Fine Mapping of Quantitative Trait Loci Using Linkage Disequilibria With Closely Linked Marker Loci

T. H. E. Meuwissen* and M. E. Goddard†

**Research Institute of Animal Science and Health, Box 65, 8200 AB Lelystad, The Netherlands and*
†*Institute of Land and Food Resources, University of Melbourne, Parkville, Victoria 3052, Australia*

Manuscript received August 24, 1999
Accepted for publication January 14, 2000

ABSTRACT

A multimarker linkage disequilibrium mapping method was developed for the fine mapping of quantitative trait loci (QTL) using a dense marker map. The method compares the expected covariances between haplotype effects given a postulated QTL position to the covariances that are found in the data. The expected covariances between the haplotype effects are proportional to the probability that the QTL position is identical by descent (IBD) given the marker haplotype information, which is calculated using the genedropping method. Simulation results showed that a QTL was correctly positioned within a region of 3, 1.5, or 0.75 cM in 70, 62, and 68%, respectively, of the replicates using markers spaced at intervals of 1, 0.5, and 0.25 cM, respectively. These results were rather insensitive to the number of generations since the QTL occurred and to the effective population size, except that 10 generations yielded rather poor estimates of the QTL position. The position estimates of this multimarker disequilibrium mapping method were more accurate than those from a single marker transmission disequilibrium test. A general approach for identifying QTL is suggested, where several stages of disequilibrium mapping are used with increasingly dense marker spacing.

LINKAGE disequilibrium mapping has been successful in mapping genetical disorders (*e.g.*, Hästbacka *et al.* 1992). The method attempts to find a chromosomal region that is identical by descent (IBD) among the diseased individuals, since such a region may carry the disease gene. The IBD region is detected by closely linked marker loci that carry identical alleles at this region in the diseased individuals (*e.g.*, Pritchard *et al.* 1991; Houwen *et al.* 1994). Hence, simultaneous linkage disequilibria between several closely linked markers and the disease gene are detected. The size of the IBD region decreases with the number of meioses since the disease mutation occurred and may be small, which leads to the detection of a small region that contains the disease gene.

A linkage analysis is often used for the mapping of quantitative trait loci (QTL), where the inheritance of chromosomal regions within the data set is traced by markers (see Hoeschele *et al.* 1997, for a review). The region whose inheritance explains most of the variance of the phenotypic records indicates the most likely position of the QTL. To position the QTL, linkage mapping uses only the recombinations that occurred within the data set, which typically contains two to three generations. With closely linked markers, there will be few recombinations between adjacent markers during these two to three generations and hence a dense marker map will provide little extra information about the position of the QTL, unless the number of individuals per generation is very large (Darvasi *et al.* 1993). In a high-resolution mapping experiment, even with the use of recombinant inbred lines, Long *et al.* (1995) could map only QTL affecting bristle numbers in Drosophila to regions of ~5–10 cM using linkage analysis. Linkage disequilibrium mapping uses all recombinations since the mutation occurred, which increases the precision of the estimate of the position. Linkage disequilibrium mapping methods seem, therefore, more useful for precise estimation of QTL positions, while linkage mapping is more useful for a genome-wide scan for QTL.

Linkage disequilibrium mapping methods for QTL that consider several markers have not been proposed in the literature, mainly because the carriers of the mutant QTL allele cannot be identified. This is the case unless the effect of the mutation is very large and the methods for qualitative traits such as disease genes can be applied. An additional complication is that the allele, which represents the most recent QTL mutation, is unknown. Hence, even if we could identify the QTL alleles, it is unknown which QTL allele should contain the identical marker alleles in a chromosome region. However, calculation of linkage disequilibria between QTL and marker loci does not require knowledge about which mutation is most recent and, therefore, can be considered for the positioning of the QTL.

The linkage disequilibrium between a single marker

Corresponding author: Theo Meuwissen, Department of Animal Breeding and Genetics, DLO-Institute for Animal Science and Health, Box 65, 8200 AB Lelystad, The Netherlands.
E-mail: t.h.e.meuwissen@id.dlo.nl

TABLE 1

Correlation matrix of haplotype effects

	Putative QTL between markers 1 & 2			
	1 q_1 1 1 1 1	2 q_2 2 2 2 2	1 q_1 1 1 2 2	2 q_2 2 2 1 1
1 q_1 1 1 1 1	1	0	1	0
2 q_2 2 2 2 2		1	0	1
1 q_1 1 1 2 2			1	0
2 q_2 2 2 1 1				1
	Putative QTL between markers 4 & 5			
	1 1 1 1 q_1 1	2 2 2 2 q_2 2	1 1 1 2 q_2 2	2 2 2 1 q_1 1
1 1 1 1 q_1 1	1	0	0	1
2 2 2 2 q_2 2		1	1	0
1 1 1 2 q_2 2			1	0
2 2 2 1 q_1 1				1

Correlation matrix of chromosomal effects of four marker haplotypes as a function of the QTL position is shown, when founder marker alleles (only alleles 1 and 2 are present in the sample of haplotypes) and founder QTL alleles existed in the base population. The double recombination rate is assumed negligible, which shows that the QTL allele numbers are equal to those of their surrounding markers. Note that the marker haplotypes are identical for both QTL positions. Founder alleles denote that all alleles were different in the original base population; *e.g.*, the N base animals had haplotypes of the type

$x\ x\ x\ x q_x x$ with $x = 1, 2, 3, \ldots, 2N$.

