Genet. Sel. Evol. 35 (2003) 605–622
© INRA, EDP Sciences, 2003
DOI: 10.1051/gse:2003042

Original article

A comparison of bivariate and univariate QTL mapping in livestock populations

Peter SØRENSEN*, Mogens Sandø LUND, Bernt GULDBRANDTSEN, Just JENSEN, Daniel SORENSEN

Danish Institute of Agricultural Sciences,
Department of Animal Breeding and Genetics,
Research Centre Foulum, PO Box 50, 8830 Tjele, Denmark

(Received 22 April 2002; accepted 6 March 2003)

Abstract – This study presents a multivariate, variance component-based QTL mapping model implemented *via* restricted maximum likelihood (REML). The method was applied to investigate bivariate and univariate QTL mapping analyses, using simulated data. Specifically, we report results on the statistical power to detect a QTL and on the precision of parameter estimates using univariate and bivariate approaches. The model and methodology were also applied to study the effectiveness of partitioning the overall genetic correlation between two traits into a component due to many genes of small effect, and one due to the QTL. It is shown that when the QTL has a pleiotropic effect on two traits, a bivariate analysis leads to a higher statistical power of detecting the QTL and to a more precise estimate of the QTL's map position, in particular in the case when the QTL has a small effect on the trait. The increase in power is most marked in cases where the contributions of the QTL and of the polygenic components to the genetic correlation have opposite signs. The bivariate REML analysis can successfully partition the two components contributing to the genetic correlation between traits.

multivariate / QTL mapping / livestock

1. INTRODUCTION

In many quantitative trait loci (QTL) mapping experiments in livestock populations, a number of phenotypic traits are recorded *e.g.* [8, 11, 26]. Usually, QTL are mapped for individual traits using single trait analyses. The traits, however, may be environmentally and genetically correlated. 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 QTLs, each affecting one trait only [5].

When a QTL has a pleiotropic effect on two or more traits, a joint analysis involving both traits can result in a higher statistical power of detecting it, and in higher precision of the estimate of its map position [14, 15].

* Corresponding author: pso@genetics.agrsci.dk

Apart from the issue of power, it is important to understand the structure of a genetic correlation between two traits. Indeed, partitioning the genetic correlation into a component due to the action of many pleiotropic genes of small effect, and another due to the effect of a pleiotropic QTL can provide relevant information, for example, for selection decisions.

Several approaches for a multivariate QTL analysis have been proposed. One is to use a canonical transformation of the original data followed by single trait analyses [16,23]. However, a transformation that uncorrelates the traits phenotypically and genetically in the transformed scale does not ensure that each QTL influences a single canonical trait only [15]. A second approach is to use multivariate least squares methods for QTL detection and location *e.g.* [3,15]. This approach was applied to a three-generation pedigree and was shown to increase the power to detect a pleiotropic QTL, and the precision of the estimate of its location, relative to a univariate approach [15]. The advantage of multivariate least squares is that it is easy to implement without using sophisticated software and the method is computationally fast. However, it is not applicable for more general pedigree structures with many different relationships and multiple generations, as found typically in livestock populations. A third approach is to use multivariate maximum likelihood (ML) methods. These have been implemented for a number of different experimental designs, such as crosses between inbred lines [14], and half-sib families [19]. The multivariate ML methods have been shown to result in estimates of parameters with improved precision and to increase the power to detect QTL. The advantage of a fully parametric ML method is that it explicitly models the number of loci, the number of alleles per locus and their frequencies and that it can be applied to general pedigrees. However, a fully parametric ML method is computationally demanding.

Here, a multivariate QTL mapping approach based on the variance component model *e.g.* [1,9,10,24] is presented. This model decomposes the overall genetic variance into a component due to the segregation of a putative QTL, and another due to the effect of a polygenic term (the collective effects of all other QTL affecting the trait). An advantage of this approach is that it can be applied to general pedigree structures and multiple generations *e.g.* [12,7]. In this study, the model is implemented *via* restricted maximum likelihood (REML). The maximization of the restricted likelihood is achieved using a novel and efficient algorithm known as average information [13].

The variance component model has previously been applied to a multivariate QTL mapping analysis, and shown to increase the statistical power to detect QTL, relative to univariate analyses [2]. However, the results from power studies for different scenarios of genetic and phenotypic relationships between traits have not been given. A more detailed simulation study is needed to evaluate the properties of the multivariate variance component-based QTL

mapping approach. This would highlight situations in which it is advantageous to use multivariate QTL analyses.

The objective of this work was to implement the multivariate variance component-based QTL mapping model *via* REML and to compare bivariate and univariate QTL mapping analyses of simulated data, with respect to the statistical power to detect a QTL and to the precision of parameter estimates. In particular, we studied genetic scenarios that lead to differences in power between univariate and multivariate analyses. The developed methodology was also applied to partition the overall genetic correlation into components due to the action of many pleiotropic genes and due to a single pleiotropic QTL.

2. METHODS

2.1. Multivariate mixed model

The multivariate mixed model with a single QTL can be written in generalized matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{q} + \mathbf{e}, \tag{1}$$

where $\mathbf{y}$ is a $n * t$ vector of n observations on t traits, $\mathbf{X}$ is a known design matrix, $\boldsymbol{\beta}$ is a vector of unknown fixed effects, $\mathbf{Z}$ is a known matrix relating records to individuals, $\mathbf{u}$ is a vector of unknown additive polygenic effects, $\mathbf{W}$ is a known matrix relating each individual record to its unknown additive QTL effect, $\mathbf{q}$ is a vector of unknown additive QTL effects of individuals and $\mathbf{e}$ is a vector of residuals. Here model (1) is considered as the full model and for tests of hypothesis described in the next section, a number of different sub models is derived from it.

