Quantitative Trait Loci Analysis for Five Milk Production Traits on Chromosome *Six* in the Dutch Holstein-Friesian Population

Richard J. Spelman,*[,1] Wouter Coppieters,[†] Latifa Karim,[†] Johan A. M. van Arendonk* and Henk Bovenhuis*

**Department of Animal Breeding, Wageningen Institute of Animal Sciences, Wageningen Agricultural University, The Netherlands and [†]Department of Genetics, University of Liege, Belgium*

Manuscript received March 8, 1996
Accepted for publication September 9, 1996

ABSTRACT

Twenty Dutch Holstein-Friesian families, with a total of 715 sires, were evaluated in a granddaughter experiment design for marker-QTL associations. Five traits—milk, fat and protein yield and fat and protein percent—were analyzed. Across-family analysis was undertaken using multimarker regression principles. One and two QTL models were fitted. Critical values for the test statistic were calculated empirically by permuting the data. Individual trait distributions of permuted test statistics differed and, thus distributions, had to be calculated for each trait. Experimentwise critical values, which account for evaluating marker-QTL associations on all 29 autosomal bovine chromosomes and for five traits, were calculated. A QTL for protein percent was identified in one and two QTL models and was significant at the 1 and 2% level, respectively. Extending the multimarker regression approach to an analysis including two QTL was limited by families not being informative at all markers, which resulted in singularity. Below average heterozygosity for the first and last marker lowered information content for the first and last marker bracket. Highly informative markers at the ends of the mapped chromosome would overcome the decrease in information content in the first and last marker bracket and singularity for the two QTL model.

USE of deoxyribonucleic acid (DNA) markers to search for loci that affect quantitative traits, known as quantitative trait loci (QTL), has become widespread in recent times. Identifying marker-QTL associations in farm animals may be undertaken in various experimental settings including the so-called "daughter" or "granddaughter" designs (GELDERMANN 1975; WELLER *et al.* 1990; VAN DER BEEK *et al.* 1995). Analytical techniques have been developed to identify marker-QTL associations (*e.g.*, WELLER 1986; LANDER and BOTSTEIN 1989; HALEY and KNOTT 1992).

KNOTT *et al.* (1994) developed a multimarker regression method to determine position and effect of QTL. The multimarker technique was demonstrated on simulated data for a half-sib population (KNOTT *et al.* 1994). The issue of calculating appropriate critical values that account for repeated testing has been addressed (*e.g.*, HALEY *et al.* 1994; JANSEN 1993; CHURCHILL and DOERGE 1994). CHURCHILL and DOERGE (1994) developed an empirical method based on the concept of the permutation test and illustrated the method on real data sets derived from F_2 and recombinant inbred plant populations and simulated data from a backcross design.

GEORGES *et al.* (1995) reported five chromosomes that gave evidence (LOD score 3) for the presence of a QTL controlling milk yield in the American Holstein population. Chromosome *six* was one of the five chromosomes identified. The QTL on chromosome *six* increased milk yield but not fat or protein yield and as a result protein and fat percent decreased. BOVENHUIS and WELLER (1994) reported an effect for fat percent that was linked to the casein locus, which is also found on chromosome *six*.

The objective of this study is to demonstrate the application of KNOTT *et al.*'s (1994) multimarker approach and CHURCHILL and DOERGE's (1994) empirical method of calculating critical values to outbred dairy population data generated from a granddaughter design. Both methods are extended; KNOTT *et al.*'s (1994) multimarker approach to a two-QTL model and CHURCHILL and DOERGE's (1994) permutation test to accommodate multiple traits in the calculation of critical values. The application of these methods is demonstrated for chromosome *six*.

MATERIALS AND METHOD

Experiment structure: Twenty Holstein-Friesian families from the Netherlands in a granddaughter experiment design were evaluated for marker-QTL associations. Average number of sons per grandsire is 36 with a range of 12–140 (Table 1). To avoid selection bias and its influence on detecting QTL (as described by MACKINNON and GEORGES 1992), selected sons were scrutinized by date of progeny testing within each

Corresponding author: Richard Spelman, Department of Animal Breeding, Wageningen Institute of Animal Sciences, Wageningen Agricultural University, P.O. Box 338, 6700 AH, Wageningen, The Netherlands. E-mail: richard.spelman@alg.vf.wau.nl

[1]On leave from Livestock Improvement Corporation, New Zealand.

TABLE 1

Experimental design and genetic markers used for chromosome *six*

	Marker										
Grandsire	1 BM1329	2 BM143	3 TGLA37	4 BM4528	5 BM415	6 KCAS	7 BM4311	8 BP7	9 BM2320	Total	Sons
A		1			1		1	1		4	13
B		1		1	1	1	1	1		6	40
C		1			1		1	1		4	22
D		1	1		1	1			1	5	12
E					1		1	1		3	16
F		1			1		1	1	1	5	32
G		1			1	1		1		4	42
H	1	1		1			1		1	5	140
I	1	1	1	1	1	1	1	1		8	20
J		1		1		1			1	4	54
K		1	1	1	1	1				5	23
L		1	1	1	1		1	1	1	7	71
M		1		1		1	1	1	1	6	26
N	1	1	1	1	1	1		1	1	8	12
O		1	1		1	1	1	1		6	75
P	1	1			1		1	1	1	6	60
Q		1	1				1			3	15
R		1		1		1	1	1	1	6	14
S				1					1	2	16
T		1	1	1	1		1	1		6	12
Total	4	18	8	11	14	10	14	14	10		
Map, cM	0	13	20	31	41	52	54	58	95		