and a QTL can be measured by estimating the effect of the marker on the quantitative trait in a regression analysis. This approach is extended to multiple marker loci by estimating the effect of marker haplotypes on the quantitative trait. The marker haplotypes that have identical marker alleles in a region surrounding the QTL are expected to show similar haplotype effects, since the identical markers indicate that the region is IBD and thus the haplotypes are expected to carry similar QTL alleles. In statistical terms, similar haplotype effects imply that the covariance between the haplotype effects is high. Whether two marker haplotypes have identical alleles in a region surrounding the QTL depends on the position of the QTL, and hence, the covariance between the haplotype effects depends on the position of the QTL. This dependence of haplotype covariances on the position of the QTL is illustrated in Table 1 for a simple situation, where marker haplotypes identify QTL alleles with certainty; *i.e.*, correlations among the haplotypes are either 0 or 1. The general principle, however, also holds under less simple conditions. The covariances between haplotype effects can thus be used to position the QTL. The aim of this article is to formally develop this method for the estimation of the QTL position using the linkage disequilibria between several closely linked markers and the QTL. The accuracy of the method in finding the correct QTL position is investigated mainly in the context of livestock populations but the method (and results) can clearly be applied also to other types of populations. Last, the precision of the method is compared to that of a single marker transmission disequilibrium test (TDT) analysis (Rabinowitz 1997).

METHODS

General: It is assumed here that a linkage mapping study has narrowed the position of the QTL down to a 5- to 20-cM region. Within this region, there are many markers available, with a between-marker spacing of typically 0.25–1 cM. Linkage disequilibrium mapping is used to find the most likely marker bracket, *i.e.*, the region between two adjacent markers, that contains the QTL. In principle the method can also be used to estimate the position of the QTL within a marker bracket,

but it seems that there is too little information for such a precise estimate.

Maximum-likelihood estimation of QTL position: It is assumed that phenotypic records from only the last generation of individuals are used in the analysis. Marker data are available on this generation and their parents so that for many individuals it is known which of the marker alleles is paternally and which is maternally derived; *i.e.*, the linkage phases of the marker alleles are known and marker haplotypes can be constructed. For example, when the individual has genotype M_1M_2/N_1N_2 and its parents have M_1M_3/N_1N_3 and M_2M_4/N_2N_4, the linkage phase of the individual is M_1N_1/M_2N_2. If the linkage phase is not known, the haplotype is missing and the individual is excluded from the analysis. However, in the case of highly polymorphic closely linked markers (such that double recombinations are unlikely), the correct linkage phase can often be assigned with a high probability.

The phenotypic records of the last generation are modeled by

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zh} + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{y}$ is the vector of records; $\boldsymbol{b}$ is the vector of fixed (nuisance) effects for which the data are to be corrected; $\boldsymbol{h} = (q \times 1)$ vector of random effects of the haplotypes; $\boldsymbol{e}$ is the vector of residuals; and $\boldsymbol{X}$ and $\boldsymbol{Z}$ are known incidence matrices for the effects in $\boldsymbol{b}$ and $\boldsymbol{h}$, respectively. The variance of the residuals is $\mathrm{Var}(\boldsymbol{e}) = \sigma_e^2 \boldsymbol{R}$, where $\boldsymbol{R}$ is assumed here to be an identity matrix, but in general $\boldsymbol{R}$ can account for covariances between residuals, which may be due to background genes and family relationships. The variance of the haplotype effects is $\mathrm{Var}(\boldsymbol{h}) = \sigma_h^2 \boldsymbol{H}_p$, where the matrix $\boldsymbol{H}_p$ yields the (co)variances of the haplotype effects up to proportionality and subscript p indicates that $\boldsymbol{H}_p$ depends on the assumed position of the QTL. The dimension of $\boldsymbol{H}_p$ is $q*q$, where q is the number of different haplotypes in the data.

Assuming multivariate normality, the residual log-likelihood of the data under the above model is

$$L(\boldsymbol{H}_p, \sigma_h^2, \sigma_e^2) \propto -0.5[\ln(|\boldsymbol{V}|) + \ln(|\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}|) + (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}})'\, \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}})] \tag{2}$$

(Patterson and Thompson 1971), where $\boldsymbol{V} = \mathrm{Var}(\boldsymbol{y}) = [\boldsymbol{Z}\boldsymbol{H}_p\boldsymbol{Z}'\,\sigma_h^2 + \boldsymbol{R}\,\sigma_e^2]$, and $\hat{\boldsymbol{b}}$ is the generalized least-squares estimate of $\boldsymbol{b}$. The term $\ln(|\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}|)$ corrects for the fact that fixed effects are estimated instead of known and is redundant when fixed effects are absent or fixed-effect classes are large; *i.e.*, estimation of fixed effects is accurate. Given a QTL position, p, *i.e.*, given $\boldsymbol{H}_p$, this likelihood is maximized to obtain estimates of the variance components $\hat{\sigma}_h^2$ and $\hat{\sigma}_e^2$ (see, *e.g.*, Henderson 1984). Algorithms that require the inverse of the $\boldsymbol{H}_p$ matrix should be avoided for the residual maximum likelihood (REML) estimation of the variance components, because $\boldsymbol{H}_p$ will be (close to) singular when the distances between the markers become small; *i.e.*, methods that require the inverse of $\boldsymbol{H}_p$ may be numerically unstable.

