

Multitrait Fine Mapping of Quantitative Trait Loci Using Combined Linkage Disequilibria and Linkage Analysis

M. S. Lund,^{*,1} P. Sørensen,[†] B. Guldbrandtsen^{*} and D. A. Sørensen^{*}

^{*}Danish Institute of Agricultural Sciences, Department of Animal Breeding and Genetics, Research Centre Foulum, DK-8830 Tjele, Denmark and [†]Department of Dairy Science, Virginia Polytechnic and State University, Blacksburg, Virginia 24061-0315

Manuscript received April 25, 2002
Accepted for publication October 14, 2002

ABSTRACT

A novel multitrait fine-mapping method is presented. The method is implemented by a model that treats QTL effects as random variables. The covariance matrix of allelic effects is proportional to the IBD matrix, where each element is the probability that a pair of alleles is identical by descent, given marker information and QTL position. These probabilities are calculated on the basis of similarities of marker haplotypes of individuals of the first generation of genotyped individuals, using "gene dropping" (linkage disequilibrium) and transmission of markers from genotyped parents to genotyped offspring (linkage). A small simulation study based on a granddaughter design was carried out to illustrate that the method provides accurate estimates of QTL position. Results from the simulation also indicate that it is possible to distinguish between a model postulating one pleiotropic QTL affecting two traits *vs.* one postulating two closely linked loci, each affecting one of the traits.

IN recent years a number of genome scans have been reported. The aim of these studies was to identify regions on the genome with genes affecting traits of interest. This is achieved by linkage analysis, which utilizes information on recombination events that occur between genetic markers when gametes are formed and transmitted from parents to offspring. The low number of recombination events between closely linked markers restricts the precision of quantitative trait loci (QTL) mapping using linkage analysis. In genome scans in livestock populations a QTL can typically be mapped to a chromosomal region of 10–30 cM. To utilize QTL in selective breeding, or to identify functional genes, a higher level of resolution of position estimates is required.

An assumption common to linkage analysis methods is that founder individuals with no parents are unrelated and noninbred. It follows from this assumption that the probability of genes in founders being identical by descent (IBD) at marker loci or QTL is zero. However, similarities in haplotypes of closely linked markers around a given position provide information on the probability of founder genes being identical by descent in this position. Because individuals that carry a mutant gene will also be IBD in a chromosomal region surrounding the gene, linkage disequilibrium between haplotypes and QTL loci can be utilized to map QTL (MEUWISSEN and GODDARD 2000). MEUWISSEN *et al.*

(2002) showed that this information from linkage disequilibrium could be combined with linkage information in a model with random QTL effects.

The advantage of this method is that it utilizes information on both the historic recombination events, since the mutation in the QTL occurred, and the recombinations observed in the recorded family structure. Therefore, closely linked markers provide much more information than pure linkage analysis and the QTL can be mapped to a region of 1–3 cM (MEUWISSEN and GODDARD 2000).

In most QTL-mapping experiments several traits are measured. Usually, QTL are mapped for individual traits using single-trait analyses. However, there are several reasons why it is important to infer QTL-related parameters using multitrait methods.

First, many of the traits are environmentally and genetically correlated. To use all information optimally the correlation structure between traits should be taken into account in the analysis. This can increase the statistical power of detection. However, analyzing many traits jointly will not necessarily lead to higher power of detection because an increased number of parameters must be inferred. At least, when only two traits are considered it has been shown, in the framework of linkage analysis, that the statistical power to detect a QTL can be increased by utilizing information from correlated traits. Such increase in power was demonstrated using regression methods (JIANG and ZENG 1995; CALINSKI *et al.* 2000; KNOTT and HALEY 2000), a maximum-likelihood method (KOROL *et al.* 1995), and variance component models (ALMASY *et al.* 1997; P. SØRENSEN, M. S. LUND, B. GULDBRANDTSEN, J. JENSEN and D. SØRENSEN, unpub-

¹Corresponding author: Danish Institute of Agricultural Sciences, Department of Animal Breeding and Genetics, Research Centre Foulum, P.O. Box 50, DK-8830 Tjele, Denmark.
E-mail: mogens.lund@agrsci.dk

lished data). It is particularly important to utilize the correlated information when mapping QTL for low heritability traits that are correlated to a trait of higher heritability (P. SØRENSEN, M. S. LUND, B. GULDBRANDTSEN, J. JENSEN and D. SØRENSEN, unpublished data). In this case, the single-trait analysis can have substantially lower statistical power compared to a multiple-trait analysis, which combines phenotypic information from both traits. This is important, because utilizing identified QTL through marker-assisted selection for traits with low heritability will have a relatively higher impact on practical breeding work.

