, E. S., and D. B. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 138:1301–1308.
- Ma, R. Z., I. Russ, C. Park, D. W. Heyen, J. E. Beever, C. A. Green, and H. A. Lewin. 1996. Isolation and characterization of 45 polymorphic microsatellites from the bovine genome. *Anim. Genet.* 27:43–47.
- Meuwissen, T. H. E., and M. E. Goddard. 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics* 155:421–430.
- 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. Communication 21–18. 7th. World Congr. Genet. Appl. Livest. Prod., Montpellier, France.
- Meuwissen, T. H. E., and M. E. Goddard. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genet. Sel. Evol.* (accepted)
- 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.
- Olsen, H. G., L. Gomez-Raya, D. I. Våge, I. Olsaker, H. Klungland, M. Svendsen, T. Adnøy, A. Sabry, G. Klemetsdal, N. Schulman, W. Krämer, G. Thaller, K. Rønningen, and S. Lien. 2002. A genome scan for quantitative trait loci affecting milk production traits in Norwegian Dairy Cattle. *J. Dairy Sci.* 85:3124–3130.
- Perez-Enciso, M. 2003. Fine mapping of complex trait genes combining pedigree and linkage disequilibrium information. A bayesian unified framework. *Genetics* 163:1497–1510.
- 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 daughter design. *Genetics* 159:727–735.
- 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.
- USMARC Genome Database. 2003. US Meat Animal Research Center, Clay Center, Nebraska, USA. Available: <http://www.marc.usda.gov/genome/genome.html>.
- Velmal, R. J., H. J. Vilkkil, 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, M. Schwerin. 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.
- Zhang, Q., D. Boichard, I. Hoeschele, C. Ernst, A. Eggen, B. Murkve, M. Pfister-Genskow, L. A. Witte, F. E. Grignola, P. Uimari, G. Thaller, and M. D. Bishop. 1998. Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. *Genetics* 149:1959–1973.