The likelihood $L(\boldsymbol{H}_p, \hat{\sigma}_h^2, \hat{\sigma}_e^2)$ can be calculated for every position of the QTL. The maximum-likelihood estimate of the QTL is the position where $L(\boldsymbol{H}_p, \hat{\sigma}_h^2, \hat{\sigma}_e^2)$ is highest. To calculate $L(\boldsymbol{H}_p, \hat{\sigma}_h^2, \hat{\sigma}_e^2)$, we need to obtain the (co)variance matrix of the haplotype effects, $\boldsymbol{H}_p$, given the position of the QTL.

Calculation of $\boldsymbol{H}_p$: The covariance between two haplotypes effects, h_i and h_j, is

$$\mathrm{Cov}(h_i, h_j) = \mathrm{Prob}(\mathrm{IBD}|\text{marker haplotypes}) \times \sigma_h^2,$$

where Prob(IBD|marker haplotypes) is the probability that the QTL locus is IBD given the marker haplotypes. The probability that a locus is IBD given the haplotypes surrounding the locus may be obtained from using the coalescence process (Hudson 1985, 1993), but this proved complex for multiple markers and the IBD probabilities were difficult to obtain. Here, we used the genedropping method (Maccluer *et al.* 1986) to obtain the IBD probabilities given the haplotype information.

In the genedropping method, markers and a putative QTL are simulated in a base generation. All $2N_e$ base generation QTL alleles, which are called founder alleles, have a unique number. The next N_G descendant generations are simulated by choosing at random parents from the previous generation and letting their N_e offspring inherit haplotypes or recombinant haplotypes according to Mendel's rules and the recombination probabilities. The parents of each progeny are randomly sampled such that the effective population size is N_e. Because the founder QTL alleles have unique numbers, any two QTL alleles with the same number in generation N_G are IBD. N_G is the number of generations that passed since the mutation occurred.

A problem with this application of the genedropping method is that after N_G generations, we will not obtain the same haplotypes as in our data set, because there are too many possible haplotypes. For instance, with 10 biallelic markers there are 1024 possible haplotypes and hence 1024×1024 covariances to be estimated. Fortunately many of these covariances are expected to be equal and so can be grouped together and a single value is estimated for the group. To assess the probability that haplotypes i and j contain QTL alleles that are IBD, we move along the chromosome away from the QTL locus and find a marker that has, say, allele M_1 for haplotype i and allele M_2 for haplotype j. The haplotypes are clearly not IBD at this marker locus, M. Hence, if there was an IBD region around the QTL, it has ended before marker M. Any equality or nonequality of alleles at markers further away from the QTL than locus M does not affect the probability that the QTL locus is IBD. Let N_l denote the number of markers for which two haplotypes have identical alleles, if we start at the QTL position and count toward the left until the first nonidentical marker alleles occur ($N_l = 0, 1, 2, \ldots$). Similarly, let N_r denote

TABLE 2

Details of the simulations

No. of generations since mutation occurred:	10, 50, <u>100</u>, 200, or 1030
Selective advantage of mutation:	0
Effective size of population:	N_e = 50, <u>100</u>, 200, 1000
No. of markers:	10
No. of alleles per marker (initial frequency of marker alleles):	2 (0.5)
Distance between adjacent markers:	1, 0.5 or 0.25 cM
Frequency of QTL	
In generation 0:	$1/2N_e$
In last generation:	>0 or <u>≥0.1</u>
Position of QTL:	In the middle between markers 5 and 6
Additive effect of one positive QTL allele:	1
Dominance effect of QTL allele:	0
Residual standard deviation:	1
Records measured on (no. of individuals):	<u>100</u> or 500
No. of replicates used to estimate $\boldsymbol{H}_p$:	100,000

The default simulation is underlined.

the number of markers that carry identical alleles on the right side of the QTL until the first nonidentical marker alleles occur. Now we can classify all haplotype pairs into groups with equal (N_l,N_r), which are expected to have equal probabilities of being IBD at the marker locus. For instance, the probability that haplotypes (11111*Q*11111) and (22211*Q*111222) are IBD at the QTL is the same as the probability that haplotypes (22222*Q*22222) and (111122*Q*22211) share a IBD QTL, because both pairs have two identical markers to the left of the QTL and three to the right of the QTL [*i.e.*, (N_l,N_r) = (2, 3)].

With the genedropping method the IBD probabilities of a pair of haplotypes can be estimated within each genedrop by dividing the number of times the QTL locus was IBD by the total number of times the haplotype pair was found. The estimates of the IBD probabilities of the haplotype pairs that belong to the same (N_l,N_r) group are averaged within a genedrop, and these averages are accumulated across 100,000 repeated genedrops to obtain the estimates of the IBD probabilities for every haplotype combination group (N_l,N_r).

Testing the linkage disequilibrium mapping method: An analysis of simulated data, where the correct position of the QTL is known, is used to test the proposed mapping method. Details of the simulation are provided in Table 2. The simulation of the base and later generations was as with the genedropping method (see previous section). In the last generation, at random, one founder QTL allele is chosen among the group of surviving founder QTL alleles, and this allele obtains a value of 1 while all others obtain a value of 0. The latter results in a frequency of the positive QTL allele that is >0 (otherwise there was no polymorphic QTL to detect). Because the QTL was previously detected by linkage analysis study, it seems reasonable to assume that the frequency of the positive QTL allele is >0.1. The requirement that the QTL allele frequency is >0.1 is studied by choosing at random one founder allele among the group of founder alleles with a frequency >0.1 and giving this chosen founder allele an effect of 1. Marker haplotypes were known without error in the simulation study. One record per individual of the last generation was obtained by adding to the sum of both QTL allele effects of the individual an environmental effect that was sampled from $N(0, 1)$. In the last generation there were 100 or 500 instead of N_e individuals to avoid confounding between the amount of information in the data set and the effective size of the population.