Second, decomposing the genetic correlation and thereby assessing the effect of the QTL on all traits in the breeding goal is important. The contribution of a QTL to the genetic correlation between two traits does not necessarily follow the overall genetic correlation. Such information is relevant when the value of the QTL for selection is assessed.

Third, knowledge of the genetic background of the QTL contribution to genetic correlations is valuable. A genetic correlation can be the result of pleiotropic effects of a single QTL affecting more than one trait or of linkage disequilibrium between two or more QTL, each affecting one trait only (FALCONER and MACKAY 1996). The important issue is to decide whether one or two QTL are involved. Specifically, if a chromosomal region is found to affect two traits, it is important to distinguish whether the association is due to a pleiotropic QTL that affects both traits or due to close linkage of two QTL, each affecting one trait. If the aim is selective breeding, and the QTL contribute to an unfavorable genetic correlation, it is crucial to know if it is due to one pleiotropic QTL or to two linked QTL. If there is one pleiotropic QTL, then the genetic correlation is fixed; if there are two linked loci, selection of individuals in which a recombination between the QTL has changed the linkage phase can change the unfavorable correlation. If the aim is to clone the functional gene, it is obviously important that the linked QTL are not modeled as one pleiotropic QTL with a position somewhere between the true QTL positions (MARTINEZ and CURNOW 1992).

The necessary tests to distinguish between one-locus or two-loci models can, in principle, be performed using multitrait linkage analysis. However, the QTL position is estimated with accuracy too low to distinguish between a single gene with pleiotropic effects or two loci, each affecting one trait. Combining fine mapping in a multitrait framework is expected to result in higher power to carry out this test.

The objectives of this study are, first, to present a method for multitrait fine mapping using combined linkage disequilibrium and linkage analysis and, second, to show via a small example that the method gives accurate position estimates of closely linked QTL, each affecting one trait. Finally, the small example provides

evidence that the proposed method can be used for the rather difficult task of distinguishing between a model postulating one pleiotropic QTL affecting two traits and one postulating two closely linked QTL, each affecting one trait.

METHODS

Model: The multivariate mixed model can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \sum_{i=1}^{n_{\text{qtl}}} \mathbf{W}\mathbf{q}_i + \mathbf{e}, \quad (1)$$

where \mathbf{y} is a $n \times t$ vector of observations on t traits, \mathbf{X} is a design matrix, $\boldsymbol{\beta}$ is a vector of fixed effects, \mathbf{Z} is a matrix relating records to individuals, \mathbf{u} is a vector of additive polygenic effects, \mathbf{W} is a matrix relating each individual's record to its QTL effect, \mathbf{q}_i is a vector of additive QTL effects corresponding to the i th QTL, and \mathbf{e} is a vector of residuals. The number of QTL, n_{qtl} , is here assumed to be equal to one or two.

The random variables \mathbf{u} , \mathbf{q} , and \mathbf{e} are assumed to be multivariate normally distributed and mutually uncorrelated. Specifically:

$\mathbf{u} \sim \text{MVN}(\mathbf{0}, \mathbf{G})$, where $\mathbf{G} = \mathbf{G}_0 \otimes \mathbf{A}$ and

$$\mathbf{G}_0 = \begin{pmatrix} \sigma_{g1}^2 & \dots & \sigma_{g1,gt} \\ \vdots & \ddots & \vdots \\ \sigma_{gt,g1} & \dots & \sigma_{gt}^2 \end{pmatrix},$$

