Least Squares Interval Mapping of Quantitative Trait Loci Under the Infinitesimal Genetic Model in Outbred Populations

Z. Liu[1] and J. C. M. Dekkers[2]

Centre for Genetic Improvement of Livestock, Department of Animal and Poultry Science, University of Guelph, Ontario, Canada N1G 2W1

Manuscript received May 20, 1997
Accepted for publication September 11, 1997

ABSTRACT

Genetic marker and phenotypic data for a quantitative trait were simulated on 20 paternal half-sib families with 100 progeny to investigate properties of within-family-regression interval mapping of a postulated single quantitative trait locus (QTL) in a marker interval under the infinitesimal genetic model, which has been the basis of the application of quantitative genetics to genetic improvement programs, and to investigate use of the infinitesimal model as null hypothesis in testing for presence of a major QTL. Genetic effects on the marked chromosome were generated based on a major gene model, which simulated a central biallelic QTL, or based on 101 biallelic QTL of equal effect, which approximated the infinitesimal model. The marked chromosome contained 0, 3.3%, 13.3%, or 33.3% of genetic variance and heritability was 0.25 or 0.70. Under the polygenic model with 3.3% of genetic variance on the marked chromosome, which corresponds to the infinitesimal model for the bovine, significant QTL effects were found for individual families. Correlations between estimates of QTL effects and true chromosome substitution effects were 0.29 and 0.47 for heritabilities of 0.25 and 0.70 but up to 0.85 with 33.3% of polygenic variance on the marked chromosome. These results illustrate the potential of marker-assisted selection even under the infinitesimal genetic model. Power of tests for presence of QTL was substantially reduced when the polygenic model with 3.3% of genetic variance on the chromosome was used as a null hypothesis. The ability to determine whether genetic variance on a chromosome was contributed by a single QTL of major effect or a large number of QTL with minor effects, corresponding to the infinitesimal model, was limited.

WITH rapid development in molecular genetics, many highly polymorphic markers have been detected in recent years, which can be used to investigate segregation of quantitative trait loci (QTL) for economic traits in livestock and plants. Most production traits do not follow simple Mendelian monogenic inheritance and genetic variation likely involves a small number of major loci, a larger number of loci with moderate effects, and a very large number of loci with minor effects (Robertson 1967). Absence of more specific knowledge of the genetic architecture of quantitative traits has resulted in use of the infinitesimal genetic model, which assumes a large number of loci with small effects (polygenes; Bulmer 1980). The infinitesimal genetic model forms the basis of quantitative genetics theories (Bulmer 1980; Falconer and Mackay 1996) that have been applied successfully to genetic improvement of livestock and that have formed the basis for development and application of sophisticated statistical models for genetic evaluation of livestock (Henderson 1988).

Corresponding author: Dr. Jack C. M. Dekkers, Department of Animal Science, 201 Kildee Hall, Iowa State University, Ames, IA, 50011. E-mail: jdekkers@iastate.edu

[1]*Present address:* Canadian Dairy Network, 150 Research Lane, Suite 307, Guelph, Ontario, Canada, N1G 4T2

[2]*Present address:* Department of Animal Science, 201 Kildee Hall, Iowa State University, Ames, IA 50011.

In recent years, statistical methods have been developed to detect QTL with the aid of genetic marker information (for a review see Van Arendonk *et al.* 1994). Most methods were initially developed for detection of QTL in line crosses but have been extended for use in outbred populations. In segregating populations that are in linkage equilibrium, QTL mapping analyses are complicated by the fact that QTL must be estimated on a within family basis (Smith and Simpson 1986) and QTL effects are specific to a family.

For the purpose of detecting QTL, it is customary to fit a single QTL in a marker interval against a background of unlinked polygenes. Although methods are available to fit more than one QTL simultaneously on a chromosome (Haley and Knott 1992; Martinez and Curnow 1992) or to account for genetic effects on the chromosome outside the marker interval by fitting other markers as cofactors in analyses (Jansen 1993; Zeng 1994), the application of these approaches to segregating populations, with typically limited and heterogeneous informativeness of markers across families, is restricted (Spelman *et al.* 1997; Vilkki *et al.* 1997), as is the ability of these approaches to distinguish between significant effects caused by a single major QTL and effects caused by clusters of multiple QTL with smaller effects.

Statistical testing in QTL analysis is complicated by the multiple tests that are conducted across many chro-

mosomal regions. As a result, test statistics do not follow standard distributions under the null hypothesis (Rebai *et al.* 1994; Weeks *et al.* 1990). False positive results created by improper statistical methods must be avoided by either setting up stringent threshold values for statistical tests by analytical approximations, or by empirically estimating significance levels for specific data structures, *e.g.*, Andersson *et al.* (1994); Churchill and Doerge (1994).

Another aspect of statistical testing involves choice of the null hypothesis. For statistical testing for presence of a QTL, a null hypothesis is usually implemented that assumes that the marked chromosome segment under study contributes no genetic variance; in other words, it is assumed in the null hypothesis that inheritance of genetic markers and phenotypes are totally independent. However, prior knowledge that (1) the trait exhibits quantitative genetic variation, (2) genes are linked on chromosomes of limited size, and (3) there is limited information on the number of loci or their distribution over the genome leads to a null hypothesis of an even distribution of genetic effects across the genome. The infinitesimal genetic model can be used as the basis of development of such a null hypothesis. In addition, the infinitesimal model can be viewed as the "worst case scenario" for QTL mapping and for the use of genetic markers in quantitative trait selection programs, *i.e.*, marker-assisted selection (Weller and Fernando 1991). Under the infinitesimal genetic model, the marked chromosome is expected to contain polygenic variation of a magnitude that is proportional to the map length of chromosome. This amounts to 3.3% of genetic variance for an average chromosome in the bovine. With the infinitesimal model as null hypothesis, the objective of QTL analysis may be to find QTL or chromosome segments that have more genetic variation than can be expected based on the infinitesimal model.

Associations between genetic markers and quantitative traits under the infinitesimal model has been investigated in few studies. Dekkers and Dentine (1991) derived that 43% of polygenic variance contributed by an average chromosome in the bovine can be traced from parent to progeny by a single polymorphic marker at the center of the chromosome. They also showed that substantial effects can be associated with a marker in some families under the infinitesimal model. Polygenic effects may explain the fact that marker-associated effects have been limited to certain families in studies in dairy cattle, *e.g.*, Bovenhuis and Weller (1994); Georges *et al.* (1995); Spelman *et al.* (1997).

