# Mapping Quantitative Trait Loci for Milk Production and Health of Dairy Cattle in a Large Outbred Pedigree

**Qin Zhang,* Didier Boichard,† Ina Hoeschele,* Cynthia Ernst,‡ Andre Eggen,‡ B. Murkve,‡ Margaret Pfister-Genskow,‡ LaRee A. Witte,‡ Fernando E. Grignola,* Pekka Uimari,* Georg Thaller* and Michael D. Bishop‡**

**Department of Dairy Science, Virginia Polytechnic Institute, Blacksburg, Virginia 24061-0315, †Station de Genetique Quantitative et Appliquee, Institut National de la Richerche Agronomique, 78352 Jouy-en-Josas, France and ‡ABS Global Inc., De Forrest, Wisconsin 53532*

Manuscript received December 21, 1997
Accepted for publication May 15, 1998

ABSTRACT

Quantitative trait loci (QTL) affecting milk production and health of dairy cattle were mapped in a very large Holstein granddaughter design. The analysis included 1794 sons of 14 sires and 206 genetic markers distributed across all 29 autosomes and flanking an estimated 2497 autosomal cM using Kosambi's mapping function. All families were analyzed jointly with least-squares (LS) and variance components (VC) methods. A total of 6 QTL exceeding approximate experiment-wise significance thresholds, 24 QTL exceeding suggestive thresholds, and 34 QTL exceeding chromosome-wise thresholds were identified. Significance thresholds were determined via data permutation (for LS analysis) and chi-square distribution (for VC analysis). The average bootstrap confidence interval for the experiment-wise significant QTL was 48 cM. Some chromosomes harbored QTL affecting several traits, and these were always in coupling phase, defined by consistency with genetic correlations among traits. Chromosome *17* likely harbors 2 QTL affecting milk yield, and some other chromosomes showed some evidence for 2 linked QTL affecting the same trait. In each of these cases, the 2 QTL were in repulsion phase in those families appearing to be heterozygous for both QTL, a finding which supports the build-up of linkage disequilibrium due to selection.

DAIRY cattle and other livestock species have undergone selection with the goal of improving economically important traits for a number of generations. Most traits of economic importance are of quantitative nature, *i.e.*, are influenced by many genes and by environmental factors. Selection has solely relied upon the collection and utilization of phenotypic and pedigree data, and on statistical tools for partitioning the phenotypic performances of individuals into their additive genetic values plus environmental contributions. At present, major collaborative projects are underway to map genes affecting traits of economic importance in several livestock species, using moderate resolution genetic marker maps. The collaborations are producing genetic maps and genotypes on the one hand and suitable statistical methods for analysis of these data on the other hand. There have been substantial advances both in the map densities and in the development of statistical methods. The latter are needed for QTL (quantitative trait loci) mapping, for genetic parameter estimation (*e.g.*, variance contributions at individual QTL), and for the estimation of additive genetic values by combining phenotypic, pedigree, and genetic marker information. The benefits resulting from the mapping collaborations include the gaining of basic scientific knowledge about the genetic basis of quantitative traits, the achievement of the necessary first step toward fine-mapping and function evaluation of important QTL, and, in the context of livestock improvement, an increase in the selection efficiency for production and health-related traits through marker-assisted selection (MAS).

The granddaughter design (GDD; Geldermann 1975; Weller *et al.* 1990; Georges *et al.* 1995) is sufficiently powerful for the detection of moderate QTL, provided that it consists of multiple grandsire families with a total of several hundred sons in these families. Georges *et al.* (1995) have previously mapped QTL in a large granddaughter design for US Holstein dairy cattle. Since the completion of their work, the same granddaughter design has been expanded to include more progeny, with 276 additional sons in the same families, two additional phenotypic traits (somatic cell score, length of productive life), and genotypes for additional markers, with all markers covering nearly 2500 cM or 100% instead of 1645 cM or 66% of the estimated length of the male genome in cattle (Logue and Harvey 1978; cited in Georges *et al.* 1995). While in the study of Georges *et al.* (1995) each of the fourteen families was analyzed separately by maximum likelihood, here we analyzed all

*Corresponding author:* I. Hoeschele, Department of Dairy Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24060. E-mail: inah@vt.edu

families jointly using two approaches, least-squares (LS) and variance components (VC) analysis. An advantage of analyzing each family separately is that there is no need for assuming a particular genetic model specifying the number of alleles at a QTL, which may be incorrect. Maximum likelihood analysis across families has been based on the assumption of a biallelic QTL (*e.g.*, Weller 1986) although other models of QTL variation should be assumed (Hoeschele *et al.* 1997). The VC analysis by residual maximum likelihood (Grignola *et al.* 1994, 1996a,b, 1997; Xu and Atchley 1995) permits an analysis across families without the need for estimating allele or genotype frequencies. Hence, the VC analysis, where QTL allelic effects have a prior normal distribution, requires fewer parametric assumptions than standard maximum likelihood, where the effects at a biallelic QTL are treated as fixed. While the VC analysis has been tested with simulated data and shown to be robust to the number of alleles at a QTL (Grignola *et al.* 1994, 1996a,b, 1997; Xu and Atchley 1995), in this study it is applied for the first time to real data in a large granddaughter design and its results are compared to those from least-squares analysis, which has been used previously to map QTL in actual data (*e.g.*, Spelman *et al.* 1996; Vilkki *et al.* 1997).

## MATERIALS AND METHODS

**Experimental design:** The granddaughter design consisted of 14 paternal half-sib families with an average number of 128 and a range of 33 to 313 sons per sire, with sire denoting the male parent common to each half-sibship. The total number of genotyped sons was 1794. The sires have more sons than those included in the analysis because semen samples from many sons that were culled after the progeny test were discarded, and genotyping was not possible. Georges *et al.* (1995) showed that for production traits the average phenotype of all sons in one of the families was clearly lower than the average of those sons with marker information. Although pedigree information was available, *i.e.*, there were (considerable) relationship ties among the 14 sires, these sires were assumed to be unrelated in the analyses for several reasons. First, previous simulation studies have shown that inclusion of relationships among sires has only minor effects on power of QTL detection and QTL parameter estimation in a GDD with similarly large numbers of sons per family (Grignola *et al.* 1996b). Second, the marker genotypes in the original granddaughter design were coded within the sire family, and the original allele sizes were not available. Third, many common ancestors of the sires do not have any genotype information at this time.

**Marker data:** Marker data were available on 246 microsatellite markers. Of these, 222 were assigned to the 29 autosomal chromosomes, while 24 markers were not assigned to a chromosome. Most likely orders of and recombination rates among markers were estimated with the CRI-MAP program (P. Green, unpublished results). Because on some chromosomes, 2 to 3 markers were located at the same position, the most informative of these markers within each family was chosen for the analysis, or these were treated as one locus. Hence, only 206 marker positions were actually used in the analyses. Using Kosambi's mapping function and summing over all linkage groups, we obtained a total of 2497 autosomal cM flanked by linked markers, which represents the estimated genetic length of the entire male genome of 2500 cM (Logue and Harvey 1978; cited in Georges *et al.* 1995). Because we employed Haldane's function to map QTL relative to the markers, the map was reconstructed using this mapping function. The Haldane map was 3104 cM in length. The average length of the linkage groups was 107 (88) cM, with a range from 26 (23) cM on chromosome *25* to 179 (138) cM on chromosome *1* (values in parentheses are for the Kosambi function). The average length of the marker intervals was 17.8 (14.3) cM. Heterozygosity measured as the percentage of sires heterozygous at a marker was 46.1% on average across all 206 markers, compared with 45.8% for the markers used in the previous study (Georges *et al.* 1995). The average number of markers per chromosome was 7.83, with a range from 4 to 10. However, the average number of informative markers per chromosome and sire was only 2.81, with a range from 1.36 to 4.64 across chromosomes (averaged over sires). Table 1 contains for each chromosome the estimated genetic length, the number of markers used, and the number of informative markers within each sire family averaged across sires.

**TABLE 1**

**Length (in Morgans, using Haldane's function) of linkage group and number of markers on 29 chromosomes and average number of informative markers across sires**

| Chromosome | Length of linkage group | No. of markers[a] | Average no. of inform. markers |
|---|---|---|---|
| *1* | 1.795 | 9 | 4.64 |
| *2* | 1.597 | 10 | 4.50 |
| *3* | 1.300 | 5 | 2.36 |
| *4* | 1.687 | 9 | 3.79 |
| *5* | 1.569 | 7 | 2.50 |
| *6* | 1.317 | 9 | 3.21 |
| *7* | 1.162 | 8 | 3.71 |
| *8* | 0.712 | 9 | 4.36 |
| *9* | 1.032 | 8 | 3.71 |
| *10* | 1.259 | 8 | 3.07 |
| *11* | 0.737 | 6 | 1.64 |
| *12* | 1.473 | 6 | 2.86 |
| *13* | 1.404 | 8 | 3.00 |
| *14* | 1.196 | 7 | 2.71 |
| *15* | 1.149 | 7 | 2.14 |
| *16* | 1.126 | 7 | 2.79 |
| *17* | 1.165 | 5 | 2.07 |
| *18* | 1.296 | 6 | 2.86 |
| *19* | 1.209 | 8 | 1.93 |
| *20* | 0.333 | 5 | 2.93 |
| *21* | 0.798 | 8 | 3.36 |
| *22* | 0.899 | 7 | 2.07 |
| *23* | 0.896 | 7 | 2.43 |
| *24* | 0.704 | 7 | 2.14 |
| *25* | 0.257 | 4 | 1.36 |
| *26* | 0.879 | 8 | 3.00 |
| *27* | 0.637 | 4 | 1.64 |
| *28* | 0.732 | 7 | 2.14 |
| *29* | 0.718 | 7 | 2.57 |
| Total | 31.038 | 206 | |
| Ave. | 1.070 | 7.1 | 2.81 |

[a] Number of markers actually used in the analysis.