The table details the markers for which grandsires are heterozygous (indicated by a 1) and the total for each grandsire, number of sons for each grandsire, number of grandsires heterozygous at each marker, and marker distances based on HALDANE's (1919) mapping function.

grandsire family. When there was DNA (semen samples) for only some sons that were progeny tested during a given period, information from this group of selected sons was not used. The grandsire family for the selected sons was not necessarily removed as there were time periods when all of the progeny tested sons had semen samples retained. Some 80 selected sons were not analyzed (selected sons are not in Table 1). If a son was not informative at any of the markers, he was still retained in the analysis as he contributed to calculation of the fixed effect of the grandsire (Equation 1).

Nine microsatellite markers were positioned and ordered on chromosome *six* with the ANIMAP programs (D. NIELSON and M. GEORGES, unpublished data) as described by GEORGES *et al.* (1995). The map for chromosome *six* is 95 cM long using HALDANE's (1919) mapping function (Table 1). For one of the nine markers, the position could not be determined unambiguously. The odds for switching marker six (casein locus) and marker seven were only 2.6:1 in favor of the order that was used in the analysis. With the exception of the orientation of marker six and seven and marker three (TGLA37, GEORGES *et al.* 1995), the map in this study corresponds to that of BISHOP *et al.* (1994).

Grandsire heterozygosity was on average 57% for the nine markers. However, there was large variation in the heterozygosity of the grandsires over the nine markers and also heterozygosity level between markers (Table 1). When the grandsire was heterozygous at a marker locus, it was, on average, known with certainty in 65% of cases which marker allele was transmitted from grandsire to son.

Five traits were analyzed for marker-QTL effects: milk, fat and protein yield (termed yield traits) and fat and protein percent (termed percentage traits). Daughter yield deviations (DYDs), weighted averages of a sire's daughter's lactation performances expressed as deviations from the population mean (VAN RADEN and WIGGANS 1991) were used as the phenotypic measurement. DYDs for the percentage traits were calculated from the yield traits. DYDs were taken from the September 1995 evaluation conducted by the Royal Dutch Cattle Syndicate.

Power of this design using the method of WELLER *et al.* (1990) was 0.6 for a bi-allelic QTL of size (half the difference in genetic value between homozygotes) 0.2 phenotypic standard deviation with equal allele frequency for a trait with heritability of 0.3 (*e.g.*, yield traits), type I error (comparisonwise) set to 0.05 and no recombination between marker and QTL with fully informative markers. Power was 0.9 for the same criteria but for a trait with heritability of 0.6 (*e.g.*, percentage traits) and a QTL effect of 0.4 phenotypic standard deviation.

One QTL analysis: Analysis was undertaken using multimarker regression principles as developed by KNOTT *et al.* (1994). Basic steps of multimarker regression were determination of the most likely haplotypes of the two grandsire gametes based on genotypes of his sons. The most likely linkage phase was taken, and when both phases were equally likely, one was selected at random. The QTL allele of interest was arbitrarily assigned to the linkage phase denoted linkage phase one. The probability of inheriting the chromosomewise segment of linkage phase one at any position was calculated for each son based on information from the closest informative flanking markers. DYDs were then regressed on this conditional probability.

Across-family analysis was undertaken by fitting a one QTL model to the data:

$$Y_{ijk} = \mu + gs_i + b_{ik}X_{ijk} + e_{ijk} \quad (1)$$

where Y_{ijk} is the DYD for the jth son of the ith grandsire at the kth chromosomewise position, μ is the overall mean, gs_i is the fixed effect of the ith grandsire, b_{ik} is the regression coefficient for the ith grandsire at the kth chromosomewise position, X_{ijk} is the probability of the jth son receiving the chromosomewise segment for gamete one from the ith grandsire at the kth position, and e_{ijk} is the random residual.

This model allows multiple QTL alleles. Each grandsire family was constrained to a bi-allelic QTL as only the transmission of marker alleles from the grandsire were considered and grandams contribution ignored. Thus over the 20 families there were, in total, 40 possible alleles. For across-family analysis, residual sums of squares (RSS) were summed across families, thus the larger grandsire families contributed to a larger extent to overall RSS. Within grandsire families the number of daughters that each son had varied from those which had only a part progeny test proof to sires that were used extensively as proven sires. Contribution of each sire was weighted according to the number of daughters contributing to the DYD. The weighting factor was based on the variance of the DYD for a son being:

$$\text{Var DYD} = \left[\frac{1 + (n-1)\frac{1}{4}h^2}{n}\right]\sigma_p^2 \quad (2a)$$

where Var DYD is the variance of son's DYD; n is the number of daughters contributing to the DYD; h^2 is the heritability, which was taken as 0.35 for yield traits and 0.75 for percentage traits (average heritabilities from VAN DER WERF and DE BOER 1989); and σp^2 = the phenotypic variance.