$\mathbf{q}_i | M, p_i \sim \text{MVN}(\mathbf{0}, \mathbf{K}_i)$, where $\mathbf{K}_i = \mathbf{K}_{0i} \otimes \text{IBD}_{i|M, p_i}$ and

$$\mathbf{K}_{0i} = \begin{pmatrix} \sigma_{q1}^2 & \dots & \sigma_{q1,qt} \\ \vdots & \ddots & \vdots \\ \sigma_{qt,q1} & \dots & \sigma_{qt}^2 \end{pmatrix},$$

and $\mathbf{e} \sim \text{MVN}(\mathbf{0}, \mathbf{R})$, where $\mathbf{R} = \mathbf{E}_0 \otimes \mathbf{I}$ and

$$\mathbf{E}_0 = \begin{pmatrix} \sigma_{e1}^2 & \dots & \sigma_{e1,et} \\ \vdots & \ddots & \vdots \\ \sigma_{et,e1} & \dots & \sigma_{et}^2 \end{pmatrix}.$$

Above, \mathbf{A} is the additive genetic relationship matrix and $\text{IBD}_{i|M, p_i}$ is the IBD matrix for the i th QTL, conditional on marker data (M) and the position (p_i) of the i th QTL on the chromosome.

IBD matrix: The following four steps were used to calculate the IBD matrices for QTL i . First, a gametic relationship matrix (FERNANDO and GROSSMAN 1989) was generated on the basis of linkage disequilibrium (see steps one and two below) and linkage information (step three below). This matrix was then transformed into the IBD matrix to diminish the dimensionality (step four below).

Step one: For each position p_i , a table of IBD probabilities, conditional on haplotype similarities, was generated. This is denoted ($P_{\text{IBD}|N_i, N_r, p_i}$). Similarities between a

pair of haplotypes were defined as the number of alleles at consecutive marker loci, to the left and to the right of p_i , that were identical by state (IBS). The similarities are denoted as (NI, Nr). If the linkage group consists of 11 marker loci and p_i is in marker bracket k ($k = 1, \dots, 10$) there are k markers to the left and $11 - k$ markers to the right of p_i . Thus, there can be $0-k$ markers IBS to the left of p_i and $0-11 - k$ markers IBS to the right of p_i . This means that the dimension of $P_{\text{IBD}|N_i, N_r, p_i}$ is $(k + 1) \times (11 - k)$. The probabilities of haplotypes being IBD at p_i for each of the possible similarities were assessed using the gene-drop method of MEUWISSEN and GODDARD (2000), using 100,000 replicated gene drops.

For each individual in the conceptual base population, two haplotypes were sampled. At each position p_i a QTL locus was assumed. All QTL alleles were unique, such that any alleles that were IBS in subsequent generations were also IBD. Subsequent generations were simulated by randomly sampling parents from the previous generation to produce N_o offspring. Marker and QTL alleles were then transmitted from parents to offspring, allowing for recombinations between each locus according to Haldane's map function (HALDANE 1919). After 100 generations the similarities of all pairs of haplotypes were assessed and the number of pairs in each group (N_{N_i, N_r, p_i}) was updated. Within each group, the number of haplotypes that were IBD in the putative QTL position p_i was assessed. This number is denoted ($N_{\text{IBD}|N_i, N_r, p_i}$). After 100,000 gene drops, the IBD probabilities at each position p_i , conditional on haplotype similarities, were calculated as $P_{\text{IBD}|N_i, N_r, p_i} = N_{\text{IBD}, N_i, N_r, p_i} / N_{N_i, N_r, p_i}$.

Step two: In this step we return to the data set to be analyzed. First, the founder haplotypes were established and the similarity of each pair of haplotypes $N_i N_r(i, j)$ was assessed. The probability of identity by descent between haplotype i and haplotype j at position p_i was found in the appropriate table that was generated in step one ($P_{\text{IBD}|N_i, N_r, p_i}$) and inserted as element (i, j) in the gametic relationship matrix.

Step three: In the case of individuals with genotyped parents, IBD probabilities were calculated on the basis of the inheritance of marker alleles from parents to offspring. The element (i, j) of the gametic relationship matrix was computed using the recurrence equations of WANG *et al.* (1995).

Step four: The IBD matrix was computed using the linear relationship between the IBD matrix and the gametic relationship matrix (GEORGE *et al.* 2000).