The random variables $\mathbf{u}$, $\mathbf{q}$ and $\mathbf{e}$ are assumed to be multivariate normally distributed and mutually uncorrelated (MVN).

Specifically, the vector $\mathbf{u}$ is MVN ($\mathbf{0}$, $\mathbf{G}_0 \otimes \mathbf{A}$), the vector $\mathbf{q}$ is MVN ($\mathbf{0}$, $\mathbf{K}_0 \otimes \mathbf{Q}_{|\mathbf{M},\mathbf{p}}$) and the vector $\mathbf{e}$ is MVN (0, $\mathbf{E}_0 \otimes \mathbf{I}$). Matrices $\mathbf{G}_0$, $\mathbf{K}_0$ and $\mathbf{E}_0$ include variances and covariances among the traits due to polygenic effects, QTL effects and residuals effects, respectively. The symbol $\otimes$ represents the Kronecker product. Matrix $\mathbf{A}$ has elements that describe additive genetic relationships among elements of $\mathbf{u}$. Matrix $\mathbf{Q}_{|\mathbf{M},\mathbf{p}}$ is the identity by descent (IBD) matrix of the QTL, and is a function of marker data ($\mathbf{M}$) and the position (p) of the QTL on the chromosome.

2.2. IBD matrix

The IBD matrix for the QTL effects, $\mathbf{Q}_{|\mathbf{M},\mathbf{p}}$, was computed constructing first the gametic relationship matrix [6], and then using the linear relationship

between the gametic relationship matrix and the IBD matrix [7]. The gametic relationship matrix describes the covariance structure among the random QTL allelic effects of all the individuals in the pedigree. The covariance between any two QTL allelic effects is proportional to the probability that the QTL alleles are identical by descent. The gametic relationship matrix of a QTL is not observable because the QTL genotype is unknown. However, transmission of linked markers can be followed from the parents to the offspring. This information is used to calculate IBD probabilities at the position of a putative QTL, thus yielding an expected gametic relationship matrix, conditional on QTL position and marker information.

In outbred populations, markers may only be partially informative. It is therefore important to use information on all markers in the linkage group. Here, information from all markers in the analysis was accounted for in a similar way as described in Yi and Xu [25]. The method is illustrated using a simple pedigree consisting of a sire with QTL alleles (g_{S1} and g_{S2}) and a single offspring with QTL alleles (g_{O1} and g_{O2}). Consider a linkage group with m marker loci. Assume that the QTL (q) is located between markers k and $k+1$ for $1 \leq k \leq m-1$. The probability that the paternal QTL allele (g_{O1}) in the offspring is identical by descent ($\equiv$) to the first QTL allele (g_{S1}) in the sire, given the inherited parental marker haplotype ($\mathbf{H}^{\text{pat}}$) can be written as

$$P(g_{O1} \equiv g_{S1}|\mathbf{H}^{\text{pat}}) = \frac{\mathrm{P}(\mathbf{H}^{\text{pat}}|g_{O1} \equiv g_{S1})\mathrm{P}(g_{O1} \equiv g_{S1})}{\mathrm{P}(\mathbf{H}^{\text{pat}}|g_{O1} \equiv g_{S1})\mathrm{P}(g_{O1} \equiv g_{S1}) + \mathrm{P}(\mathbf{H}^{\text{pat}}|g_{O1} \equiv g_{S2})\mathrm{P}(g_{O1} \equiv g_{S2})}, \quad (2)$$

where $\mathrm{P}(g_{O1} \equiv g_{S1})$ and $\mathrm{P}(g_{O1} \equiv g_{S2})$ are the prior distribution of the IBD state for the QTL which are equal to 0.5. The conditional probability of the inherited haplotype in the offspring, given the inheritance of the first QTL allele from the sire, can then be computed as [25]

$$\mathrm{P}(\mathbf{H}^{\text{pat}}|g_{O1} \equiv g_{S1}) = \begin{pmatrix}1\\1\end{pmatrix}^{\mathrm{T}} \mathbf{N}_1\mathbf{R}_{1,2}\ldots\mathbf{N}_k\mathbf{R}_{k,q}\begin{pmatrix}1&0\\0&0\end{pmatrix}\mathbf{R}_{q,k+1}\mathbf{N}_{k+1}\ldots\mathbf{N}_{m-1}\mathbf{R}_{m-1,m}\mathbf{N}_m\begin{pmatrix}1\\1\end{pmatrix}$$

and similarly for the second allele (g_{S2})

$$\mathrm{P}(\mathbf{H}^{\text{pat}}|g_{O1} \equiv g_{S2}) = \begin{pmatrix}1\\1\end{pmatrix}^{\mathrm{T}} \mathbf{N}_1\mathbf{R}_{1,2}\ldots\mathbf{N}_k\mathbf{R}_{k,q}\begin{pmatrix}0&0\\0&1\end{pmatrix}\mathbf{R}_{q,k+1}\mathbf{N}_{k+1}\ldots\mathbf{N}_{m-1}\mathbf{R}_{m-1,m}\mathbf{N}_m\begin{pmatrix}1\\1\end{pmatrix}.$$

The matrix $\mathbf{R}_{k,k+1} = \begin{pmatrix}1-r_{k,k+1} & r_{k,k+1}\\ r_{k,k+1} & 1-r_{k,k+1}\end{pmatrix}$ is computed using the recombination fraction $r_{k,k+1}$ between loci k and $k+1$. The matrix, $\mathbf{N}_k =$

$\begin{pmatrix} \mathrm{P}(m^k_{O1} \equiv m^k_{S1}|\mathbf{M}^k) & 0 \\ 0 & \mathrm{P}(m^k_{O1} \equiv m^k_{S2}|\mathbf{M}^k) \end{pmatrix}$ is computed using the probabilities that the paternal marker allele (m^k_{O1}) in the offspring, is IBD with the first (m^k_{S1}) or second (m^k_{S2}) marker allele in the sire, at the marker locus k. If the marker information is complete, then one of the diagonal elements of $\mathbf{N}_k$ is equal to 1 and the other diagonal element is equal to zero. In the absence of marker information, the diagonal elements of $\mathbf{N}_k$ are equal to 0.5. Equation (2) was used to compute the IBD elements in the gametic relationship matrix for a given position of the QTL, using a recursive algorithm [22] and assuming the most likely linkage phase is the true linkage phase in the sire.