Assuming equal phenotypic variance for all observations the weighting factor (w) is:

$$w = \left[\frac{1 + (n-1)\frac{1}{4}h^2}{n}\right] \quad (2b)$$

The weighted residual sums of squares is:

$$\sum_{j=1}^{n}\frac{1}{w_{ij}}(y_{ijk} - \mu - gs_i - b_{ik}X_{ijk})^2 \quad (3a)$$

and the sums of squares explained by the QTL fitted in the model is:

$$\sum_{j=1}^{n}\frac{1}{w_{ij}}(y_{ijk} - \mu - gs_i)^2 - \sum_{j=1}^{n}\frac{1}{w_{ij}} \times (y_{ijk} - \mu - gs_i - b_{ik}X_{ijk})^2 \quad (3b)$$

where w_{ij} is the weighting factor (Equation 2b) for the jth son of the ith grandsire. Equation 3b is equivalent to R(QTL$|\mu$, gs_i) (reduction in residual sums of squares) where QTL represents the QTL fitted (*i.e.*, b_{ik} and X_{ijk}).

Test statistics were calculated similar to a F statistic but were not termed as such because the distribution of the test statistics did not follow a F distribution. Test statistics were calculated every centiMorgan over the mapped chromosome.

Critical values: Test statistic critical values were calculated empirically from the permutation method outlined by CHURCHILL and DOERGE (1994). In brief, the permutation test was undertaken by repeatedly randomly shuffling the phenotypic data (DYDs with their weighting factors) within each family and calculating test statistics for each shuffle. The conditional probabilities (X_{ijk}'s) that the DYDs are regressed on were not shuffled. Critical values were calculated from the distribution of test statistics. Comparisonwise, chromosomewise and experimentwise critical values were calculated. Comparisonwise values were calculated each centiMorgan and provided critical values for that point but did not account for repeated testing over the genome or for the five different traits. Chromosomewise values accounted for the multiple, dependent, testing on chromosome *six* and the five traits analyzed. The experimentwise critical values accounted for evaluation of marker-QTL associations on 29 autosomal bovine chromosomes and also the five traits analyzed.

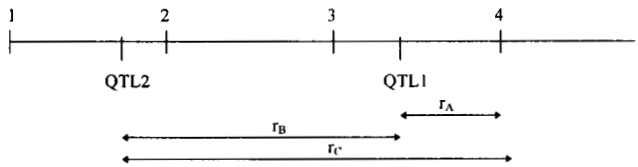

FIGURE 1.—Example of fitting a two-QTL model. Numbers 1–4 are the position of the markers and *QTL1* and *QTL2* are the positions of the postulated QTLs. Recombination rate between informative marker 4 and *QTL1* and *QTL2* are denoted r_A and r_C, respectively. Recombination rate between QTLs is r_B.

Two-QTL analysis: A two-QTL model was fitted to the data by extending the multimarker regression one-QTL model. The two-QTL model is:

$$Y_{ijk1k2} = \mathrm{m} + gs_i + b_{ik1}X_{ijk1} + b_{ik2}X_{ijk2} + \mathrm{e}_{ijk} \quad (4)$$

where k_1 and k_2 refer to the position of the first and second QTL. Other terms are as in (1).

The two QTL model was fitted by grid searching *i.e.*, each combination of 1-cM positions was evaluated. However, HALEY and KNOTT (1992) observed that QTLs 20 cM apart could not be differentiated. To ensure that the two postulated QTL had some distance between them, it was decided that only those positions where they were separated by an empty marker bracket would be evaluated. Having an empty marker bracket between postulated QTL was in agreement with ZENG (1993). He reported that two sample partial coefficients are generally uncorrelated unless the two markers are adjacent markers. An empty marker bracket between postulated QTL was not possible for all families as they were not informative at all marker loci (Table 1). For example, when a QTL was fitted in marker bracket one and the second QTL in marker bracket three, some families did not have an empty marker bracket between both QTL (Figure 1) because they were not informative at one or both of markers 2 and 3. Further, if the family was not informative at markers 1, 2 and 3, the two QTL were placed to the left of the first informative marker, marker 4. The probabilities of transmission of the QTL were calculated from information derived from marker 4 (Figure 1) using the technique of KNOTT *et al.* (1994). Thus there are only two groups of progeny, depending on the allele that they inherit at marker 4. The probability for inheriting a given allele at each of the QTL is the same for all individuals within a group. This can be demonstrated mathematically. Using HALDANE's (1919) equation we know that:

$$r_C = r_A + r_B - 2r_Ar_B \quad (5)$$

and if the probability of receiving QTL1 (X_{ijk1}) is:

$$\text{probability QTL1} = (1 - r_A) \quad (6)$$

and probability for QTL2 (X_{ijk2}) is:

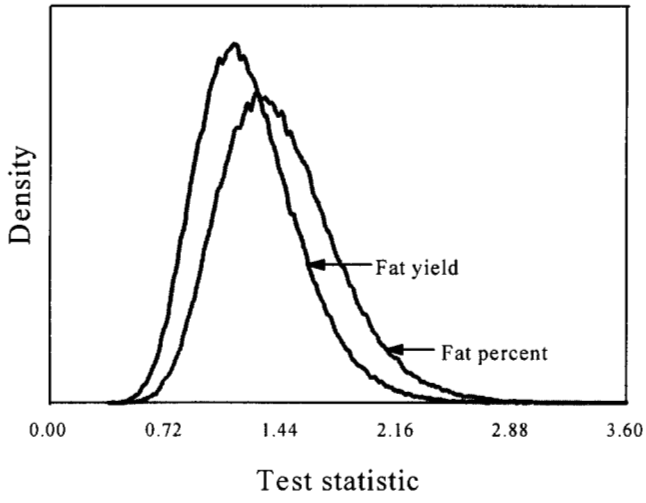

FIGURE 2.—Approximate density function of test statistics for fat yield and fat percent derived from permutation test (150,000 shuffles).