**Phenotypic data:** Seven traits [MY, milk yield; FY, fat yield; PY, protein yield; F%, fat percentage; P%, protein percentage;

PL, length of productive life (VanRaden and Klaaskate 1993); and SCS, somatic cell score (Schutz 1994)] were considered in the analysis. The phenotypic unit of measurement was daughter yield deviation (DYD; VanRaden and Wiggans 1991), which is the average of the phenotypes of the daughters of a son adjusted for systematic environmental effects and the additive genetic values of the daughters' dams. We analyzed DYDs, because a simulation study had shown that there is virtually no difference in the accuracy of estimation of QTL locations and variance contributions when analyzing DYD or phenotypes of individual daughters with the VC method (Q. Zhang and I. Hoeschele, unpublished results). Also available were the numbers of daughters per son and trait, and the reliabilities (REL) for the estimated additive genetic values of the sons and their sires and dams, with REL representing the squared correlation between true and predicted values. Average REL and number of daughters by trait were 0.85 and 1200 for MY, FY and F%, 0.83 and 1146 for PY and P%, 0.67 and 1048 for SCS, and 0.63 and 1058 for PL. All the phenotypic information was provided by the Animal Improvement Programs Laboratory of the Agricultural Research Service of the U.S. Department of Agriculture, Beltsville, MD.

**Statistical methods:** The statistical methods employed for QTL mapping were least-squares (LS) analysis and variance components (VC) analysis. One chromosome was analyzed at a time, using all markers available on that chromosome and with the genetic variation contributed by all other chromosomes accounted for via polygenic effects in the VC analyses or included in the error in the LS analysis.

*Least-squares analysis:* This method is described in detail by Knott *et al.* (1994), Spelman *et al.* (1996), and Uimari *et al.* (1996). The model of analysis was

$$\mathrm{DYD}_{jl} = \mu + s_j + \sum_{i=1}^{t} b_{ikj} P_{ikjl} + e_{jl}, \qquad (1)$$

where $\mu$ is an overall mean, $s_j$ is fixed effect of sire $j$, $b_{iki}$ is the regression coefficient for QTL $i$ nested within sire $j$ at QTL position $k_i$, $P_{ik}$ is the probability of inheriting a QTL allele from sire $j$ for son $l$ and for QTL $i$ at position $k_i$, $t$ ($t = 1$ or 2) is the number of QTL fitted, and $e_{jl}$ is a residual with variance approximately equal to $\sigma_e^2/\mathrm{REL}_{jl}$, where $\mathrm{REL}_{jl}$ is reliability of son $jl$ due to his daughters only, which can be computed as described by VanRaden and Wiggans (1991) and Georges *et al.* (1995). A weighted LS analysis was conducted, with the weights equal to $1/\mathrm{REL}_{jl}$.

Using model (1) with $t = 1$, an "F" statistic for testing $H_0$ ("all $b$'s are zero") versus $H_A$ ("some $b$'s are nonzero") was calculated using the standard type III sums of squares at QTL position intervals of 1 cM on a chromosome. The calculations were conditional on the most likely linkage phase of the sires (see below), and were computed using all markers on the chromosome simultaneously and by including not only offspring with known but also those with unknown marker allelic inheritance. The distribution of the test statistic was obtained empirically using data permutation as described below. It happened rarely that a sire did not have any informative markers for a particular chromosome. This sire family was then deleted for analyses of that chromosome.

For $t = 2$ (two-QTL model), a two-dimensional search was performed, *i.e.*, all combinations of the positions of the two QTL (in 1-cM intervals) were evaluated. However, to ensure estimability of both QTL positions and regression coefficients (Zeng 1993; Whittaker *et al.* 1996), only those combinations of QTL positions that were separated by two markers were considered. For any combination, those sire families that did not have an informative, empty marker interval between the two QTL were discarded. Tests of one QTL versus two QTL were performed by computing the F statistic ($\mathrm{SSE}_{\mathrm{reduced}} - \mathrm{SSE}_{\mathrm{full}}$)/($q\mathrm{MSE}_{\mathrm{full}}$), where SSE (MSE) is the residual sum of squares (mean-square) at the two positions yielding the smallest $\mathrm{MSE}_{\mathrm{full}}$ under the two-QTL model, and $q$ is the number of sire families used for this set of QTL positions. In the reduced model, the regression coefficients for the first or the second QTL were set to zero, and both resulting test statistics were used in an intersection-union test (Berger 1997; Grignola *et al.* 1997). Rejection of the null hypothesis of the one-QTL model required both test statistics to be significant.

To investigate a potential increase in power for detecting QTL, a few additional analyses were conducted, where a chromosome was searched for a single QTL while fitting a second QTL on another chromosome at the position with the highest significance among all experiment-wise significant (see below) QTL positions for the same trait, which were previously identified in the one-QTL analyses. In these analyses, sire families with no informative markers on either of the two chromosomes were discarded.

*Variance components analysis:* The VC analysis is described in detail by Grignola *et al.* (1996a,b) for the single-QTL model and for the two-QTL model by Grignola *et al.* (1997), who referred to it as an approximate Residual Maximum Likelihood (AREML) method. For analysis of the two-QTL model, a more efficient search strategy than the two-dimensional search in LS based on cyclic optimization of the QTL positions was employed. In cyclic optimization, the first QTL was fixed at its current most likely position while the position of the second QTL was optimized, subsequently the second QTL was fixed while the position of the first QTL was optimized, etc. Testing for linkage (one versus no QTL on the chromosome) was based on a likelihood ratio test as described in Grignola *et al.* (1996a), and testing for one *vs.* two QTL was performed as described by Grignola *et al.* (1997) and based on the intersection-union test as outlined for the LS analysis. The model for DYD in the VC analysis was

$$\mathrm{DYD}_{jl} = \mu + u_{jl} + \sum_{i=1}^{t} (v_{ijl}^{1} + v_{ijl}^{2}) + e_{jl},$$

$$\mathrm{var}(e_{jl}) = \frac{1 - \mathrm{REL}_{jl}}{\mathrm{REL}_{jl}} \sigma_e^2 = w_{jl}\sigma_e^2, \qquad (2)$$

where $u_{jl}$ is the polygenic effect of son $j$ of sire $l$, and $v_{ijl}^{k}$ is effect of allele $k$ ($k = 1, 2$) at QTL $i$ in son $l$ of sire $j$. As pointed out by Grignola *et al.* (1996b), when DYD is analyzed with the weights $1/w_{jl}$, it has an expected heritability of 0.5. Hence, there is the option of treating heritability as known.

*Linkage phases:* Probabilities of all possible linkage phases of the sires were computed as

$$P(\boldsymbol{G}_i|\boldsymbol{M}) = \frac{P(\boldsymbol{G}_i|\boldsymbol{M}_i) \prod_{j=1}^{n_i} \sum_{\boldsymbol{H}_{ij}^{s}} P(\boldsymbol{H}_{ij}^{s}|\boldsymbol{G}_i, \boldsymbol{M}_{ij}) P(\boldsymbol{H}_{ij}^{d}(\boldsymbol{H}_{ij}^{s}))}{\sum_{\boldsymbol{G}_i} \text{numerator}}, \qquad (3)$$

where $\boldsymbol{G}_i$ is a particular multi-locus, phase-known marker genotype of sire $i$, $\boldsymbol{M}$ is the marker information for the sire and its sons, $\boldsymbol{M}_i$ is the marker information on the sire, $\boldsymbol{H}_{ij}^{s}$ is a multi-locus haplotype that son $j$ inherited from sire $i$, $\boldsymbol{H}_{ij}^{d}$ is the haplotype inherited from the dam that is determined by the haplotype inherited from the sire (dam not genotyped), $\boldsymbol{M}_{ij}$ is the marker information on son $j$ of sire $i$, and $n_i$ is the number of sons of the sire. While the first term in the summation over $\boldsymbol{H}_{ij}^{s}$ in (3) depends on the recombination rates among loci, the second term is a function of the marker allelic frequencies, which were estimated within sire families as described in Georges *et al.* (1995). All multi-locus genotypes and haplotypes include only those loci for which the sire is heterozygous.

*Significance thresholds:* QTL findings from the one-QTL model analyses across families are reported in three ways, by

listing (1) all locations significant at the chromosome-wise $\alpha_c = 0.05$ type-I error level [with six ($= 116 \times 0.05$) type-I errors expected by chance under the null hypothesis of no QTL segregating, see below], (2) all locations of suggestive significance (with one type-I error expected by chance), and (3) all locations of experiment-wise significance. For experiment-wise significance, the type-I error ($\alpha_c$) for each chromosome by trait combination was determined from the equation $\alpha_{exp} = 1 - (1 - \alpha_c)^n$ (Weir 1990), where $\alpha_{exp}$ represents the experiment-wise type-I error set equal to 0.05, and $n$ is the number of independent tests in the entire experiment. Approximately, $n\alpha_c = 0.05$. A canonical transformation (Weller *et al.* 1996), based on estimates of the genetic correlations taken from the literature (Pearson *et al.* 1990; Schutz *et al.* 1990; Weigel *et al.* 1997), produced four uncorrelated factors that accounted for 94% of the total variation. Hence, the number of independent subexperiments was equal to the number of chromosomes times the number of uncorrelated factors, or $n = 29 \times 4 = 116$, resulting in $\alpha_c = 0.0004421$. A similar approach was used by Spelman *et al.* (1996) and Uimari *et al.* (1996). The type-I error for suggestive significance (Lander and Kruglyak 1995) was determined similarly, but with $\alpha_c$ calculated from the equation $n\alpha_c = 1$, which is the expected number of type-I errors in the experiment when the null hypotheses of no QTL segregating is true, yielding $\alpha_c = 0.008621$. Each subexperiment consisted of multiple dependent tests along a chromosome, with dependence among these tests accounted for via data permutation (see below).