2.3. AI-REML analysis

Conditional on the IBD matrix for the QTL effects, $\mathbf{Q}_{|\mathbf{M},\mathbf{p}}$, the restricted likelihood [18] of the multivariate mixed model, assuming a single QTL, is given by

$$L(\boldsymbol{\theta}|\mathbf{K}'\mathbf{y}, \mathbf{Q}_{|\mathbf{M},\mathbf{p}}) \propto \iint \mathrm{p}(\mathbf{K}'\mathbf{y}|\mathbf{u}, \mathbf{q}, \mathbf{E}_0 \otimes \mathbf{I})\mathrm{p}(\mathbf{u}|\mathbf{G}_0 \otimes \mathbf{A})\mathrm{p}(\mathbf{q}|\mathbf{K}_0 \otimes \mathbf{Q}_{|\mathbf{M},\mathbf{p}})\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{q}, \tag{3}$$

where $\boldsymbol{\theta} = \left(\mathrm{vech}(\mathbf{G}_0)' \ \mathrm{vech}(\mathbf{K}_0)' \ \mathrm{vech}(\mathbf{E}_0)'\right)$ 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 with respect to the variance components ($\mathbf{G}_0$, $\mathbf{K}_0$ and $\mathbf{E}_0$) using the AI-REML algorithm [13]. Preceding the AI-REML analysis and using only marker data, the IBD matrix $\mathbf{Q}_{|\mathbf{M},\mathbf{p}}$ is computed, conditional on the QTL position p, on the chromosome. 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 [13]. It was implemented by combining it with the Expectation Maximization (EM) algorithm [4], to ensure that parameter estimates stay within the parameter space [13]. 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. This has to be accounted for in order to detect convergence, which 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 the 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 the first derivatives is not necessarily zero. Therefore the other convergence

criterion requires that changes in estimates of the (co)variance components between successive rounds approach zero.

2.4. Simulation

A granddaughter design with 20 unrelated grandsires, each having 50 sons, was simulated. Each son produced 100 daughters, and dams of sons were assumed to be unrelated. The structure and size of this design resembles that of a current experiment involving the Danish Holstein population [11].

2.4.1. Genetic scenarios

To compare univariate and bivariate QTL mapping analyses, a number of different genetic scenarios were simulated (Tab. I). All the simulations mimic a situation where two traits are affected by a single pleiotropic QTL, in addition to polygenic and residual effects. The QTL was placed at a map position of 34 cM from the start of the linkage group. In order to evaluate the robustness of the method to changes in the number of QTL alleles, the QTL was simulated using either a biallelic or a multiallelic QTL model. The variance ratios (λ_1 and λ_2) involving the proportion of genetic variance explained by the QTL, were 15% for trait 1 and 5% for trait 2. In all scenarios, the total phenotypic variance was 100 for each trait, and the polygenic heritabilities (h_1^2 and h_2^2) were 0.3 and 0.14 for traits 1 and 2, respectively. The simulated scenarios differed in the correlations between traits due to the QTL ($r_{\mathbf{K}}$), polygenes ($r_{\mathbf{G}}$) and residuals ($r_{\mathbf{E}}$). In Table I, each alternative is characterized by three signs indicating a characteristic of the correlation between the QTL effects, the polygenic effects and the residual effects, in this order. A "+" indicates that the correlation is positive, a "−" that it is negative, and a "0" that it is zero. Specifically, the QTL correlation was 0.5 in the multiallelic case and 1.0 in the biallelic case. The polygenic and residual correlations were zero in the "+00" scenario. The polygenic correlation was 0.5 in the "+ + +" and "+ + −" scenarios and −0.5 in the "+ − +" and "+ − −" scenarios. The residual correlation was 0.5 in the "+ + +" and "+ − +" scenarios, and −0.5 in the "+ + −" and "+ − −" scenarios. The analyses presented are based on 200 replicated simulations.

2.4.2. Marker and QTL genotypes

The simulated linkage group was 80 cM long. It consisted of five markers which were positioned at 0, 20, 40, 60 and 80 cM. Founder alleles (*i.e.* alleles in grandsires and all maternally inherited alleles) were sampled from a base population which was assumed to be in Hardy Weinberg and linkage equilibrium. Five alleles with equal frequencies were simulated for each marker, whereas the simulation of the QTL was biallelic with equal frequencies. In the case of the multiallelic QTL model, all founder QTL alleles were assumed to be different.

Table I. Parameters in the designs simulated with a multiallelic or a biallelic QTL model.