$$\text{probability QTL2} = (1 - r_C) \tag{6a}$$

then utilizing Equation 5:

$$\begin{aligned}\text{probability QTL2} &= 1 - (r_A + r_B - 2r_Ar_B)\\ &= (1 - r_A) - r_B + 2r_Ar_B\\ &= \text{probability QTL1} - r_B\\ &\quad + 2\, r_Ar_B\end{aligned} \tag{6b}$$

which is equivalent to:

$$\text{probability QTL2} = \text{probability QTL1} + \text{constant}. \tag{6c}$$

This results in singularity. Three of the 20 families (E, K, S; Table 1) were noninformative for the first or last three markers and thus excluded from the two QTL across family analysis.

Two test statistics were calculated for the two-QTL model. One test statistic compared the fit of the two-QTL model to that corresponding with the position of the highest test statistic in the one-QTL model (Equation 7c). The second test statistic determined if neither, one or both positions explained a significant amount of the variance in the two QTL model by the following method (Equation 7, a and b). The following reduction in sums of squares were calculated.

$$R(X_{ijk1} \mid \mu, gs_i, X_{ijk2}) \tag{7a}$$

$$R(X_{ijk2} \mid \mu, gs_i, X_{ijk1}) \tag{7b}$$

$$R(X_{ijk1}, X_{ijk2}, \mu, gs_i) - R(X_{ijk}, \mu, gs_i) \tag{7c}$$

where k corresponds to the position with the highest explained sums of squares for the one-QTL model. If neither Equation 7a nor 7b is significant, then neither of the two QTL positions are significant, otherwise at least one of the two positions is significant.

RESULTS

Permutation test: Distribution of test statistics between traits differed quite markedly. For example, fat percent had a larger proportion of higher permuted test statistics than fat yield (Figure 2). This is demonstrated numerically by the critical value at the 1% threshold level for fat yield being nearly equal to that at the 5% level for fat percent (Table 2). The mean of the test statistic distributions in Figure 2 are not one as would be expected with a F distribution. This is due to that the test statistic distributions in Figure 2 account for repeated testing across the chromosome and thus are not comparisonwise test statistics. CHURCHILL and DOERGE (1994) would refer to these distributions as chromosomewise, but in this paper, they are not as they do not account for repeated tests on the five correlated traits.

To account for the five traits being analyzed, the highest permuted value for the five traits from each shuffle (traits shuffled together) was used and combined critical values calculated (Table 2). However, with between trait differences, the highest value for each shuffle were dominated by traits that had higher absolute critical values, *i.e.*, fat percent in this study. Thus combined trait critical values were not applicable to the individual traits especially those traits with lower distribution of critical values.

The approach taken in this study was to estimate the equivalent number of independent traits tested. This was calculated by factor analysis (using SAS 1985) on a genetic correlation matrix for the five traits (VAN DER WERF and DE BOER 1989) and on the experimental phenotypic data. It was calculated that two factors account for ~90% of the variation and three factors account for some 99% on both the correlation matrix and DYDs. This was checked by analyzing each shuffle of the permuted test statistics for the five traits. In each shuffle, it was determined whether the permuted test statistics for each trait was significant at a certain threshold level using individual trait critical values. It was assumed that if the data was equivalent to x independent traits, then at the 10% threshold level (for a single trait) from 10,000 shuffles, there would be $10{,}000 \cdot 0.1^x$ occurrences where all five traits in the one shuffle were significant. Solving for x at the 10% threshold level with 10,000 permuted F values, gave 2.8 independent traits, which agrees with the results from the factor analysis. Based on these considerations, three independent traits were taken.

Equation 8a calculates the probability (P) of false positives at a given type I error (α) with n independent tests:

$$p = 1 - (1 - \alpha)^n \tag{8a}$$

which can be rearranged to:

$$\alpha = 1 - \exp\frac{\log(1 - \gamma)}{n} \tag{8b}$$

where α is the threshold level to ensure γ significance level over the n independent tests.

Equation 8b is equivalent to the standard Bonferroni correction for multiple testing. The correction factor is applied to all five traits (Table 2).

To account for testing on 29 autosomal chromosomes, experimentwise critical values were calculated for each trait. It was assumed that the distribution of test statistics seen for chromosome *six* were very similar for all of the other chromosomes. This assumption was based on the knowledge that the length of chromosome *six* is representative of the average length of the 29 autosomal chromosomes and thus representative of the amount of repeated testing across a chromosome. Using Equation 8b with $n = 87$ (three independent traits analyzed on 29 independent chromosomes), experimentwise critical levels were calculated (Table 3).

Comparisonwise critical values (not reported) were similar to tabulated F values. Comparisonwise critical values were relatively constant over the chromosome that is in agreement with the findings of CHURCHILL and DOERGE (1994) and VILKKI *et al.* (1996).