For LS analysis, the significance thresholds were determined by the permutation method of Churchill and Doerge (1994). A total of 100,000 permutations of the DYD and REL of the sons relative to their marker genotypes were performed within sire families for each chromosome by trait combination. From each permutation, the largest value of the test statistic within a chromosome by trait combination was retained. The threshold was then set equal to the $(1 - \alpha_c)$ percentile of these 100,000 values. For VC analysis, performing 100,000 permutations for each chromosome by trait combination was too CPU-time-consuming. Grignola *et al.* (1996b) showed that the distribution of the VC likelihood ratio statistic for zero versus one QTL linked is in between a 1- and a 2-d.f. chi-square distribution. Therefore, we are reporting QTL with test statistics exceeding the 2-d.f. chi-square threshold, as well as QTL with test statistics in between the 1- and 2-d.f. thresholds.

For the tests of one *vs.* two QTL, the significance thresholds in the LS analysis were obtained from the F distribution. For the VC analysis, a chi-square distribution with 1 d.f. was used, as Grignola *et al.* (1997) showed that using the 1-d.f. chi-square distribution made the empirical type-I error level under the null hypothesis of one QTL close to the nominal value, while using 2 or 3 d.f. made it too conservative. No adjustment for multiple testing was performed for these tests.

*Confidence intervals:* Confidence intervals (CIs) for the QTL position were calculated with the LOD drop-off method (Lander and Botstein 1989) and with bootstrapping (Visscher *et al.* 1996). The LOD drop-off method finds the location at either side of the estimated QTL position that corresponds to a decrease of 0.83 units in the LOD score, yielding a 95% CI (Visscher *et al.* 1996). The LS and REML test statistics were transformed to LOD scores for this purpose. For REML, the test statistic is a likelihood ratio (LR), which can be directly transformed to a LOD score. For LS analysis, when errors are independent and normally distributed, the likelihood ratio statistic can be computed as (Aitkin *et al.* 1989)

$$LR = n \ln(SSE_{reduced} / SSE_{full}),$$

where $n$ is the number of observations, and $SSE_{reduced}$ and $SSE_{full}$ are residual sums of squares of the reduced model (no QTL) and the full model (one QTL), respectively.

Bootstrap CIs were calculated for those QTL locations exceeding experiment-wise thresholds. For bootstrapping, estimates of QTL position from $n$ samples (with replacement) were obtained and their 2.5th and 97.5th percentiles determined for an empirical 95% CI. Three bootstrapping methods were employed. In the first method, all samples were used and all families in the original data set were retained. In a second method, only those families showing evidence for segregation of the QTL were retained for sampling. In the third method, all families were used, but only those bootstrap samples where the QTL was significant were retained. Selective bootstrapping has recently been proposed by Lebreton and Visscher (1997).

## RESULTS

**LS analysis:** *Single QTL analysis:* Figures 1 and 2 depict the test statistic profiles for all chromosome $\times$ trait (C $\times$ T) combinations with profiles exceeding or nearly exceeding the experiment-wise and suggestive significance thresholds, respectively, somewhere on the chromosome. The significance thresholds determined by permutation varied among traits, as observed by Spelman *et al.* (1996), and also among chromosomes. Generally, the variation among traits was smaller than that among chromosomes. The profiles in Figure 1 provide strong evidence for the presence of at least one QTL on chromosome (C) *3* for P%, on C*6* for MY, F% and P%, and on C*20* for F% and P%. The C*6* effects on F% and P% may be pleiotropic effects of the same QTL (same estimated map position) or may be due to closely linked QTL. The C*6* effect on MY appears to have a different map position, although bootstrap CIs are wide and overlap (see Table 2). The QTL for F% and P% on C*20* have nearly the same estimated map position, indicating the possibility of a pleiotropic effect. But unexpectedly the effect on the percentage traits was not accompanied by an effect on MY as opposed to the findings for C*6*. The segregation of additional QTL is suggested (Figure 2) for C*4* and SCS, C*9* and FY, C*13* and SCS, C*14* and SCS, C*17* and MY, C*26* and F% and SCS, and C*28* and F% and P%. Maximum and mean values of the estimates of the within-sire QTL substitution effects (regression coefficients) are given in Table 3 for the QTL of experiment-wise significance.

*Two-QTL analyses:* Test statistics and corresponding QTL map positions under the two-QTL model can be found in Table 4 for those C $\times$ T combinations with experiment-wise significance from the single-QTL analysis and for other combinations with both test statistics exceeding the 5% significance threshold for the two-QTL analyses (only C*17* and MY). C*17* probably harbors two QTL for MY located at the two opposite ends of the linkage group. C*6* also showed some evidence for two QTL for F%, but the two QTL positions were very close to each other and the significance was not very

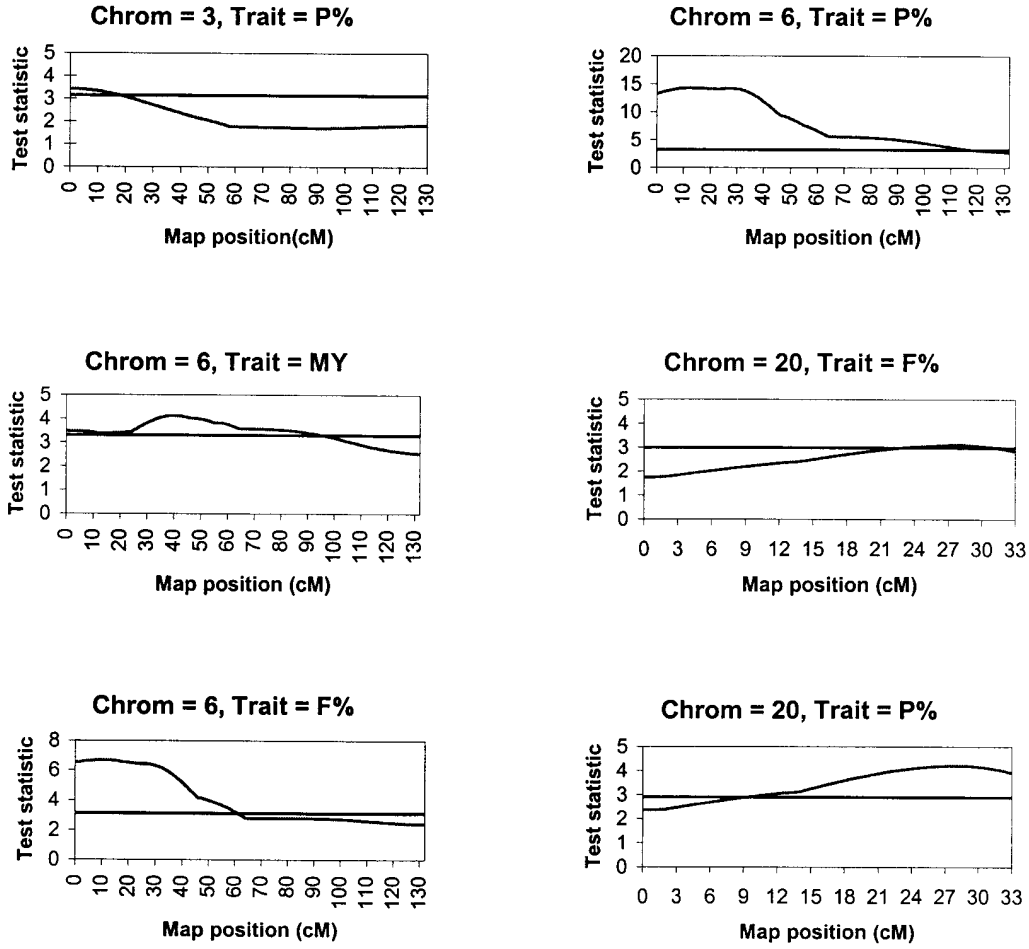


Figure 1.—Test statistic profiles from LS analysis. The straight lines represent experiment-wise significance threshold levels from permutations.

strong. To investigate a potential increase in power for the suggestive QTL, these were reanalyzed by also fitting a second QTL on another chromosome, which had experiment-wise significance in the single-QTL analysis for the same phenotypic trait. Some QTL positions were different from those of the single QTL analysis, possibly because in the two-unlinked-QTL analyses, families without informative markers on either chromosome were discarded. For the same reason, there was no clear increase in the test statistics due to fitting an unlinked QTL of relatively large effect (results not shown).

**VC analysis:** *Single QTL analysis:* Table 5 contains parameter estimates and test statistics from VC analysis with heritability ($h^2$) estimated or fixed at 0.5, and for comparison, LS position estimates and test statistics. The results in Table 5 pertain to those QTL above or near experiment-wise or suggestive thresholds from the LS or VC analyses. Estimates of the QTL parameters $v^2$ (ratio of QTL allelic to total additive genetic variance) and $d$ (QTL map position) and likelihood ratios from both VC analyses (with $h^2$ estimated or fixed) were very similar except for the QTL position on chromosome *28* for trait F%, where the likelihood ratio profile was very flat, and the likelihood ratios at the two positions estimated (0.19 and 0.73) were very similar. The C × T combinations with an experiment-wise significant QTL from the VC analysis were exactly the same as those from the LS analysis. Most of the suggestive findings were also in agreement with those from the LS analysis. There were some minor discrepancies, because for C*1* and PY, for C*9* and MY, and for a few other combinations, LS statistics were not significant while VC statistics were. For C*28* and P%, however, the LS analysis identified a significant QTL, while VC did not.

Likelihood ratio profiles from VC analyses with heritability fixed at 0.5 are depicted in Figures 3 and 4, for the same C × T combinations as those shown in Figures 1 and 2 for LS analyses. The LS and VC estimates of the QTL positions were generally in close agreement (Table 5), a finding that is consistent with the agreement between the LS and VC test statistic profiles, and given the width of the CIs (see Table 2). The largest difference in the estimates for QTL position from the LS and VC analyses was found for chromosome *4* and trait SCS and amounted to 43 − 0 cM (11 cM) = 43 cM (32 cM), with the numbers in parentheses pertaining to VC analysis with heritability fixed.

