

Complex Genetic Effects in Quantitative Trait Locus Identification: A Computationally Tractable Random Model for Use in F₂ Populations

Daisy Zimmer, Manfred Mayer¹ and Norbert Reinsch

Leibniz Institute for Farm Animal Biology, Research Unit Genetics and Biometry, 18196 Dummerstorf, Germany

Manuscript received August 17, 2010
Accepted for publication October 7, 2010

ABSTRACT

Methodology for mapping quantitative trait loci (QTL) has focused primarily on treating the QTL as a fixed effect. These methods differ from the usual models of genetic variation that treat genetic effects as random. Computationally expensive methods that allow QTL to be treated as random have been explicitly developed for additive genetic and dominance effects. By extending these methods with a variance component method (VCM), multiple QTL can be mapped. We focused on an F₂ crossbred population derived from inbred lines and estimated effects for each individual and their corresponding marker-derived genetic covariances. We present extensions to pairwise epistatic effects, which are computationally intensive because a great many individual effects must be estimated. But by replacing individual genetic effects with average genetic effects for each marker class, genetic covariances are approximated. This substantially reduces the computational burden by reducing the dimensions of covariance matrices of genetic effects, resulting in a remarkable gain in the speed of estimating the variance components and evaluating the residual log-likelihood. Preliminary results from simulations indicate competitiveness of the reduced model with multiple-interval mapping, regression interval mapping, and VCM with individual genetic effects in its estimated QTL positions and experimental power.

MAPPING procedures often treat the effects of quantitative trait loci (QTL) as fixed, in particular the maximum likelihood-based method of interval mapping (IM) of LANDER and BOTSTEIN (1989) and the least-squares regression interval mapping (RIM) of HALEY and KNOTT (1992) and MARTÍNEZ and CURNOW (1992).

Single-QTL approaches with fixed effects were later extended to multiple QTL to avoid the so-called “ghost-QTL” phenomenon (*e.g.*, HALEY and KNOTT 1992) and to improve the power to detect linked QTL in repulsion (*e.g.*, KAO 2000) as well as epistatic QTL (*e.g.*, JANNINK and JANSEN 2001; CARLBORG and HALEY 2004). The multiple-interval mapping (MIM) approach of KAO and ZENG (1997) and KAO *et al.* (1999) as an extension of IM considers fixed additive genetic, dominance, and epistatic QTL effects as parts of the likelihood function for a mixture model in experimental populations. Both MIM and RIM are known to be powerful and well suited to identifying multiple, possibly interacting QTL in mapping experiments. However, the accuracy of the estimates of the positions and effects of the QTL from RIM is less compared with MIM in some situations [*e.g.*, QTL in repulsion (KAO 2000; MAYER *et al.* 2004; MAYER 2005)].

Supporting information is available online at <http://www.genetics.org/cgi/content/full/genetics.110.122333/DC1>.

¹Corresponding author: Leibniz Institute for Farm Animal Biology, Research Unit Genetics and Biometry, Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany. E-mail: mmayer@fbn-dummerstorf.de

Considering QTL effects as random in a linear mixed model (LMM) leads to the variance component method (VCM) for QTL mapping. This is often applied in scenarios with a large number of small families as is frequently found in humans (*e.g.*, HASEMAN and ELSTON 1972; XU and ATCHLEY 1995) or in livestock (*e.g.*, GRIGNOLA *et al.* 1996), where a mixture of families with parents of different QTL genotypes is expected to occur. Experiments with multiple line crosses, *e.g.*, F₂, are often advocated because of their potential to avoid non-detection of QTL by representing genetic variability of a population by only a few lines—the so-called “genetic drift error” (XU 1996). Although fixed effect approaches are equivalent in power, at least in situations with a single QTL, VCM are easier to implement and have computational advantages in this context (XU 1998). Rules for setting up the required QTL allelic relationship matrices from marker data were given by WANG *et al.* (1995) and ABDEL-AZIM and FREEMAN (2001). Marker-based relationship matrices for QTL with additive genetic and nonadditive genetic (dominance, epistasis) gene action in noninbred populations were applied by LIU *et al.* (2002).

The focus of XIE *et al.* (1998) was on backcross (BC) and F₂ designs descending from inbred lines. For these types of experiments additive genetic and dominance relationship matrices can be calculated from conditional QTL genotype probabilities (given the flanking

marker genotypes) for all individuals of the mapping population (as used as regressor variables in RIM). CREPIEUX *et al.* (2004) provided a general extension to any type of multicross designs from inbred parents. Furthermore, LI and CUI (2009) demonstrated how VCM can be employed for mapping imprinted QTL in a combination of different BC populations derived from inbred lines.

In this article we first propose extensions of the variance component approach of XIE *et al.* (1998) to multiple interacting QTL with pairwise epistatic effects. Then, maintaining the focus on inbred line-derived F₂ populations, a reduced model is suggested, in which individual genetic effects are replaced by average genetic effects for different marker classes. The covariance matrix of the phenotypes is approximated in different ways, leading to less computational effort.

THEORY

Linear mixed model: From an F₂ generation derived from a cross between inbred lines, one observation per individual is considered. The vector of phenotypes \mathbf{Y} (length n) is modeled with respect to additive genetic, dominance, and pairwise epistatic effects of the QTL, whose total number is v . A pair of QTL is indexed by l and k . The LMM in matrix notation is given as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \sum_{l=1}^v \mathbf{Z}_l(\mathbf{u}_{a_l} + \mathbf{u}_{d_l}) + \sum_{l=1}^{v-1} \sum_{k=l+1}^v \mathbf{Z}_{lk}(\mathbf{u}_{aa_{lk}} + \mathbf{u}_{ad_{lk}} + \mathbf{u}_{da_{lk}} + \mathbf{u}_{dd_{lk}}) + \mathbf{e}. \quad (1)$$

The vector of fixed effects $\boldsymbol{\beta}$ has the related design matrix \mathbf{X} . The random vectors \mathbf{u}_τ with $\tau \in \{a_b, d_b, aa_{lk}, ad_{lk}, da_{lk}, dd_{lk}\}$ denote the additive genetic, the dominance, and the four pairwise epistatic effects (first-order interactions) at QTL l and k . For each τ the length of \mathbf{u}_τ equals the number of F₂ individuals n ; *i.e.*, all QTL effects differ between individuals. The incidence matrices \mathbf{Z}_l and \mathbf{Z}_{lk} with $\dim(\mathbf{Z}_l) = \dim(\mathbf{Z}_{lk}) = n \times n$ relate the observations to genetic effects. The residuals are assumed to be independently and identically normally distributed with $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, where \mathbf{I} is the identity matrix of order n and σ_e^2 is the residual variance. The covariances between normally distributed random genetic effects \mathbf{u}_τ and the residuals \mathbf{e} are assumed zero as well as the covariances between different types of genetic effects \mathbf{u}_τ . The expectations of the QTL effects are $E(\mathbf{u}_\tau) = \mathbf{0}$ and the variances are $\text{Var}(\mathbf{u}_\tau) = \mathbf{V}_\tau\sigma_\tau^2$, where σ_τ^2 is the related QTL variance and \mathbf{V}_τ is the corresponding expected QTL relationship matrix conditional on the marker genotypes. The phenotypic vector therefore follows a multivariate normal distribution with $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$. The covariance matrix \mathbf{V}

is derived conditional on the observed marker genotypes and can be written as

$$\mathbf{V} = \sum_{l=1}^v \mathbf{Z}_l(\mathbf{V}_{a_l}\sigma_{a_l}^2 + \mathbf{V}_{d_l}\sigma_{d_l}^2)\mathbf{Z}'_l + \sum_{l=1}^{v-1} \sum_{k=l+1}^v \mathbf{Z}_{lk}(\mathbf{V}_{aa_{lk}}\sigma_{aa_{lk}}^2 + \mathbf{V}_{ad_{lk}}\sigma_{ad_{lk}}^2 + \mathbf{V}_{da_{lk}}\sigma_{da_{lk}}^2 + \mathbf{V}_{dd_{lk}}\sigma_{dd_{lk}}^2)\mathbf{Z}'_{lk} + \mathbf{I}\sigma_e^2. \quad (2)$$

Calculation of covariance matrices: We follow the approach of XIE *et al.* (1998) and derive the required genetic covariance matrices of (2) from conditional QTL genotype probabilities and elementary covariance matrices.