	Design									
	Multiallelic					Biallelic				
Parameter	+00*	+++	++−	+−+	+−−	+00	+++	++−	+−+	+−−
p			0.34					0.34		
λ_1			0.15					0.15		
λ_2			0.05					0.05		
h_1^2			0.3					0.3		
h_2^2			0.14					0.14		
r_K	0.5	0.5	0.5	0.5	0.5	1.0	1.0	1.0	1.0	1.0
r_G	0	0.5	0.5	−0.5	−0.5	0	0.5	0.5	−0.5	−0.5
r_E	0	0.5	−0.5	0.5	−0.5	0	0.5	−0.5	0.5	−0.5

* The three symbols describe the correlations between the QTL effects, the polygenic effects and the residual effects, respectively. "+" indicates that the correlation is positive; "−" that it is negative; and "0" that is null. p is the position of the QTL in cM; $\lambda_1 = \dfrac{\sigma_{q1}^2}{\sigma_{q1}^2 + \sigma_{u1}^2}$ is the variance ratio due to the QTL for trait 1; λ_2 is the variance ratio due to the QTL for trait 2; h_1^2 and h_2^2 are the heritabilities for trait 1 and 2; r_K, r_G and r_E are correlations due to QTL, polygenes and residual.

Alleles were transmitted from parents to offspring according to the Haldane mapping function. Marker genotypes were simulated for grandsires and their sons, while QTL genotypes were simulated for grandsires, sons and daughters.

2.4.3. Phenotypes

For each son, a daughter yield deviation (DYD) based on 100 daughters was simulated. DYD is an average of the phenotypes of the daughters adjusted for the fixed effects and genetic values of the daughters' dams [21]. For the *i*th son, the phenotype was simulated as a sum of the effects due to the QTL, the polygenes and the residuals, using the following model:

$$\mathbf{DYD}_i = \frac{1}{n_i}\sum_{j=1}^{n} \mathbf{q}_{ij} + \mathbf{u}_i + \mathbf{e}_i,$$

where $\mathbf{DYD}_i = \begin{pmatrix} \mathrm{DYD}_{i1} \\ \mathrm{DYD}_{i2} \end{pmatrix}$ is a vector of daughter yield deviations for trait 1 and 2 for son i, n_i is the number of daughters of son i, $\mathbf{q}_{ij} = \begin{pmatrix} q^p_{ij1} \\ q^p_{ij2} \end{pmatrix}$ is a vector of the paternal (p) QTL allelic effects for trait 1 and 2 in daughter j of son i, $\mathbf{u}_i = \begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix}$ is a vector of polygenic effects and $\mathbf{e}_i = \begin{pmatrix} e_{i1} \\ e_{i2} \end{pmatrix}$ is a vector of residual effects.

The QTL effects were sampled as follows. For the biallelic QTL model with alleles Q and q, genotypes QQ, Qq, and qq were assigned the effects $a_1(a_2)$, 0(0), and $-a_1(-a_2)$ for trait 1(2). For example, if the individual i genotype is QQ, then the QTL effect for trait 1 is $q_{i1} = a_1$. The total variance explained by the QTL is $2\sigma^2_{q1} = 2p_Q(1-p_Q)a_1^2$ for trait 1, and $2\sigma^2_{q2} = 2p_Q(1-p_Q)a_2^2$ for trait 2, respectively, where p_Q is the frequency of the Q allele. The covariance between the traits due to the QTL is $2\sigma_{q1q2} = 2p_Q(1-p_Q)a_1a_2$. Therefore the correlation between the traits is unity.

In the multiallelic QTL model the QTL effects for founder alleles were drawn from MVN $(0, \mathbf{K}_0)$, where $\mathbf{K}_0 = \begin{pmatrix} \sigma^2_{q1} & \sigma_{q1q2} \\ \sigma_{q2q1} & \sigma^2_{q1} \end{pmatrix}$ is the 2×2 (co)variance matrix of the QTL effects. Under both QTL models, sampling of the daughters' QTL generated the contribution of the QTL to the DYD. This sampling of the QTL effects ensures that the variance between DYD among the daughters of a heterozygous son, is larger than the corresponding variance associated with a homozygous son.

The polygenic effects $\mathbf{u}_i$ were sampled from MVN $(\mathbf{0}, \mathbf{G}_0 \otimes \mathbf{A})$, where $\mathbf{G}_0 = \begin{pmatrix} \sigma^2_{u1} & \sigma_{u1u2} \\ \sigma_{u2u1} & \sigma^2_{u1} \end{pmatrix}$ is the 2×2 additive genetic (co)variance matrix between

traits and $\mathbf{A}$ is the relationship matrix. Specifically, the polygenic effects for the grandsire were generated from MVN(0,$\mathbf{G}_0$), and for a son, from MVN($0.5u_{\text{sire}}$, $0.75\mathbf{G}_0$), where u_{sire} is the polygenic effect for the sire of the son, and $0.75\mathbf{G}_0$ is the sum of the genetic variance from unknown dams and the Mendelian sampling term.

The residual effects, $\mathbf{e}_i$, were sampled from MVN $\left(\mathbf{0}, \frac{1}{n_i}(0.5\mathbf{G}_0 + \mathbf{E}_0)\right)$, where $\mathbf{E}_0 = \begin{pmatrix} \sigma^2_{e1} & \sigma_{e1e2} \\ \sigma_{e2e1} & \sigma^2_{e1} \end{pmatrix}$ is the 2×2 residual (co)variance matrix between the traits.

2.5. Hypotheses testing

Hypothesis testing for the presence of a QTL can be based on a single trait analysis, or on a joint analysis including several traits. Here, the joint analysis involves only two traits. The hypothesis tests in the univariate and bivariate testing procedures are performed using the likelihood ratio test statistic, $LRT = -2\ln(L_{\text{reduced}} - L_{\text{full}})$, where L_{reduced} and L_{full} are the maximized likelihoods under the reduced model and full model, respectively. The data analyzed in the tests described below, were simulated using model (1).