Table 6 contains test statistics, estimates of QTL locations and estimates of QTL variance contributions from VC analysis for QTL exceeding chromosome-wise 1-d.f.

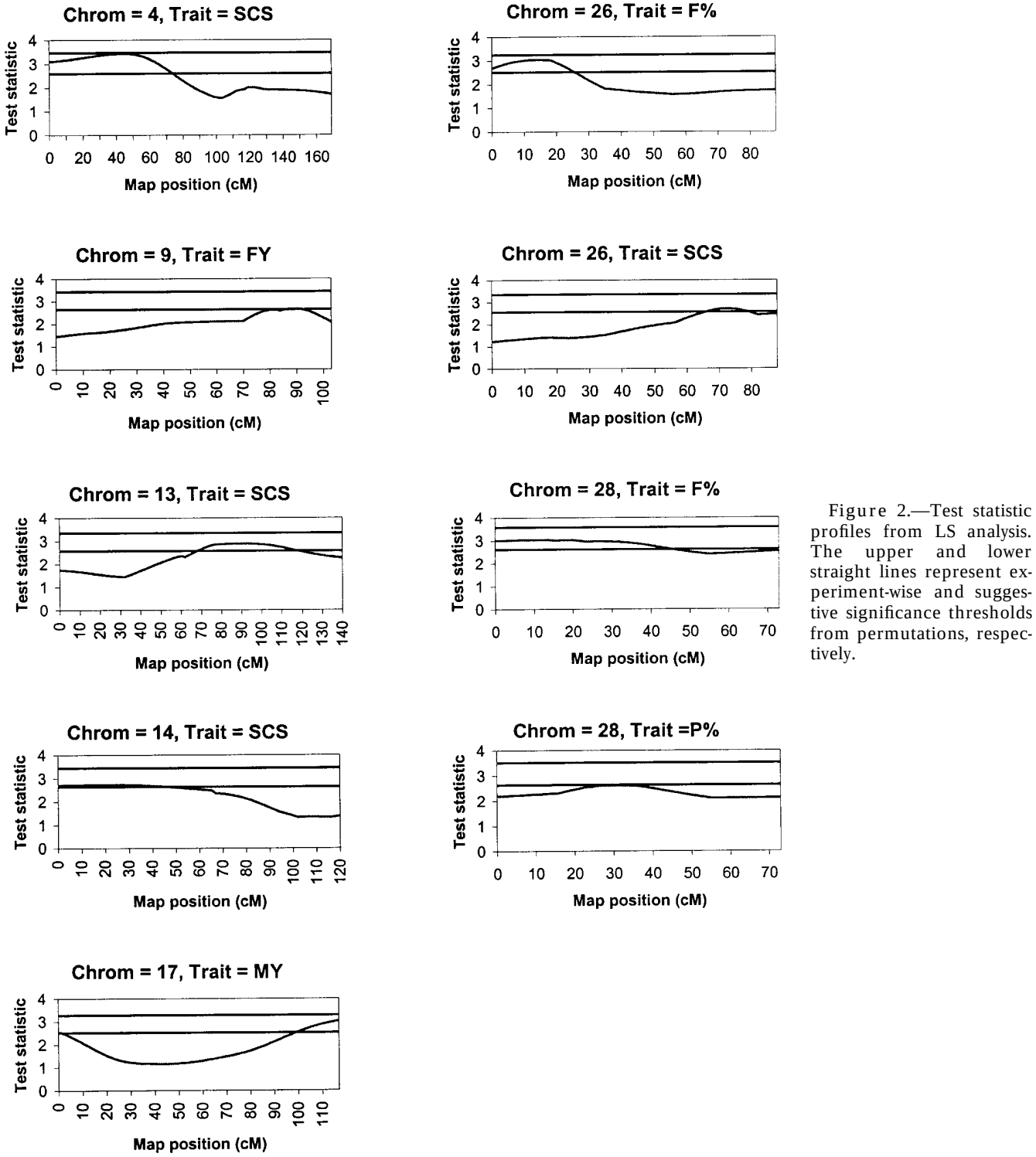


Figure 2.—Test statistic profiles from LS analysis. The upper and lower straight lines represent experiment-wise and suggestive significance thresholds from permutations, respectively.

chi-square thresholds with $\alpha_c = 0.05$ but not reaching suggestive thresholds. Again, for comparison, the corresponding LS test statistics and estimates of QTL locations were also given. Ten such QTL were found with an expected number of six type-I errors under the null hypothesis of no QTL among the 10 + 18 + 6 = 34 QTL positions exceeding the chromosome-wise thresholds, with 18 QTL also exceeding suggestive thresholds and 6 exceeding experiment-wise thresholds.

For those QTL positions with test statistics exceeding experiment-wise significance thresholds, the LOD drop-off and bootstrap confidence intervals with 95% coverage are given in Table 2. The bootstrap CIs are considerably larger than the LOD drop-off ones. The bootstrap CIs in Table 2 were calculated by using all families in the original data and all bootstrap samples. When all samples but only those families that appeared to be segregating in the original data were used, virtually the

**TABLE 2**

**LOD drop-off and bootstrap confidence intervals of QTL positions for QTL above experiment-wise thresholds: first row, LS analysis; second row, VC analysis with heritability fixed at 0.5**

| | | | 95% interval | |
|---|---|---|---|---|
| Chromosome | Trait | QTL position | LOD drop-off | Bootstrap |
| *3* | P% | 0.00 | [0.00, 0.17] | [0.00, 1.30] |
| | | 0.01 | [0.00, 0.16] | [0.00, 0.93] |
| *6* | MY | 0.40 | [0.31, 0.54] | [0.00, 0.86] |
| | | 0.41 | [0.29, 0.63] | [0.02, 0.87] |
| | F% | 0.11 | [0.00, 0.27] | [0.00, 0.33] |
| | | 0.04 | [0.00, 0.24] | [0.00, 0.24] |
| | P% | 0.12 | [0.06, 0.31] | [0.04, 0.31] |
| | | 0.22 | [0.15, 0.29] | [0.00, 0.31] |
| *20* | F% | 0.28 | [0.21, 0.33] | [0.13, 0.33] |
| | | 0.28 | [0.23, 0.31] | [0.04, 0.32] |
| | P% | 0.28 | [0.22, 0.32] | [0.20, 0.33] |
| | | 0.28 | [0.24, 0.31] | [0.07, 0.32] |

same CIs were obtained with the exception of C*3* and P%, where the interval was only nearly half as wide, but still wider than the LOD drop-off interval. When only those bootstrap samples where the QTL was significant were used, again virtually the same intervals were obtained, as the QTL was significant in over 90% of the samples.

Table 7 presents the variance explained by QTL exceeding experiment-wise or suggestive significance thresholds for each trait, obtained by adding the variance estimates for individual QTL in Table 5 for the same trait. The largest fractions of the additive genetic variance attributed to QTL were found for F% and SCS. In a simulation study with a single QTL and analysis across families by the VC method (results not shown), we found that for true QTL variance ranging from 50 to 5% of the additive genetic variance, overestimation of this parameter increased from 0 to 114% for those replicates where the QTL was significant at the experiment-wise threshold. This finding is in agreement with Georges *et al.* (1995), who simulated a single family and used a LOD score of 3. As a consequence, the QTL variance contributions in Table 7 are expected to overestimate the true variance explained by the identified QTL. However, our analysis did not include those sons in the same families that were not genotyped, and most of these sons were culled after progeny test. In this case, an underestimation of QTL variance is expected from theory (Im *et al.* 1989; Mackinnon and Georges 1992) and has been verified via simulation (Mackinnon and Georges 1992).

*Two-QTL analyses:* Test statistics and the corresponding QTL map positions under the two-QTL model are also given in Table 4 together with the results from LS analysis. None of these combinations seemed to have a second significant QTL, with the exception of the two MY QTL on chromosome *17*. There was good agreement in the QTL positions under the two-QTL model between the LS and VC analyses, except for C*6* and F% and C*6* and P%. The VC analysis was also run with a

**TABLE 3**

**Estimates of within-sire QTL substitution effects (*b*) and their standard errors (in parentheses) from LS analysis**

| | | Maximum significant $\|b\|$ | | Mean significant $\|b\|$ | | No. of significant families[b] |
|---|---|---|---|---|---|---|
| Chromosome | Trait | $\|b\|$ | $\|b\|/\sigma_g$[a] | $\|b\|$ | $\|b\|/\sigma_g$[a] | |
| *3* | P% | 0.047 (0.018) | 0.222 (0.085) | 0.038 | 0.179 | 3 |
| *6* | MY | 1699 (658) | 1.188 (0.460) | 780 | 0.546 | 5 |
| | F% | 0.145 (0.021) | 0.489 (0.071) | 0.132 | 0.445 | 2 |
| | P% | 0.090 (0.009) | 0.424 (0.042) | 0.059 | 0.278 | 4 |
| *20* | F% | 0.092 (0.048) | 0.310 (0.162) | 0.069 | 0.231 | 4 |
| | P% | 0.046 (0.021) | 0.217 (0.099) | 0.034 | 0.162 | 7 |

[a] $\sigma_g$: Additive genetic standard deviation (MY:1430, FY:53, PY:41, F%:0.2967, P%:0.2121).
[b] Total number of families was 14 for chromosomes *3*, *6*, and *20*.