Conditional QTL genotype probabilities: For a particular QTL the F₂ generation can be partitioned into nine different marker classes (see Table 2 column headings) conditional on the observed genotype of the flanking markers. QTL alleles originating from the first line are denoted by uppercase letter indexes (Q, H) and those from the second line by lowercase indexes (q, h), and for marker alleles the respective line origins are indicated by numbers (1 and 2). Conditional QTL genotype probabilities depend on flanking marker genotypes and the recombination rates between the markers and QTL and can be derived as described by, *e.g.*, CARBONELL *et al.* (1992, Table 1). We allow for double recombinations and assume Haldane's mapping function (HALDANE 1919).

Probabilities for the genotypes G_{QQ} , G_{Qq} and G_{qq} of an individual at the l th QTL conditional on flanking marker information M_i can be collected in a row vector \mathbf{I}_i^l with

$$\mathbf{I}_i^l = \begin{pmatrix} \Pr(G_{QQ} | M_i) \Pr(G_{Qq} | M_i) \Pr(G_{qq} | M_i) \\ = (p_i^{QQ} \ p_i^{Qq} \ p_i^{qq}), \end{pmatrix}$$

where M_i denotes the observed flanking marker genotype $i \in \{1, \dots, 9\}$ of an individual. We assume that in each marker interval either no or only a single QTL exists. The joint conditional probability for two linked QTL is just the product of both single probabilities if at least one completely informative marker is in between (RÖNNEGÅRD *et al.* 2008). Thus, the probability of a two-locus QTL genotype, *e.g.* G_{QQHh} , given the particular marker genotypes M_i and N_j ($i, j \in \{1, \dots, 9\}$) at QTL l and k , respectively, is defined as $\Pr(G_{QQHh} | M_i, N_j) = \Pr(G_{QQ} | M_i) \Pr(G_{Hh} | N_j)$. We define \mathbf{I}_{ij}^{lk} as the row vector with all joint conditional QTL genotype probabilities for a pairwise epistatic effect at QTL l and k .

Elementary covariance matrices: As a second ingredient we need elementary covariance matrices between all possible QTL genotypes G_{QQ} , G_{Qq} and G_{qq} in the F₂ populations. The elementary matrices for additive genetic QTL effects \mathbf{A} (XIE *et al.* 1998) and dominance QTL effects \mathbf{D} (SMITH 1984; XIE *et al.* 1998) are

$$\mathbf{A} = \begin{matrix} & G_{QQ} & G_{Qq} & G_{qq} \\ G_{QQ} & \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \end{matrix} \text{ and } \mathbf{D} = \begin{matrix} & G_{QQ} & G_{Qq} & G_{qq} \\ G_{QQ} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{matrix}.$$

We use the Kronecker product (symbol \otimes) of \mathbf{A} and \mathbf{D} to compute the four different 9×9 elementary matrices, $\mathbf{A} \otimes \mathbf{A}$, $\mathbf{A} \otimes \mathbf{D}$, $\mathbf{D} \otimes \mathbf{A}$, and $\mathbf{D} \otimes \mathbf{D}$, which include covariances between pairwise epistatic effects and correspond to nine genotypes (G_{QQHH} , G_{QQHh} , G_{QQhh} , G_{QqHH} , G_{QqHh} , G_{Qqhh} , G_{qqHH} , G_{qqHh} , and G_{qqhh}) for pairwise QTL combinations.

QTL relationship matrices: The $n \times n$ additive genetic, dominance, and pairwise epistatic relationship matrices for all F_2 individuals can be set up for a putative QTL position or combinations thereof with conditional QTL genotype probabilities ($\mathbf{1}_i^l$ and $\mathbf{1}_{ij}^k$ vectors) and elementary matrices (XIE *et al.* 1998). Relationship coefficients are averages of possible QTL genotype combinations. For the additive genetic relationship matrix $\mathbf{V}_a = \{a_{st}^l\}_{s,t=1}^n$ we get diagonal elements

$$a_{ss}^l = \mathbf{1}_i^l \text{diag}(\mathbf{A}) = 2p_i^{QQ} + p_i^{Qq} + p_i^{qq} \quad (3)$$

and off-diagonals

$$a_{st}^l = \mathbf{1}_i^l \mathbf{A} (\mathbf{1}_i^l)' = 2p_i^{QQ} p_j^{QQ} + p_i^{Qq} p_j^{QQ} + p_i^{QQ} p_j^{Qq} + p_i^{Qq} p_j^{Qq} + p_i^{qq} p_j^{Qq} + p_i^{Qq} p_j^{qq} + 2p_i^{qq} p_j^{qq} \quad (4)$$

at the l th QTL. If both individuals s and t belong to the same marker class i , then a_{st}^l can be simplified to

$$a_{st}^l = 2(p_i^{QQ})^2 + 2p_i^{QQ} p_i^{Qq} + (p_i^{Qq})^2 + 2p_i^{Qq} p_i^{qq} + 2(p_i^{qq})^2, \quad (5)$$

because the conditional probabilities are equal. The dominance relationship matrix $\mathbf{V}_d = \{d_{st}^l\}_{s,t=1}^n$ is set up equivalently, but instead of \mathbf{A} the elementary matrix \mathbf{D} is used, *i.e.*, $d_{ss}^l = \mathbf{1}_i^l \text{diag}(\mathbf{D})$ and $d_{st}^l = \mathbf{1}_i^l \mathbf{D} (\mathbf{1}_i^l)'$.

We suggest that the pairwise epistatic relationship matrices \mathbf{V}_{aak} , \mathbf{V}_{adk} , \mathbf{V}_{dak} , \mathbf{V}_{ddk} at the l th and k th QTL are computed analogously to \mathbf{V}_a using the appropriate Kronecker product of elementary matrices (*e.g.*, $\mathbf{A} \otimes \mathbf{A}$). Computation of matrix elements is done as in Equations 3 and 4, employing corresponding row vectors $\mathbf{1}_{ij}^k$. Note that this is equivalent to using Hadamard products of QTL relationship matrices \mathbf{V}_a and \mathbf{V}_d given that there is at least one completely informative marker between both QTL or no linkage between them (RÖNNEGÅRD *et al.* 2008), which is always fulfilled by our assumptions. To ensure positive definiteness of covariance matrices, we assume that locations of putative QTL and markers do not coincide.