In the bivariate testing procedure, initially the null hypothesis "there is no QTL affecting the traits" was tested against the hypothesis "there is a QTL affecting both traits". This test was performed using the test statistic $LRT_{\text{B12}} = -2\ln(L_{\text{B0}} - L_{\text{B12}})$, where L_{B0} is the maximum likelihood for a bivariate model with no QTL affecting the traits and L_{B12} is the maximum likelihood for a bivariate model with a single pleiotropic QTL affecting both traits. This is a joint test for the combined effect of the QTL on both traits and, therefore, does not test whether each trait is significantly affected by the QTL. When the joint test was significant the two following trait specific tests were performed: First, the null hypothesis "there is a QTL affecting trait 1" was tested against the hypothesis "there is a QTL affecting both traits" using the test statistic $LRT_{\text{B1}} = -2\ln(L_{\text{B1}} - L_{\text{B12}})$. Second, the null hypothesis "there is a QTL affecting trait 2" was tested against the hypothesis "there is a QTL affecting both traits" using the test statistic $LRT_{\text{B2}} = -2\ln(L_{\text{B2}} - L_{\text{B12}})$. L_{B1} (L_{B2}) is the maximum likelihood for a bivariate model with a QTL affecting only trait 1 (trait 2).

In the univariate testing procedure, each trait was analyzed separately and the null hypothesis "there is no QTL affecting the trait" was tested against the hypothesis "there is a QTL affecting the trait" using the test statistic $LRT_{\text{U1}} = -2\ln(L_{\text{U0_1}} - L_{\text{U1}})$ for trait 1 and $LRT_{\text{U2}} = -2\ln(L_{\text{U0_2}} - L_{\text{U2}})$ for trait 2. $L_{\text{U0_1}}$ ($L_{\text{U0_2}}$) is the maximum likelihood for a univariate model with no QTL and L_{U1} (L_{U2}) is the maximum likelihood for a univariate model with a single QTL affecting trait 1 (trait 2).

For each test the likelihood ratio test statistic was calculated and compared with the empirically derived significance threshold (below we explain how this threshold was obtained).

The comparison in terms of power to detect a QTL *via* the univariate *versus* the bivariate QTL mapping approaches was as follows. In the bivariate QTL mapping approach, the power of detecting a QTL affecting trait 1 (B1) or the power of detecting a QTL affecting trait 2 (B2) was computed as the proportion out of the total number of replicates in which LRT_{B12} was larger than the threshold and where LRT_{B1} for trait 1 or LRT_{B2} for trait 2 was larger than the threshold. The overall power of detecting a QTL (B12) in the bivariate analyses was computed as the proportion out of the total number of replicates in which LRT_{B12} was larger than the threshold. In the univariate QTL mapping approach, the power of detecting a QTL for trait 1 (U1) or the power of detecting a QTL for trait 2 (U2) was computed as the proportion out of the total number of replicates in which the test statistics LRT_{U1} or LRT_{U2} was larger than the threshold. The overall power to detect a QTL in the univariate analyses (U12) was computed as the proportion out of the total number of replicates in which either the test statistics LRT_{U1} or LRT_{U2} was larger than the threshold.

2.5.1. Distribution of the test statistics

Under regularity conditions, the asymptotic distribution of the likelihood ratio test statistic follows a χ^2 distribution, with degrees of freedom equal to the difference in the number of independent parameters between the models tested [20]. However, in the context of gene mapping, the null hypothesis "there is no QTL affecting the trait(s)" places parameters on the boundary of the parameter space, and therefore the asymptotic distribution of the likelihood ratio test statistic has a non-standard form. Here, the empirical distribution of the test statistics was found by simulation of data under the specific null hypothesis used in the test. This approach also accounts for the large number of correlated tests along the chromosome [15].

2.5.2. Significance thresholds and models under the null hypothesis

In both the bivariate and the univariate testing procedure the thresholds under the null hypothesis "there is no QTL affecting the trait(s)" were obtained by simulating individuals using the same design and marker information as above, but with phenotypes depending on polygenic and residual effects only. In the bivariate testing procedure the thresholds under the null hypothesis "there is a QTL affecting trait 1" or the null hypothesis "there is a QTL affecting trait 2" were obtained with phenotypes depending on polygenic and residual effects in addition to a biallelic QTL affecting trait 1 or trait 2, respectively.

The test statistic was calculated for each replicate, and the 5% significance threshold was obtained from its distribution over 500 replicates. For the univariate tests (U1 and U2), to account for the fact that two traits were being analysed we used the 2.5% significance threshold value for each trait, which is equivalent to 5% over both traits (following a Bonferroni adjustment). For the bivariate test (B1, B2 and B12), the significance threshold already accounts for the fact that two traits are being analysed [15].

3. RESULTS

3.1. Power to detect QTL

The power of the univariate and bivariate tests to detect QTL in different simulated designs is shown in Table II. The results are based on the significance thresholds obtained by simulating data under the null hypothesis as explained above. The overall power of detecting a QTL using the bivariate joint test (B12) was generally higher than that obtained in the univariate tests (U12). The power of using the bivariate test (B2), for the case of a QTL affecting trait 2 that had a small effect on the phenotype, was generally 1.5–3.2 times higher than that of the univariate tests (U2). The highest increase in power, using the bivariate test (B1) compared to the univariate test (U1), was found when the QTL was biallelic and the QTL correlation and the polygenic correlation had opposite signs.

3.2. Estimation of parameters

Position estimates agreed well with the simulated values in both univariate and bivariate analyses (Tab. III). In the univariate analyses, estimates of the QTL position (p_1 and p_2) had a lower standard deviation for the trait when the QTL had a large effect. The bivariate analyses generated a single position estimate (p), which had a smaller standard deviation than that obtained from the univariate analyses, especially when the QTL correlation and the polygenic correlation had opposite signs. The type of QTL model (multiallelic or biallelic) did not affect the precision of the position estimates.

