Reduced Variance Component Model To Map QTL

D. Zimmer[*], M. Mayer[*] and N. Reinsch[*]

Introduction

In mapping experiments it is often desirable to fit multiple quantitative trait loci (QTL) with additive genetic, dominance and epistatic effects simultaneously. One reason is that some QTL can only be identified if their interactions with other loci (epistasis) are considered. Especially, if two linked loci have opposed effects, i. e. QTL in repulsion, epistatic effects have to be considered. A variance component method (VCM) was suggested by Xie et al. (1998) for detecting single QTL in an F_2 design from a cross of two divergent parental inbred lines. They developed an efficient approach to set up the additive genetic and dominance relationship matrices with conditional QTL genotype probabilities and elementary QTL relationship matrices in a single QTL analysis. We extended the approach of Xie et al. (1998) to multiple QTL and pairwise epistatic effects. Furthermore, we developed a reduced model, which estimates average genotypic effects for flanking marker genotypes instead of individual effects, in order to substantially decrease the computational requirements.

Theory

Linear mixed model. The linear mixed model (LMM) with respect to additive genetic, dominance and pairwise epistatic effects of the QTL in an F_2 population is

$$\mathbf{y} = \mathbf{X}\beta + \sum_{\ell=1}^{\nu} \mathbf{Z}_\ell(\mathbf{u}_{a_\ell} + \mathbf{u}_{d_\ell}) + \sum_{\ell=1}^{\nu-1} \sum_{k=\ell+1}^{\nu} \mathbf{Z}_{\ell k}(\mathbf{u}_{aa_{\ell k}} + \mathbf{u}_{ad_{\ell k}} + \mathbf{u}_{da_{\ell k}} + \mathbf{u}_{dd_{\ell k}}) + \mathbf{e}\,. \quad (1)$$

The vector of phenotypes is $\mathbf{y}$ (order p), where p is the number of individuals with an observation. In total ν QTL are considered. A pair of QTL is indicated by ℓ and k. Let β be the vector of fixed effects and $\mathbf{X}$ the related design matrix. The vectors $\mathbf{u}_w$ with $w \in \{a_\ell, d_\ell, aa_{\ell k}, ad_{\ell k}, da_{\ell k}, dd_{\ell k}\}$ denote the additive genetic, dominance and the four pairwise epistatic effects. The expectation of a QTL effect is $E(\mathbf{u}_w) = \mathbf{0}$ and the covariance matrix is $V(\mathbf{u}_w) = \mathbf{V}_w\sigma^2_w$, where σ^2_w is the related QTL variance component and $\mathbf{V}_w$ is the corresponding relationship matrix. The design matrices $\mathbf{Z}_\ell$ with $\dim(\mathbf{Z}_\ell) = p \times n_\ell$ and $\mathbf{Z}_{\ell k}$ with $\dim(\mathbf{Z}_{\ell k}) = p \times n_{\ell k}$ relate the records to genetic effects. The parameters n_ℓ and $n_{\ell k}$ denote the number of effects to be estimated for the ℓ th QTL and for a pair of QTL, respectively. The residuals are independently and identically distributed with $\mathbf{e} \sim MVN(\mathbf{0}, \mathbf{I}\sigma^2_e)$, where $\mathbf{I}$ is the identity matrix and σ^2_e is the residual variance component.

[*]Leibniz Institute for Farm Animal Biology, Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany

Conditional QTL genotype probabilities. The QTL genotype probabilities are conditional probabilities, which depend on flanking marker genotypes and the recombination rate between markers and QTL. Taking the flanking markers of a QTL, nine different marker genotypes can be distinguished and the conditional QTL genotype probabilities can be derived as described e. g. by Haley and Knott (1992). It is assumed that there is at most a single QTL within a marker interval. The QTL alleles are denoted by Q and q. The conditional probabilities for the genotypes QQ, Qq and qq at the ℓ th QTL are written in the columns of a matrix $\mathbf{L}_\ell$ with $\dim(\mathbf{L}_\ell) = n_\ell \times 3$. For pairwise QTL combinations the matrix $\mathbf{L}_{\ell k}$ with $\dim(\mathbf{L}_{\ell k}) = n_{\ell k} \times 9$ contains the joint probabilities for both QTL genotypes given the marker genotypes of each individual. Because there is a completely informative marker between them, the joint probability is the product of both individual conditional probabilities. Double recombination events are fully considered. Any of the numerous programs available to calculate these conditional probabilities for F_2 line-cross experiments can be applied for our purposes.

Elementary QTL relationship matrices. The elementary QTL relationship matrices assume that the marker location and position of the QTL coincide. In such a case the QTL genotype can be clearly derived from the marker genotype. The elementary matrices for additive QTL effects $\mathbf{A}$ (Xie et al., 1998) and dominance QTL effects $\mathbf{D}$ (Smith, 1984; Xie et al., 1998) are

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \quad \text{and} \quad \mathbf{D} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The rows and columns of $\mathbf{A}$ and $\mathbf{D}$ correspond to the QTL genotypes QQ, Qq and qq. We use the Kronecker product (symbol $\otimes$) of the elementary matrices to compute the four different elementary QTL interaction matrices (order nine), $\mathbf{A}\otimes\mathbf{A}$, $\mathbf{A}\otimes\mathbf{D}$, $\mathbf{D}\otimes\mathbf{A}$, $\mathbf{D}\otimes\mathbf{D}$, including covariances of pairwise epistatic effects between two loci.