Equivalent model with average genetic effects: What we have outlined so far is termed “individual model,” because each individual receives its own genetic effects for the different kinds of genetic components. For a

particular QTL l the LMM of (1) with only additive genetic effects becomes

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_l \mathbf{u}_a + \mathbf{e}, \quad (6)$$

with covariance matrix of the phenotypes conditional on the observed marker genotypes

$$\mathbf{V} = \mathbf{Z}_l \mathbf{V}_a \mathbf{Z}_l' \sigma_a^2 + \mathbf{I} \sigma_e^2. \quad (7)$$

A model equivalent to (6) is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \tilde{\mathbf{Z}}_l \tilde{\mathbf{u}}_a + \mathbf{m}_a + \mathbf{e}, \quad (8)$$

where a vector $\tilde{\mathbf{u}}_a$ with length $n_l = 9$ (number of different marker classes) of average additive genetic effects for all possible marker genotype classes is considered. An additional random effect \mathbf{m}_a of length n appears, termed “additive genetic sampling effect,” and it describes the deviations of the individual additive genetic effects from the average additive genetic effects of marker classes. The dimension of $\tilde{\mathbf{Z}}_l$ is $n \times n_l$. Accordingly, the covariance matrix of the phenotypes can be expressed as

$$\mathbf{V} = \tilde{\mathbf{Z}}_l \tilde{\mathbf{V}}_a \tilde{\mathbf{Z}}_l' \sigma_a^2 + \mathbf{V}_{m_a} \sigma_a^2 + \mathbf{I} \sigma_e^2, \quad (9)$$

where $\tilde{\mathbf{V}}_a = \{\tilde{a}_{ij}^l\}_{i,j=1}^{n_l}$ denotes the reduced $n_l \times n_l$ relationship matrix of the average additive genetic effects at the QTL. The additive genetic variance of the individual model (7) is σ_a^2 , which is identical to σ_a^2 in (9). The variance of the additive genetic sampling effect is $\text{Var}(\mathbf{m}_a) = \mathbf{V}_{m_a} \sigma_a^2$, where \mathbf{V}_{m_a} denotes the relationship matrix of the additive genetic sampling effect. There are n_i^l individuals with the same marker genotype i at the QTL. The variance of the average additive genetic effect of a certain marker class i , averaged over n_i^l individuals, is given in the reduced model as

$$\tilde{a}_{ii}^l = \begin{cases} \mathbf{1}_i^l \text{diag}(\mathbf{A}) & \text{if } n_i^l = 0, \\ \frac{1}{n_i^l} \mathbf{1}_i^l \text{diag}(\mathbf{A}) + \left(1 - \frac{1}{n_i^l}\right) \mathbf{1}_i^l \mathbf{A} (\mathbf{1}_i^l)' & \text{else.} \end{cases} \quad (10)$$

Equation 10 is valid, because there are n_i^l diagonal elements and $(n_i^l)^2 - n_i^l$ off-diagonal elements in the relationship matrix of the individual additive genetic effects.

The three possible cases appearing in the additive genetic relationship matrix of the individual model are further investigated (see Equations 3–5). First, the variance of an individual additive genetic effect with marker class i is $2p_i^{QQ} + p_i^{Qq} + 2p_i^{qq} := \tilde{v}_{ii}^l$ and second, the covariance between two additive genetic effects with the same marker class i is $2(p_i^{QQ})^2 + 2p_i^{QQ} p_i^{Qq} + (p_i^{Qq})^2 + 2p_i^{Qq} p_i^{qq} + 2(p_i^{qq})^2 := v_{ii}^l$. Then the element \tilde{a}_{ii}^l for $n_i^l > 0$ can be written as

$$\tilde{a}_{ii}^l = \frac{1}{n_i^l} \tilde{v}_{ii}^l + \left(1 - \frac{1}{n_i^l}\right) v_{ii}^l. \quad (11)$$

The variance of the average additive genetic effect is asymptotically equal to the covariance between individ-

ual additive genetic effects of the same marker class i ; i.e., $\lim_{n_i^l \rightarrow \infty} \tilde{a}_{ii}^l = v_{ii}^l$. Third, the covariance of additive genetic effects with marker classes i and j is $2p_i^{QQ} p_j^{QQ} + p_i^{Qq} p_j^{QQ} + p_i^{Qq} p_j^{Qq} + p_i^{qq} p_j^{Qq} + p_i^{qq} p_j^{qq} + 2p_i^{Qq} p_j^{qq} := v_{ij}^l$. Now, the covariance of the average additive genetic effects of marker genotypes i and j ($i \neq j$) can be expressed as

$$a_{ij}^l = \mathbf{I}_i^l \mathbf{A} (\mathbf{I}_j^l)' = v_{ij}^l. \quad (12)$$

This is equal to the covariance among the two individual additive genetic effects of marker classes i and j .

The relationship matrix of the additive genetic sampling effects $\mathbf{V}_{m_{a_i}}$ can be determined as the difference between the relationship matrices of additive genetic effects from the individual model (individual genetic effects) and the reduced model (average genetic effects), which are inferred from Equations 7 and 9; i.e., $\mathbf{V}_{m_{a_i}} = \mathbf{Z}_l \mathbf{V}_a \mathbf{Z}_l' - \tilde{\mathbf{Z}}_l \tilde{\mathbf{V}}_a \tilde{\mathbf{Z}}_l'$. Generally, $\mathbf{V}_{m_{a_i}}$ (order n) can be written as

$$\mathbf{V}_{m_{a_i}} = \begin{pmatrix} \mathbf{M}_{11}^{a_i} & \mathbf{M}_{12}^{a_i} & \dots & \mathbf{M}_{19}^{a_i} \\ \mathbf{M}_{21}^{a_i} & \mathbf{M}_{22}^{a_i} & \dots & \mathbf{M}_{29}^{a_i} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{M}_{91}^{a_i} & \mathbf{M}_{92}^{a_i} & \dots & \mathbf{M}_{99}^{a_i} \end{pmatrix} \quad (13)$$

if the individuals are arranged by marker class. To study the matrices $\mathbf{M}_{ij}^{a_i}$ we assume that each marker genotype appears at least once; i.e., $n_i^l \geq 1$.

Concerning the third case, the additive genetic covariance between a pair of individuals s and t with different marker genotypes i and j equals the difference of (4) and (12): $m_{ij}^{a_i} = v_{ij}^l - v_{ij}^l = 0$. Therefore, for $i \neq j$ $\mathbf{M}_{ij}^{a_i} = \mathbf{0}$ in (13) and $\mathbf{V}_{m_{a_i}}$ is a block diagonal matrix if the observations are ordered by marker genotypes. The diagonal block $\mathbf{M}_{ii}^{a_i}$ corresponding to marker class i has the order n_i^l and can be expressed as

$$\mathbf{M}_{ii}^{a_i} = \begin{pmatrix} \tilde{m}_{ii}^{a_i} + m_{ii}^{a_i} & m_{ii}^{a_i} & \dots & m_{ii}^{a_i} \\ m_{ii}^{a_i} & \tilde{m}_{ii}^{a_i} + m_{ii}^{a_i} & \dots & m_{ii}^{a_i} \\ \vdots & \vdots & \ddots & \vdots \\ m_{ii}^{a_i} & m_{ii}^{a_i} & \dots & \tilde{m}_{ii}^{a_i} + m_{ii}^{a_i} \end{pmatrix}.$$

The covariance $m_{ii}^{a_i}$ of the additive genetic sampling effects of two individuals s and t given the same marker genotype i (second case) is the difference of (5) and (11):

$$m_{ii}^{a_i} = v_{ii}^l - \left(\frac{1}{n_i^l} \tilde{v}_{ii}^l + \left(1 - \frac{1}{n_i^l} \right) v_{ii}^l \right) = -\frac{1}{n_i^l} (\tilde{v}_{ii}^l - v_{ii}^l). \quad (14)$$