RESULTS

Number of generations since mutation and effective size are 100: Table 3 shows the estimated positions of the QTL, when the QTL mutation occurred 100 generations ago and the effective population size was 100, which was also used to calculate $\boldsymbol{H}_p$. In 31–35 out of 50 replicated simulations, the estimated position was in the correct marker bracket or the neighboring one. This number decreased somewhat with a decreasing size of the marker brackets. However, with smaller bracket sizes the position estimates expressed in centimorgans became more precise. It should be noted also that the prior estimate of the QTL position was more precise with the smaller bracket sizes. Prior to the analysis, it was assumed known that the QTL was somewhere between the 10 markers, which span a region of 9, 4.5, and 2.25 cM if the distances between the markers are 1, 0.5, and 0.25 cM, respectively. The requirement that the allele frequency of the QTL in the last generation exceeds 0.1, which accounts for the fact that detected QTL are probably not rare, increased the precision of the estimates somewhat.

The estimated position was more often four brackets

TABLE 3

Precision of QTL position estimates

Between-marker distance (cM)	Deviation of estimated from correct position[a]					Total
	0	1	2	3	4	
	Replicates with frequency of QTL > 0.1					
1.0	12	23	6	6	3	50
0.5	14	17	11	3	5	50
0.25	14	20	7	3	6	50
	Replicates with frequency of QTL > 0					
1.0	12	23	5	2	8	50
0.5	12	16	9	3	10	50
0.25	5	22	6	8	9	50

The estimated position of the QTL relative to the correct position, when the QTL occurred 100 generations ago, is shown. The effective size of the population was 100, and the residual variance was 1. The number of replicated simulations in which the indicated position was estimated is presented. The total number of replicates is 50.

[a] Measured in no. of marker brackets; *i.e.*, 0 indicates that the estimated position was in the correct marker bracket, 1 indicates that the estimated position was in the bracket next to the correct position, etc.

away from the correct position than three (Table 3), which seems counterintuitive. However, a close examination of some replicates whose estimates deviated four brackets from the correct position showed that the QTL carrying haplotypes often had identical alleles at markers 1–5 (or 6–10). Hence, the QTL appears to be somewhere between markers 1 and 5 (or 6 and 10). The method predicts the highest covariances between marker haplotypes, which share alleles at markers 1–5 when the QTL is in the leftmost bracket. Hence, the QTL is predicted in the leftmost bracket, which deviates four brackets from the correct bracket. The situation that all QTL carrying haplotypes have identical marker alleles at positions 1–5 occurs more often when there are few QTL carrying haplotypes, *e.g.*, in replicates with a low frequency of the QTL (Table 3).

Figure 1 shows the average log-likelihood curve of the 50 replicated simulations as a deviation from the base likelihood, where the base likelihood is calculated using Equation 2 but with $\sigma_h^2 = 0$, *i.e.*, without fitting haplotype effects. The average likelihood shows a bell-shaped curve with the peak at the QTL position. The symmetry of the curve suggests that the estimation of the QTL position is approximately unbiased. The maximum of the likelihood increases with smaller marker bracket sizes and the curve becomes more peaked. The likelihood at the brackets that are adjacent to the QTL-carrying bracket is, however, not much lower than that at the QTL-carrying bracket, which indicates that there will be little information to distinguish the correct bracket from its neighboring brackets. The difference in likelihood between the model with a QTL and the model without a QTL anywhere in the marked region is evidence for the existence of the QTL. Since this evidence will usually be independent of the linkage analysis that originally mapped the QTL, this likelihood-ratio test, based on linkage disequilibrium, will help to confirm the existence of the QTL.

In Table 3, the actual number of individuals, and thus phenotypic records, was equal to the effective number (100). In practice, the actual number of individuals may be much larger than the effective number. Hence, the phenotypic number of records can also exceed the effective number of animals, which results in more records per haplotype, *i.e.*, more accurate estimates of haplotype effects. If there are 500 animals with records in the last generation, the number of position estimates that are in the correct or a neighboring bracket is increased to 38–39 out of 50 replicated simulations when the bracket size is 1 or 0.5 cM (Table 4). However, if the bracket size is 0.25 cM, the mapping precision is not increased compared to the situation with 100 recorded animals.

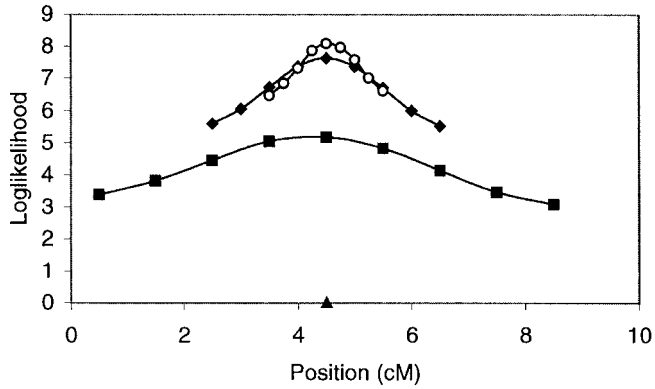

Figure 1.—The log-likelihood of a QTL minus that of having no QTL averaged over 50 replicated simulations, with marker bracket sizes of 1, 0.5, and 0.25 cM. The number of generations since the QTL mutation and the effective population size were both 100.