TABLE 2

Chromosomewise threshold levels

Threshold level (%)	Milk	Fat	Protein	Fat (%)	Protein (%)	Combined	*F* values
0.1	2.95	2.67	2.79	3.11	3.00	3.04	2.33
1	2.57	2.32	2.41	2.65	2.58	2.61	2.13
5	2.26	2.03	2.11	2.32	2.26	2.28	1.82
10	2.11	1.89	1.98	2.16	2.11	2.12	1.69

The chromosomewise critical values for the five milk production traits account for repeated testing over chromosome *six* and on the five correlated traits or equivalently three independent traits (150,000 shuffles). The critical values in the combined column are when the highest test value is taken from each shuffle of the five traits. The *F* values are tabulated values that have been adjusted with Bonferroni correction for testing on three independent traits.

For the rest of the paper, experimentwise critical values are used for across-family analysis unless stated otherwise. At chromosomewise areas of interest, based on significance levels, within family critical levels are tabulated *F* values unless stated otherwise. *F* values were chosen for ease of computation as the issue of repeated testing had been accounted in the across-family analysis, and comparisonwise values were similar to *F* values. All additive genetic effects are reported as half the difference in genetic value between homozygotes.

One-QTL model: Across-family analysis for the five production traits revealed a possible QTL for protein percent positioned at 13 cM, *i.e.*, the location of the second marker (Figure 3). The test statistic was significant at the 1% level for protein percent. The yield traits showed little indication of a QTL on chromosome *six* (Figure 4).

Two families were identified as having significant effects for protein percent at the mapped position of marker 2 (Table 4). The test statistics were significant at the 0.1% level for both families. Point estimates for the QTL effect for the two families were 1.12 and 0.68 genetic standard deviation, when using an estimate of protein percent genetic standard deviation of 0.136 (VAN DER WERF and DE BOER 1989) (Table 4).

Absolute marker readings for grandsires at marker position 2 (location of QTL) revealed that both families received a common marker allele (denoted as *X*). Grandsire B is one of grandsire A's six sons, which are grandsires in the experiment. Grandsire B was the only son that received marker allele *X*. One other family in the experiment also had marker allele *X*. This family, distantly related to families A and B, had no significant QTL effect for any of the traits. Marker allele *X* is associated with lower protein percent compared with the other marker allele for both families.

Grandsire A had a significant effect for milk (1%) and protein yield (5%). The effect was an increase of 555 kg of milk and an increase of 8.46 kg of protein in DYDs for sons that received marker allele *X*. The corresponding increase in milk is approximately double that expected for an increase of 8.46 kg of protein based on average protein percent of 3.46% in the Netherlands (AGRA EUROPE 1995). Grandsire B had a significant effect for protein yield (5%) and no significant effect for milk yield. Protein yield DYDs were 5.36 kg less for sons that received marker allele *X* while there was no difference in milk yield.

TABLE 3

Experimentwise threshold levels for the five traits (150,000 shuffles)

Threshold level (%)	Milk	Fat	Protein	Fat (%)	Protein (%)
1	3.05	2.86	2.92	3.29	3.17
5	2.85	2.60	2.69	2.99	2.88
10	2.73	2.48	2.59	2.87	2.77
15	2.67	2.41	2.52	2.76	2.69

Information content: Seventy percent of peak test statistics derived in across and within-family one QTL analysis occurred at the position of a marker locus (105 observations: 20 families and one across-family analysis for five traits). KNOTT and HALEY (1992) reported that when considering only flanking markers, the QTL position can be biased and placed in the marker brackets with higher information content. HALEY *et al.* (1994) reported that this problem can be overcome with the use of multimarker approach. However, the change in information content in HALEY's *et al.* (1994) simulation of outbred line crosses was not as marked as seen in this study.

If true descent (maternal or paternal) of every centiMorgan of DNA was known, the distribution of the QTL conditional probabilities would be for both 0 and 1. This distribution has mean 0.5 and variance 0.25. True descent is generally unknown and has to be inferred from informative flanking markers. Following a similar application by KRUGLYAK and LANDER (1995), variance of QTL conditional probabilities was calculated for each centiMorgan and is reported in Figure 5 as a fraction of maximum variance (0.25), which is used as a measure of information content.

Two-QTL model: The two-QTL models for all traits were not significant at the 15% threshold level when compared to the best one QTL model (test statistic 1, Table 5). However, using test statistic 2, there was a significant effect (2%) for one of the two positions for protein percent. The significant position for protein percent was at 1 cM whereas in the one QTL model it was at 13 cM (Figure 3).

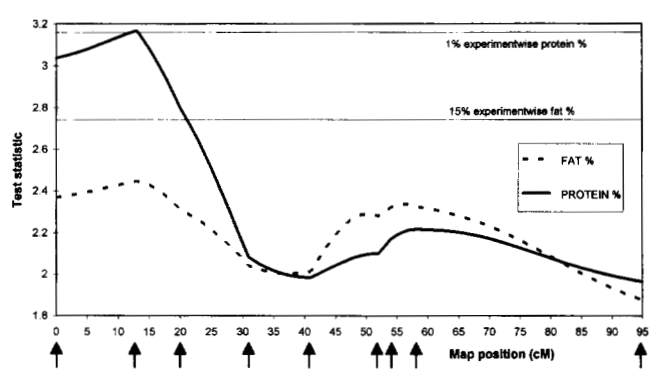

FIGURE 3.—Test statistics for different positions on chromosome *six* from an across-family analysis for protein and fat percent (arrows indicate position of markers).