For $n_i^l \geq 1$ $m_{ii}^{a_i} \in [-0.5, 0.0]$. The variance $\tilde{m}_{ii}^{a_i} + m_{ii}^{a_i}$ of the additive genetic sampling effect given the

marker genotype i (first case) is the difference of (3) and (11),

$$\tilde{m}_{ii}^{a_i} + m_{ii}^{a_i} = \tilde{v}_{ii}^l - \left(\frac{1}{n_i^l} \tilde{v}_{ii}^l + \left(1 - \frac{1}{n_i^l} \right) v_{ii}^l \right) = \left(1 - \frac{1}{n_i^l} \right) (\tilde{v}_{ii}^l - v_{ii}^l) \quad (15)$$

with $\tilde{m}_{ii}^{a_i} \in [0.0, 0.5]$. Note that the elements $\tilde{m}_{ii}^{a_i} = \tilde{v}_{ii}^l - v_{ii}^l$ are independent of n_i^l . However, $\tilde{m}_{ii}^{a_i}$ and $m_{ii}^{a_i}$ depend on conditional genotype probabilities. From (14) and (15) it is obvious that $\tilde{m}_{ii}^{a_i}$ is a function of the covariance of the additive genetic sampling effects from the same marker class i and the corresponding number n_i^l of observations, $\tilde{m}_{ii}^{a_i} = -n_i^l m_{ii}^{a_i}$.

The calculation of the relationship matrix of the additive genetic effect of the individual model (6) and the reduced model (8) as well as the additive genetic sampling relationship matrix $\mathbf{V}_{m_{a_i}}$ is summarized in Table 1.

If model (6) includes not only additive genetic but also dominance effects, the genetic parameters for average dominance effects and dominance sampling terms can be obtained analogously. The genetic sampling relationship matrices of the pairwise epistatic effects can also be calculated similarly to the additive genetic and dominance effects, but the row vectors \mathbf{I}_{ij}^{lk} that considered the joint conditional QTL genotype probabilities of the l th and k th QTL have to be used. Then n_{lk} different marker classes have to be considered, where $n_{lk} = 27$ if the QTL are in two adjacent marker intervals and $n_{lk} = 81$ otherwise.

If we assume that the number of F_2 individuals approaches infinity ($n \rightarrow \infty$), then the number of individuals per marker class i also increases ($n_i^l \rightarrow \infty$). The diagonal elements \tilde{a}_{ii}^l as well as $m_{ii}^{a_i}$ depend on n_i^l , where $1/n_i^l$ tends to zero for $n \rightarrow \infty$. Hence $\lim_{n_i^l \rightarrow \infty} \mathbf{V}_{m_{a_i}} = \mathbf{\Delta}_{a_i}$, where $\mathbf{\Delta}_{a_i}$ is a diagonal matrix of order n of elements $\tilde{m}_{ii}^{a_i}$. Therefore, the covariance matrix of the additive genetic sampling effects is asymptotically diagonal.

Reduced model: Instead of an individual model we developed a reduced model approach, which is an approximation of model (8), with decreased dimension of the relationship matrices. The LMM is $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \tilde{\mathbf{Z}}_l \tilde{\mathbf{u}}_a + \boldsymbol{\varepsilon}$, where the residuals are assumed to be independently and identically normally distributed with $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma_\varepsilon^2)$. Here the F_2 individuals are grouped according to their marker genotypes and average genetic effects are estimated for marker classes instead of individual genetic effects, as described in (8). The dimension of the relationship matrices depends on the number of marker classes (n_l and n_{lk}), but not on the experiment size n . We call this procedure the reduced model (vs. the individual model).

In general, the reduced model with respect to additive genetic, dominance, and pairwise epistatic effects is

TABLE 1

Correspondence of elements of additive genetic relationship matrices in the individual model and the equivalent model with additive genetic sampling effects

Case	Individual model	Equivalent model	
	$(\mathbf{Z}_l \mathbf{V}_a \mathbf{Z}_l')_{st}$	$(\tilde{\mathbf{Z}}_l \tilde{\mathbf{V}}_a \tilde{\mathbf{Z}}_l')_{st}$	$(\mathbf{V}_{m_i})_{st}$
1	\tilde{v}_{ii}^j	$\frac{1}{n_i^j} \tilde{v}_{ii}^j + \left(1 - \frac{1}{n_i^j}\right) v_{ii}^j$	$\left(1 - \frac{1}{n_i^j}\right) (\tilde{v}_{ii}^j - v_{ii}^j)$
2	v_{ii}^j	$\frac{1}{n_i^j} \tilde{v}_{ii}^j + \left(1 - \frac{1}{n_i^j}\right) v_{ii}^j$	$-\frac{1}{n_i^j} (\tilde{v}_{ii}^j - v_{ii}^j)$
3	v_{ij}^l	v_{ij}^l	0

Each variable in the second column (individual model) is the sum from the two expressions of the third and fourth columns (equivalent model). Case 1: diagonal elements for marker class $i \in \{1, \dots, 9\}$; case 2: two individuals with equal marker class i ; case 3: two individuals with different marker classes i and j

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \sum_{l=1}^v \tilde{\mathbf{Z}}_l (\tilde{\mathbf{u}}_{a_l} + \tilde{\mathbf{u}}_{d_l}) + \sum_{l=1}^{v-1} \sum_{k=l+1}^v \tilde{\mathbf{Z}}_{lk} (\tilde{\mathbf{u}}_{aa_{lk}} + \tilde{\mathbf{u}}_{ad_{lk}} + \tilde{\mathbf{u}}_{da_{lk}} + \tilde{\mathbf{u}}_{dd_{lk}}) + \boldsymbol{\varepsilon}, \tag{16}$$

where the residuals are again assumed to be independently and identically normally distributed with $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma_\varepsilon^2)$. The vectors $\tilde{\mathbf{u}}_\tau$ with $\tau \in \{a_l, d_l, aa_{lk}, ad_{lk}, da_{lk}, dd_{lk}\}$ consider the average additive genetic, dominance, and pairwise epistatic effects of length n_l and n_{lk} .

The calculation of the reduced dominance relationship matrix $\tilde{\mathbf{V}}_{d_l}$ at the l th QTL is done similarly to the notes above, but \mathbf{A} has to be replaced by \mathbf{D} . Both $\tilde{\mathbf{V}}_{a_l}$ and $\tilde{\mathbf{V}}_{d_l}$ are matrices of order n_b where $n_l = 9$ if the flanking markers are fully informative. The reduced epistatic relationship matrices $\tilde{\mathbf{V}}_{aa_{lk}}, \tilde{\mathbf{V}}_{ad_{lk}}, \tilde{\mathbf{V}}_{da_{lk}}, \tilde{\mathbf{V}}_{dd_{lk}}$ of the l th and k th QTL are computed analogously to $\tilde{\mathbf{V}}_{a_l}$ from (10) and (12), but the corresponding Kronecker product is used instead of \mathbf{A} and the row vector $\mathbf{1}_{ij}^{lk}$ for the i th and j th marker class is applied.

The difference $\tilde{v}_{ii}^j - v_{ii}^j$ (asymptotic variance $\tilde{m}_{ii}^{a_i}$) between the variance of an individual additive genetic effect and the covariance between two additive genetic

effects of the same marker class decreases as the distance between flanking markers becomes smaller. Decreasing QTL effects and genetic variances lead to the same effect. In the extreme case, when the marker location and the position of the QTL coincide, the difference $\tilde{v}_{ii}^j - v_{ii}^j$ is zero and therefore $\mathbf{V}_{m_{a_i}} = \mathbf{0}$. In this case the covariances of the phenotypes in the reduced and the individual model are identical. Therefore, approximating $\mathbf{V}_{m_{a_i}} \sigma_{a_i}^2 + \mathbf{I}\sigma_e^2$ (or its multilocus equivalent) by $\mathbf{I}\sigma_\varepsilon^2$ seems to be a reasonable choice. Note that XU and ATCHLEY (1995) and XU (1998) investigated the inflation of the residual variance through the within-marker genotype QTL variance in the RIM, which is similar to our genetic sampling effects.