(■) 1-cM brackets, (◆) 0.5-cM brackets, (○) 0.25-cM brackets, (▲) QTL position.

TABLE 4

Precision of QTL position estimates when the number of records was increased

Between-marker distance (cM)	Deviation of estimated from correct position (no. of brackets)					
	0	1	2	3	4	Total
1.0	18	20	7	3	2	50
0.5	20	19	5	1	5	50
0.25	13	19	5	6	7	50

The estimated position of the QTL, when the number of recorded individuals was 500 while the effective population size was 100 and the mutation occurred 100 generations ago, is shown. The frequency of the positive QTL allele was >0.1 in all 50 replicated simulations.

The latter may be because there are too few (detectable) recombinations close to the QTL in the case of 0.25-cM brackets, such that an increased accuracy of the estimation of the effect of a haplotype hardly improves the accuracy of the position estimates.

Different numbers of generations since the mutation: Table 5 shows the fractions of the QTL position estimates that are in the correct or in the neighboring bracket when the time since the mutation is varied (N_G = 200, 100, 50, or 10). Generally, a N_G of 100 yielded the most precise QTL estimates. The QTL estimates with N_G = 200 were slightly less precise probably because the inbreeding was getting slightly too high; *i.e.*, the number of segregating haplotypes was reduced. After 200 generations at an effective size of 100 the inbreeding coefficient is 0.63. A N_G of 50 yielded QTL position estimates when the bracket size was 0.25 cM much less precise than the higher N_G values. This is probably because there have been too few recombinations between adjacent markers at this small bracket size. When N_G is only 10 generations the position estimates become no better than chance, which is probably again due to too few recombinations.

In previous simulations it was assumed that the number of generations since the QTL mutation occurred was known when calculating $\boldsymbol{H}_p$, whereas this parameter is unknown in practice. However, when $\boldsymbol{H}_p$ is calculated assuming a N_G of 100, the precision of the QTL position estimates decreases only slightly (Table 5). In fact, for N_G = 10 the precision is higher when N_G = 100 instead of the true value is used to calculate $\boldsymbol{H}_p$. It seems counterintuitive that using a wrong N_G for estimating $\boldsymbol{H}_p$ results in more precise QTL estimates here. However, this is probably because the contrasts between the haplotype effects that result in estimating the QTL position are much more pronounced in the $\boldsymbol{H}_p$ matrix with N_G = 100 than in that with N_G = 10, which may lead to little covariance between different haplotypes since recombination probabilities are small. Hence, the $\boldsymbol{H}_p$ assuming N_G = 100 may have contrasted the small differences between haplotypes that result in QTL position estimates more, while the differences in likelihood with an $\boldsymbol{H}_p$ matrix assuming N_G = 10 may be more affected by sampling errors on the estimates of $\boldsymbol{H}_p$ matrices at different positions. In general, however, Table 5 shows that assuming N_G = 100 when estimating $\boldsymbol{H}_p$ results in close to optimal position estimates.

Different numbers of effective population sizes: Table 6 shows the results when the effective population size, N_e, was either 50 or 200. The results at N_e equal to 50 or 200 are similar to those at the N_e of 100, except at the marker bracket size of 0.25 cM, where the N_e =

TABLE 5

Precision of QTL position estimates with varying numbers of generations since the mutation

	$\boldsymbol{H}_p(N_G)$[a]			$\boldsymbol{H}_p(100)$[a]		
Bracket size (cM):	1	0.5	0.25	1	0.5	0.25
N_G = 200	0.62	0.64	0.62	0.60	0.62	0.68
100	0.70	0.62	0.68		Same as $\boldsymbol{H}(N_G)$	
50	0.66	0.56	0.34	0.70	0.60	0.30
10	0.32	0.30	0.28	0.48	0.50	0.36

The fraction of the replicated simulations, where the estimate of the QTL position was in the correct bracket or in a bracket next to the correct bracket, is shown. The number of generations since the mutation at the QTL occurred (N_G) is varied; the effective size of the population was 100; and the frequency of the QTL in the last generation was >0.1.

[a] $\boldsymbol{H}_p(x)$ denotes that the matrix of (co)variances of haplotype effects is estimated assuming that the mutation occurred x generations ago.

TABLE 6

Precision of QTL position estimates with varying effective population sizes

	$H_p(N_e)^b$			$H_p(100)^b$		
Bracket size (cM):	1	0.5	0.25	1	0.5	0.25
N_e = 200	0.64	0.60	0.44	0.70	0.60	0.44
50	0.68	0.70	0.46	0.62	0.68	0.50
1000/100[a]	0.48	0.52	0.60	0.50	0.58	0.58

The fraction of the replicated simulations, where the estimate of the QTL position was in the correct bracket or in a bracket next to the correct bracket, is shown. The effective size of the population is varied; the number of generations since the mutation at the QTL occurred is 100 or 1030; and the frequency of the QTL in the last generation was >0.1.

[a] In this case, there are 1000 generations at N_e =1000 and the last 30 generations are at N_e = 100, which may be realistic for livestock populations.