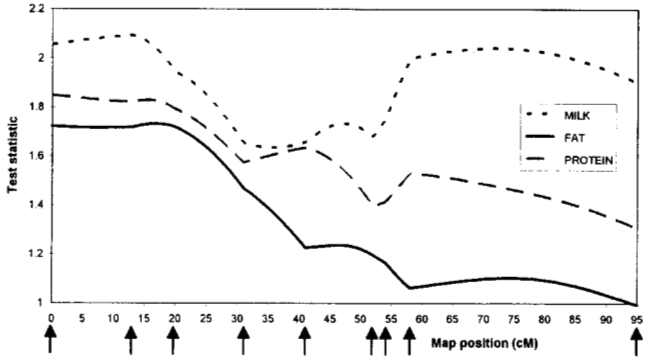

FIGURE 4.—Test statistics for different positions on chromosome *six* from an across family analysis for milk, fat and protein yield (arrows indicate position of markers).

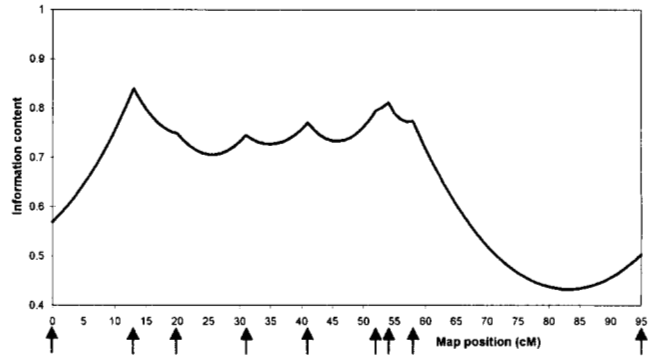

FIGURE 5.—Information content derived from chromosome *six*. Information content calculated from variance of QTL conditional probabilities at each centiMorgan as a proportion of the variance when true descent is known (arrows indicate position of markers).

DISCUSSION AND CONCLUSIONS

Permutation test: The permutation test is a quick method of calculating critical levels that takes into account repeated measures over the genome for individual traits. Individual trait distributions for test statistics differed and thus distributions had to be calculated for each trait. Different trait test statistic distributions are caused by differences in phenotypic distributions as the marker data is the same for all traits. The degree of nonnormality of the individual traits did not seem to have a direct link with the observed test statistic distribution differences. However, normality was calculated for each trait over all families, whereas the RSS are calculated within each family and then summed across all families. Therefore the normality of the phenotypic distributions within family may be the cause. Degree of normality was not determined as it would be calculated on <30 observations for over half of the families.

To account for repeated testing of the five traits analyzed the number of independent traits were calculated using factor analysis and analyzing the permuted test statistics. The chromosomewise critical levels were considerably higher than tabulated *F* values (Table 2); this reflects the repeated testing over the chromosome is accounted for in the permuted values.

The other method to account for testing on correlated traits, shuffling each trait and then taking the highest permuted test statistic from the five traits, had the effect that the combined test statistics were dominated by the traits with higher absolute values. Test statistics calculated in this manner were not applicable to the traits especially those with lower distributions of critical values. This is demonstrated by the combined critical values at the 5% level being equivalent to the critical value at the 1% level for fat yield (Table 2).

Experimentwise critical values were calculated on the assumptions that chromosome *six* was representative in length of the bovine chromosomes and thus the degree of repeated testing, and that the marker data has very little effect on critical values. The latter assumption was justified based upon the result that when altering the marker density for chromosome *six* to the extremes likely to be seen for the other chromosomes in this experiment (five and 12 markers per chromosome), only minor differences in critical values occurred. CHURCHILL and DOERGE (1994) using simulated data found differences in the distributions of test statistics for a 100 cM chromosome of "high marker density" (50 markers) compared with "low marker density" (10 markers). The difference in the finding of this study and that of CHURCHILL and DOERGE (1994) may be due to the influence and peculiarities of real data and the smaller contrast in marker density in this study.

Experimentwise critical values were chosen as all autosomal chromosomes will be analyzed for marker-QTL associations in this experimental design. However, the determination of which threshold level should be used is uncertain. If the objective of the experiment is to identify QTL that will be subsequently confirmed in a second study, an appropriate threshold level may be 15–20% on an experimentwise basis to ensure QTL are not missed. The experiment objective and the effect of using a false positive will determine appropriate threshold levels (for further discussion see LANDER and SCHORK 1994; LANDER and KRUGLYAK 1995).

Information content: Information content as measured by the variance of QTL conditional probabilities was not constant over the chromosome. Low heterozygosity at marker one affected information content for the first marker bracket. The large distance for the last

TABLE 4

QTL effect for protein percent in families A and B at position 13 cM

	Family A	Family B
Number of sons	13	40
F value	24.52[a]	15.85[a]
QTL effect (%)	15.25 ± 3.57	9.29 ± 1.94
QTL effect (σ_G)	1.12 ± 0.26	0.68 ± 0.14

[a] Significance at 0.1% based on tabulated *F* values.