The approximation of the individual model by the reduced model relies on two different aspects. First, the covariances $m_{ii}^{a_i}$ between genetic sampling effects (deviation of individual genetic effects from average genetic effects of marker classes) are assumed to be zero. Second, the asymptotic variances $\tilde{m}_{ii}^{a_i}$ of the additive genetic sampling effects are treated as equal for all marker classes. Covariances $m_{ii}^{a_i}$ between additive genetic sampling effects of individuals sharing the same marker class i are shown in Table 2 for an additive QTL in the middle of a 10-cM marker interval in dependence on sample size. The elements $m_{ii}^{a_i}$ were calculated using

TABLE 2

Covariances $m_{ii}^{a_i}$ of additive genetic sampling effects within marker class i for different numbers (n) of F_2 individuals: QTL in the middle of a 10-cM marker interval

n	$m_{ii}^{a_i}$								
	$i = 1:$	$i = 2:$	$i = 3:$	$i = 4:$	$i = 5:$	$i = 6:$	$i = 7:$	$i = 8:$	$i = 9:$
	$G_{11/11}$	$G_{11/12}$	$G_{11/22}$	$G_{12/11}$	$G_{12/12}$	$G_{12/22}$	$G_{22/11}$	$G_{22/12}$	$G_{22/22}$
500	0.000	-0.012	-0.487	-0.012	0.000	-0.012	-0.487	-0.012	0.000
1000	0.000	-0.006	-0.243	-0.006	0.000	-0.006	-0.243	-0.006	0.000
2000	0.000	-0.003	-0.122	-0.003	0.000	-0.003	-0.122	-0.003	0.000
3000	0.000	-0.002	-0.081	-0.002	0.000	-0.002	-0.081	-0.002	0.000

Flanking marker genotypes $G_{i,j}$ are indexed by their alleles for each i .

TABLE 3

Asymptotic variances \tilde{m}_{ii}^a of additive genetic sampling effects within marker class i for differently sized marker intervals and different QTL positions within marker intervals (cM)

Marker interval	Position of QTL	\tilde{m}_{ii}^a								
		$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$	$i = 9$
0	0	0.00								
10	1	0.00	0.09	0.18	0.09	0.00	0.09	0.18	0.09	0.00
10	2	0.00	0.16	0.32	0.16	0.01	0.16	0.32	0.16	0.00
10	3	0.00	0.21	0.42	0.21	0.01	0.21	0.42	0.21	0.00
10	4	0.00	0.24	0.48	0.24	0.01	0.24	0.48	0.24	0.00
10	5	0.00	0.25	0.50	0.25	0.01	0.25	0.50	0.25	0.00
20	10	0.02	0.26	0.50	0.26	0.04	0.26	0.50	0.26	0.02
30	15	0.04	0.27	0.50	0.27	0.08	0.27	0.50	0.27	0.04
40	20	0.07	0.29	0.50	0.29	0.13	0.29	0.50	0.29	0.07

the number of expected proportions for each marker genotype according to Equation 14. To make sure that $n_i^l \geq 1$, we used $n \geq 500$. For 500 F_2 individuals this covariance m_{ii}^a is $\leq 1\%$ of the QTL variance and shows a further decline when the sample size increases. Only for marker classes $G_{11/22}$ and $G_{22/11}$ is there a very high (negative) covariance (48.7% of the additive genetic variance) and an experiment with >2000 F_2 individuals would be required to reach a value $<10\%$. These marker genotypes are rare, we expect these marker genotypes to occur twice in total among 500 F_2 genotypes. Therefore, omitting these covariances m_{ii}^a has little effect on the likelihood.

The asymptotic variances \tilde{m}_{ii}^a of the additive genetic sampling effects for different marker classes are, however, larger than their corresponding covariances and, more importantly, they show considerable variation between more frequent marker classes. The sixth line of Table 3 shows the genetic sampling variances for all marker classes, again for an additive QTL in the middle of a 10-cM marker interval. For the three most frequent marker classes, the genetic sampling variance is at $\leq 1\%$ of the additive genetic variance (classes 1, 5, and 9) and for another four marker classes it equals 25% (classes 2, 4, 6, and 8), while a 50% value occurs only in the very rare classes (3 and 7). Note that the genetic sampling variances become smaller when the QTL is located

closer to the boundary of the marker interval. The genetic sampling effects completely vanish if marker locations and positions of the QTL coincide (Table 3, first line). In such cases, the covariances of the genetic sampling effects are zero and the assumption $\text{Var}(\boldsymbol{\varepsilon}) = \mathbf{I}\sigma_e^2$ of the reduced model (16) is exact.

The latter considerations suggest, as a further alternative, a weighted approach, where the second part of the approximation inherent in the reduced model, *i.e.*, equal genetic sampling variances for all marker classes, is skipped, while the assumption (first part) of zero covariances for genetic sampling effects within marker class is maintained. For a single additive QTL this results in the following mixed model equations (MME):

$$\begin{pmatrix} \mathbf{X}'\mathbf{W}\mathbf{X} & \mathbf{X}'\mathbf{W}\tilde{\mathbf{Z}}_l \\ \tilde{\mathbf{Z}}_l'\mathbf{W}\mathbf{X} & \tilde{\mathbf{Z}}_l'\mathbf{W}\mathbf{Z}_l + \tilde{\mathbf{V}}_a\boldsymbol{\lambda} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \tilde{\mathbf{u}}_a \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{W}\mathbf{y} \\ \tilde{\mathbf{Z}}_l'\mathbf{W}\mathbf{y} \end{pmatrix},$$

where $\boldsymbol{\lambda} = \sigma_e^2/\sigma_a^2$. The variance of the residuals is $\text{Var}(\boldsymbol{\varepsilon}) = \text{Var}(\mathbf{m}_a + \mathbf{e}) = \mathbf{W}^{-1}\sigma_e^2$, where σ_e^2 is defined as in (8) and all other symbols as in (1) and (16). The diagonal matrix \mathbf{W} of order n has the entries $w_{ss} = \sigma_e^2(\tilde{m}_{ii}^a\sigma_a^2 + \sigma_e^2)^{-1}$, which differ between observations from different marker classes and are equal for observations from the same marker class i . If more QTL and nonadditive genetic gene actions are considered in the model, then the genetic sampling variances for

TABLE 4

Brief summary of simulated scenarios: the number of QTL ν , length of the chromosome l_c (cM), QTL positions P_1 and P_2 (cM), marker positions (cM), residual variance σ_e^2 , additive genetic effects (a_1 , a_2), and additive-by-additive genetic effects aa_{12} as well as the relative QTL variance R^2 (%)

Scenario	ν	l_c	P_1	P_2	Marker positions	σ_e^2	a_1	a_2	aa_{12}	R^2
1	1	50	35	—	0, 10, 20, 30, 40, 50	9.529	1.0	—	—	5.0
2	1	50	35	—	0, 10, 20, 30, 40, 50	1.000	1.0	—	—	33.3
3	2	50	25	35	0, 10, 20, 30, 40, 50	0.181	1.0	-1.0	—	50.0
4	2	80	35	45	0, 40, 80	1.000	1.0	-1.0	1.0	30.0