[b] $\boldsymbol{H}_p(x)$ denotes that the matrix of (co)variances of haplotype effects is estimated assuming an effective size of x.

50 or 200 schemes yielded somewhat less precise estimates. At N_e = 50 genetic drift is high, which makes the frequency of large chromosomal segments drift rapidly before recombination can reduce their sizes. The latter hampers mainly the mapping precision when small bracket sizes are used. At N_e = 200 there is little genetic drift of haplotypes, such that a steady-state linkage disequilibrium due to drift may not have occurred yet. The establishing of a steady-state linkage disequilibrium is probably slower when recombination frequencies are small; hence, the mapping precision is more reduced at small marker bracket sizes.

Also in Table 6, a situation in which a QTL mutation occurred 1030 generations ago is investigated, where the effective size was 1000 during the first 1000 generations and 100 during the last 30 generations. In livestock populations this may be realistic, where the recent bottleneck is due to the introduction of herd books. The bottleneck seems to have favored the mapping precision at the small bracket size of 0.25 cM while the precision at the larger bracket sizes is reduced. The long time period since the mutation probably resulted in a small IBD region around the QTL, such that the further away markers will show too little linkage disequilibrium with the QTL (although this linkage disequilibrium is increased by the recent bottleneck).

Comparison to single-marker disequilibrium mapping: Table 7 compares position estimates when the last generation consists of a half-sib family structure, which is obtained by mating 10 randomly chosen males each to 10 different females. This half-sib family structure was also analyzed using the transmission disequilibrium test (Rabinowitz 1997). The TDT was applied by fitting a model with a fixed half-sib family effect and the effect of a single marker to the data, where the half-sib family effect ensures that the TDT uses only within-half-sib family deviations for the estimation of the marker effect. All markers were fitted in turn and the marker that resulted in the highest likelihood of the data was expected to be closest to the QTL. The half-sib family structure yields somewhat less precise QTL position estimation than the unstructured population of Table 3. This may be in part due to the use of the $\boldsymbol{H}_p$ matrices of the unstructured population for the analysis in Table 7. A more correct analysis with the multimarker method would require the use of a $\boldsymbol{H}_p$ matrix within and across

TABLE 7

Precision of QTL position estimates when single or multiple markers are used

	No. of markers used simultaneously	
Bracket size (cM)	1	All
1.0	0.44	0.58
0.5	0.56	0.64
0.25	0.42	0.56

The fraction of the replicated simulations, where the multimarker disequilibrium mapping estimate of the QTL position was in the correct bracket or in a bracket next to the correct bracket, and where the single-marker estimate was not more than two markers away from the QTL position, is shown. In the last generation a half-sib design was obtained by mating 10 randomly chosen males each to 10 females, yielding a total of 100 offspring. The effective size of the population and the number of generations since the mutation were each 100.

half-sib families. Despite the use of the slightly wrong $\boldsymbol{H}_p$ matrices, the multimarker method yielded a 13–33% higher probability of positioning the QTL in the correct or an adjacent bracket than the TDT. It may be noted that the position estimates of the TDT method are at the marker position instead of in a bracket, which complicates an exact comparison of the precision of the position estimates somewhat. In the comparison of Table 7, the TDT position estimates that are at the boundaries of the brackets that are next to the QTL bracket are also counted as within these brackets.

DISCUSSION

Relations with other methods and extensions: A multiple marker-QTL linkage disequilibrium mapping method that often positioned a QTL within one marker bracket from its correct position was presented, where the size of the brackets was 1, 0.5, or 0.25 cM. Hence, the QTL is probably in the estimated bracket or in the bracket next to it, which implies a region of 3, 1.5, or 0.75 cM, respectively. Reducing the size of the marker brackets can thus reduce the size of this region (see also Figure 1). The method may be seen as an extension of the multipoint linkage disequilibrium mapping method for discrete traits (Terwilliger 1995) toward QTL. However, Terwilliger's discrete trait mapping method was a full maximum-likelihood method, whereas the present QTL mapping method is approximately maximum likelihood, due to the multivariate normality assumption involved in likelihood (2). This assumption implies that the method uses only the first two moments of the data for the estimation of the QTL position, which may be satisfactory in many situations because the higher moments of the data are often unknown and often contain little extra information.

The method presented in this article is also an extension of the single-marker-QTL TDT method (Rabinowitz 1997) toward using all markers simultaneously for the estimation of the QTL positions. This extension to more markers increased the accuracy of the position estimates substantially (Table 7). In the comparison of Table 7 the TDT method had a disadvantage because it used only within-half-sib family deviations for the estimation of the marker effects. This avoids family effects causing bias in the estimated marker effects in the TDT. In the simulations there were no family effects other than those due to the QTL, but in practice these would be likely to occur. Hence, in practice, the multimarker mapping method also has to correct for the effects of polygenetic and environmental family effects. Possible environmental family effects are easily included in the fixed-effect structure of model (1). A polygenic term should be included in the random effects of model (1) to account for the background genes.