Statistical analysis: Inferences were drawn using restricted maximum likelihood (PATTERSON and THOMPSON 1971). Conditional on the IBD matrix for the QTL effects, IBD_{iM, p_i} , the restricted likelihood of the multivariate mixed model, assuming a single QTL, is

$$L(\boldsymbol{\theta}|\mathbf{K}'\mathbf{y}, \mathbf{Q}_{iM, p_i}) \propto \int \int p(\mathbf{K}'\mathbf{y}|\mathbf{u}, \mathbf{q}, \mathbf{E}_0 \otimes \mathbf{I}) p(\mathbf{u}|\mathbf{G}_0 \otimes \mathbf{A}) \\ \times p(\mathbf{q}|\mathbf{K}_0 \otimes \text{IBD}_{iM, p_i}) d\mathbf{u} d\mathbf{q}$$

where $\boldsymbol{\theta} = (\text{vech}(\mathbf{G}_0)' \text{vech}(\mathbf{K}_0)' \text{vech}(\mathbf{E}_0)')$ is the vector containing the N unique elements of the symmetric matrices \mathbf{G}_0 , \mathbf{K}_0 , and \mathbf{E}_0 , and $\mathbf{K}'\mathbf{y}$ is the vector of "error contrasts." The restricted likelihood was maximized using the average information restricted maximum-likelihood (AI-REML) algorithm (JENSEN *et al.* 1997). The AI-REML algorithm maximizes the restricted likelihood with respect to the variance components (\mathbf{G}_0 , \mathbf{K}_0 , and \mathbf{E}_0). Preceding the AI-REML analysis and using only marker data, the IBD matrix IBD_{iM, p_i} is computed, conditional on the QTL position p_i . Maximizing a sequence of restricted likelihoods over a grid of specific positions yields a profile of the restricted likelihood of the QTL position.

The AI-REML algorithm is based on first and second derivatives of the restricted log likelihood (JENSEN *et al.* 1997). It was implemented by combining it with the expectation-maximization (EM) algorithm (DEMPSTER *et al.* 1977), to ensure that parameter estimates stay within the parameter space (JENSEN *et al.* 1997). There are cases, however, when estimates of the elements of \mathbf{K}_0 are expected to fall at the boundary of the parameter space. Specifically, if a biallelic QTL has a pleiotropic effect on two or more traits, then the QTL correlation between the traits is unity, which has to be taken into account to check for convergence. This was achieved here using two different criteria. One of these checked for small values of the vector of first derivatives of the restricted log likelihood. If the algorithm converges to a point inside the parameter space, then the values of the vector of first derivatives of the restricted log likelihood should approach zero. However, if the estimates are at the boundary of the parameter space, then the vector of first derivatives is not necessarily zero. Therefore the other convergence criterion requires that changes in estimates of the (co)variance components between successive rounds are smaller than a minimum, well-tuned value.

Simulation of data: The proposed method was tested using simulated data sets based on a granddaughter design (WELLER *et al.* 1990), which is presently the most commonly used design in QTL mapping in dairy populations. It is assumed that a previous linkage analysis study had identified a 10-cM region including one or two QTL that affect two traits. The aim is to test if there is information to distinguish between one pleiotropic QTL affecting both traits and two QTL, each affecting one trait.

Pedigree: The simulation was built on a founder population created 100 generations ago with a population size of 100 (50 males and 50 females). For each of the subsequent 100 generations, 50 males and 50 females were produced by randomly sampling parents from the previous generation. After the first 100 generations, individuals of the genotyped population were sampled. Ten sires and 200 dams (later referred to as grandsires and granddams) were produced by sampling parents

TABLE 1
Input parameters for simulated data sets

Effect	Variance		Correlation
	Trait 1	Trait 2	
Polygenic	20	20	0.5
QTL 1	12–20	0	0
QTL 2	0	12–20	0
Residuals	60	60	–0.5

randomly from the 100th generation. Each sire was mated to 20 dams and each mating produced one son.