TABLE 5

Average estimates (mean) for QTL positions (P_1, P_2) with associated root mean squared error (RMSE) and quantiles together with mean estimates of the residual variance $\hat{\sigma}_r^2$ and the observed power (%) for different scenarios: 200 F_2 individuals per simulated experiment and 1000 replications per scenario

	Reduced model		Individual model		RIM		MIM	
	P_1	P_2	P_1	P_2	P_1	P_2	P_1	P_2
Scenario 1								
Mean	32.58		32.53		32.73		32.40	
RMSE	10.71		10.79		10.78		10.89	
5% quantile	9.00		9.00		9.00		9.00	
95% quantile	48.00		48.00		48.00		48.00	
$\hat{\sigma}_r^2$	9.49		9.38		9.50		9.37	
Power	81.70		81.80		81.70		81.90	
Scenario 2								
Mean	34.82		34.87		34.88		34.68	
RMSE	2.80		2.66		2.68		2.57	
5% quantile	31.00		31.00		31.00		31.00	
95% quantile	39.00		39.00		39.00		39.00	
$\hat{\sigma}_r^2$	1.04		0.99		1.04		0.99	
Power	100.00		100.00		100.00		100.00	
Scenario 3								
Mean	21.67	38.38	23.70	36.70	22.22	37.82	24.70	34.89
RMSE	4.20	4.17	2.03	2.02	4.29	4.25	1.35	1.38
5% quantile	17.00	35.00	22.00	34.00	17.00	32.00	23.00	33.00
95% quantile	25.00	42.00	26.00	38.00	28.00	42.00	27.00	37.00
$\hat{\sigma}_r^2$	0.27		0.18		0.28		0.19	
Power	100.00		100.00		100.00		100.00	
Scenario 4								
Mean	34.03	46.22	33.56	46.70	34.13	46.09	32.29	44.44
RMSE	5.36	5.50	5.26	5.47	5.51	5.55	6.47	3.72
5% quantile	24.50	41.00	24.00	41.00	24.00	41.00	21.00	41.00
95% quantile	39.00	56.50	39.00	57.00	39.00	57.00	39.00	52.00
$\hat{\sigma}_r^2$	1.21		0.93		1.23		1.04	
Power	99.20		99.40		99.60		100.00	

different QTL and different kinds τ of genetic effects have to be summed to get the entire genetic sampling variance of an observation and w_{ss} (sth individual given the marker class i) becomes

$$w_{ss} = \frac{\sigma_e^2}{\sum_{\tau} \tilde{m}_{ii}^{\tau} \sigma_{\tau}^2 + \sigma_e^2}, \tag{17}$$

where $\tau \in \{a_b, d_b, aa_{lk}, ad_{lk}, da_{lk}, dd_{lk}\}$. This weighted version of the reduced model retains the advantage of a reduced dimension of the QTL relationship matrices as in the reduced model, but may provide a better approximation of the exact residual log-likelihood-ratio test (RLRT) statistics. If marker location and position of QTL coincide, the weights of (17) are one and \mathbf{W} is an identity matrix. The weights of (17) are similar to the weights in the weighted least-squares method of QTL mapping as shown by XU and ATCHLEY (1995) and XU (1998).

Coincidence of markers and QTL results in singularity of $\tilde{\mathbf{V}}_a$ (identical to \mathbf{A} in this case) and was not further considered here. However, this situation can be treated,

e.g., by regularization [adding a small quantity to the diagonal elements of $\tilde{\mathbf{V}}_a$ (NEUMAIER 1998)], which has little effect on the test statistics and is easy to implement, by including allelic effects in the model instead of genotypic effects, or by replacing $\tilde{\mathbf{V}}_a$ by a reduced rank approximation (RÖNNEGÅRD *et al.* 2007) obtained by spectral decomposition.

SIMULATIONS

First, a single F_2 family as the simplest case of a combination of multiple line crosses was considered to demonstrate the properties of the reduced model in comparison to the individual model (XIE *et al.* 1998) and the fixed-effects methods MIM (KAO and ZENG 1997; KAO *et al.* 1999) and RIM (HALEY and KNOTT 1992; MARTÍNEZ and CURNOW 1992). Experiments from four different scenarios were simulated with 1000 replications per scenario and $n = 200$ F_2 individuals per experiment. Scenarios 1 and 2 consisted of a single additive genetic QTL at 35 cM on a single chromosome

TABLE 6

Average estimates (mean) for QTL positions (P_1 , P_2) with associated root mean squared error (RMSE) and quantiles together with mean estimates of the residual variance $\hat{\sigma}_r^2$ and the observed power (%) for the third scenario with 50 F_2 individuals for each of the four families per simulated experiment (1000 replications per scenario)

	Reduced model		Individual model	
	P_1	P_2	P_1	P_2
Mean	22.38	35.99	23.00	35.59
RMSE	6.77	4.78	6.26	4.27
5% quantile	8.00	28.00	9.00	28.00
95% quantile	28.00	45.00	28.00	44.00
$\hat{\sigma}_r^2$	0.21		0.18	
Power	99.50		99.60	

of 50 cM length, whereas in the other scenarios (3 and 4) there were two linked QTL with equally sized QTL effects in repulsion. In the fourth scenario chromosome length was extended to 80 cM and an interaction effect was included. For further characteristics of all scenarios see Table 4. The observations were simulated using Cockerham's F_2 -metric model (COCKERHAM 1954; KAO and ZENG 2002, Table 3). The relative QTL variance R^2 is the proportion of the phenotypic variance σ_p^2 explained by the QTL and is $R^2 = \sigma_{QTL}^2 / \sigma_p^2$.

In the second part our small simulation study focused on the performance of the reduced *vs.* the individual model in a situation with multiple families. Four independent F_2 families, each with 50 progeny ($n = 200$), were derived from a population consisting of four different inbred lines, representing all pairwise combinations of QTL genotypes (G_{QQHH} , G_{QQhh} , G_{qqHH} , G_{qqhh}). For each family F_1 individuals were generated from a random pair of inbred lines. Markers were always assumed to be fully informative. In the LMM family means were treated as fixed. Remaining parameters were chosen as previously described for the third scenario (Table 4). For each genetic effect a single (population-specific) variance was assumed. The simulated data can be found in File S1.

Significance thresholds for the null hypothesis of no linked QTL were determined by simulating 1000 experiments of the same size for each scenario, where QTL with the same kind and size of effects were present, but unlinked to the markers. After analyzing these experiments, the 95% quantile of the maximum values of the test statistic from all replications was taken as a significance threshold, specific for each scenario and method, which allowed the determination of experimental power. We performed the residual log-likelihood-ratio test for the reduced and the individual model, the log-likelihood-ratio test for MIM, and the F -test for RIM. Mean QTL positions, root mean squared error (RMSE) of the QTL positions, and their 5% and 95% quantiles were evaluated to characterize the precision of location

estimates. For each replication we analyzed positions or combinations thereof, where marker locations and QTL positions did not coincide (step width 1 cM, both QTL in different marker intervals). Therefore, we applied RIM and MIM with the same restrictions as the VCM. Our analyses used the true genetic model for testing for segregating QTL; *i.e.*, the model included only the simulated effects of QTL and no model selection was performed. All calculations were done with self-written Fortran 95 programs in combination with ASReml (GILMOUR *et al.* 2008) for estimation of variance components and evaluation of the restricted maximum-likelihood function (PATTERSON and THOMPSON 1971).