QTL relationship matrices for the full model. The required relationship matrices for F_2 individuals at an assumed QTL position are set up with conditional QTL genotype probabilities and elementary QTL relationship matrices (Xie et al., 1998). The matrices are positive definite at non-marker positions. If the QTL genotypes of two individuals s and t are known, the relationship coefficient for all possible nine QTL genotype combinations may be directly taken from $\mathbf{A}$. If the recombination rates are different from zero the relationship coefficient is a weighted average of the QTL genotype combinations. The additive genetic and dominance relationship matrices $\mathbf{V}_{a_\ell}$ and $\mathbf{V}_{d_\ell}$ at the ℓ th QTL ($\ell = 1, \ldots, \nu$) include one row and column for each individual (order p). The elements of $\mathbf{V}_{a_\ell}$ are set up analogous to Xie et al. (1998) by $a^\ell_{st} = (\mathbf{L}_\ell \mathbf{A} \mathbf{L}'_\ell)_{st}$ and $a^\ell_{ss} = (\mathbf{L}_\ell \operatorname{diag}(\mathbf{A}))_s$ with $s, t = 1, \ldots, p$. The vector of diagonal elements of $\mathbf{A}$ is denoted by $\operatorname{diag}(\mathbf{A})$. Analogously, the elements of $\mathbf{V}_{d_\ell}$ and the four pairwise epistatic relationship matrices are calculated, but $\mathbf{A}$ is substituted by the corresponding dominance and pairwise epistatic elementary QTL relationship matrices.

Following Xie et al. (1998) the full model approach, termed long VCM, considers conditional genotypic effects for each individual, i. e. $n_\ell = n_{\ell k} = p$, and therefore $\mathbf{Z}_\ell$ and $\mathbf{Z}_{\ell k}$ are equal to an identity matrix $\mathbf{I}$. The size of the relationship matrices, as well as the computational demand, rises considerably with the number of F_2 individuals. Hence, we suggest a reduced model, where average genetic effects for each marker class are estimated instead of individual effects.

QTL relationship matrices of the reduced VCM. An equivalent model to (1) is

$$\mathbf{y} = \mathbf{X}\beta + \sum_{\ell=1}^{\nu} \widetilde{\mathbf{Z}}_\ell(\widetilde{\mathbf{u}}_{a_\ell} + \widetilde{\mathbf{u}}_{d_\ell}) + \sum_{\ell=1}^{\nu-1} \sum_{k=\ell+1}^{\nu} \widetilde{\mathbf{Z}}_{\ell k}(\widetilde{\mathbf{u}}_{aa_{\ell k}} + \widetilde{\mathbf{u}}_{ad_{\ell k}} + \widetilde{\mathbf{u}}_{da_{\ell k}} + \widetilde{\mathbf{u}}_{dd_{\ell k}}) + \epsilon\,. \quad (2)$$

In the reduced VCM the F_2 individuals are grouped according to their flanking marker genotypes and average conditional genetic effects $\widetilde{\mathbf{u}}_w$ for each marker class are estimated. The vector of residuals ϵ of (2) are the sum of deviations of individual genetic effects from average effects and the residual deviation $\mathbf{e}$ of (1). In a simplifying manner - justified by arguments of asymptotic - the covariance structure of ϵ is treated as $\mathbf{I}\sigma^2_\epsilon$. Taking account of the observed number n_i^ℓ for a given marker genotype i, the elements of the reduced additive relationship matrix, denoted as $\widetilde{\mathbf{V}}_{a_\ell}$, at the ℓ th QTL become

$$\begin{aligned} \widetilde{a}^\ell_{ii} &= \begin{cases} \left(\widetilde{\mathbf{L}}_\ell \text{diag}(\mathbf{A})\right)_{i} & \text{if} \quad n_i^\ell = 0\,, \\ \frac{1}{n_i^\ell}\left((\widetilde{\mathbf{L}}_\ell \text{diag}(\mathbf{A}))_{i} + (n_i^\ell - 1)(\widetilde{\mathbf{L}}_\ell \mathbf{A} \widetilde{\mathbf{L}}'_\ell)_{ij}\right) & \text{else}\,, \end{cases} \\ \widetilde{a}^\ell_{ij} &= (\widetilde{\mathbf{L}}_\ell \mathbf{A} \widetilde{\mathbf{L}}'_\ell)_{ij}\,, \end{aligned}$$