Marker and QTL alleles: Marker alleles were sampled for the 11 biallelic loci placed with a distance of 1 cM between each locus. Two QTL were positioned 5 cM apart in marker bracket 3 and marker bracket 8. In the founder population, the two alleles of each marker locus were sampled with equal probability, and all QTL alleles were assigned a unique number. Marker and QTL alleles were transmitted from parents to offspring during a period spanning 100 generations. Recombinations were sampled according to Haldane's mapping function (HALDANE 1919).

From among the QTL alleles that were still present after 100 generations, one was sampled at random for each QTL as being the mutant gene, while all other QTL alleles were assumed to be of the wild type. The frequency of the mutant gene was restricted to be between 0.2 and 0.8. This resulted in different QTL variances between replicates (0.12–0.2) as indicated in Table 1.

Phenotypes: For each of the 200 sons in the last generation of the pedigree, a daughter yield deviation (DYD; VAN RADEN and WIGGANS 1991) was simulated on the basis of 100 daughters. Phenotypic records were simulated for two traits using the following model for the DYD of the i th son,

$$\begin{bmatrix} \text{DYD}_{1i} \\ \text{DYD}_{2i} \end{bmatrix} = \frac{1}{n_j} \sum_{j=1}^n \begin{bmatrix} a_{ij}^h \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ a_{2ij}^h \end{bmatrix} + \begin{bmatrix} u_{1i} \\ u_{2i} \end{bmatrix} + \begin{bmatrix} e_{1i} \\ e_{2i} \end{bmatrix},$$

where a_{ij}^h are the effects of paternally transmitted alleles on trait h in daughter j of son i , u_{hi} are the polygenic effects on trait h of son i , and e_{hi} are the residual effects on trait h of son i . Polygenic effects for grandsires were sampled from

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \sim \text{MVN}\left(\mathbf{0}, \begin{pmatrix} \sigma_{u1}^2 & \sigma_{u1,u2} \\ \sigma_{u1,u2} & \sigma_{u2}^2 \end{pmatrix}\right),$$

and for son i from

$$\begin{bmatrix} u_{1i} \\ u_{2i} \end{bmatrix} \sim \text{MVN}\left(0.5 \begin{bmatrix} u_{1GS} \\ u_{2GS} \end{bmatrix}, 0.75 \begin{pmatrix} \sigma_{u1}^2 & \sigma_{u1,u2} \\ \sigma_{u1,u2} & \sigma_{u2}^2 \end{pmatrix}\right),$$

where u_{iGS} is the polygenic value of the grandsire for trait h . Residual effects were sampled from

$$\begin{bmatrix} e_{1i} \\ e_{2i} \end{bmatrix} \sim \text{MVN}\left(\mathbf{0}, \frac{1}{n} \left[0.5 \begin{pmatrix} \sigma_{e1}^2 & \sigma_{e1,e2} \\ \sigma_{e1,e2} & \sigma_{e2}^2 \end{pmatrix} + \begin{pmatrix} \sigma_{e1,e2} & \sigma_{e1,e2} \\ \sigma_{e1,e2} & \sigma_{e2}^2 \end{pmatrix} \right]\right).$$

The parameters used for simulation of phenotypes are shown in Table 1.

Analysis of simulated data: When the IBD matrices were calculated, marker haplotypes were assumed known for grandsires and unknown for granddams. This means that the maternally inherited haplotypes in the sons were assumed to be founder haplotypes and to contribute with linkage disequilibrium information accordingly.

To mimic a situation where linkage analysis has identified a 10-cM region including one or two QTL that affect two traits, each trait of the simulated data was analyzed initially using a single-trait model (*i.e.*, model (1) with $t = 1$). A significant QTL was identified if the null hypothesis with no QTL was rejected. If a significant QTL was detected for both traits, the data were analyzed subsequently using a pleiotropic model (*i.e.*, model (1) with $t = 2$, $n_{qtl} = 1$, and

$$K_0 = \begin{bmatrix} \sigma_{q1}^2 & \sigma_{q1,q2} \\ \sigma_{q1,q2} & \sigma_{q2}^2 \end{bmatrix}$$

and a two-QTL model (*i.e.*, model (1) with $t = 2$, $n_{qtl} = 2$,

$$K_{01} = \begin{bmatrix} \sigma_{q1}^2 & 0 \\ 0 & 0 \end{bmatrix},$$

and

$$K_{02} = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{q2}^2 \end{bmatrix}.$$

The test for pleiotropy *vs.* two linked loci was based on a test statistic involving the ratio of the log likelihood of both models, evaluated at their maximum-likelihood estimates. The null hypothesis poses that the model contains one pleiotropic segregating QTL. The test was performed using Monte Carlo-derived significant thresholds, as explained below. The power of the test was calculated as the proportion of times, out of 100 replicates, that the null hypothesis was rejected.