Visscher and Haley (1996) investigated the impact of the infinitesimal genetic model on QTL mapping in line crosses. Use of the infinitesimal model for hypothesis testing and tests to distinguish effects because of single QTL from polygenic effects were investigated. They found that, given an assumed genetic difference between the two lines caused by polygenes under the infinitesimal model, spurious effects of a QTL with large effect could be detected frequently and that test statistics were inflated. The effectiveness of tests to detect whether differences resulted from polygenic effects was limited. In the present study, implications of the infinitesimal model as a basis for mapping QTL will be investigated for outbred populations in linkage disequilibrium. In such populations, the difference in genetic value between two homologue chromosomes in a parent (chromosome substitution effect), which can give rise to detection of major QTL, as in Visscher and Haley (1996), is random under the polygenic model, differs from parent to parent, and relates to the Mendelian sampling variance on the chromosome (Dekkers and Dentine 1991).

Alternative methods have been developed for mapping QTL based on genetic marker data (see Van Arendonk *et al.* 1994). In recent years, least-squares-regression interval mapping (Haley and Knott 1992) has gained wide acceptance because of its simplicity and limited sacrifice in accuracy relative to more complex maximum likelihood methods, *e.g.*, Liu *et al.* (1994). The least squares regression method has been used for mapping QTL in line cross data (Haley and Knott 1992; Martinez and Curnow 1992; Andersson *et al.* 1994) and for half-sib families from outbred populations (Knott *et al.* 1996; Spelman *et al.* 1997; Vilkki *et al.* 1997).

The objective of this study was to investigate (1) properties of within-family-regression interval mapping of a postulated QTL in outbred populations for traits that are governed by the infinitesimal genetic model; (2) comparisons of estimates for a postulated QTL when effects on the chromosome are polygenic *vs.* caused by a single major QTL and, thereby, the ability to distinguish effects caused by a major gene *vs.* effects caused by polygenes; (3) the effect of using the infinitesimal model as null hypothesis on significance thresholds and power of detecting significant QTL effects; and (4) the magnitude, distribution, and accuracy of estimated QTL effects under the infinitesimal model and implications for marker-assisted selection. These objectives were addressed through analysis of data that were simulated according to alternative genetic models. The present study focused on the use of two flanking genetic markers on a chromosome.

METHODS

Genetic models and data simulation: Data on a quantitative trait were generated for 20 paternal half-sib families with 100 progeny by stochastic simulation based on additive genetic models in which breeding values were composed of genetic effects on a marked chromosome and joint polygenic effects for QTL on other chromosomes. Each data set was replicated 10,000 times. The marked chromosome was 100 cM long and contained two highly polymorphic genetic markers, which were located 20 cM apart and centered on the chromosome.

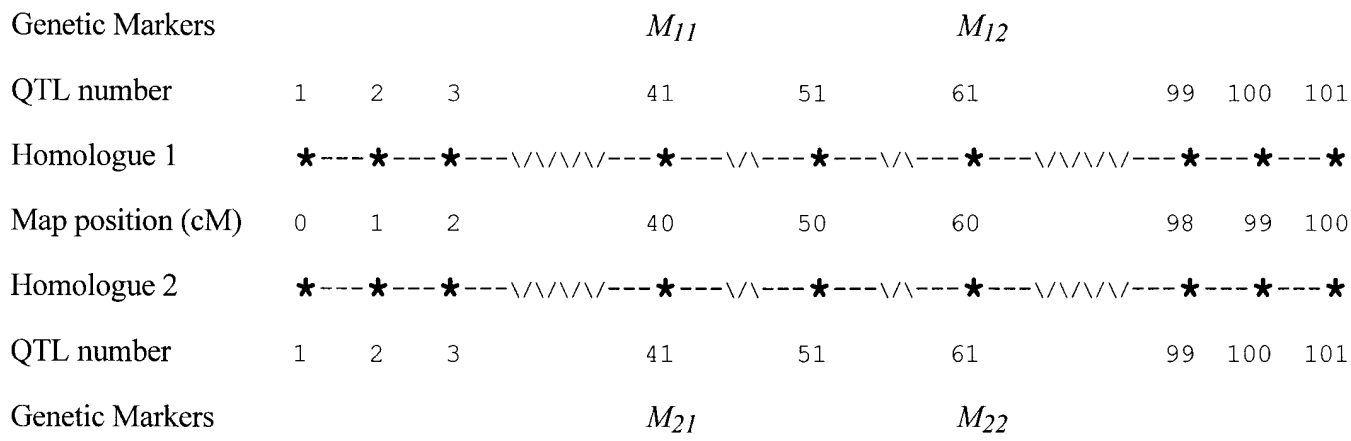

Figure 1.—Simulation of chromosomes with genetic markers and quantitative trait loci under the infinitesimal genetic model.

Two models were considered to simulate genetic effects on the marked chromosome: a major gene model and a polygenic model, which approximated the infinitesimal model. Under the major gene model, genetic variance on the marked chromosome was contributed by a single QTL. The QTL had two alleles with equal frequency and was located at the center of the marker interval. Under the polygenic model, polygenic effects on the marked chromosome were simulated by 101 biallelic QTL of equal effect that were distributed evenly spaced across the chromosome, as illustrated in Figure 1, and in linkage equilibrium in the population.

For both models, situations were considered in which genetic effects on the marked chromosome contributed a fraction (f) 0, 3.3%, 13.3%, or 33.3% of total additive genetic variance in the population. The model with no genetic variance on the chromosome will be referred to as the unlinked polygenic model and corresponds to the model that is often used for hypothesis testing in QTL mapping analyses. The polygenic model with 3.3% of genetic variance on the marked chromosome corresponds to the infinitesimal model for the bovine. Polygenic models with 13.3% and 33.3% of genetic variance on the chromosome represent situations in which the chromosome contributes substantially larger than average genetic variance but through polygenic effects.

Phenotypic variance for the quantitative trait was set to 1 and overall heritability (h^2) was set equal to 0.25 or 0.70. For milk production traits in dairy cattle, a heritability of 0.25 corresponds to analysis of phenotypic records in a daughter design and a heritability of 0.70 represents analysis of progeny averages of sons in a granddaughter design (Weller *et al.* 1990). For the latter, phenotypic variance (=1) corresponds to the variance of sire progeny averages.

Marker loci were highly polymorphic such that all sires were double heterozygotes: $M_{i11}M_{i12}/M_{i21}M_{i22}$, where M_{i12} is the allele for the second marker on homologue number one for the marked chromosome pair of sire i (see Figure 1). Because of the high degree of polymorphism, marker transmission from sire to progeny could be determined with certainty for all progeny. It was also assumed that linkage phase of markers for the sire could be determined with certainty based on the marker genotypes of 100 progeny.

The genetic value of sire i for the quantitative trait was generated as:

$$g_i = c_{i1} + c_{i2} + s_i \quad (1)$$

where c_{i1} (c_{i2}) is the genetic value of homologue number 1 (2) of the marked chromosome pair and s_i is the sire's cumulative value of QTL on unmarked chromosomes. Polygenic effects s_i were simulated according to a Normal distribution with mean 0 and variance $(1-f)h^2$. Genetic effects c_{i1} and c_{i2} were generated as described below for the two genetic models (generating data for the marked chromosome).