TABLE 5

Results for the two-QTL analysis applied to the five milk production traits

	Milk		Fat yield		Protein yield		Fat (%)		Protein (%)	
Positions (cM)	36	63	25	65	36	65	13	58	1	61
Test statistic 1	1.37		1.07		1.08		1.63		1.57	
Test statistic 2	1.72	2.07	1.48	1.28	1.48	1.47	1.98	1.63	3.16[a]	1.72

QTL positions are where the lowest RSS occurred. Test statistic one is comparing the two-QTL model to the best one-QTL model. Test statistic two is when the other QTL in the two-QTL model position has been accounted for in a one-QTL model. Significance levels have been calculated from critical values presented in Table 3.

[a] Significance at the 2% level.

marker bracket combined with below average heterozygosity of the last marker also resulted in lower information content in the last marker bracket. As a result of the below average heterozygosity at the chromosomewise ends, nine of the 20 families could not have QTL position and effect separated in the first and last marker brackets. The information content peak at marker two was because 18 of the 20 families were informative at that position (Table 1). The information content peak at 50–60 cM was due to the high density of markers in that region. Improvement in information content will be achieved when the dam allele frequencies are used to calculate probabilities for animals in which transmission of alleles is uncertain.

The approach of having evenly spaced markers (*e.g.*, DARVASI and SOLLER 1994) is not be the best approach to have information content equitable over the chromosome. Once the postulated QTL is positioned beyond the last informative marker, information is coming from only a single marker and thus information content decreases. Highly informative markers at the end of the mapped chromosomes would overcome the decrease in information content at the boundaries. However, it is not possible to know marker heterozygosity before the experiment and thus the use of two markers closely positioned at either end of the chromosome may increase the heterozygosity and information content. Increased heterozygosity with closely placed markers is seen with markers six and seven (Table 1). The four families homozygous at marker seven are all heterozygous for marker six. Increased heterozygosity at chromosomewise ends will ensure estimates of position and QTL effect can be separated for most families in the first and last marker bracket. In addition, increased heterozygosity at chromosomewise ends will overcome the singularity problem for the two QTL model.

The observation that some 70% of peak test statistics occurred at the marker positions is mostly derived from within family analysis. The information content for each family will differ. The across-family information content has an averaging effect on information content in the individual families. The information content for an individual family will fluctuate more than that shown for across family. Local information content peaks at marker positions may be the cause of location of peak test statistics occurring at the marker.

Two-QTL model: Extending KNOTT's *et al.* (1994) multimarker regression approach to a two-QTL analysis was limited as families were not informative at all markers. Homozygosity at the start or end of the mapped chromosome resulted in fitting two QTL using information from only one of the flanking markers. This resulted in singularity and therefore three families being excluded from across family analysis. The approach of fitting two QTL is similar to that of using of markers as cofactors in the analysis of inbred crosses as described by JANSEN (1993) and ZENG (1994). These authors in addition to marker genotypes use trait phenotypic values in assigning conditional probabilities and also weight the probability of QTL phase in contrast to the KNOTT *et al.* (1994) approach where phase is assumed to be known with certainty. However, not assigning probabilities to phase was not critical in this study, as in most cases (80%) the probability of chosen phase was >70%. The exceptions to this was for the last marker bracket where the distance is large and for the smaller families. Using all markers together instead of an individual marker haplotype may improve the determination of phase for the smaller families. The approach of JANSEN (1993) and ZENG (1994) may overcome the singularity problem due to using information in addition to that from the single marker and thus breaking the complete colinearity between postulated QTL.

The use of markers as cofactors (JANSEN 1993; ZENG 1994) in outbred populations may not be possible as markers are not uniformly informative in all families as found in crosses of inbred lines. The approach of fitting postulated QTLs as cofactors, within families, on the same and other chromosomes may overcome this.

Two test statistics for comparisonwise of a two-QTL model to a one QTL were used. Comparing the two-QTL model to the best one-QTL model had the bias that the comparisonwise between models was for different QTL positions. The one-QTL model may detect a ghost QTL in between the two QTLs (MARTINEZ and

CURNOW 1992; HALEY and KNOTT 1992). If the two QTLs are in phase and of the same effect, the variance explained by a ghost QTL will be inflated and therefore not a good comparisonwise for the two QTL model. Fitting the two-QTL model and then determining if neither, one or both positions explained a significant amount of the variance in the two-QTL model was the preferred option for this study, as the comparisonwise is then between a two- and one-QTL model for the same QTL positions. However, it is acknowledged that for the second test statistic that when two QTL are in phase and of the same effect, this will inflate the variance explained at both positions in the one QTL model. This will also reduce the significance of the two QTL model when compared with the one-QTL models. Further research is needed in this area.

Casein: The findings of earlier studies for effects at and linked to the casein locus (summarized by BOVENHUIS *et al.* 1992) were not confirmed in this study. Nonsignificant peaks for the test statistic near the casein loci (marker six) were found for fat percent in the one-QTL model and for all traits for one of the two locations identified in the two-QTL model.

Protein percent QTL: A QTL for protein percent was identified in the across family study with a one-QTL model and was significant at the 1% level. Location of the protein percent QTL at marker two (13 cM) is practically the end of the mapped chromosome as marker one was informative in only four of the 20 grandsire families (Table 1). Families A and B were not informative at the first marker. Therefore QTL location and effect can not be separated for a QTL located in the first marker bracket for these two families. Informative markers to the left of marker two may change the mapped position of the QTL.

Allele *X* at marker position two was associated with the change in protein percent. Relative to the other marker allele the effect was a decrease in protein percent. The protein percent effect was caused by an increase in milk yield in family A and a decrease in protein yield in family B.