Significance thresholds: First, the significance threshold for the single-trait tests was found by simulation. Five hundred replicates were simulated under the null hypothesis of no QTL segregating. Each replicate was simulated with the same structure, the same genetic variance, and the same environmental variance as the data sets including a QTL. The simulated data sets were then analyzed using a single-trait model with and without a QTL fitted. For each replicate, the test statistic based on the ratio of the log likelihoods of the models was calculated, and the significance threshold was set such that 5% of the test statistics were higher than the threshold.

Second, the significance threshold for the tests for pleiotropy *vs.* linkage was found by simulation as follows. Two hundred and fifty replicates were simulated under the null hypothesis, in which a single pleiotropic QTL

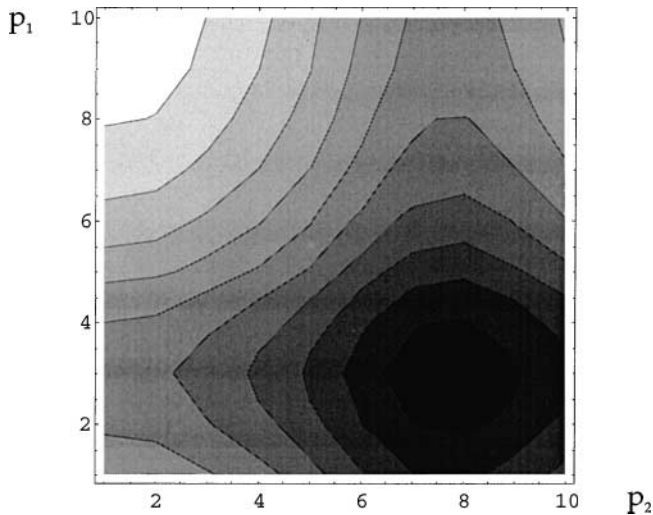


FIGURE 1.—Contour plot of mean likelihood of positions of the QTL affecting trait 1 (p_1) and the QTL affecting trait 2 (p_2) over 100 replicates.

was assumed to affect both traits. The effect of the pleiotropic QTL on each trait was the same as the effect of the two linked QTL. In each replicate, both traits were analyzed using a single-trait model. If both tests were significant, the traits were analyzed jointly using the pleiotropic model and the two-QTL model. The significance threshold was assessed on the basis of the empirical distribution of the test statistic (ratio of log likelihoods) over replicates.

RESULTS

Figure 1 shows the mean profile log likelihood for the QTL positions, over 100 replicates. On average, the ridge of the log-likelihood curve indicates that the estimates of the positions are at interval 3 for the QTL affecting trait 1 and interval 8 for the QTL affecting trait 2. These average estimates agree well with the input parameters.

Figure 2 shows the number of times the different intervals had the highest log-likelihood value for the QTL affecting trait 1 (light gray columns) and for the QTL affecting trait 2 (dark gray columns). For both QTL the most frequently estimated position was the correct input parameter used in the simulations. The histograms show some skewness: This is because the QTL were placed at the end of a confined interval and because of the small size of the experiment.

When data were simulated on the basis of the two-QTL model, the null hypothesis, which assumes that the state of nature is represented by a single-QTL model, was rejected in 56% of the replicates.