Phenotype for daughter j of sire i was generated as:

$$y_{ij} = c_{ij} + \tfrac{1}{2}s_i + e_{ij} \quad (2)$$

where c_{ij} is the genetic effect of the marked chromosome received through the paternal gamete (generated as described below) and e_{ij} is the residual effect, which includes the effect of all genes received from the dam, the random segregation effect for the polygenic effect of the sire, and a random environmental effect. Effects e_{ij} were generated by sampling from a Normal distribution with mean 0 and variance $1-(1+f)h^2/4$.

Generating data for the marked chromosome: Under the unlinked polygenic model, the marked chromosome contains no genetic effects and c_i, c_{ij}, and f in Equations 1 and 2 are zero. Under the major gene model, the effects of the two alleles at the major gene (equal frequency) are equal to plus and minus $\alpha = (\tfrac{1}{2}fh^2)^{1/2}$. The genotype of sire i for the QTL (c_{i1} and c_{i2} in Equation 1) was generated at random based on QTL allele frequencies of 0.5. Gametes produced by sire i were generated by simulating crossovers between the two markers and the QTL based on a recombination rate of 0.0906 in each interval, which corresponds to a map distance of 10 cM according to the mapping function of Haldane (1919). Each progeny received a random gamete, which generated the paternal marker haplotype of the progeny and the term c_{ij} (= c_{i1} or c_{i2}) in Equation 2.

For the linked polygenic model, sire genotypes for the marked chromosome were simulated as outlined in Figure 1, with 101 QTL evenly spaced over the 100 cM chromosome and two markers at 40 and 60 cM from the centromere, which was located at position 0 of the chromosome (telomeric). Each QTL had two alleles with equal frequencies in the population and effects of plus and minus $(\tfrac{1}{2}fh^2/101)^{1/2}$. Genotypes at each QTL for each sire were generated at random. Effects c_{i1} and c_{i2} in Equation 1 are equal to the sum over loci of allele effects assigned to homologues 1 and 2, respectively.

To generate gametes produced by sire i, crossovers were simulated in each 1 cM interval between two loci, starting at the centromere (position 0), based on a recombination rate of 0.0099 between adjacent loci. Marker loci were treated together with QTL at positions 40 and 60. After simulation of crossovers, one of the two recombined chromosomes was allocated at random to a progeny. This process was repeated for

each progeny. The genetic effect of the marked paternal chromosome (c_{ij} in Equation 2) was equal to the sum of QTL effects on the recombined chromosome.

Statistical analysis: A model with a postulated single QTL in the marker interval was used to analyze data generated under the two genetic models based on within family regression mapping (Knott *et al.* 1996):

$$y_{ij} = u_i + p_{ij}\alpha_i + e_{ij} \quad (3)$$

where y_{ij} is the phenotype of progeny j of sire i, u_i is the main effect of sire i, e_{ij} is a random residual effect, assumed distributed Normal with mean zero and variance σ_e^2, α_i is the gene substitution effect of the postulated QTL for sire i, and p_{ij} is the probability that progeny j received the QTL allele associated with marker haplotype $M_{i1}M_{i2}$ in sire i, *i.e.*, the QTL allele associated with homologue one of the sire.

Probabilities p_{ij} were based on marker linkage phase of the sire and the paternal marker haplotype received by the progeny, and were conditional on a postulated position of the QTL in the marker interval (Haley and Knott 1992). In the model of analysis, the effect of between sire differences in probabilities p_{ij} and in QTL substitution effects on residual variance was ignored, and the usual assumption with implementation of least-squares interval mapping of homogeneous residual variance was applied (Knott *et al.* 1996).

Conform the method of regression interval mapping, least-squares equations were solved for varying positions of the postulated QTL, by moving the QTL from one marker locus to the other in steps of 1 cM. The QTL position that led to the smallest residual sum of squares was accepted as the most likely position of the QTL, and its corresponding estimates of sire effects, QTL substitution effects, and residual variance were accepted as the best estimates. Prediction error variances of estimates of QTL substitution effects were estimated based on diagonal elements of the inverse of the coefficient matrix of the least squares equations (d_i^{-1}) and averaged across q sires within a replicate as:

$$\mathrm{PEV}_\alpha = [\Sigma_i d_i^{-1}\sigma_e^2]/q \quad (4)$$

For each replicate, genetic variance contributed by the postulated QTL was estimated based on:

$$\sigma_\alpha^2 = V_\alpha - \mathrm{PEV}_\alpha \quad (5)$$

where V_α is the variance among estimates of QTL substitution effects for the 20 sires. Equation 5 is based on the fact that the fixed-effect estimate of QTL substitution effects is equal to its true value plus its prediction error; therefore, the variance of estimates is equal to $V_\alpha = \sigma_\alpha^2 + \mathrm{PEV}_\alpha$.

Hypothesis testing: To test for presence of a QTL in the marked interval, a reduced model without QTL effects was fitted. A likelihood ratio test statistic was used to test the hypothesis of presence of a linked QTL:

$$\mathrm{LR} = N^*\log_e(\mathrm{RSS}_{\mathrm{reduced}}/\mathrm{RSS}_{\mathrm{full}}) \quad (6)$$

where $\mathrm{RSS}_{\mathrm{reduced}}$ and $\mathrm{RSS}_{\mathrm{full}}$ are residual sums of squares of the reduced- and full-regression models, respectively. LR was used to test against the following two null hypotheses: $H_{o\text{-}unl}$ = no effects at the postulated QTL in the marker bracket, *i.e.*, $\alpha_i = 0$ for all sires. This null hypothesis is what is used most frequently in existing studies. $H_{o\text{-}inf}$ = effects at the postulated QTL in the marker bracket follow expectations from the infinitesimal genetic model. This hypothesis can be used to determine whether the postulated QTL explains greater effects than can be expected based on the infinitesimal model.

Because the distribution of LR is not well defined for either null hypothesis, threshold values of LR at 95% and 99% significance levels were determined empirically from data generated under the unlinked polygenic model (= $H_{o\text{-}unl}$) and under the linked polygenic model with $f = 0.033$ (= $H_{o\text{-}inf}$), which corresponds to the infinitesimal model for an average chromosome in the bovine.

If $H_{o\text{-}unl}$ has been rejected, it is necessary to test for which sires the estimates of the QTL-substitution effect are significant. This was done by comparing individual QTL estimates to 95% or 99% confidence intervals of sire QTL-effect estimates under the null hypothesis. Confidence ranges were derived empirically based on the unlinked polygenic model.