Estimates of the variance ratios (λ_1 and λ_2) involving the proportion of genetic variance explained by the QTL were close to the simulated values for both traits (Tab. IV). The estimates of heritabilities (h_1^2 and h_2^2) were slightly higher than the simulated heritabilities in both the univariate and bivariate analysis. There was no apparent difference in the standard deviation of the estimates between the two analyses.

Table V shows estimates of the correlation between the QTL effects on the two traits (r_K), correlation between polygenic effects on the two traits (r_G) and the correlation between residual effects on the two traits (r_E). When the QTL

Table II. Power of test to detect a QTL affecting one of two traits (1 or 2), and overall power (12) for the designs with a multiallelic or a biallelic QTL model.

	Design									
	Multiallelic					Biallelic				
	+00*	+++	++−	+−+	+−−	+00	+++	++−	+−+	+−−
Bivariate										
B12	0.81	0.89	0.85	0.93	0.96	0.85	0.75	0.83	0.98	0.96
B1	0.81	0.89	0.85	0.92	0.96	0.85	0.75	0.83	0.97	0.96
B2	0.29	0.31	0.22	0.29	0.38	0.48	0.42	0.44	0.56	0.55
Univariate										
U12	0.75	0.80	0.80	0.82	0.82	0.84	0.78	0.84	0.85	0.84
U1	0.71	0.72	0.76	0.78	0.77	0.81	0.74	0.81	0.82	0.79
U2	0.17	0.21	0.20	0.23	0.19	0.23	0.23	0.15	0.25	0.21

* The three symbols describe the correlations between the QTL effects, the polygenic effects and the residual effects, respectively. "+" indicates that the correlation is positive; "−" that it is negative; and "0" that is null.

Table III. Estimates [1] of position (p, p_1 and p_2) for the designs with a multiallelic or a biallelic QTL model.

	Design									
	Multiallelic					Biallelic				
	+00*	+++	++−	+−+	+−−	+00	+++	++−	+−+	+−−
Bivariate										
p	37.4 (13.7)	33.8 (12.4)	36.2 (12.9)	34.6 (10.9)	35.9 (7.6)	35.7 (12.2)	36.8 (14.8)	34.2 (12.8)	34.1 (8.2)	35.5 (7.9)
Univariate										
p_1	37.3 (12.8)	34.9 (14.5)	36.0 (12.2)	34.9 (13.5)	35.6 (13.5)	37.1 (12.1)	36.9 (12.9)	35.3 (12.3)	33.1 (11.9)	36.2 (13.3)
p_2	39.1 (23.0)	34.6 (21.8)	39.4 (19.6)	34.9 (22.6)	36.9 (19.9)	38.1 (22.9)	35.8 (21.8)	36.1 (20.9)	39.2 (21.0)	35.6 (22.7)

* The three symbols describe the correlations between the QTL effects, the polygenic effects and the residual effects, respectively. "+" indicates that the correlation is positive; "−" that it is negative; and "0" that is null. [1] SD in parentheses (200 replicates per design).

Table IV. Estimates[1] of variance ratios (λ_1 and λ_2) for the QTL and heritabilities (h_1^2 and h_2^2) for the designs with a multiallelic or a biallelic QTL model.

	Design									
	Multiallelic					Biallelic				
	+00*	+++	++−	+−+	+−−	+00	+++	++−	+−+	+−−
Bivariate										
λ_1	0.12 (0.07)	0.13 (0.08)	0.14 (0.08)	0.13 (0.08)	0.13 (0.07)	0.12 (0.07)	0.13 (0.06)	0.13 (0.07)	0.15 (0.07)	0.14 (0.07)
λ_2	0.06 (0.05)	0.06 (0.05)	0.06 (0.05)	0.05 (0.04)	0.07 (0.06)	0.08 (0.06)	0.07 (0.05)	0.06 (0.05)	0.06 (0.04)	0.08 (0.06)
h_1^2	0.36 (0.24)	0.36 (0.23)	0.37 (0.25)	0.37 (0.20)	0.34 (0.22)	0.40 (0.29)	0.37 (0.22)	0.36 (0.21)	0.33 (0.18)	0.38 (0.27)
h_2^2	0.18 (0.13)	0.16 (0.08)	0.16 (0.07)	0.19 (0.14)	0.16 (0.09)	0.15 (0.08)	0.17 (0.12)	0.15 (0.08)	0.17 (0.09)	0.17 (0.12)
Univariate										
λ_1	0.13 (0.09)	0.14 (0.07)	0.14 (0.07)	0.13 (0.08)	0.13 (0.07)	0.14 (0.06)	0.14 (0.06)	0.14 (0.06)	0.15 (0.06)	0.14 (0.07)
λ_2	0.06 (0.05)	0.07 (0.05)	0.07 (0.05)	0.06 (0.05)	0.08 (0.06)	0.08 (0.05)	0.08 (0.05)	0.07 (0.05)	0.06 (0.04)	0.09 (0.07)
h_1^2	0.36 (0.24)	0.36 (0.24)	0.39 (0.33)	0.38 (0.21)	0.34 (0.24)	0.39 (0.29)	0.36 (0.22)	0.36 (0.22)	0.35 (0.21)	0.38 (0.25)
h_2^2	0.18 (0.14)	0.16 (0.08)	0.16 (0.07)	0.19 (0.17)	0.16 (0.09)	0.15 (0.08)	0.17 (0.12)	0.14 (0.10)	0.17 (0.10)	0.17 (0.12)

* The three symbols describe the correlations between the QTL effects, the polygenic effects and the residual effects, respectively. "+" indicates that the correlation is positive; "−" that it is negative; and "0" that is null. [1] SD in parentheses (200 replicates per design).

Table V. Estimates[1] of the correlations (r_K, r_G and r_E) for the designs with a multiallelic or a biallelic QTL model.