DISCUSSION

We have presented a multitrait method for fine-mapping detection of QTL, based on combined linkage

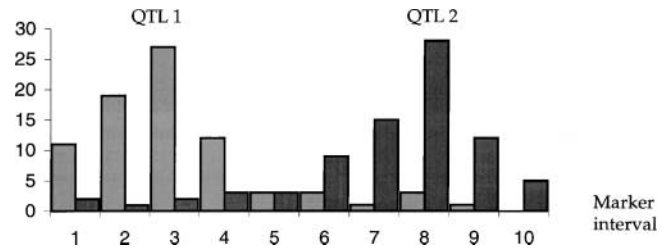


FIGURE 2.—Distribution of QTL position estimates. The x-axis is the marker interval and the y-axis the number of times the QTL affecting trait one (light gray bars) and the QTL affecting trait two (dark gray bars) were estimated in the different marker intervals.

disequilibria and linkage analysis, extending the work of MEUWISSEN and GODDARD (2000). The performance of the method was illustrated using simulated data. The results (Figure 2) showed that the position of two closely linked loci, each affecting one trait, could be estimated correctly on average, with the proposed method. The two histograms of the estimated positions of the QTL affecting trait one and trait two were well separated and overlap only in areas where the probability mass was small. The results of MEUWISSEN and GODDARD (2000) suggested that the precision in terms of marker intervals was fairly constant over different marker densities. It is therefore expected that QTL that are more closely linked than those in the present study (*e.g.*, 1 cM) can still be distinguished from a pleiotropic QTL if a denser marker map is used (*e.g.*, 0.25 cM). Furthermore, in the simulations, random drift over the 100 generations might have caused a proportion of the markers to become monomorphic in the genotyped individuals. This might have adversely affected the method to distinguish between one pleiotropic QTL and two QTL each affecting one trait.

More precise inferences about the position of one pleiotropic QTL affecting two traits using multitrait linkage analysis, compared to single-trait analyses, have been reported in the literature (JIANG and ZENG 1995; ALMASY *et al.* 1997; P. SØRENSEN, M. S. LUND, B. GULDBRANDTSEN, J. JENSEN and D. SØRENSEN, unpublished data). Similar results are expected to hold for the method proposed here as well, even though we did not investigate this problem.

An important and difficult question that we addressed is whether the proposed method can distinguish between a model postulating one pleiotropic QTL and another postulating two closely linked QTL, each affecting one trait. This is a problem of model choice, in a situation where models are not nested and have the same number of parameters. This issue was discussed from a frequentist perspective by COX (1961, 1962). Within the constraints of the paradigm, we have avoided a frontal comparison of the models and instead posed the question as follows. When data are simulated under the two-QTL model, in what proportion of times over conceptual replications is the null hypothesis postulat-

ing one pleiotropic QTL rejected? The test statistic was the ratio of the log likelihoods of both models, whose small sample distribution was disclosed via Monte Carlo simulation. With a choice of a type I error rate of 5% and given the size of the present simulation experiment and the parameters used to simulate the data, the power of the test was 56%—a long way from the minimum possible power of 5%, which would result if the models were indistinguishable. It is a rather subtle task to discern between the two models, and a larger experiment is needed to arrive at more conclusive results. It is not the purpose of this article to carry out in-depth simulation studies. Via the small example presented, we wish to illustrate that the two models can in principle be distinguished by the proposed method. The test performed herein was constructed to answer the specific question above. If real data are analyzed a more appropriate testing scheme might include a number of nested tests resulting from stepwise reductions starting from a full model with two pleiotropic QTL.

When applied to real data more information can be expected compared to that in the simulation in this study. This is because larger granddaughter designs and more informative markers are available. However, this could be counterbalanced by lower frequencies of QTL alleles or different patterns of linkage disequilibrium. If a larger part of the linkage disequilibrium were due to very strong selection or recent population admixture, larger areas of linkage disequilibrium would be expected. This may impair the possibility of fine mapping the QTL. It would be useful to study further the properties of the proposed method in such situations.

The method can be used in other designs involving more complex pedigrees. The main requirement is to be able to establish the haplotypes of the first-generation individuals of the genotyped population. In livestock populations, there is often sufficient information to reconstruct haplotypes very accurately and this requirement can be fulfilled. In species with small offspring groups, the method should be modified to take account of the uncertainty of haplotypes.

Currently, the age of mutation and the structure of the historic population are used as fixed parameters. MEUWISSEN and GODDARD (2000) found that the method was rather insensitive to changes in these parameters, except when the mutation was recently introduced into the population. Uncertainty about the age of mutation could be taken into account by making analyses conditional on different ages and producing a profile likelihood over ages.