DISCUSSION

Results for all simulated single-QTL scenarios are summarized in Table 5. The experimental power was 100% (scenarios 2 and 3) or nearly so (scenario 4), with the exception of scenario 1, where the experimental power was uniformly at 82% for all methods. There was almost no variation between methods in the mean estimated position in the single-QTL scenarios (1 and 2); even the distributions of the estimates showed identical 5% and 95% quantiles. Differences between methods became, however, apparent in the two-QTL scenarios. For scenario 3 (two QTL in repulsion, no interactions), MIM resulted in average estimated QTL positions at 24.7 and 34.9 cM, nearly identical to the simulated values at 25 and 35 cM. The RMSEs for positions of the QTL were <1.4 cM for both QTL for MIM and ~ 2.0 cM for the individual model, while the reduced model and RIM performed very similarly with RMSEs of ~ 4.2 cM. In scenario 3 the reduced model, the individual model, and RIM on average placed the QTL somewhat more toward the ends of the chromosome compared to MIM and the true values, resulting in an overestimation of the distance (true distance: 10 cM) between both QTL, ranging from 2.7 cM (individual model) to 6.7 cM (reduced model). For scenario 4 (two QTL in repulsion with interactions) this overestimation of the distance between the QTL was, however, very similar for all methods at ~ 2.0 – 3.1 cM. The RMSEs for estimated positions of the QTL were between 5.3 and 5.6 cM with little difference between the first and second QTL for RIM as well as the reduced and the individual model. However, the RMSE of MIM at the same time showed the highest deviation of 6.5 cM for the first and the smallest deviation of 3.7 cM for the second QTL.

Note that MIM was applied according to the original approach of KAO and ZENG (1997) and KAO *et al.* (1999), which ignores double recombination events (complete interference) within the marker interval. However, double recombinations were taken into account for RIM and the VCM.

As theory indicated, estimated residual variance components from methods coping better with genetic

deviations from the mean of a marker class (MIM, individual model) were smaller in the two-QTL scenarios compared to RIM and the reduced model, where the genetic sampling variance (QTL genotype variability within marker genotype) is part of the residual variance.

The results of the analysis of the multiple families are shown in Table 6. The accuracy of the estimated QTL positions of the individual model under consideration of four families was slightly better than that of the reduced model. However, when multiple families were considered, the difference between both models (reduced and individual model) was less than that of the single family (scenario 3). The RMSEs for positions of the QTL as shown in Table 6 were increased compared to the RMSEs of the third scenario of Table 5, because not all families are fully informative. The observed power of the individual and the reduced model again almost reached 100%. As expected, the estimated residual variance was inflated by the within-marker genotype QTL variance.

The required CPU time for ASREML (GILMOUR *et al.* 2008) of the reduced and the individual model was 26.7 and 80.1 sec for each repetition recorded on an HP DL380 G6 (72 GB RAM, 2× XEON X5570, 2.93 GHz, multiuser environment) in a two-QTL scenario with only additive genetic effects (four families); *i.e.*, the individual model required threefold more computing time. The run time required for the evaluation of a single QTL (scenario 1 or 2) was sevenfold for the individual model compared with the reduced model for each repetition. If the number of individuals and the number of variance components increase, the speed gain of the reduced model relative to the individual model is expected to increase.

Average RLRT profiles from the reduced and the individual model were almost identical for the first scenario with a single QTL (Figure 1A). For two QTL in scenario 3 (Figure 1B), the shapes of the RLRT surfaces from both methods were again very similar, but the average size of the maximum was higher for the individual model (60.62 compared to 44.52). The RLRT surfaces of scenario 4 of the reduced and the individual model as well as the weighted reduced model are nearly identical (results not shown). The likelihood profile of the weighted approach was smaller than that of the reduced model, but QTL positions seemed to be estimated more accurately.

The considerable advantage of the reduced model with respect to computing time is achieved by a smaller number of genetic effects accompanied by a smaller dimension of their associated covariance matrices. Moreover, this dimension does not depend on the size of the experiment, in contrast to the individual model. The amount of savable computing time can be expected to vary somewhat between different REML algorithms. Average information (AI) REML (GILMOUR *et al.* 1995; JOHNSON and THOMPSON 1995) may be implemented

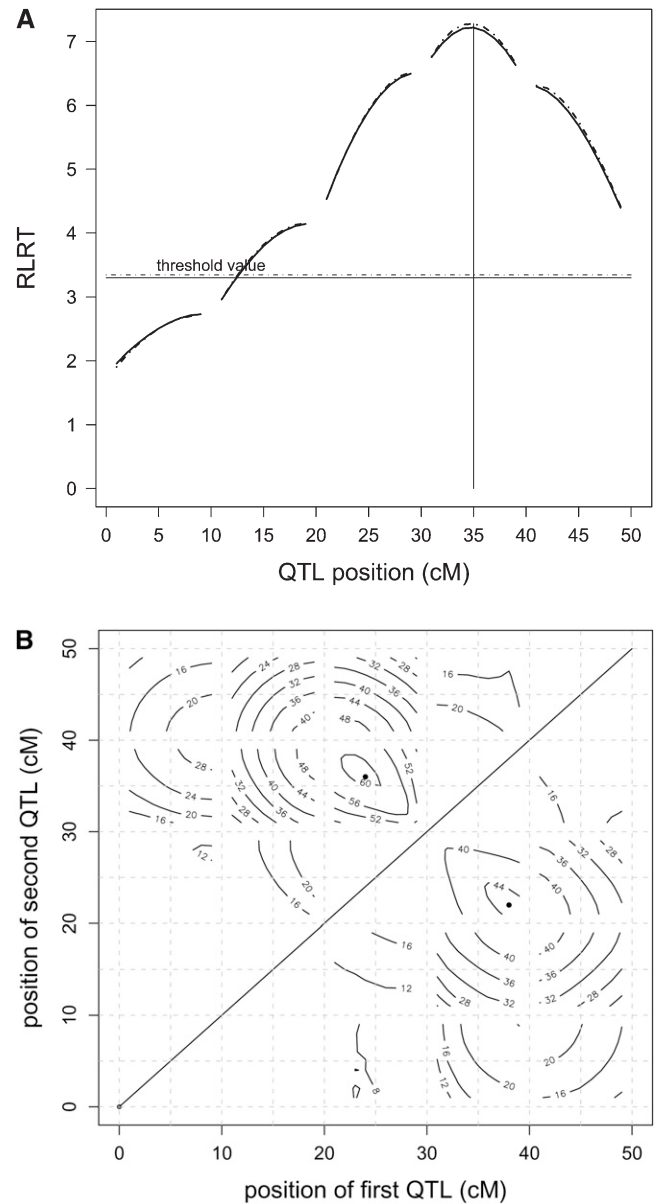


FIGURE 1.—For a single QTL (scenario 1) average RLRT profiles (A) of the individual model (dashed line) and the reduced model (solid line) nearly coincide, and so do their significance thresholds. When two QTL were present (scenario 3), contour plots (B) of the RLRT surfaces from the reduced model (below diagonal) and the individual model (above diagonal) showed a similar shape, but different absolute heights (respective RLRT maxima 44.52 and 60.62). Averaging was over 1000 replications.

either in an MME-based version or as a variant requiring the inversion of the covariance matrix \mathbf{V} of phenotypes, termed the “direct method” by LEE and VAN DER WERF (2006). These authors recommend the direct method if genetic covariance matrices are dense because of both speed and numerical stability. Application of the Sherman–Morrison–Woodbury matrix identity (*e.g.*, HENDERSON and SEARLE 1981; XU 1998) to determine the inverse of \mathbf{V} results in

$$\mathbf{V}^{-1} = (\mathbf{HGH}' + \mathbf{R})^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{H}(\mathbf{G}^{-1} + \mathbf{H}'\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}'\mathbf{R}^{-1},$$

where \mathbf{R} denotes the covariance matrix of residuals, \mathbf{G} is the covariance matrix of all genetic effects (block diagonal), and \mathbf{H} is the corresponding incidence matrix. To obtain \mathbf{V}^{-1} the inversion of a dense matrix of the same order as \mathbf{G} is required, which usually is considerably smaller than the number of observations for the reduced model (e.g., $\dim(\mathbf{G}) = 9 \times 9$ for a single QTL with additive genetic effects and $\dim(\mathbf{G}) = 36 \times 36$ for two QTL with additive genetic and dominant effects). In conclusion, the increase in computing speed obtained by the reduced model may differ between algorithms, but is substantial when compared with the individual model, thus broadening the general applicability of the VCM for mapping purposes.