	Design									
	Multiallelic					Biallelic				
	+00	+++	++−	+−+	+−−	+00	+++	++−	+−+	+−−
Full model										
$r_{K\,(\mathrm{QTL})}$	0.48 (0.58)	0.32 (0.65)	0.39 (0.64)	0.50 (0.61)	0.49 (0.53)	0.81 (0.32)	0.86 (0.36)	0.87 (0.30)	0.85 (0.25)	0.76 (0.40)
$r_{G\,(\mathrm{polygenic})}$	0.02 (0.13)	0.51 (0.09)	0.52 (0.13)	−0.49 (0.11)	−0.49 (0.08)	0.03 (0.11)	0.50 (0.09)	0.53 (0.10)	−0.51 (0.12)	−0.50 (0.10)
$r_{E\,(\mathrm{residual})}$	−0.02 (0.37)	0.45 (0.30)	−0.44 (0.36)	0.40 (0.30)	−0.51 (0.31)	−0.08 (0.40)	0.49 (0.33)	−0.41 (0.33)	0.44 (0.31)	−0.42 (0.37)
Reduced model										
$r_{G0\,(\mathrm{polygenic})}$	0.04 (0.10)	0.50 (0.07)	0.49 (0.10)	−0.41 (0.10)	−0.41 (0.09)	0.09 (0.10)	0.53 (0.08)	0.55 (0.10)	−0.38 (0.10)	−0.38 (0.09)
$r_{E0\,(\mathrm{residual})}$	−0.01 (0.35)	0.43 (0.33)	−0.42 (0.35)	0.41 (0.32)	−0.49 (0.32)	−0.07 (0.40)	0.45 (0.35)	−0.44 (0.32)	0.42 (0.35)	−0.45 (0.34)

* The three symbols describe the correlations between the QTL effects, the polygenic effects and the residual effects, respectively. "+" indicates that the correlation is positive; "−" that it is negative; and "0" that is null. [1] SD in parentheses (200 replicates per design).

was multiallelic, estimates of the QTL correlation ranged from 0.32 to 0.52. There was a tendency of estimates being smaller than the simulated value of 0.5 when the QTL correlation had the same sign as the polygenic correlation. When the QTL was biallelic, estimates of the QTL correlation ranged from 0.76 to 0.87 and were generally biased downwards, compared to the simulated correlation (1.0), as expected. The precision of the QTL correlation was higher when the QTL was biallelic. Estimates of the polygenic correlation agreed well with the simulated value and were in general estimated with a higher precision than the QTL and residual correlations.

Table V also shows the estimates of correlation between polygenic effects on the two traits (r_{G0}) and the correlation between residual effects on the two traits (r_{E0}) when data was analyzed using the fully reduced model. Compared to estimates from the full model, the polygenic correlations differed when the QTL correlation was positive and the polygenic correlation was negative. As expected, there was no difference in the estimates of the residual correlations.

4. DISCUSSION

In this study we implemented a multivariate variance component-based QTL mapping model *via* REML, and compared bivariate and univariate QTL mapping analyses with respect to the statistical power to detect a QTL and the precision of parameter estimates. The simulation study showed that when a QTL has a pleiotropic effect on two traits, a bivariate QTL mapping analysis leads generally to a gain in power, and to higher precision of estimates of QTL positions. Additionally, the bivariate REML analysis can successfully partition the two components contributing to the genetic correlation between traits.

The gain in statistical power using a bivariate analysis was observed across different scenarios of genetic and phenotypic relationships between traits. However, it is important to detect the kind of genetic structures where a joint analysis can be beneficial. This is so, because joint analyses of several traits result typically in a larger number of parameters to be estimated. This can have the effect of reducing power and increasing sampling variances of estimates of parameters of interest. The simulation study indicated that it is particularly beneficial to do a joint analysis involving two traits, when the QTL affects a trait with low heritability. This is important because marker assisted selection is especially relevant for such traits. Furthermore, the joint analysis was beneficial when the QTL is biallelic, has a small effect on the phenotypes and when the QTL correlation and the polygenic correlation have opposite signs. A previous study reported that the gain in power was the highest when the QTL correlation and the residual correlation had opposite signs [14]. On the contrary to the work reported here, this result was obtained using a model without a polygenic component to the genetic correlation.

The improved precision of position estimates with a bivariate model was also observed using a similar variance component model applied on human pedigrees [2]. Similar results were obtained in multivariate regression models applied to a granddaughter design [15], an F2 design [14] and a backcross design [19]. Thus, improved precision of the position estimates seems to be a feature of multivariate QTL mapping across methods and different designs.

The higher precision of position estimates observed when the QTL explained a larger proportion of the total genetic variance (15% *versus* 5%) was expected. In this situation, more information is available for inferring the position, and is a result found in other simulation studies [8].

The bivariate REML analysis can successfully partition the genetic correlation into a component due to the action of many pleiotropic genes of small effect, and another due to the effect of a pleiotropic QTL. In cases where the polygenic and QTL correlations have opposite signs the partioned genetic correlation can provide relevant information, especially, for selection decisions.

The bivariate test for detecting a QTL does not distinguish between a pleiotropic QTL having an effect on both traits, and two closely linked QTL each having an effect on only one trait. Distinguishing between these two situations can be important from several perspectives. From an animal breeding perspective it is important to know whether the genetic correlation due to a QTL can be broken. This could be possible in the case of two linked QTL, but not if the correlation is due to a single pleiotropic QTL. For finemapping [17] and positional cloning of gene(s), it is obviously important to know if one or two genes are responsible for the genetic correlation.