A disadvantage of the multimarker mapping method in the comparison of Table 7 is that the $\boldsymbol{H}_p$ matrices were not adapted to the half-sib family structure. Estimation of the $\boldsymbol{H}_p$ matrices within and across families would overcome this problem. However, in practice the family structure will often be more complex than that of full- or half-sib families. This may be overcome by fitting haplotype × individual effects in model (1); *i.e.*, there will be two haplotype effects estimated for every individual, and the IBD probabilities have to be calculated conditional on both the markers and the pedigree. The latter is an extension of the model of Fernando and Grossman (1989), who fitted two QTL effects per individual and used a covariance matrix of these QTL effects that was proportional to IBD probabilities. In the model of Fernando and Grossman, identity by descent, however, occurred only when a common ancestor was found within the marked pedigree, not before pedigree recording started. The latter makes Fernando and Grossman's method useful for linkage analysis mapping (see Hoeschele *et al.* 1997). We are in the process of combining the current IBD probabilities with those of Fernando and Grossman, which will combine the information from linkage disequilibrium mapping with that of linkage analysis.

The information from linkage disequilibria: Hill and Weir (1994) showed that the variance of the linkage disequilibrium between a closely linked marker and a QTL is large, such that the disequilibrium cannot be used for the precise mapping of the QTL. The situation investigated here differs from that of Hill and Weir in two respects. First, we use the disequilibria between all markers and the QTL simultaneously, which seems to avoid the problem of a high variability of a single linkage disequilibrium. Second, Hill and Weir assumed that the linkage disequilibrium was caused by random genetic drift, whereas, in the present study, the disequilibrium was initially created by the occurrence of a mutation in a single haplotype, which is probably less variable. Hence, the present population was not (yet) in equilibrium. Kaplan *et al.* (1995) investigated linkage disequilibrium mapping of disease genes and concluded that this method can be very precise in locating a gene in nonequilibrium populations.

There are three factors that limit the accuracy with which the method can map a QTL. First, if there are too few animals, the haplotype effects are estimated too poorly. Second, if chromosome segments surrounding the QTL are too big, the linkage disequilibrium will be uniform throughout the marker region. In that case we will be able to detect the presence of the QTL but not position it accurately. Third, if the chromosome segments surrounding the QTL are smaller than the marker brackets, then we may not detect any disequilibrium between the markers and the QTL. The chromosome segments will be large if either N_G or N_e is small. If N_G is small, there will be too few opportunities for recombination to break up the chromosome segment that contained the original QTL mutation. If N_e is small,

then all existing copies of a QTL allele are likely to trace back to a common ancestor a small number of generations ago, again limiting the opportunities for recombination.

The situation with a population at a size of 1000 individuals for 1000 generations and a size of 100 individuals for the last 30 generations (Table 7) showed a high fraction of correct marker brackets at a bracket size of 0.25 cM while this fraction reduced when the bracket size became bigger. This suggests that the linkage disequilibrium between the markers, which were distanced at 1 cM, eroded during the 1000 generations at the high population size, while the 30 generations at a population size of 100 were not long enough to regenerate these linkage disequilibria.

The initial linkage disequilibrium due to the new mutation breaks down at a rate of $(1 - r)$ per generation, where r is the recombination between the marker and the QTL. When the distance between the marker and the QTL is 1 or 0.25 cM, then 37 or 78%, respectively, of the disequilibrium remains after 100 generations. After 200 generations these figures are 14 and 60%, respectively. Hence, after many generations the initial disequilibrium disappears unless the marker is very close to the QTL. However, in Table 6, marker brackets of 1 cM still achieved a reasonable accuracy even when the QTL mutation occurred 200 generations ago. This suggests that the disequilibrium due to random drift was also used by the method to estimate the position of the QTL. Hence, disequilibrium due to finite size is building up as that caused by the initial mutation is eroding. Goddard (1991) showed that finite population size alone could generate substantial linkage disequilibrium between a QTL and a marker haplotype and this may explain why the method is not highly sensitive to the number of generations since the mutation.

Conditions affecting mapping precision: The results in Tables 6 and 7 showed that it is difficult to predict which population size should be used for the mapping of a QTL, especially since the age of the mutation is generally unknown. The combination of 100 generations at an effective size of 100 seemed to result in good QTL position estimates for the marker bracket sizes that were investigated here. However, other effective population sizes and numbers of generations since the mutation did not reduce the precision markedly, except that 10 generations seem to be too short to yield good QTL position estimates with linkage disequilibrium mapping. Mutations that occurred just 10 generations ago might not be so likely, but the QTL may have been introduced into the population by migration of an individual. Multiple mutations at the QTL can also generate substantial linkage disequilibrium in multimarker haplotypes because every mutation will be introduced in a more or less unique haplotype, and its effect will cause a covariance pattern as indicated by the $\boldsymbol{H}_p$ matrix. Hence, the precision of the multimarker linkage disequilibrium mapping method will hardly be affected by multiple QTL mutations if the haplotypes surrounding the mutations are sufficiently unique.

Mutations at the markers were ignored here, but their effect is similar to that of recombination between the markers; namely, the haplotype that is associated with the QTL is altered. Since mutations occur much less frequently than recombination, it seems sufficient to account for only the recombination between the markers. Also, the statistical model for the estimation of the haplotype effects assumes that the effects of the paternal and maternal haplotypes are additive. This does not imply that the effect of the QTL alleles needs to be additive, but only the average effect of the haplotypes across all genotypes will be used by the analysis. Strong dominance deviations may reduce or increase this average effect and can therefore affect the power of the analysis. This reduction or increase of the average effects depends also on the frequencies of the QTL alleles (Falconer 1989).