**TABLE 4**

**Tests of one QTL versus two QTL from least-squares analysis (first row) and VC analysis with heritability fixed at 0.5 (second row)**

| Chromosome | Trait | Positions (M) | | Test statistics | | No. of sire families |
|---|---|---|---|---|---|---|
| | | QTL 1 | QTL 2 | QTL 1 | QTL 2 | |
| *3* | P% | 0.00 | 0.58 | 3.60* | 1.24 | 10 |
| | | 0.00 | 0.46 | 15.41* | 1.49 | |
| *6* | MY | 0.45 | 0.63 | 1.61 | 0.73 | 7 |
| | | 0.05 | 0.60 | 6.16* | 3.56 | |
| | F% | 0.00 | 0.12 | 2.60* | 2.96* | 5 |
| | | 0.04 | 0.65 | 49.50* | 0.09 | |
| | P% | 0.32 | 0.57 | 16.01* | 2.13 | 7 |
| | | 0.05 | 0.27 | 1.30 | 6.27* | |
| *20* | F% | 0.13 | 0.28 | 1.22 | 2.12 | 12 |
| | | 0.00 | 0.28 | 0.01 | 14.20* | |
| | P% | 0.02 | 0.28 | 0.99 | 3.18* | 12 |
| | | 0.03 | 0.28 | 0.66 | 18.59* | |
| *17* | MY | 0.00 | 1.08 | 4.25* | 4.71* | 7 |
| | | 0.00 | 1.16 | 7.16* | 14.39* | |

* Above F-threshold (LS analysis) or $\chi^2$ (d.f. = 1)-threshold (VC analysis) at 0.05 level.

second, unlinked and experiment-wise significant QTL fitted. Results (not reported here) were similar to the LS findings of very little increase in the test statistics.

## DISCUSSION

The breakthrough study of Georges *et al.* (1995) demonstrated that QTL affecting milk production can be identified in current populations using a granddaughter design. In this study, we have analyzed the same data set augmented by additional sons, additional traits, and additional markers. While Georges *et al.* (1995) analyzed 1518 sons and 159 markers with 138 markers assigned to 27 linkage groups, an estimated coverage of 1645 cM and an average bracket size of 14.8 cM (based on the Kosambi function), we analyzed 1794 sons (of the same 14 families) and 206 markers, with several markers assigned to each of the 29 autosomes, an estimated coverage of 2497 cM (using Kosambi's mapping function), and an average bracket size of 14.3 cM (Kosambi). There is also a major difference in the statistical methods employed for QTL mapping. Georges *et al.* (1995) used maximum likelihood with fixed QTL effects, analyzed each family separately, and employed a rather stringent significance threshold, not accounting, however, for the actual marker map and correlation structure of the phenotypic traits analyzed. Here, we analyzed all families jointly using least-squares and variance components methods. We also used data permutation techniques and canonical transformation to determine an "effective number of independent traits," leading to (approximate) experiment-wise significance thresholds accounting for multiple tests per chromosome, multiple chromosomes, and multiple traits. We computed and compared confidence intervals with LOD drop-off and bootstrap, and estimated the variances explained by individual QTL.

The VC analysis with random QTL allelic effects requires fewer parametric assumptions than an ML analysis across families with fixed QTL effects, as the number of alleles at a QTL does not need to be specified, and there is no need for estimating allelic or genotypic frequencies. The VC method is therefore particularly suited for analyses of segregating livestock or human pedigrees.

In this study, LS and VC analyses gave similar estimates of QTL locations. This finding is to be expected for a half-sib design with very large families, similar to the one analyzed here. An advantage of the LS analysis is that it is computationally feasible to perform a permutation test so that the distribution of test statistics can be determined empirically. Advantages of the VC analysis are that it provides an estimate of the additive genetic variance in the population attributable to a QTL, rather than only estimates of QTL substitution effects for specific sires, and that it is applicable to any design or pedigree. The VC analysis is therefore capable of utilizing all the available information, rather than only the half-sib relationships, for example. Since the LS analysis treats gene substitution effects as fixed rather than random as in the VC analysis, the estimates are sufficiently accurate only in very large families. Estimates of gene substitution or allelic effects and of QTL variance contribution are important for marker-assisted selection.

Our analysis of 29 chromosomes and 7 quantitative traits of economic importance in a very large US Holstein granddaughter design reveals 6 QTL with experi-

**TABLE 5**

**Test statistics and estimates of QTL positions from LS and VC analyses and estimates of QTL variance ratios from VC analysis for QTL above or near experiment-wise or suggestive thresholds**

| | | Test statistic | | QTL positions (M) | | |
|---|---|---|---|---|---|---|
| Chromosome | Trait | LS[a] | VC[b] | Positions[c] | Marker bracket[d] | Var. ratio[e] |
| *3* | P% | 3.44* | 18.89* | 0.00 | TGLA263(0.00) | 0.054 |
| | | | 20.55* | 0.00/0.00 | AGLA247(0.47) | 0.059 |
| *6* | MY | 4.13* | 19.13* | 0.40 | BM143(0.24) | 0.048 |
| | | | 21.69* | 0.41/0.40 | RM28(0.46) | 0.054 |
| | F% | 6.69* | 51.48* | 0.11 | BM1329(0.01) | 0.124 |
| | | | 58.13* | 0.04/0.04 | TGLA37(0.12) | 0.139 |
| | P% | 14.29* | 136.13* | 0.12 | TGLA37(0.12) | 0.129 |
| | | | 147.09* | 0.23/0.22 | BM143(0.24) | 0.138 |
| *20* | F% | 3.12* | 18.57* | 0.28 | TGLA126(0.14) | 0.045 |
| | | | 19.65* | 0.28/0.28 | TGLA153(0.28) | 0.047 |
| | P% | 4.21* | 29.97* | 0.28 | | 0.075 |
| | | | 26.88* | 0.28/0.28 | | 0.068 |
| *1* | PY | 2.22 | 9.57** | 0.00 | AGLA17(0.00) | 0.062 |
| | | | 9.93** | 0.21/0.27 | TGLA57(0.89) | 0.092 |
| *2* | F% | 2.76 | 7.28**** | 0.46 | TGLA377(0.37) | 0.028 |
| | | | 7.21**** | 0.41/0.42 | ETH121(0.49) | 0.031 |
| *3* | MY | 2.18 | 8.96**** | 0.07 | TGLA263(0.00) | 0.041 |
| | | | 7.74**** | 0.10/0.08 | TGLA247(0.47) | 0.040 |
| *4* | SCS | 3.44** | 13.86*** | 0.43 | RM188(0.00) | 0.272 |
| | | | 12.07** | 0.00/0.11 | TGLA116(0.46) | 0.158 |
| *6* | FY | 2.41 | 6.85 | 0.24 | TGLA37(0.12) | 0.026 |
| | | | 8.06**** | 0.12/0.12 | BM143(0.24) | 0.031 |
| *9* | MY | 2.47 | 9.52** | 0.83 | TGLA427(0.77) | 0.029 |
| | | | 12.94*** | 0.90/0.90 | TGLA73(0.91) | 0.039 |
| | FY | 2.67** | 6.46 | 0.89 | | 0.030 |
| | | | 7.30**** | 0.90/0.80 | | 0.036 |
| | PY | 2.51 | 10.39** | 0.91 | | 0.036 |
| | | | 13.65*** | 0.90/0.90 | | 0.048 |
| *13* | SCS | 2.89** | 7.26**** | 0.91 | TGLA381(0.78) | 0.054 |
| | | | 10.05** | 0.91/0.82 | AGLA232(1.40) | 0.059 |
| *14* | F% | 2.55 | 12.32** | 0.00 | ILSTS11(0.00) | 0.237 |
| | | | 11.67** | 0.00/0.00 | BM302(0.65) | 0.219 |
| | FY | 2.25 | 9.16**** | 0.00 | | 0.211 |
| | | | 9.44**** | 0.00/0.00 | | 0.207 |
| | SCS | 2.74** | 14.91*** | 0.21 | ILSTS11(0.00) | 0.149 |
| | | | 14.09*** | 0.35/0.30 | BM302(0.65) | 0.152 |
| *17* | MY | 3.02** | 12.71*** | 1.17 | TGLA170(0.68) | 0.123 |
| | | | 11.51** | 1.16/1.16 | TGLA322(1.17) | 0.125 |
| *23* | FY | 2.39 | 9.53** | 0.66 | MGTG7(0.65) | 0.027 |
| | | | 9.45**** | 0.65/0.65 | AGLA212(0.74) | 0.033 |
| *26* | F% | 3.04** | 13.12*** | 0.15 | TGLA22(0.00) | 0.036 |
| | | | 13.29*** | 0.14/0.14 | BM4505(0.56) | 0.038 |
| | SCS | 2.70** | 6.94**** | 0.72 | TGLA429(0.71) | 0.030 |
| | | | 10.56** | 0.72/0.71 | BM804(0.82) | 0.044 |
| *28* | F% | 3.04** | 7.61**** | 0.12 | TGLA82(0.12) | 0.029 |
| | | | 6.86 | 0.19/0.73 | TGLA306(0.37) | 0.037 |
| | P% | 2.62** | 3.91 | 0.31 | | 0.021 |
| | | | 3.20 | 0.22/0.21 | | 0.019 |

[a] Threshold values were from permutation test.

[b] Threshold values were from chi-square distribution with 1 or 2 d.f. First row, heritability estimated; second row, heritability fixed at 0.5.

[c] Distances from origin of linkage groups: first row, LS estimates; second row, VC estimates (heritability estimated/fixed at 0.5).

[d] Pair of markers from published maps (BovMAP, MARC) flanking the estimated QTL position; values in parentheses are distances from origin of linkage groups.

[e] Ratio of QTL allelic to additive genetic variance for VC analysis. First row, heritability estimated; second row, heritability fixed at 0.5.

* Above experiment-wise significance threshold at 0.05 level.