RESULTS

Figure 2 gives the distribution of true sire chromosome substitution effects under the linked polygenic model (thick lines). True chromosome substitution effects are the difference in value between the two marked homologues of a sire ($c_{i1} - c_{i2}$ from Equation

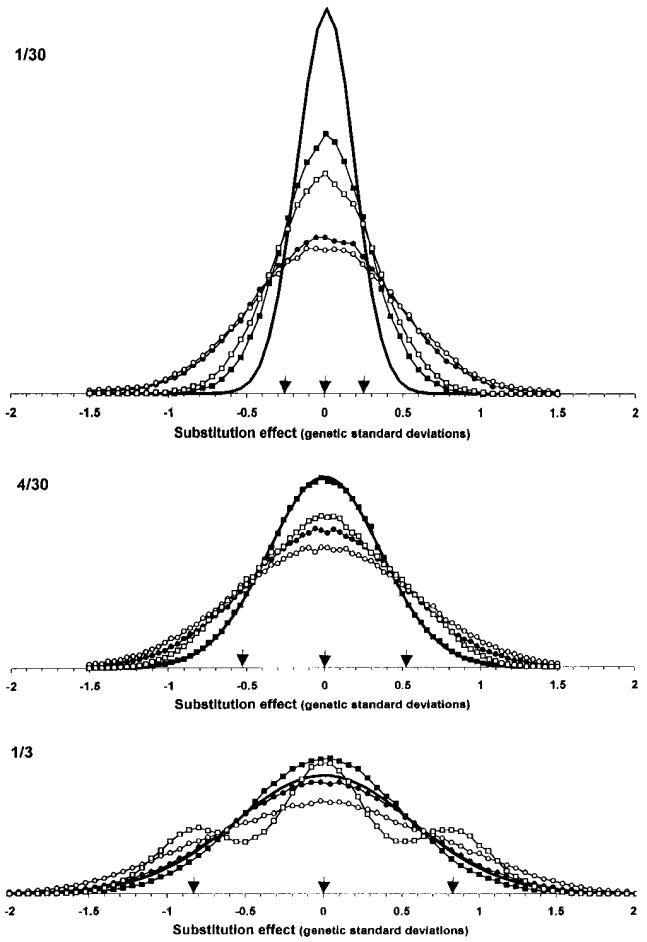

Figure 2.—Distribution of estimates of a postulated single QTL in a marker interval. Results are shown for the linked polygenic model with heritability equal to 0.25 (shaded circles) or 0.70 (shaded squares) and for the linked major gene model with heritability equal to 0.25 (open circles) or 0.70 (open squares). The three figures show results for marked chromosomes that contribute a fraction of 3.3%, 13.3%, or 33.3% of total genetic variance. Thick solid lines show the distribution of true chromosome substitution effects under the linked polygenic model. Arrows indicate the magnitude of true genotypic effects under the major gene model.

1). Theoretically, under the infinitesimal genetic model in a population that is in linkage disequilibrium, true chromosome substitution effects follow a Normal distribution with mean zero and variance equal to the variance contributed by the chromosome. For an average bovine chromosome that contributes 3.3% of genetic variance, close to 3% of individuals had true chromosome substitution effects greater than 0.4 genetic standard deviations. With 13.3% of genetic variance on the chromosome, this percentage increased to close to 28%. Therefore, large chromosome substitution effects are not rare under the linked polygenic model, even for an average bovine chromosome (Dekkers and Dentine 1991).

Estimates of QTL position: Table 1 gives average estimates for the position of a putative QTL in the marker interval under the three genetic models (unlinked, major gene, and linked polygenic models) for chromosomes that contributed 3.3%, 13.3%, or 33.3% of genetic variance. Average estimates were close to the center of the marker interval for all three models. This is as expected because the marker interval was centered on the chromosome, both markers were highly polymorphic, and, in the case of the major gene model, the QTL was located at the center. Standard deviations of QTL position estimates were, however, large and were highest for the unlinked model and lowest for the linked major gene model (Table 1). Standard deviations reflect the accuracy with which the position of the putative QTL can be estimated. Standard deviations decreased with increasing variance on the chromosome and were lower when heritability of the trait was higher.

Figure 3 shows the distribution of estimated QTL positions for the linked polygenic model and the major gene model when the marked chromosome contributes 3.3% or 33.3% of genetic variance. Results for 13.3% of genetic variance on the chromosome were intermediate. When the marked chromosome contributed 3.3% of genetic variance, the QTL was mapped to one of the markers in 65–70% of cases. When the chromosome contributed 33.3% of genetic variance, the percentage of cases in which the QTL was mapped to

TABLE 1

Least squares regression mapping of a postulated QTL under alternative genetic models

Genetic variance on marked chromosome (% of σ^2)	Estimate of QTL position[a] (cM)	Variance of QTL estimates (% of σ^2)	Estimate of QTL variance (% of σ^2)	Correlation of QTL estimates with true sire effects for:		Regression of QTL estimates on true sire effects for:	
				Marker bracket	Chromosome	Marker bracket	Chromosome
Unlinked polygenic model; heritability = 0.25							
0	10.0 (9.0)[b]	18.1 (5.7)	2.4 (5.3)	—	—	—	—
Linked major gene model; heritability = 0.25							
3.3	10.0 (8.6)	21.1 (6.9)	5.1 (6.3)	0.36	0.36	0.93	0.93
13.3	9.9 (7.1)	30.7 (10.0)	14.0 (9.3)	0.62	0.62	0.96	0.96
33.3	10.0 (6.5)	50.7 (15.4)	33.2 (14.8)	0.78	0.78	0.97	0.97
Linked polygenic model; heritability = 0.25							
3.3	9.8 (8.8)	19.9 (6.4)	4.0 (5.9)	0.18	0.29	0.96	0.70
13.3	9.8 (8.1)	25.7 (8.5)	9.4 (8.5)	0.31	0.52	0.96	0.71
33.3	9.9 (6.9)	37.7 (12.8)	20.9 (12.0)	0.42	0.69	0.97	0.73
Unlinked polygenic model; heritability = 0.70							
0	10.0 (9.0)	5.7 (2.0)	0.8 (1.9)	—	—	—	—
Linked major gene model; heritability = 0.70							
3.3	9.9 (7.5)	8.8 (2.9)	3.6 (2.7)	0.55	0.55	0.95	0.95
13.3	10.0 (4.4)	18.8 (5.5)	13.3 (5.4)	0.80	0.80	0.99	0.99
33.3	10.0 (2.5)	38.7 (10.2)	33.2 (10.1)	0.91	0.91	1.00	1.00
Linked polygenic model; heritability = 0.70							
3.3	9.8 (8.3)	7.6 (2.5)	2.5 (2.3)	0.29	0.47	0.96	0.71
13.3	9.9 (6.5)	13.6 (4.6)	8.3 (4.4)	0.44	0.73	0.98	0.73
33.3	9.9 (4.9)	25.8 (8.7)	20.4 (8.5)	0.52	0.85	0.99	0.75

The trait has genetic variance σ^2 and heritability 0.25 or 0.70. Data are generated based on the unlinked polygenic model, the linked major gene model, or the linked polygenic model and for three levels of genetic variance contributed by QTL on the marked chromosome.

[a] Estimated position of QTL in cM from first marker.

[b] Standard deviation based on 10,000 replicates.