Another effect that limits the precision of estimates of QTL position is the calculation of $\boldsymbol{H}_p$. If the distance between the markers is small, the fraction of the replicates where recombination between two particular markers occurs also becomes small, which implies that many replicates are needed. In the present simulations 100,000 replicated simulations were sufficient to estimate IBD probabilities for marker bracket sizes down to 0.25 cM. Although modern computers are fast, the number of replicated simulations needed to calculate $\boldsymbol{H}_p$ may limit the minimum distance between the markers. Accurate deterministic methods to assess $\boldsymbol{H}_p$, which may be based on coalescence processes (Hudson 1985, 1993), will be useful to improve the precision of the present fine mapping technique. On the other hand, the estimation of $\boldsymbol{H}_p$ by simulation is probably more flexible than a deterministic prediction and thus can be easily adjusted to any known (changes in) structure of the real population.

Tables 3 and 4 show that the accuracy of the estimates of QTL positions is only moderately increased when there are more records available in a situation with equal effective population size. The benefit from increasing the number of animals would probably increase if the size of the QTL effect relative to the error variance was reduced. However, the results show that in some cases the structure of the population is the limiting factor in the precision of position estimation, *e.g.*, too few recombinations since the QTL occurred.

A multi-stage approach to QTL mapping: A practical method for estimating the position of a QTL would be to start first with a genome-wide scan for QTL using linkage analysis methods. This would result in a region of, say, 20 cM, which most likely contains the QTL. Next, this region will be covered with markers that are 1 cM apart, and the present linkage disequilibrium mapping

method will probably narrow the position of the QTL down to 3 cM. It will also be useful to test for the existence of the QTL using the likelihood-ratio test to confirm that there is a QTL in this region. This linkage disequilibrium-based test uses different information than the tests that were performed in the linkage analysis. In the following step, this 3-cM region is covered by markers that are 0.25 cM apart, and the position of the QTL is narrowed down to a 0.75-cM region. Within this region, even smaller marker brackets may be used to narrow down further the position of the QTL. Finally, comparative mapping (O'Brien *et al.* 1993), evolutionary tree mapping (Templeton *et al.* 1987), and/or positional cloning (*e.g.*, Wallace *et al.* 1990) may be used to identify the QTL.

LITERATURE CITED

Darvasi, A., A. Weinreb, V. Minke, J. I. Weller and M. Soller, 1993 Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics **134:** 943–951.

Falconer, D. S., 1989 *An Introduction to Quantitative Genetics*. Longman Group, Essex, United Kingdom.

Fernando, R. L., and M. Grossman, 1989 Marker-assisted selection using best linear unbiased prediction. Genet. Sel. Evol. **21:** 246–477.

Goddard, M. E., 1991 Mapping genes for quantitative traits using linkage disequilibrium. Genet. Sel. Evol. **23** (Suppl. 1): 131s–134s.

Hastbacka, J., A. De La Chapelle, I. Kaitila, P. Sistonen, A. Waever *et al.*, 1992 Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nat. Genet. **2:** 204–211.

Henderson, C. R., 1984 *Applications of Linear Models in Animal Breeding.* University of Guelph, Guelph, Ontario, Canada.

Hill, W. G., and B. S. Weir, 1994 Maximum-likelihood estimation of gene location by linkage disequilibrium. Am. J. Hum. Genet. **54:** 705–714.

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

Houwen, R. H. L., S. Baharloo, K. Blankenship, P. Raeymaekers, J. Juyn *et al.*, 1994 Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. Nat. Genet. **8:** 380–386.

Hudson, R. R., 1985 The sampling distribution of linkage disequilibrium under an infinite alleles model without selection. Genetics **109:** 611–631.

Hudson, R. R., 1993 The how and why of generating gene genealogies, pp. 23–36 in *Mechanics of Molecular Evolution*, edited by N. Takahata and A. G. Clarck. Sinauer, Sunderland, MA.

Kaplan, N. L., W. G. Hill and B. S. Weir, 1995 Likelihood methods for locating disease genes in nonequilibrium populations. Am. J. Hum. Genet. **56:** 18–32.

Long, A. D., S. L. Mullaney, L. A. Reid, J. D. Fry, C. H. Langley *et al.*, 1995 High resolution mapping of genetic factors affecting abdominal bristle number in Drosophila melanogaster. Genetics **139:** 1273–1291.

Maccluer, J. W., J. L. Vandeberg, B. Read and O. A. Ryder, 1986 Pedigree analysis by computer simulation. Zoo Biol. **5:** 147–160.

O'Brien, S. J., J. E. Womack, L. A. Lyons, K. J. Moore, N. A. Jenkins *et al.*, 1993 Anchored reference loci for comparative genome mapping in mammals. Nat. Genet. **3:** 103–112.

Patterson, H. D., and R. Thompson, 1971 Recovery of inter-block information when block sizes are equal. Biometrika **58:** 545–554.

Pritchard, C., D. R. Cox and R. M. Myers, 1991 The end in sight for Huntington disease? Am. J. Hum. Genet. **49:** 1–6.

Rabinowitz, D., 1997 A transmission disequilibrium test for quantitative trait loci. Hum. Hered. **47:** 342–350.

Templeton, A. R., E. Boerwinkle and C. F. Sing, 1987 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in Drosophila. Genetics **117:** 343–351.

Terwilliger, J. D., 1995 A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. Am. J. Hum. Genet. **56:** 777–787.

Wallace, M. R., D. A. Marchuck, L. B. Anderson, R. Letcher, H. M. Odeh *et al.*, 1990 Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. Science **249:** 181–186.

Communicating editor: C. Haley