A Bayesian implementation of this method would be desirable. While not exempt from problems, the Bayesian approach is conceptually more transparent for model comparison than the one presented here. This is accomplished conditional on the available data, via the posterior probability of the respective models, for example, as in SILLANPÄÄ and ARJAS (1998) and YI and XU (2000).

Daniel Sorensen acknowledges funding for research visits from the Wellcome Trust Biomedical Research Collaboration Grant 056266/Z/98/Z.

LITERATURE CITED

- ALMASY, L., T. D. DYER and J. BLANGERO, 1997 Bivariate quantitative trait linkage analysis: pleiotropy versus co-incident linkages. *Genet. Epidemiol.* **14**: 953–958.
- CALINSKI, T., Z. KACZMAREK, P. KRAJEWSKI, C. FROVA and M. SARI-GORLA, 2000 A multivariate approach to the problem of QTL localization. *Heredity* **84**: 303–310.
- COX, D. R., 1961 Tests of separate families of hypotheses. *Proceedings of the 4th Berkeley Symposium*, Berkeley, CA, Vol. 1, pp. 105–123.
- COX, D. R., 1962 Further results on tests of separate families of hypotheses. *J. R. Stat. Soc. B* **24**: 406–424.
- DEMPSTER, A. P., N. M. LAIRD and D. B. RUBIN, 1997 Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* **39**: 1–39.
- FALCONER, D. S. and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Longman Group, Essex, UK.
- FERNANDO, R. L., and M. GROSSMAN, 1989 Marker-assisted selection using best linear unbiased prediction. *Genet. Sel. Evol.* **21**: 467–477.
- GEORGE, A. W., P. M. VISSCHER and C. HALEY, 2000 Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. *Genetics* **156**: 2081–2092.
- HALDANE, J. B. S., 1919 The combination of linkage values and the calculation of distances between the loci of linked factors. *J. Genet.* **8**: 299–309.
- JENSEN, J., E. A. MÄNTYSAARI, P. MADSEN and R. THOMPSON, 1997 Residual maximum likelihood estimation of (co)variance components in multivariate mixed linear models using average information. *J. Indian Soc. Agric. Stat.* **49**: 215–236.
- JIANG, C., and Z-B. ZENG, 1995 Multiple trait analysis of genetic mapping of quantitative trait loci. *Genetics* **140**: 1111–1127.
- KNOTT, S. A., and C. S. HALEY, 2000 Multitrait least squares for quantitative trait loci detection. *Genetics* **156**: 899–911.
- KOROL, A. B., Y. I. RONIN and V. M. KIRZHNER, 1995 Interval mapping of quantitative trait loci employing correlated trait complexes. *Theor. Appl. Genet.* **92**: 998–1002.
- MARTINEZ, O., and R. N. CURNOW, 1992 Estimating the location and size of the effect of quantitative trait loci using flanking markers. *Theor. Appl. Genet.* **85**: 480–488.
- MEUWISSEN, T. H. E., and M. E. GODDARD, 2000 Fine mapping of quantitative traits using linkage disequilibria with closely linked marker loci. *Genetics* **155**: 421–430.
- MEUWISSEN, T. H. E., A. KARLSEN, S. LIEN, I. OLSAKER and M. E. GODDARD, 2002 Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* **161**: 373–379.
- PATTERSON, H. D., and R. THOMPSON, 1971 Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**: 545–554.
- SILLANPÄÄ, M. J., and E. ARJAS, 1998 Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics* **148**: 1373–1388.
- VAN RADEN, P. M., and G. R. WIGGANS, 1991 Derivation, calculation, and use of lactation animal-model information. *J. Dairy Sci.* **74**: 2737–2746.
- WANG, T., R. H. FERNANDO, S. VAN DER BEEK and J. A. M. VAN ARENDONK, 1995 Covariance between relatives for a marked quantitative trait locus. *Genet. Sel. Evol.* **27**: 251–274.
- 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.
- YI, N., and S. XU, 2000 Bayesian mapping of quantitative trait loci under the identity-by-descent-based variance component model. *Genetics* **156**: 411–422.