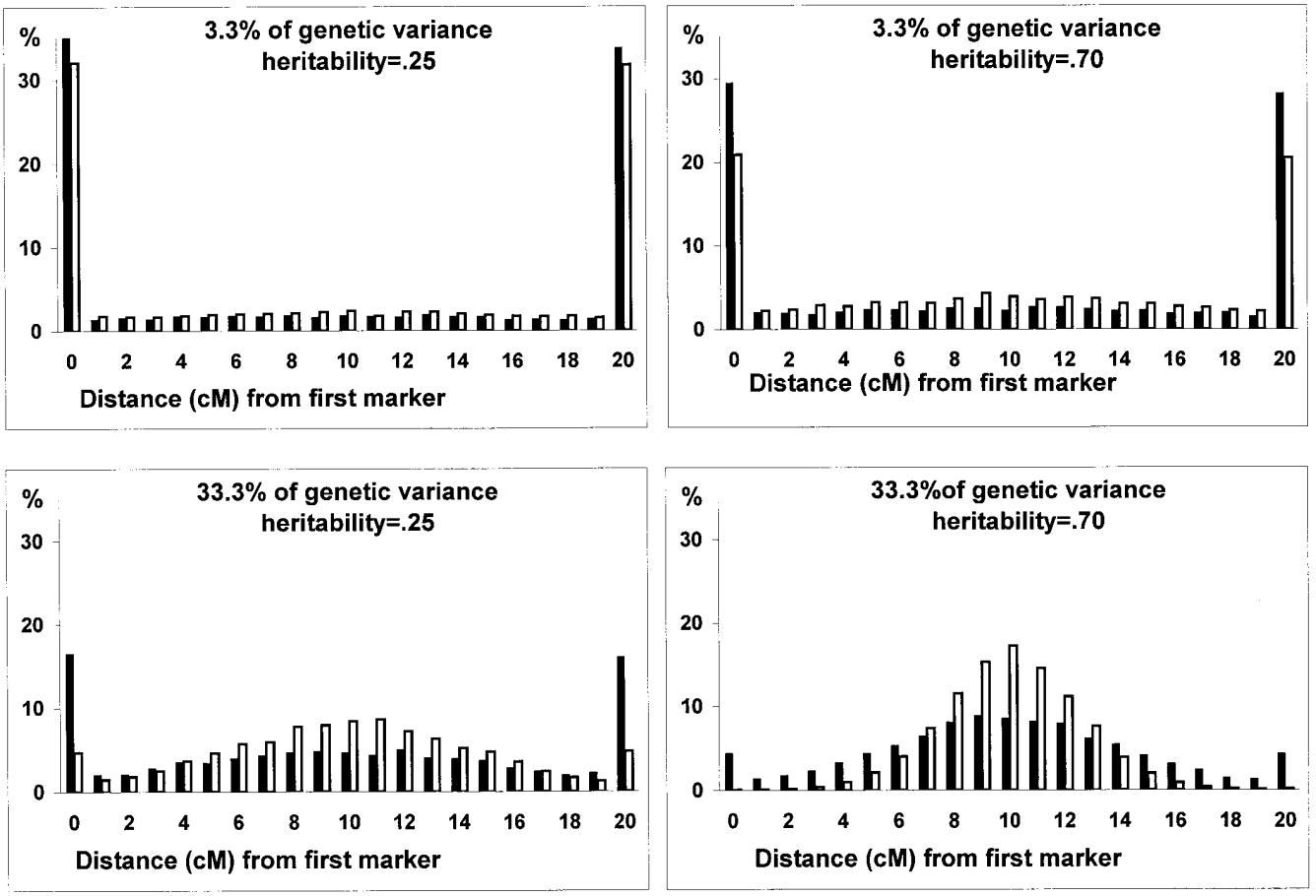

Figure 3.—Frequency distribution of estimates of position of a postulated QTL. The QTL is mapped to an interval of 20 cM that is flanked by two genetic markers on a chromosome with a single major gene at the center of the chromosome (open bars) or with polygenic variability (solid bars). Results are shown for two levels of genetic variability on the marked chromosome (3.3% or 33.3% of total genetic variance) and two levels of heritability (0.25 and 0.70).

one of the markers was reduced to 35% for the linked polygenic model and to less than 10% for the major gene model. It is clear that the proportion of QTL that were mapped to one of the markers would have been reduced if additional markers were present on the chromosome and used in the analysis.

Estimates of QTL effects: Table 1 also shows estimates of genetic variance contributed by the putative QTL. Genetic variance contributed by the QTL was overestimated for the unlinked model (2.4% instead of 0% for heritability = 0.25) and for major genes with small effect (Table 1). The method that was used for estimation of variance resulted in negative estimates for several replicates. For example, under the unlinked polygenic model, 35.7% and 36.2% of estimates of QTL variance were negative for heritabilities of 0.25 and 0.7, respectively. Negative estimates were included in the summary statistics presented in Table 1.

Estimates of variance contributed by the QTL were unbiased for major genes with large effect (Table 1). For the linked polygenic model, variance contributed by a putative single QTL underestimated total variance contributed by the chromosome but significantly overestimated variance contributed by genes located within the marker bracket. This indicates that effects of genes outside the marker interval also contributed in part to estimated effects for the putative QTL. The variance of QTL estimates was substantially greater than the estimated QTL variance (Table 1) because of prediction error variances (see Equation 5). Prediction error variances of QTL estimates were significantly smaller when heritability was greater (0.70 *vs.* 0.25) because of lower residual variance.

Table 1 also shows correlations and regressions of estimates of QTL effects for individual sires on true substitution effects. For the major gene model, these correlations and regressions reflect precision and bias of estimates of QTL effects, respectively. For the linked polygenic model, these parameters reflect the degree to which the effect of the putative QTL represents the cumulative effects of polygenic QTL on the chromosome. For the linked polygenic model, correlations and

regressions are shown in relation to both polygenic effects within the marker bracket and to polygenic effects for the whole chromosome.

For the same genetic variance contributed by the chromosome, correlations of estimates for the putative QTL with true chromosome-substitution effects were greater for the major gene model than for the linked polygenic model (Table 1). Correlations were moderately high for the linked polygenic model, *e.g.*, 0.29 and 0.47 with 3.3% of genetic variance on the chromosome and heritabilities of 0.25 and 0.70, respectively. Correlations increased with heritability for both genetic models, which reflects the additional precision obtained when residual variance is reduced. For the linked polygenic model, correlations of estimates were greater with true substitution effects for the whole chromosome than with true substitution effects within the marker bracket.

Regression coefficients of QTL substitution effect estimates on true substitution effects were slightly less than one for the major QTL model (Table 1), especially when the effect of the major gene was small. This reflects a slight underestimation of the effect of the QTL. For the linked polygenic model, regressions were close to one for the substitution effect within the marker bracket but between 0.70 and 0.75 for regression on the substitution effect for the whole chromosome. This indicates that the putative QTL included a more or less unbiased estimate of combined polygenic effects within the marker bracket but an underestimate of the cumulative effect of QTL outside the marker bracket, which is the result of recombination. Regression coefficients were marginally affected by heritability (Table 1).