with $i, j = 1, \ldots, n_\ell$ and $i \neq j$. The matrix $\widetilde{\mathbf{L}}_\ell$ now contains a single row per each possible marker genotype. The calculation of $\widetilde{\mathbf{V}}_{d_\ell}$ at the ℓ th QTL is done similarly to the notes above, but $\mathbf{A}$ has to be substituted by $\mathbf{D}$. For $\widetilde{\mathbf{V}}_{a_\ell}$ and $\widetilde{\mathbf{V}}_{d_\ell}$ $n_\ell = 9$ given that the markers are fully informative and we are going to estimate an average genotypic effect for each marker genotype. The epistatic relationship matrices $\widetilde{\mathbf{V}}_{aa_{\ell k}}$, $\widetilde{\mathbf{V}}_{ad_{\ell k}}$, $\widetilde{\mathbf{V}}_{da_{\ell k}}$, $\widetilde{\mathbf{V}}_{dd_{\ell k}}$ at the ℓ th and k th QTL are computed analogously to $\widetilde{\mathbf{V}}_{a_\ell}$ using the corresponding Kronecker product instead of $\mathbf{A}$. In total we estimate $n_{\ell k}$ levels of the epistatic effects, where $n_{\ell k} = 27$ if the QTL are in two adjacent marker intervals, and $n_{\ell k} = 81$ otherwise.

Simulation study. The ability of the reduced VCM in terms of estimating the QTL positions was studied by a small simulation. The result of the reduced VCM was compared to the long VCM. Experimental size (F_2) was 200 for each of 1000 replicates. A chromosome of 80 cM length with three markers at 0, 40 and 80 cM was simulated with two QTL at 35 and 45 cM. The additive genetic effects were set to $a_1 = -a_2 = 1$ (QTL in repulsion) and the additive by additive interaction was $aa_{12} = 1$. The population mean was set to zero and $\sigma^2_e = 1$. The genetic effects were simulated using Cockerham's model (Kao and Zeng, 2002). The genetic variance components were estimated by REML. The detection of QTL was achieved by a residual likelihood ratio test, computed with the software ASReml (Gilmour et al., 2006). The test statistic was RLRT $= 2\,(\mathcal{L}_{H_A} - \mathcal{L}_{H_0})$, where $\mathcal{L}_{H_A}$ and $\mathcal{L}_{H_0}$ were the logarithmic residual likelihood under H_A (at least one genetic variance component > 0) and H_0 (all genetic variance components zero), respectively. A chromosome-wide significance threshold (5%) was determined with 1000 additional simulations, where no linked QTL was segregating.

Results and discussion

The criteria of comparison were the accuracy of the mean, the empirical standard error (sd) and the mean squared error (mse) of the estimated QTL positions. From Table 1 it is obvious

that the parameter estimates of both methods were very similar. The power of the reduced and the long VCM was $99.2\,\%$ and $99.4\,\%$, respectively. Both methods identified nearly the same positions of the QTL in all runs. The average likelihood ratio profiles over 1000 replications, each with 200 individuals, of the reduced and the long VCM are very similar (results not shown) with maximum RLRT $= 20.2$ and maximum RLRT $= 21.5$, respectively. It is obviously, that the reduced VCM produced somewhat flatter average likelihood profiles. However, the reduced VCM needed essentially less computing time than the long VCM. The average estimates of the residual variance component of the reduced and long VCM was $\hat{\sigma}^2_\epsilon = 1.21$ and $\hat{\sigma}^2_e = 0.93$, respectively. Note that the estimated residual variance components of both methods are not directly comparable.

Table 1: Estimated positions of QTL (P_1, P_2) over 1000 replications.

	reduced VCM		long VCM	
	P_1	P_2	P_1	P_2
mean	34.03	46.22	33.56	46.70
sd	5.28	5.37	5.06	5.20
mse	28.75	30.27	27.62	29.92

The reduced relationship matrices depend on the exact number of observations per marker genotype i (n_i^ℓ), affecting the variance of the average genetic effects. Alternatively, the expected number of observations per marker genotype could be used to compute the corresponding row in the relationship matrix.

The reduced VCM is a flexible approach to map multiple linked QTL regarding multiple marker intervals. A major advantage of the presented method is that it is easy to implement and it is a useful genome scan method for detecting QTL. The proposed method can currently be used for F_2 populations from inbred parental populations or lines which are homozygous at the QTL. The reduced VCM is computationally fast and numerically stable. The advantage of the reduced VCM increases with growing marker density and experimental size. Therefore, this method is expected to have a practical importance for future QTL analyses. Further investigations comparing the performance of the reduced and the long VCM are, however, necessary.

References

Gilmour, A. R., Gogel, B. J., Cullis, B. R., and Thompson, R. (2006). VSN International Ltd, Hemel Hempstead, UK.

Haley, C. S. and Knott, S. A. (1992). *Heredity*, 69(4):315–324.

Kao, C.-H. and Zeng, Z.-B. (2002). *Genetics*, 160(3):1243–1261.

Smith, S. P. (1984). Mimeo, Dep. Dairy Sci., The Ohio State Univ., Columbus.

Xie, C., Gessler, D. D., and Xu, S. (1998). *Genetics*, 149(2):1139–1146.