As described, family A and B are related. This is one of many relationships that exist within the data set but not used in this study. Accounting for the relationships within an animal model setting would most probably increase the power of the design. Methods to use these relationships are being investigated.

GEORGES *et al.* (1995) identified a QTL in one family on chromosome *six* that appeared to increase milk yield but not fat or protein yield and, as a result, fat and protein percent decreased. This family had two informative markers. The location of the QTL in GEORGES *et al.* (1995) is some 5–10 cM to the left of marker three used in this study. This is nearly the same QTL location found in this study. Family A and the family identified by GEORGES *et al.* (1995) have a common ancestor two and three generations back, respectively. The QTL found in this study and GEORGES *et al.* (1995) is very likely to be the same. A similar finding has been made in the Finnish Ayrshire population (R. VELMALA, personal communication).

Although the QTL has the same effect on protein percent in both studies and all three families, the effect on the yield traits differ between families. This may reflect the power of the respective studies. Further investigation through additional markers and more genotyping in the identified region may increase our understanding of the identified QTL on chromosome *six*.

We thank SARA KNOTT for the multimarker program. We are grateful to Livestock Improvement and Holland Genetics for financial support of the experiment and data access. The authors acknowledge the excellent genotyping work of LAURANCE MOREAU, FABIENNE MARCQ and ANDREINA SHOEBERLEIN and CHRIS SCHROOTEN for providing data.

LITERATURE CITED

AGRA EUROPE, 1995 Dairy Review. Special Report No. 82.

BISHOP, M. D., S. M. KAPPES, J. W. KEELE, R. T. STONE, S. L. SUNDEN *et al.*, 1994 A genetic linkage map for cattle. Genetics **136:** 619–639.

BOVENHUIS, H., J. A. M. VAN ARENDONK and S. KORVER, 1992 Associations between milk protein polymorphisms and milk production traits. J. Dairy Sci. **75:** 2549–2559.

BOVENHUIS, H., and J. I. WELLER, 1994 Mapping and analysis of dairy cattle quantitative trait loci by maximum likelihood methodology using milk protein genes as genetic markers. Genetics **137:** 267–280.

CHURCHILL, G. A., and R. W. DOERGE, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963–971.

DARVASI, A., and M. SOLLER, 1994 Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. Theor. Appl. Genet. **89:** 351–357.

GELDERMANN H., 1975 Investigations on inheritance of quantitative characters in animals by gene markers. I. Methods. Theor. Appl. Genet. **46:** 319–330.

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.

HALDANE, J. B. S., 1919 The combination of linkage values and the calculation of distances between the loci of linked factors. J. Genet. **2:** 3–19.

HALEY, C. S., and S. A. KNOTT, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity **69:** 315–324.

HALEY, C. S., S. A. KNOTT and J. ELSEN, 1994 Mapping quantitative trait loci in crosses between outbred lines using least squares. Genetics **136:** 1195–1207.

JANSEN, R. C., 1993 Interval mapping of multiple quantitative trait loci. Genetics **135:** 205–211.

KNOTT, S. A., J. M. ELSEN and C. S. HALEY, 1994 Multiple marker mapping of Quantitative trait loci in half sib populations. Proceedings of the 5th World Congress on Genetics applied to Livestock Production, Guelph, Canada. **21:** 33–36.

KNOTT, S. A., and C. S. HALEY, 1992 Maximum likelihood mapping of quantitative trait loci using full-sib families. Genetics **132:** 1211–1222.

KRUGLYAK, L., and E. S. LANDER, 1995 Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am. J. Hum. Genet. **57:** 439–454.

LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

LANDER, E., and L. KRUGLYAK, 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat. Genet. **11:** 241–247.

LANDER, E. S., and N. J. SCHORK, 1994 Genetic dissection of complex traits. Science **265:** 2037–2048.

MACKINNON, M. J., and M. A. J. GEORGES, 1992 The effects of selection on linkage analysis for quantitative traits. Genetics **132:** 1177–1185.

MARTINEZ, O., and R. N. CURNOW, 1992 Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. **85:** 480–488.

SAS INSTITUTE INC., 1985 *SAS User's Guide: Statistics*, Version 5. Cary, N.C.

VAN DER BEEK, S., J. A. M. VAN ARENDONK and A. F. GROEN, 1995 Power of two- and three-generation QTL mapping experiments in an outbred population containing full-sib or half-sib families. Theor. Appl. Genet. **91:** 1115–1124.

VAN DER WERF, J. H. J., and W. DE BOER, 1989 Estimation of genetic parameters in a crossbred population of black and white dairy cattle. J. Dairy Sci. **72:** 2615–2623.

VAN RADEN, P. M., and G. R. WIGGANS, 1991 Derivation, Calculation, and use of National Animal Model Information. J. Dairy Sci. **74:** 2737–2746.

VILKKI, H. J., D. J. DEKONING, K. ELO, R. VELMALA, and A. MAKI-TANILA, 1996. Multiple marker mapping of quantitative trait loci of Finnish dairy cattle by regression. J. Dairy Sci. (in press).

WELLER, J. I., 1986 Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics **42:** 627–640.

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.

ZENG, Z. B., 1993 Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA **90:** 10972–10976.

ZENG, Z. B., 1994 Precision mapping of quantitative trait loci. Genetics **136:** 1457–1468.

Communicating editor: B. S. WEIR