The amount of possible improvement of the reduced model obtained by accounting for genetic sampling variation within marker classes remains to be investigated. A more comprehensive comparison of methods than presented here is underway to obtain a more complete picture. Despite the limited number of scenarios in our simulations, it can already be concluded that the proposed reduced model may be competitive with other standard methods for mapping of (multiple) QTL not only in terms of computing time, but also in terms of detection power and precision of estimated positions of the QTL.

The authors thank the reviewers for their helpful comments and suggestions. This research was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, MA 1553/3-1).

LITERATURE CITED

- ABDEL-AZIM, G., and A. E. FREEMAN, 2001 A rapid method for computing the inverse of the gametic covariance matrix between relatives for a marked quantitative trait locus. *Genet. Sel. Evol.* **33**: 153–173.
- CARBONELL, E. A., T. M. GERIG, E. BALANSARD and M. J. ASINS, 1992 Interval mapping in the analysis of nonadditive quantitative trait loci. *Biometrics* **48**: 305–315.
- CARLBORG, Ö., and C. S. HALEY, 2004 Epistasis: Too often neglected in complex trait studies? *Nat. Rev. Genet.* **5**: 618–625.
- COCKERHAM, C. C., 1954 An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**: 859–882.
- CREPIEUX, S., C. LEBRETON, B. SERVIN and G. CHARMET, 2004 Quantitative trait loci (QTL) detection in multicross inbred designs: recovering QTL identical-by-descent status information from marker data. *Genetics* **168**: 1737–1749.
- GILMOUR, A. R., R. THOMPSON and B. R. CULLIS, 1995 Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* **51**: 1440–1450.
- GILMOUR, A. R., B. J. GOGEL, B. R. CULLIS and R. THOMPSON, 2008 *ASReml User Guide Release 3.0*. VSN International, Hemel Hempstead, UK.
- GRIGNOLA, F. E., I. HOESCHELE and B. TIER, 1996 Mapping quantitative trait loci in outcross populations via residual maximum likelihood. I. Methodology. *Genet. Sel. Evol.* **28**: 479–490.
- 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.
- 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.

- HASEMAN, J. K., and R. C. ELSTON, 1972 The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* **2**: 3–19.
- HENDERSON, H. V., and S. R. SEARLE, 1981 On deriving the inverse of a sum of matrices. *SIAM Rev. Soc. Ind. Appl. Math.* **23**: 53–60.
- JANNINK, J.-L., and R. JANSEN, 2001 Mapping epistatic quantitative trait loci with one-dimensional genome searches. *Genetics* **157**: 445–454.
- JOHNSON, D. L., and R. THOMPSON, 1995 Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. *J. Dairy Sci.* **78**: 449–456.
- KAO, C.-H., 2000 On the differences between maximum likelihood and regression interval mapping in the analysis of quantitative trait loci. *Genetics* **156**: 855–865.
- KAO, C.-H., and Z.-B. ZENG, 1997 General formulas for obtaining the MLEs and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. *Biometrics* **53**: 653–665.
- KAO, C.-H., and Z.-B. ZENG, 2002 Modeling epistasis of quantitative trait loci using Cockerham's model. *Genetics* **160**: 1243–1261.
- KAO, C. H., Z. B. ZENG and R. D. TEASDALE, 1999 Multiple interval mapping for quantitative trait loci. *Genetics* **152**: 1203–1216.
- LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- LEE, S. H., and J. H. J. VAN DER WERF, 2006 An efficient variance component approach implementing an average information REML suitable for combined LD and linkage mapping with a general complex pedigree. *Genet. Sel. Evol.* **38**: 25–43.
- LI, G., and Y. CUI, 2009 A statistical variance components framework for mapping imprinted quantitative trait locus in experimental crosses. *J. Probab. Stat.* **2009**: 1–27.
- LIU, Y., G. B. JANSEN and C. Y. LIN, 2002 The covariance between relatives conditional on genetic markers. *Genet. Sel. Evol.* **34**: 657–678.
- MARTÍNEZ, 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.
- MAYER, M., 2005 A comparison of regression interval mapping and multiple interval mapping for linked QTL. *Heredity* **94**: 599–605.
- MAYER, M., Y. LIU and G. FREYER, 2004 A simulation study on the accuracy of position and effect estimates of linked QTL and their asymptotic standard deviations using multiple interval mapping in an F_2 scheme. *Genet. Sel. Evol.* **36**: 455–479.
- NEUMAIER, A., 1998 Solving ill-conditioned and singular linear systems: a tutorial on regularization. *SIAM Rev. Soc. Ind. Appl. Math.* **40**: 636–666.
- PATTERSON, H. D., and R. THOMPSON, 1971 Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**: 545–554.
- RÖNNEGÅRD, L., K. MISCHENKO, S. HOLMGREN and Ö. CARLBORG, 2007 Increasing the efficiency of variance component quantitative trait loci analysis by using reduced-rank identity-by-descent matrices. *Genetics* **176**: 1935–1938.
- RÖNNEGÅRD, L., R. PONG-WONG and Ö. CARLBORG, 2008 Defining the assumptions underlying modeling of epistatic QTL using variance component methods. *J. Hered.* **99**: 421–425.
- SMITH, S. P., 1984 *Dominance Relationship Matrix and Inverse for an Inbred Population*. Mimeo, Department of Dairy Science, Ohio State University, Columbus, OH.
- WANG, T., R. L. FERNANDO, S. VAN DER BEEK, M. GROSSMAN and J. A. M. VAN ARENDONK, 1995 Covariance between relatives for a marked quantitative trait locus. *Genet. Sel. Evol.* **27**: 251–274.
- XIE, C., D. D. GESSLER and S. XU, 1998 Combining different line crosses for mapping quantitative trait loci using the identical by descent-based variance component method. *Genetics* **149**: 1139–1146.
- XU, S., 1996 Mapping quantitative trait loci using four-way crosses. *Genet. Res.* **68**: 175–181.
- XU, S., 1998 Mapping quantitative trait loci using multiple families of line crosses. *Genetics* **148**: 517–524.
- XU, S., and W. R. ATCHLEY, 1995 A random model approach to interval mapping of quantitative trait loci. *Genetics* **141**: 1189–1197.

GENETICS

Supporting Information

<http://www.genetics.org/cgi/content/full/genetics.110.122333/DC1>

**Complex Genetic Effects in Quantitative Trait Locus Identification:
A Computationally Tractable Random Model for
Use in F₂ Populations**

Daisy Zimmer, Manfred Mayer and Norbert Reinsch

Copyright © 2011 by the Genetics Society of America
DOI: 10.1534/genetics.110.122333

FILE S1**Supporting Data**

File S1 is available for download as a compressed file at <http://www.genetics.org/cgi/content/full/genetics.110.122333/DC1>.

The folder contains:

ReadMe.pdf
Data_scenario_one.txt
Data_scenario_two.txt
Data_scenario_three.txt
Data_scenario_four.txt
Data_multiple_families.txt