** Above suggestive significance threshold.

*** Between 1 d.f. and 2 d.f. $\chi^2$ experiment-wise significance thresholds.

**** Between 1 d.f. and 2 d.f. $\chi^2$ suggestive significance thresholds.

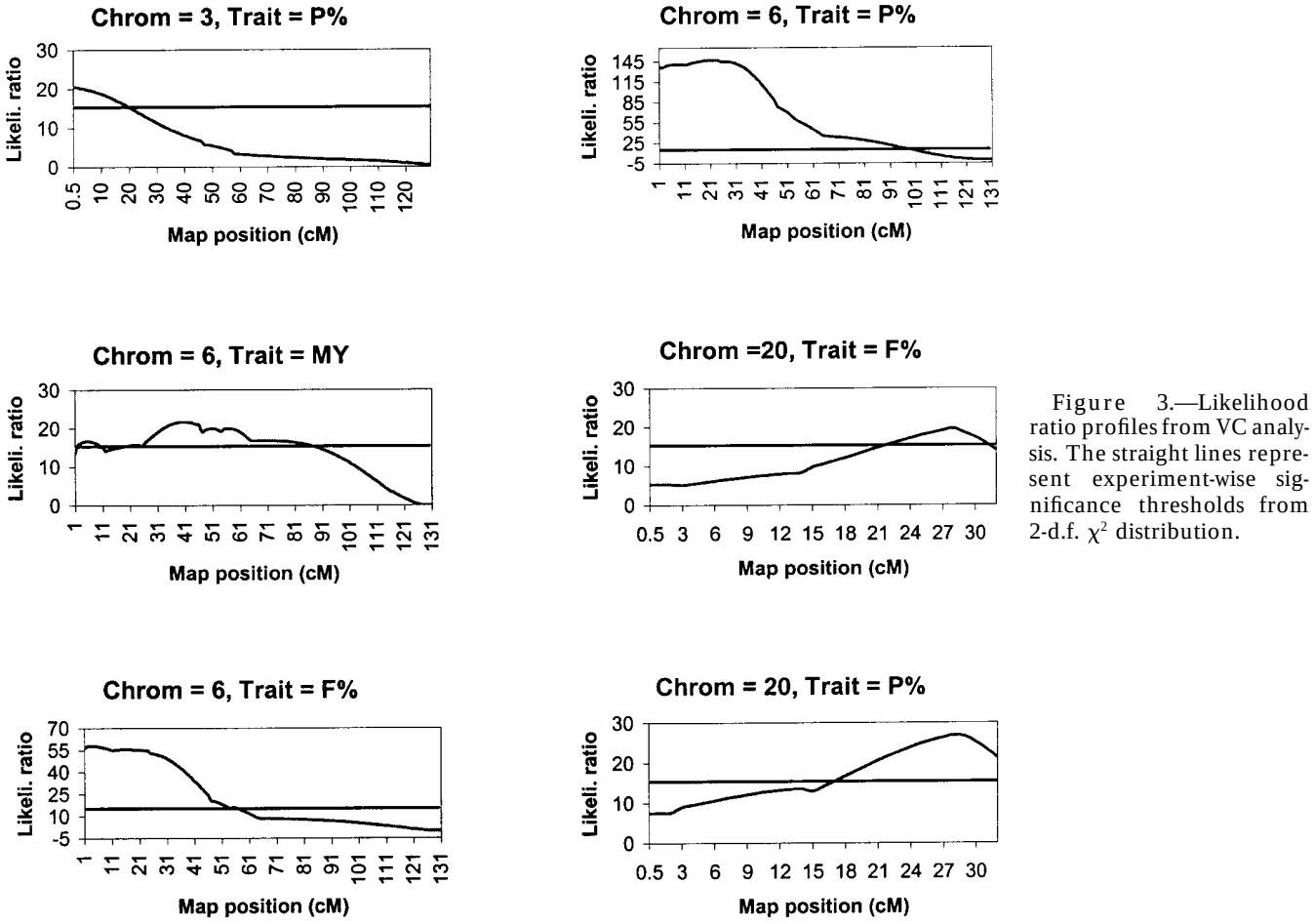


Figure 3.—Likelihood ratio profiles from VC analysis. The straight lines represent experiment-wise significance thresholds from 2-d.f. $\chi^2$ distribution.

ment-wise significance, 18 suggestive QTL locations (1 expected by chance out of 24), and 10 QTL achieving chromosome-wise significance (6 expected by chance out of 34). These findings provide very strong evidence for the segregation of QTL in our pedigree, despite the coarse nature of the current marker map. Estimates of QTL positions were consistent across both methods of analysis. However, CI were large, and bootstrap CIs were considerably larger than LOD drop-off CIs, emphasizing the need for using the bootstrap. For VC analysis with heritability fixed at 0.5 and for those QTL surpassing the experiment-wise significance thresholds, the average bootstrap CI (95%) was 47.7 cM (Haldane), the average LOD drop-off CI was 17.2 cM, and the minimum and maximum bootstrap CIs were 25 and 93 cM, respectively, using Haldane's mapping function.

This analysis involved the problem of multiple correlated testing. We followed the approach of Spelman *et al.* (1996) and Uimari *et al.* (1996) to transform 7 × 29 multiple correlated tests to 4 × 29 independent tests. Very recently, Weller *et al.* (1997) studied the empirical type-I error rates in QTL mapping for multiple correlated traits and found that in their case the empirical type-I error rates for seven correlated traits were in general between the theoretical type-I error rates for six and seven uncorrelated traits. We checked our results with thresholds calculated by assuming seven uncorrelated traits. The significance for the experiment-wise QTL remained the same, but four suggestive significant QTL became nonsignificant at the suggestive level. These are C*2* and F%, C*9* and FY, C*28* and F%, and C*28* and P%.

The current marker structure does not yet permit precision mapping of QTL for several reasons. First, the average CI (bootstrap) was much wider than a desired range of 10 to 20 cM. Second, with the exception of chromosome *17*, none of the analyses under the two-linked-QTL model revealed a significant second QTL, most likely because the number of families with several informative markers per chromosome was too small (for LS analysis, two informative markers are needed between the QTL positions to ensure the estimability of QTL positions and effects; Zeng 1993; Whittaker *et al.* 1996; therefore, those sire families, which did not have an informative, empty marker interval between the two QTL, were discarded). Future genotyping efforts must concentrate on placing more informative markers in the regions identified to harbor QTL and on increasing the number of individuals genotyped. Only after CIs have been reduced to the desired 10- to 20-cM re-

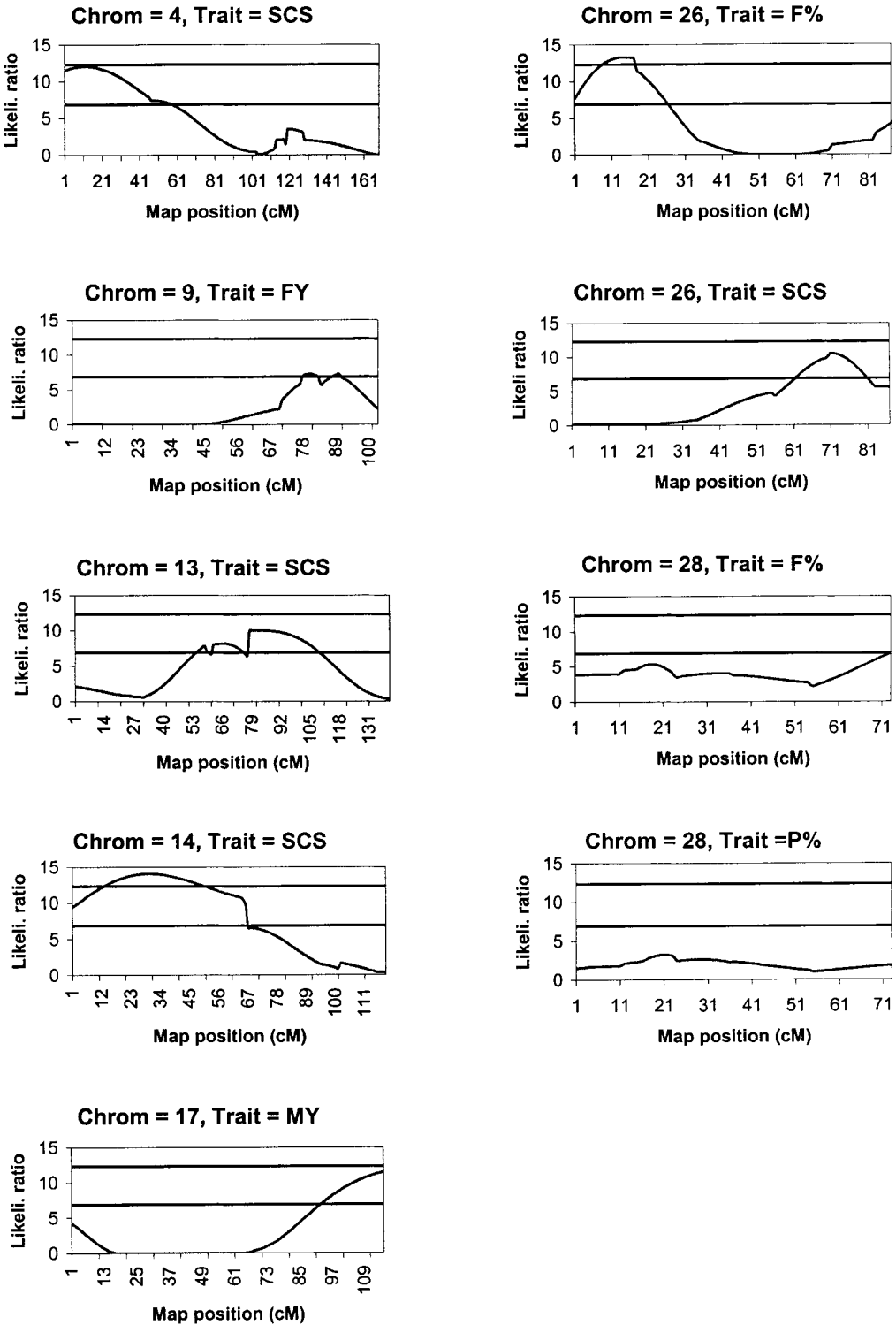


Figure 4.—Likelihood ratio profiles from VC analysis. The upper and lower straight lines represent experiment-wise and suggestive thresholds from 1-d.f. $\chi^2$, respectively.

gion, the next step of saturating a region with highly informative markers to fine-map and clone a QTL can be undertaken.