Figure 2 shows distributions of true and estimated chromosome substitution effects of individual sires under the linked polygenic model and the major gene model, with 33.3%, 13.3%, or 33.3% of genetic variance contributed by the marked chromosome. Compared with the true chromosome substitution effects,

TABLE 2

Likelihood ratio and power of tests for presence of a postulated QTL in the marker interval under alternative genetic models and against two null hypotheses

Genetic variance on marked chromosome (% of σ^2)	Likelihood ratio		Percentage significant QTL effects at 5% level		Percentage significant QTL effects at 1% level	
	Mean	Standard Deviation	$H_{o\text{-}unl}$	$H_{o\text{-}inf}$	$H_{o\text{-}unl}$	$H_{o\text{-}inf}$
Unlinked polygenic model; heritability = 0.25						
0	23.2	6.4	5	—	1	—
Linked major gene model; heritability = 0.25						
3.3	26.5	7.4	13	7	4	2
13.3	36.7	10.3	54	42	33	20
33.3	57.8	15.6	95	92	87	79
Linked polygenic model; heritability = 0.25						
3.3	25.2	7.0	9	5	3	1
13.3	31.7	9.0	33	23	16	8
33.3	44.8	13.1	77	68	59	45
Unlinked polygenic model; heritability = 0.70						
0[a]	23.1	6.4	5	—	1	—
Linked major gene model; heritability = 0.70						
3.3	33.9	9.5	43	12	22	3.3
13.3	67.6	17.9	98	91	95	77
33.3	138.0	33.3	100	100	100	100
Linked polygenic model; heritability = 0.70						
3.3[b]	29.9	8.5	26	5	11	1
13.3	50.8	14.9	87	61	73	39
33.3	94.0	28.0	100	98	99	95

The trait has genetic variance σ^2, heritability 0.25 or 0.70, data are generated based on the unlinked polygenic model, the linked major gene model, or the linked polygenic model and for three levels of genetic variance contributed by QTL on the marked chromosome. Tests are with regard to two null-hypotheses: $H_{o\text{-}unl}$; no genetic variance on the marked chromosome $H_{o\text{-}inf}$ infinitestimal model genetic variance on the marked chromosome (linked polygenic model with 3.3% of genetic variance on marked chromosome.

[a] Null hypothesis for $H_{o\text{-}unl}$.

[b] Null hypothesis for $H_{o\text{-}inf}$.

TABLE 3

Percentage of significant QTL estimates for individual sires, and averages of absolute values of QTL estimates and true chromosome substitution effects of significant results under the linked polygenic model

Genetic variance on marked chromosome (% of σ^2)	Percentage of sires with significant QTL estimates[a]		Average absolute value for sires with significant[a,b] QTL estimates (genetic standard deviation units)					
			QTL estimate		True chromosome substitution effect		True marker bracket substitution effect	
	$P < 0.05$	$P < 0.01$	$P < 0.05$	$P < 0.01$	$P < 0.05$	$P < 0.01$	$P < 0.05$	$P < 0.01$
Heritability = 0.25								
3.3	6.3	1.4	1.004	1.246	0.170	0.188	0.070	0.074
13.3	12.1	3.1	1.044	1.280	0.422	0.498	0.154	0.164
33.3	17.5	7.5	1.118	1.342	0.742	0.880	0.254	0.276
Heritability = 0.70								
3.3	9.0	2.6	0.580	0.711	0.206	0.238	0.075	0.080
13.3	20.4	9.5	0.643	0.770	0.473	0.562	0.161	0.176
33.3	35.7	22.6	0.742	0.862	0.727	0.840	0.249	0.268

[a] For null hypothesis of zero chromosome substitution effect for sire.
[b] Thresholds for significant sire QTL estimates at $P < 0.05$ and $P < 0.01$ are 0.83 and 1.10 genetic standard deviation units for heritability = 0.25 and equal to 0.47 and 0.62 genetic standard deviation units for heritability = 0.70.

estimates of chromosome substitution effects had larger variance because of prediction errors, especially when heritability and proportion of variance contributed by the chromosome were small. Although true substitution effects for the major gene models fell in three categories (-2α, 0, and $+2\alpha$) with frequencies of 0.25, 0.50, and 0.25, estimates of chromosome substitution effects followed a continuous distribution that appeared Normal (Figure 2). Distributions of estimates for the major gene model were similar to distributions observed for the linked polygenic model but with a somewhat larger standard deviation. The only exception was the situation with high heritability and 33.3% variance on the chromosome, in which case three distinct distributions could be distinguished (Figure 2).

Statistical tests: Table 2 gives the mean and standard deviation of the likelihood ratio statistic for presence of a putative QTL under the three genetic models. Under the unlinked model, the LR had a higher mean than expected for a central χ^2 distribution with 20 degrees of freedom (23.2 *vs.* 20) but a nearly equal standard deviation (Table 2). Correspondingly, the threshold value at 95% significance was greater for the unlinked model than for the central χ^2 distribution (34.7 *vs.* 31.4).

Under the linked polygenic model with an average bovine chromosome (3.3% of genetic variance on the marked chromosome), the LR had a larger mean, standard deviation, and threshold value than the LR under the unlinked model (Table 2). When the unlinked model was used as null hypothesis ($H_{o\text{-}unl}$), percent significant results at $P < 5\%$ and $P < 1\%$ were 9% and 3% for a heritability of 0.25 and 26 and 11% for a heritability of 0.70. The proportion of significant results increased with increasing polygenic variance on the chromosome.

For the same genetic variance contributed by the chromosome, mean LR, percent significant results, and, therefore, power to detect significant effects, were greater under the major gene model than under the linked polygenic model (Table 2). Under the linked polygenic model, power for detecting effects at the putative QTL was reduced because of crossovers outside the marker interval. The difference in power between the two genetic models was reduced as chromosome variance increased (Table 2).

Under the major gene model, testing against the infinitesimal model as the null hypothesis ($H_{o\text{-}inf}$), instead of against the unlinked model ($H_{o\text{-}unl}$), substantially reduced the percent-significant results (Table 2), especially when power was low. Testing against the infinitesimal model tests whether the putative QTL contributes greater effects than can be expected based on the infinitesimal model.

Characteristics of significant estimates of QTL effects for individual sires: Table 3 shows parameters for QTL effects for individual sires that had estimates outside a 95% or 99% confidence range for estimates of QTL effects under the unlinked model. Only results from replicates that resulted in significant effects for the QTL, using the unlinked polygenic model as null hypothesis, were used. Even for an average bovine chromosome and $P < 5\%$ and $P < 1\%$, respectively, 6.3% and 1.4% of sire chromosome-substitution effect estimates were found significant when heritability was 0.25 and 9.0, and 2.6% when heritability was 0.70.