In a bivariate analysis, two closely linked QTL in linkage disequilibrium each with an effect on only one trait may behave like one pleiotropic QTL. This is because the correlation between the QTL effects is proportional to the degree of linkage disequilibrium [5]. This complicates the interpretation of pleiotropy versus co-incident linkage in a bivariate analysis. The test for pleiotropy versus close linkage may be most relevant in combination with finemapping methods [17]. Research in this area is in progress and will be reported elsewhere.

The results presented in this study are based on a situation with two traits and a single QTL affecting both traits. However, there are frequently more than two traits recorded and each trait may be affected by several QTL. Therefore we have implemented the AI-REML algorithm for a general multivariate variance component-based QTL mapping model. Different genetic models (multiple QTL, close linkage, pleiotropic) can easily be specified for each trait, but further research is needed to develop efficient strategies for the testing procedures involved with more complex multivariate models, especially in cases where the null hypothesis no longer is "there is no QTL affecting the traits" [15].

REFERENCES

[1] Almasy L., Blangero J., Multipoint quantitative-trait linkage analysis in general pedigrees, Am. J. Hum. Genet. 62 (1998) 1198–1211.

[2] Almasy L., Dyer T.D., Blangero J., Bivariate quantitative trait linkage analysis: pleiotropy *versus* co-incident linkages, Genet. Epidemiol. 14 (1997) 953–958.

[3] Calinski T., Kaczmarek Z., Krajewski P., Frova C., Sari-Gorla M., A multivariate approach to the problem of QTL localization, Heredity 84 (2000) 303–310.

[4] Dempster A.P., Laird N.M., Rubin D.B., Maximum likelihood from incomplete data *via* the EM algorithm, J. Roy. Statist. Soc. B 39 (1977) 1–39.

[5] Falconer D.S., Mackay T.F.C., Introduction to quantitative genetics, 4th edn., Longman Group Ltd., Essex, 1996.

[6] Fernando R.L., Grossman M., Marker-assisted selection using best linear unbiased prediction, Genet. Sel. Evol. 21 (1989) 467–477.

[7] George A.W., Visscher P.M., Haley C.S., Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach, Genetics 156 (2000) 2081–2092.

[8] Georges M., Nielsen D., Mackinnon M., Mishra A., Okimoto R., Pasquino A.T., Sargeant L.S., Sorensen A., Steele M.R., Zhao X., Womack J.E., Hoeschele I., Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing, Genetics 139 (1995) 907–920.

[9] Grignola F.E., Hoeshele I., Tier B., Mapping quantitative trait loci *via* residual maximum likelihood: I. Methodology, Genet. Sel. Evol. 28 (1996) 479–490.

[10] Grignola F.E., Hoeshele I., Zhang Q., Thaller G., Mapping quantitative trait loci *via* residual maximum likelihood: I. Methodology, Genet. Sel. Evol. 28 (1996) 491–504.

[11] Guldbrandtsen B.G., Lund M.S., Dunø M., Nielsen V.H., Jensen H., Svendsen S., Jensen J., Sorensen D.A., Bendixen C.B., QTL mapping for health, fertility and production traits in Danish Holstein cattle, in: Proc. 52nd Europ. Assoc. Anim. Prod., Amsterdam, NL, 2000, pp. G5.13.

[12] Hoeschele I., Uimari P., Grignola F.E., Zhang Q., Gage K.M., Advances in statistical methods to map quantitative trait loci in outbred populations, Genetics 147 (1997) 1445–1457.

[13] Jensen J., Mantysaari E., Madsen P., Thompson R., Residual maximum likelihood estimation of (co)variance components in multivariate mixed linear models using average information, J. Indian Soc. Agric. Stat. 49 (1997) 215–236.

[14] Jiang C., Zeng Z.B., Multiple trait analysis of genetic mapping for quantitative trait loci, Genetics 140 (1995) 1111–1127.

[15] Knott S.A., Haley C.S., Multitrait least squares for quantitative trait loci detection, Genetics 156 (2000) 899–911.

[16] Mangin B., Thoquet P., Grimsley N., Pleiotropic QTL analysis, Biometrics 54 (1998) 88–99.

[17] Meuwissen T.H., Goddard M.E., Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci, Genetics 155 (2000) 421–430.

[18] Patterson H.D., Thompson R., Recovery of inter-block information when block sizes are unequal, Biometrika 58 (1971) 545–554.

[19] Ronin Y.I., Kirzhner V.M., Korol A.B., Linkage between loci of quantitative traits and marker loci: multitrait analysis with a single marker, Theor. Appl. Genet. 90 (1995) 776–786.
[20] Searle S.R., Linear models for unbalanced data, 2nd edn., John Wiley & Sons, New York, 1987.
[21] Van Raden P.M., Wiggans G.R., Derivation, calculation and use of national animal model information, J. Dairy Sci. 74 (1991) 2737–2746.
[22] Wang T., Fernando R.L., van der Beek S., Grossman M., Covariance between relatives for a marked quantitative trait locus, Genet. Sel. Evol. 27 (1995) 251–274.
[23] Weller J.I., Wiggans G.R., Van Raden P.M., Ron M., Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multitrait experiment, Theor. Appl. Genet. 92 (1996) 998–1002.
[24] Xu S., Atchley WR., A random model approach to interval mapping of quantitative trait loci, Genetics 141 (1995) 1189–1197.
[25] Yi N., Xu S., Bayesian mapping of quantitative trait loci under the identity-by-descent-based variance component model, Genetics 156 (2000) 411–422.
[26] Zhang Q., Boichard D., Hoeschele I., Ernst C., Eggen A., Murkve B., Pfister-Genskow M., Witte L.A., Grignola F.E., Uimari P., Thaller G., Bishop MD., Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree, Genetics 149 (1998) 1959–1973.

To access this journal online:
www.edpsciences.org