The analyses conducted here represent an initial genome scan and could be improved upon in several ways. We expect to reanalyze the most promising regions in the future, in particular after additional informative markers have been added to the data. Improved analyses will include the fitting of multiple linked QTL and accounting for heterogeneous variances within families, due to different unlinked QTL segregating in these families, by fitting residual variances within families or by fitting multiple unlinked QTL. Further improvements of the analysis might result from including those sons that have not been genotyped (their data were not available to us), and from conducting a full pedigree

**TABLE 6**

**Estimates of QTL positions and test statistics for LS and VC analyses (heritability fixed at 0.5) and estimates of QTL variance ratios from VC analysis for QTL above thresholds for 0.05 chromosome-wise type-I error level**

| | | Test statistic | | QTL positions | | |
|---|---|---|---|---|---|---|
| Chromosome | Trait | LS[a] | VC[b] | Positions[c] | Marker bracket[d] | Var. ratio[e] |
| *2* | FY | 2.35 | 5.57 | 0.67 | ETH121(0.49) | 0.052 |
| | | | | 0.70 | TGLA226(0.122) | |
| *3* | PY | 2.00 | 5.01 | 0.18 | TGLA263(0.00) | 0.039 |
| | | | | 0.18 | TGLA247(0.47) | |
| *5* | FY | 2.46 | 5.19 | 0.22 | TGLA124(0.08) | 0.031 |
| | | | | 0.28 | AGLA254(0.32) | |
| *8* | F% | 2.34 | 4.48 | 0.58 | TGLA339(0.39) | 0.024 |
| | | | | 0.40 | TGLA341(0.71) | |
| *11* | SCS | 2.48 | 4.61 | 0.46 | TGLA340(0.46) | 0.034 |
| | | | | 0.47 | TGLA58(0.54) | |
| *16* | PL | 2.19 | 4.20 | 0.29 | TGLA245(0.02) | 0.067 |
| | | | | 0.31 | TGLA53(0.51) | |
| *17* | PY | 2.63 | 5.90 | 0.00 | TGLA26(0.00) | 0.038 |
| | | | | 0.01 | TGLA231(0.34) | |
| *17* | PL | 2.50 | 5.41 | 1.17 | TGLA170(0.68) | 0.192 |
| | | | | 1.03 | TGLA322(1.17) | |
| *27* | F% | 2.89 | 5.47 | 0.14 | RM209(0.14) | 0.025 |
| | | | | 0.52 | BM1857(0.53) | |
| *27* | FY | 2.28 | 4.59 | 0.55 | BM1857(0.53) | 0.021 |
| | | | | 0.52 | BM203(0.64) | |

[a] Threshold values were from permutation test.
[b] Threshold values were from chi-square distribution with 1 or 2 d.f.
[c] Distances from origin of linkage groups: first row, LS estimates; second row, VC estimates.
[d] Pair of markers from published maps (BovMAP, MARC) flanking the estimated QTL position; values in parentheses are distances from origin of linkage groups.
[e] Ratio of QTL allelic to additive genetic variance for VC analysis.

analysis (possible only when marker alleles have unique codes across families), in which all paternal and maternal relationships are used.

Several chromosomes, in particular C*6*, C*9*, C*14*, C*20*, and C*26*, are likely to harbor QTL affecting more than one trait. This finding is expected due to the genetic correlations among the traits. The QTL on C*6* (for F% and P%), C*9* (for MY, FY, and PY), C*14* (for F% and FY), and C*20* (for F% and P%) have very similar position estimates, while on C*6* the estimated position of the MY QTL is clearly different from those for F% and P%, on C*14* the position of the SCS QTL is different from those for F% and FY, and on C*26* the QTL for F% and SCS have different estimated locations. Given the sizes of the bootstrap CIs, however, the positions of QTL on the same chromosome cannot be declared different with certainty. For the remaining chromosomes, only QTL affecting a single trait were identified, most likely due to the stringent thresholds for experiment-wise and suggestive significance.

Georges *et al.* (1995) identified the following: a QTL on C*1* for MY, which was not confirmed in this study; a QTL on C*1* for PY, which was confirmed at the suggestive level; QTL on C*6* for MY, F% and P%, which were all confirmed at the experiment-wise level; QTL on C*9* for FY and PY, which were confirmed at or near the suggestive level; a QTL on C*10* for FY, which was not confirmed; and a QTL on C*20* for P%, which was confirmed at the experiment-wise level. All the confirmed QTL have similar estimated positions in both analyses.

Ron *et al.* (1994) selected 10 microsatellite markers to search for QTL affecting milk production traits in a GDD consisting of seven Israeli Holstein families. These authors identified one marker on chromosome *21* associated with significant effects on MY and PY in one family. This finding was not confirmed in our study.

Weller *et al.* (1995) analyzed 11 microsatellite markers and the Dairy Bull DNA Repository (DBDR) and concluded that several markers were associated with significant effects on milk production and health, including C*2* for FY and F%, C*4* for herdlife, C*7* for SCS, and C*15* for FY and F%. Of the seven effects found, four are expected by chance. Our study confirmed only that the QTL on C*2* were at the chromosome-wise and suggestive levels, respectively, and that the estimated positions of both QTL were close to the marker ETH121, which was also the marker with significant allele effects for FY and F% in their study. On C*4* we

**TABLE 7**

**Proportions of additive genetic variance explained by all significant QTLs or by all significant and suggestive QTLs within trait**

| % | MY | FY | PY | F% | P% | SCS | PL |
|---|---|---|---|---|---|---|---|
| Proportion significant | 11 | 0 | 0 | 37 | 53 | 0 | 0 |
| Proportion significant and suggestive | 52 | 61 | 28 | 102 | 57 | 83 | 0 |

Results are from analysis of the actual data.

found a QTL for SCS instead of productive life at the suggestive significance level, which has the same marker (RM188) association as that for herdlife in their study.

A similar study using DBDR families was carried out by Ashwell *et al.* (1997). These authors reported ten significant effects with five expected by chance: C*18* for SCS (not confirmed here), C*21* for MY and FY (not confirmed), C*23* for SCS, FY, and PL (only FY confirmed at suggestive level), C*26* for FY and F% (both confirmed at suggestive level), and C*27* for PY and P% (not confirmed; instead we identified QTL for F% and FY at chromosome-wise significance level). Because the marker system in their study was quite different from that in our study, a comparison of the QTL positions is not possible.

Spelman *et al.* (1996) identified a P% QTL on C*6* in a Dutch Holstein-Friesian GDD, which was confirmed in this study at the experiment-wise level with almost the same position estimate.

Vilkki *et al.* (1997) analyzed a GDD in Finnish Ayrshires. Although these authors did not obtain any significant results, they reported some evidence in favor of a QTL on C*9* for MY and PY. Both effects were confirmed here at the suggestive level.

The latest QTL findings were reported at the 6th World Congress on Genetics Applied to Livestock Production. Ron *et al.* (1998) presented results from an analysis of DBDR families. They reported a significant effect of a marker on C*3* for MY (confirmed at the suggestive level), F% (not confirmed), and P% (confirmed at the experiment-wise level), a marker on C*10* for P% (not confirmed), and a marker on C*14* for FY and F% (both confirmed at the suggestive level). Maki-Tanila *et al.* (1998) reported a QTL on C*1* for MY, a QTL on C*6* for MY and P%, a QTL on C*20* for MY, P%, and FY, and a QTL on C*23* for P% in Finnish Ayrshire dairy cattle. Only the QTL on C*6* for MY and P% and C*20* for P% are confirmed at the experiment-wise level here. Gomez-Raya *et al.* (1998) reported a QTL on C*6* for MY in Norwegian dairy cattle, which is confirmed at the experiment-wise level with similar position estimate. Reinsch *et al.* (1998) found significant QTL effects on C*23*, C*8*, and C*1* for SCS in three major cattle breeds in Germany, which are not confirmed in our study.

The main reasons why some of the findings in other studies were confirmed here while others were not are that at least some of the families in our study are not the same as those in the other studies, and that some of the other studies employed significance thresholds for which fairly large numbers of false positives are expected under the null hypothesis of no QTL segregating.

The fraction of the additive genetic variance explained by all QTL exceeding experiment-wise and suggestive significance thresholds was highest for the traits F% and SCS. No QTL was identified at the experiment-wise or suggestive levels for PL, which is a composite trait and expectedly less appropriate for QTL mapping than biological or component traits. Our simulation study reconfirmed the expected result that QTL variance contributions are overestimated on average for those QTL that surpass rather stringent significance thresholds. To obtain unbiased or less biased estimates, one might consider shrinkage estimation of the variances (those estimates with the least information have the largest positive errors on average but will be shrunken the most) using an informative Bayesian prior distribution with minor more likely than major variances (Hoeschele and VanRaden 1993), increasing the number of genotyped offspring in the same families, or re-estimating the QTL variance contributions in a different population.

For those chromosomes likely to harbor QTL affecting several traits (C*6*, C*9*, C*14*, C*20*, and C*26*), we investigated the direction of the substitution effects in those sires with significant regressions for more than one trait. In every single case, the direction was consistent with the genetic correlations among traits. As an example, for C*6* and two families, the allele increasing MY decreased F%. A different situation was found when looking at two QTL situated on the same chromosome and affecting the same trait. For C*17* and one family, the substitution effects for the two MY loci had opposite signs, *i.e.*, were linked in repulsion phase. We then looked at all other two-linked-QTL analyses that had not quite reached significance but showed some evidence in favor of the two-QTL model. In all cases including C*17*-PY, C*6*-F%, C*6*-P%, C*6*-FY, and C*6*-PY, the two loci were linked in repulsion phase in those families where two QTL appeared to be segregating. Dairy cattle popula-

tions are undergoing selection for the milk production traits, and under selection negative covariances among loci are built up. This finding supports the need for adding additional, informative markers in the regions of interest, and for analyses that account for multiple QTL on the same chromosome, either by use of marker cofactors or by fitting several QTL simultaneously, as linked QTL in repulsion phase should be more difficult to detect in single QTL analyses.