Large chromosome substitution effects were detected under the linked polygenic model even with an average bovine chromosome. Therefore, detection of significant major QTL effects may not exclude presence of linked polygenes that conform to the infinitesimal model. The mean of significant QTL estimates was substantially higher than the mean of the corresponding true chromosome substitution effects, especially when heritability and variance contributed by the chromosome was low (Table 3).

DISCUSSION AND CONCLUSIONS

One of the main focuses of this study was to compare results from least-squares-regression interval mapping of a postulated single QTL on a marked chromosome in a segregating population under linkage disequilibrium when genetic effects on the chromosome are polygenic *vs.* the results of a single major gene. The interval mapping procedure that was used herein is currently used extensively in mapping QTL in outbred populations similar to those simulated here. One of the main conclusions of this study is that it is difficult to distinguish between effects caused by a single major gene and polygenic effects on the chromosome that conform to the infinitesimal model. As illustrated in Figure 2, for a given amount of genetic variance contributed by the chromosome, distributions of estimates of QTL effects were very similar for the major QTL model and the linked polygenic model, except when the effect of the major QTL was extremely large. The ability to distinguish effects of a major gene from polygenic effects will be hampered further when considering major genes with more than two segregating alleles.

Several methods to improve the ability to distinguish effects caused by a major gene from effects caused by polygenes can be explored. In the present model, with use of only two markers, polygenic effects outside the marker bracket contributed to the estimated effect at the postulated QTL (Table 1). Fitting markers outside the marker interval as cofactors in the model, as proposed by Jansen (1993) and Zeng (1994), can be used to account for polygenic effects outside the marker bracket and reduce the impact of such effects on estimates of QTL within the marker bracket. Visscher and Haley (1996) found, for line crosses, that use of markers as cofactors in QTL mapping reduced but did not completely eliminate overestimation of test statistics and QTL effects under the infinitesimal model. The impact of fitting markers as cofactors in outbred populations, with often differing and limited informativeness of genetic markers across families (Spelman *et al.* 1997), requires further investigation.

The second focus of this paper was use of the infinitesimal genetic model as the null hypothesis in detecting QTL with major effect in outbred populations in linkage disequilibrium. Use of such a null hypothesis would be appropriate if the objective of QTL mapping is to identify genetic effects within a marker bracket that are greater than can be expected based on prior knowledge of the heritability of the trait; a null hypothesis that conforms to the inifinitesimal genetic model represents the worst case scenario for mapping QTL for a trait that is known to be heritable. The main conclusion regarding this objective was that power of detecting greater effects than can be expected based on the infinitesimal genetic model, *i.e.*, testing against $H_{o\text{-}inf}$, is substantially lower than power of detecting any genetic effect, *i.e.*, testing against $H_{o\text{-}unl}$. The difference in power between testing against $H_{o\text{-}inf}$ and $H_{o\text{-}unl}$ can be reduced by use of other markers on the chromosome as cofactors in regression interval mapping, as discussed previously. Use of cofactors will, however, also reduce the absolute power of detecting a single QTL, as observed by Visscher and Haley (1996) for line crosses. Its impact on power in outbred populations will require further investigation.

For the major gene model, the least squares regression procedures for mapping QTL in outbred populations used in this study resulted in unbiased estimates of QTL position (Table 1), although these results could be because both the QTL and the marker bracket were centered on the chromosome. When results from all replicates were considered, regardless of significance of QTL effects, estimates of QTL effects had a slight downward bias, as indicated by the less- than-unity coefficients of regression of estimates on true values (Table 1). This is likely because the search among QTL was limited to the marker interval (Wang 1995). In most practical applications, information from multiple markers across the chromosome will be available, which will enable the QTL to be positioned across all marked chromosome segments. For replicates in which "by chance" the QTL would have been mapped in the adjacent marker bracket if the search had been across the chromosome, the QTL was mapped at one of the markers in the present study. Mapping a QTL at the marker, if the best estimate of its location is away from the marker, results in lower estimates for the QTL effect in those replicates, which results in a downward bias. The downward bias was greatest for major genes with small effect (Table 1).

In this study, an *ad hoc* method was used to estimate variance contributed by the postulated QTL (Equation 5). The method overestimated genetic variance when genetic effects on the marked chromosome were absent or small (Table 1). However, this did not affect hypotheses tests, which were based on empirical thresholds, nor did it effect other results presented. In literature, superior methods for estimation of variances associated with QTL are available, *e.g.*, Xu and Atchley (1995).

The final objective of this study was to investigate properties of estimates of the least-squares interval mapping method under the infinitesimal model and

implications for marker-assisted selection, for which the infinitesimal model can be considered the "worst case scenario." Results showed that under the infinitesimal model, the least-squares interval mapping analysis was able to detect significant effects associated with genetic markers within individual sire families. Although these differences will not be useful for mapping of QTL, this information can be used for marker-assisted selection, as discussed below.

Under the infinitesimal model, the difference in genetic value between two homologue chromosomes of a pair in an individual is distributed Normal in a segregating population under linkage equilibrium with mean zero and variance equal to the variance contributed by the chromosome (Figure 2). Because of linkage among loci on a chromosome, progeny inherit chromosomes that resemble chromosomes that are present in their parents, apart from recombination. With no interference, the number of crossover events per meiosis on a chromosome of 100 cM follows a Poisson distribution with mean and standard deviation equal to one. Thirty-seven percent of gametes are produced without crossover. Transmission of parental chromosome with no or limited crossovers to progeny can simulate segregation of a QTL of large effect for sires that have large chromosome substitution effects, which can be traced by genetic markers (Dekkers and Dentine 1991). This forms the basis of the use of genetic markers in marker-assisted selection (Weller and Fernando 1991).

Dekkers and Dentine (1991) found that theoretically up to 43% of polygenic variance contributed by a chromosome of 100 cM can be traced from parent to progeny by a single highly polymorphic marker at the center of the chromosome. The upper limit assumes that effects associated with markers or with postulated QTL can be estimated without error (infinite family size). In the present study, the use of two polymorphic genetic markers for mapping a postulated single QTL on a chromosome was investigated within the context of the infinitesimal model. Estimates of QTL variance in Table 1 indicate the amount of genetic variance that can be traced by two polymorphic genetic markers 20 cM apart on a chromosome of 100 cM with infinite family size. As discussed previously, these variance estimates in Table 1 may be biased upward because of underestimation of prediction error variances. To verify the amount of genetic variance on a polygenic chromosome that can be attributed to the effect of a single postulated QTL in a marker interval, the QTL regression mapping procedure for the simulated data was repeated with as dependent variable the genetic value of the paternal chromosome inherited by a progeny (c_{ij} in Equation 2), instead of progeny phenotype. Estimates of genetic variance contributed by the postulated QTL were between 61% and 62% of chromosome variance for all situations.

Recombination causes the average value of recombined gametes associated with given paternal-marker alleles to be regressed toward the mean genetic value of the two homologues in the sire. For the polygenic model, coefficients of regression of estimates of QTL effects on true chromosome substitution effects were between 0.70 and 0.75 (Table 1). Crossovers outside the marker interval were mainly responsible for the regression of marker differences toward the mean, as indicated by the fact that the regression of QTL estimates on true sire effects within the marker interval was close to unity (Table 1).

The above discussion on variance on a polygenic chromosome that can be traced from parent to progeny assumes that effects associated with markers or with postulated QTL can be estimated without error (infinite size of families). Some of the factors that affect accuracy of estimates of postulated QTL effects are illustrated in Table 1 in terms of the correlation of estimates with true chromosome substitution effects. Accuracy was affected by heritability of the trait and by the magnitude of genetic effects on the marked chromosome. In addition, accuracy will be greatly affected by family size and by the polymorphism information content (Botstein *et al.* 1980) of marker alleles. Of interest was that, even for an average chromosome in the bovine ($f = 0.033$), accuracy of estimates of the effect of postulated QTL was close to 50% under the granddaughter design ($h^2 = 0.70$) with 100 informative progeny per sire (Table 1). Accuracy increased with a greater proportion of genetic variance on the marked chromosome. For a given amount of variance on the marked chromosome, accuracy was lower for the polygenic model than for the linked major gene model (Table 1), although differences between the two genetic models were not large. For a heritability of 0.25, limited accuracy of estimates of postulated QTL effects was obtained (Table 1). The design for this level of heritability reflects a small-scale daughter design. With greater family sizes, accuracies equivalent to those found for the granddaughter design can be achieved.

The main conclusion of the results discussed above is that, even under the infinitesimal genetic model, genetic markers can trace substantial amounts of genetic variance from parents to progeny. Although recombination between polygenic effects and genetic markers will require regular reestimation of substitution effects within families, such information can be used for marker-assisted selection. Statistical models have been developed to continuously reestimate effects associated with genetic markers within the context of best linear unbiased prediction of breeding values from an animal model (Fernando and Grossman 1989; Goddard 1992).

Gametic phase equilibrium was assumed throughout this study. In populations under selection, linkage disequilibrium builds up between QTL, even if they are

unlinked (Bulmer 1971). This linkage disequilibrium implies that, among selected animals and their progeny, the frequency of animals that contain a positive effect for one QTL and a negative allele for another QTL is increased. For pairs of linked QTL, Hospital and Chevalet (1996) recently showed that selection increases the frequency of alleles that appear in the repulsion phase. This will affect the variance of the difference between two homologues of a chromosome pair. The impact on polygenic variance that can be traced by genetic markers requires further investigation.

This research was supported by the Natural Sciences and Engineering Research Council of Canada.

LITERATURE CITED

Andersson, L., C. S. Haley, H. Ellegren, S. A. Knott, M. Johansson *et al.*, 1994 Genetic mapping of quantitative trait loci for growth and fatness in pigs. Science **263:** 1771–1774.

Botstein, D., R. L. White, M. Skolnick and R. W. Davis, 1980 Constructions of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Human Genet. **32:** 314–331.

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.

Bulmer, M. G., 1971 The effect of selection on natural variability. Am. Naturalist **105:** 314–331.

Bulmer, M. G., 1980 *The Mathematical Theory of Quantitative Genetics.* Oxford University Press, Oxford.

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

Dekkers, J. C. M., and M. R. Dentine, 1991 Quantitative genetic variance associated with chromosomal markers in segregating populations. Theor. Appl. Genet. **81:** 212–220.

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics.* Longman, Harlow, Essex, U.K.

Fernando, R. L., and M. Grossman, 1989 Marker assisted selection using best linear unbiased prediction. Genet. Sel. Evol. **21:** 467

Georges, M., D. Nielsen, M. J. 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.

Goddard, M. E., 1992 A mixed model for analyses of data on multiple genetic markers. Theor. Appl. Gen. **139:** 907.

Haldane, J. B. S., 1919 The combination of linkage values and the calculation of distance between loci of linked factors. J. Genet. **8:** 299–309.

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.

Henderson, C. R., 1988 Theoretical basis and computational methods for a number of different animal models. J. Dairy Sci. **71:** (Suppl. 2): 1.

Hospital, F., and C. Chevalet, 1996 Interactions of selection, linkage and drift in the dynamics of polygenic characters. Genet. Res. **67:** 77–87.

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

Knott, S. A., J.-M. Elsen, and C. S. Haley, 1996 Multiple marker mapping of quantitative trait loci in half-sib populations. Theor. Appl. Genet. **93:** 71–80.

Liu, Z. T., H. Simianer, and H. Haussmann, 1994 Marker assisted complex segregation analysis of milk production traits in dairy cattle. Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, Guelph, Canada, pp. 307–310.

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.

Rebai, A., B. Goffinet, and B. Mangin, 1994 Approximate thresholds of interval mapping tests for QTL detection. Genetics **138:** 235–240.

Robertson, A., 1967 The nature of quantitative variation, p. 265 in *Heritage from Mendel,* edited by R. A. Brink and E. D. Styles. University of Wisconsin Press, Madison, WI.

Smith, C., and S. P. Simpson, 1986 The use of genetic polymorphisms in livestock improvement. J. Anim. Breed. Genet. **103:** 205–217.

Spelman, R. J., W. Coppieters, L. Karim, J. A. M. Van Arendonk and H. Bovenhuis, 1997 Quantitative trait loci analysis for five milk production traits on chromosome 6 in the Dutch Holstein-Friesian population. Genetics **144:** 1799–1818.

Van Arendonk, J. A. M., H. Bovenhuis, S. Van Der Beek and A. F. Groen, 1994 Detection and exploitation of markers linked to quantitative traits in farm animals. Proceedings of the 5th World Congress on Genetics applied to Livestock Production, Guelph, Canada, pp. 193–200.

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., and C. S. Haley, 1996 Detection of quantitative trait loci in line crosses under infinitesimal genetic models. Theor. Appl. Genet. **93:** 691–702.

Wang, Y., 1995 Detection and estimation of associations between genetic markers and quantitative trait loci in segregating populations. Ph.D. Thesis, University of Guelph, Canada.

Weeks, D. E., T. Lehner, E. Squires-Wheeler, Ch. Kaufmann and J. Ott, 1990 Measuring the inflation of the Lod Score due to its maximization over model parameter values in human linkage analysis. Genet. Epidemiol. **7:** 237–243.

Weller, J. I., and R. L. Fernando, 1991 Strategies for the improvement of animal production using marker-assisted selection, p. 305 in *Gene Mapping: Strategies, Techniques and Applications,* edited by L. B. Schook, H. A. Lewin and D. G. McLaren. Marcel Dekker, New York.

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.

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

Communicating editor: Z-B. Zeng