In dairy cattle, implementations of marker-assisted selection for selection of young sires before progeny testing and for selection in nucleus breeding schemes have been shown to potentially produce additional genetic and economic gains (Meuwissen and van Arendonk 1992; Brascamp *et al.* 1993; Mackinnon and Georges 1997). Hence, there is scope for QTL identification from an industry viewpoint. Application of MAS would be more efficient if essentially nonrecombining marker haplotypes bracketing the QTL could be identified. Furthermore, fine-mapping is a prerequisite for cloning QTL and for function evaluation. We confirmed using simulation that adding genotypes at reasonably polymorphic markers to the current GDD will reduce CIs to the desired 10- to 20-cM range. For example, we simulated a chromosome of length 1 M with 11 markers spaced 10 cM apart and 5 alleles per marker at equal frequencies. The bootstrap CIs for a QTL explaining 25 and 12.5% of the additive genetic variance were 13 and 17 cM, respectively. Other techniques will be then needed to reduce CIs to $\leq$1 cM in outbred populations, and these need to be investigated.

Financial support provided by the National Science Foundation (DBI-9696265/9740519), the US Department of Agriculture (92-01732), the US Holstein Associations and ABS Global Inc. to I. Hoeschele is gratefully acknowledged. Q. Zhang is supported in part by the China National Foundation of the Natural Sciences.

## LITERATURE CITED

Aitkin, M., D. Anderson, B. Francis and J. Hinde, 1989 *Statistical Modeling in GLIM*. Oxford University Press, Oxford.

Ashwell, M. S., C. E. Rexroad, R. H. Miller, P. M. VanRaden and Y. Da, 1997 Detection of loci affecting milk production and health traits in an elite US Holstein population using microsatellite markers. Anim. Genet. **28:** 216–222.

Berger, R., 1997 Likelihood ratio test and intersection-union tests, in *Advances in Statistical Decision Theory*, edited by N. Balakreshnan and S. Panchapahesan. Birkhauser, Boston (in press).

Brascamp, E. W., J. A. M. van Arendonk and A. F. Groen, 1993 Economic appraisal of the utilization of genetic markers in dairy cattle breeding. J. Dairy Sci. **76:** 1204–1213.

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

Geldermann, H., 1975 Investigations on inheritance of quantitative characters in animals by gene markers. 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.

Gomez-Raya, L., H. Klungland, D. I. Vage, I. Olsaker, E. Fimland *et al.*, 1998 Mapping QTL for milk production traits in Norwegian cattle. Proc. 6th World Congr. Genet. Appl. Livest. Prod. **26:** 429–432.

Grignola, F. E., I. Hoeschele and K. Meyer, 1994 Empirical best linear unbiased prediction to map QTL. Proc. 5th World Congr. Genet. Appl. Livest. Prod., **21:** 245–248.

Grignola, F. E., I. Hoeschele and B. Tier, 1996a Mapping quantitative trait loci in outcross populations via Residual Maximum Likelihood. I. Methodology. Genet. Sel. Evol. **28:** 479–490.

Grignola, F. E., I. Hoeschele, Q. Zhang and G. Thaller, 1996b Mapping quantitative trait loci in outcross populations via Residual Maximum Likelihood. II. A simulation study. Genet. Sel. Evol. **28:** 491–504.

Grignola, F. E., Q. Zhang and I. Hoeschele, 1997 Mapping linked quantitative trait loci in outcross populations via Residual Maximum Likelihood. Genet. Sel. Evol. **29:** 529–544.

Hoeschele, I., and P. M. VanRaden, 1993 Bayesian analysis of linkage between genetic markers and quantitative trait loci. I. Prior knowledge. Theor. Appl. Genet. **85:** 953–960.

Hoeschele, I., P. Uimari, F. E. Grignola, Q. Zhang and K. Gage, 1997 Advances in statistical methods to map quantitative trait loci in outbred populations. Genetics **147:** 1445–1457.

Im, S., R. L. Fernando and D. Gianola, 1989 Likelihood inferences in animal breeding under selection: A missing data theory viewpoint. Genet. Sel. Evol. **21:** 399–414.

Knott, S. A., J. M. Elsen and C. S. Haley, 1994 Multiple marker mapping of quantitative trait loci in halfsib populations. Proc. 5th World Congr. Genet. Appl. Livest. Prod. **21:** 33–36.

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

Lander, E. S., and L. Kruglyak, 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genet. **11:** 241–247.

Lebreton, C. M., and P. M. Visscher, 1997 Empirical non-parametric bootstrap strategies in QTL mapping: conditioning on the genetic model. Genetics **148:** 525–536.

Logue, D. N., and M. J. A. Harvey, 1978 Meiosis and spermatogenesis in bulls heterozygous for a presumptive 1/29 Robertsonian translocation. J. Reprod. Fertil. **54:** 159–165.

Mackinnon, M., and M. Georges, 1992 The effects of selection on linkage analysis for quantitative traits. Genetics **132:** 1177–1185.

Mackinnon, M., and M. Georges, 1997 A bottom-up approach to marker-assisted selection. Livest. Prod. Sci. (in press).

Maki-Tanila, A., D.-J. de Koning, K. Elo, S. Moisio, R. Velmala *et al.*, 1998 Mapping of multiple quantitative trait loci by regression in half sib designs. Proc. 6th World Congr. Genet. Appl. Livest. Prod. **26:** 269–272.

Meuwissen, T. H. E., and J. A. M. van Arendonk, 1992 Potential improvements in rate of genetic gain from marker-assisted selection in dairy cattle breeding schemes. J. Dairy Sci. **75:** 1651–1659.

Pearson, R. E., W. E. Vinson and T. R. Meinert, 1990 The potential for increasing productivity through selection for increased milk and component yields. Proc. 4th World Congr. Genet. Appl. Livest. Prod. **14:** 104–113.

Reinsch, N., N. Xu, H. Thomsen, C. Looft, E. Kalm *et al.*, 1998 First results on somatic cell count loci from the ADR bovine mapping project. Proc. 6th World Congr. Genet. Appl. Livest. Prod. **26:** 426–428.

Ron, M., M. Band, A. Yanai and J. I. Weller, 1994 Mapping quantitative trait loci with DNA microsatellites in a commercial dairy cattle population. Anim. Genet. **25:** 259–264.

Ron, M., D. W. Heyen, J. I. Weller, M. Band, E. Feldmesser *et al.*, 1998 Detection and analysis of a locus affecting milk concentration in the US and Israeli dairy cattle populations. Proc. 6th World Congr. Genet. Appl. Livest. Prod. **26:** 422–425.

Schutz, M. M., 1994 Genetic evaluation of somatic cell scores for United States dairy cattle. J. Dairy Sci. **77:** 2113–2129.

Schutz, M. M., L. B. Hansen and G. R. Steuenagel, 1990 Genetic parameters for sometic cells, protein, and fat in milk of Holsteins. J. Dairy Sci. **73:** 494–502.

Spelman, R. J., W. Coppieters, L. Karim, J. A. M. van Arendonk and H. Bovenhuis, 1996 Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. Genetics **144:** 1799–1808.

Uimari, P., Q. Zhang, F. E. Grignola, I. Hoeschele and G. Thaller, 1996 Analysis of QTL workshop I granddaughter design data

using Least-Squares, Residual Maximum Likelihood, and Bayesian methods. J. Quant. Trait Loci **2**(7).

VanRaden, P. M., and G. R. Wiggans, 1991 Derivation, calculation, and use of national animal model information. J. Dairy Sci. **74:** 2737–2746.

VanRaden, P. M., and E. J. H. Klaaskate, 1993 Genetic evaluation of length of productive life including predicted longevity of live cows. J. Dairy Sci. **76:** 2758–2764.

Vilkki, H. J., D.-J. de Koning, K. Elo, R. Velmala and A. Maki-Tanila, 1997 Multiple marker mapping of Quantitative Trait Loci of Finnish dairy cattle by regression. J. Dairy Sci. **80:** 198–204.

Visscher, P. M., R. Thompson and C. S. Haley, 1996 Confidence intervals in QTL mapping by bootstrapping. Genetics **143:** 1013–1020.

Weigel, D. J., B. G. Cassell and R. E. Pearson, 1997 Prediction of transmitting abilities for productive life and lifetime profitability from production, somatic cell count, and type traits in milk markets for fluid milk and cheese. J. Dairy Sci. **80:** 1398–1405.

Weir, B. S., 1990 *Genetic Data Analysis.* Sinauer Associates, Inc. Publishers, Sunderland, MA.

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.

Weller, J. I., A. Yanai, Y. Blank, E. Feldmesser, H. Lewin *et al.*, 1995 Detection of individual loci affecting somatic cell concentration in the U.S. Holstein population with the aid of DNA microsatellites. Proceedings of the Third International Mastitis Seminar, Tel Aviv, Israel **1:** 3–13.

Weller, J. I., G. R. Wiggans, P. M. VanRaden and M. Ron, 1996 Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multi-trait experiment. Theor. Appl. Genet. **92:** 998–1002.

Weller, J. I., J. Z. Song, Y. I. Ronin and A. B. Korol, 1997 Designs and solutions to multiple trait comparisons. Anim. Biotech. **8:** 107–122.

Whittaker, J. C., R. Thompson and P. M. Visscher, 1996 On the mapping of QTL by regression of phenotype on marker-type. Heredity **77:** 23–32.

Xu, S., and W. R. Atchley, 1995 A random model approach to interval mapping of quantitative trait loci. Genetics **141:** 1189–1197.

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.

Communicating editor: P. D. Keightley