Iranian Society of Animal Science

Journal of Livestock Science and Technologies

Shahid Bahonar University of Kerman

ISSN: 2322-3553 (Print) ISSN: 2322-374X (Online)

Genetic mapping of multiple pleiotropic quantitative trait loci in livestock exploiting a multiplicative mixed model

Ali Esmailizadeh * and Vahideh Rezaei

Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran.

Abstract A multiple marker analysis approach in the framework of the mixed-effects model was developed, allowing all markers of the entire genome to be included simultaneously in the analysis. The approach was extended to multi-trait situations. The proposed method is a one-stage process, which simultaneously models the residuals and genetic effects. In addition, it can easily accommodate co-variates, extra sources of variation, fixed or random including polygenic effects and it can easily be generalized to experimental and crossing designs commonly used. The developed approach considered an unstructured co-variance model for the traits residuals and fitted a multiplicative model for the trait by marker effects. The particular multiplicative model considered herein was the factor analytic model. This provided a parsimonious model specification to limit the number of parameters to be estimated. It was shown through the simulation study that modelling multiple phenotypes in a single linkage analysis simultaneously could markedly increase the power, compared with modelling of each phenotype separately. Correlations among phenotypes can arise from several different causal processes, which may have different implications for the power and performance of the multivariate linkage analysis. Obviously, further studies using the approach suggested herein for multitrait quantitative trait loci (QTL) mapping that specifically consider different situations, should be undertaken. Furthermore, the efficiency of the model to distinguish between a pleiotropic QTL and closely linked QTL affecting different traits is another area that needs more investigation.

Keywords: multiplicative mixed model, pleiotropy, quantitative trait loci

Paper type: Research Paper

***Corresponding author,**
E-mail address:
aliesmaili@uk.ac.ir

Received: **25 Sep. 2020,**
Accepted: **11 Nov. 2020,**
Published online: **01 Dec. 2020.**

Introduction

A gene may affect more than one trait and this is termed pleiotropic effect (Sohrabi et al., 2014). Pleiotropic effect of the loci is common in livestock species (Esmailizadeh and Mohammadabadi, 2010). For example, Esmailizadeh et al. (2008) found multiple effects of the myostatin F94I substitution on beef traits. It is common in gene mapping experiments to measure a large number of traits since genotyping costs are essentially fixed and additional phenotyping is a way of value-adding that investment (Moradian et al., 2015). Univariate analysis of relationships between genetic variants (e.g. micro-satellite allele, single nucleotide polymorphism, haplotype) and phenotypes are commonly conducted using regression of the phenotype on the marker (Moradian et al., 2014). However, multiple trait analysis is seldom per-

formed. Multivariate quantitative trait loci (QTL) mapping allows the detection of any possible pleiotropic effects and linked QTL, while exploiting the information from genetic and phenotypic correlations between traits (Korol et al., 1995; Knott and Haley, 2000; Gilbert and Le Roy, 2003). This provides potentially more insight into the nature of genetic correlations between different traits. In addition, multivariate approaches can increase the power of the test and the precision of parameter estimates (Jiang and Zeng, 1995; Korol et al., 1995; Gilbert and Le Roy, 2004; Meuwissen and Goddard, 2004). Jiang and Zeng (1995) showed that if the true model is a pleiotropic QTL, then analyzing the multiple affected traits simultaneously by fitting a pleiotropic QTL increases power and improves the resolution in mapping the QTL.

The simplest way to deal with multivariate data is by mapping individual traits and assessing whether the confidence intervals for QTL overlap for some combinations of traits. However, several approaches have been adopted to handle multivariate data collected in gene mapping experiments. In almost all the approaches, multivariate traits are often condensed to allow univariate analysis. One approach is to select one of the traits as the primary one and considering the remaining traits as co-variates, modifying the mean behavior of the primary trait. Alternatively, the multivariate trait is replaced by one or more linear combinations of the underlying univariate traits through traditional principal component analysis or factor analysis (Weller et al., 1996; Gilbert and Le Roy, 2003; Gilbert and Le Roy, 2004; Stearns et al., 2005). However, neither of these approaches is satisfactory. In the first approach, the traits are treated asymmetrically, with one trait arbitrarily designated as primary. For instance in mapping genes for carcass fatness, treating carcass weight as a co-variate runs the risk of masking linkage evidence for genes that impact both traits (Knott and Haley, 2000; Knott, 2005). In essence, information on the variance and co-variances displayed by traits is lost when they are viewed as co-variates.

Transforming the original traits into new linear combinations has been approached in several ways. For example, Weller et al. (1996) considered principal component analysis, while Gilbert and Le Roy (2003) considered discriminate analysis. Korol et al. (2001) used a transformation of the trait space followed by single-trait analysis and subsequent back transformation.

A possible disadvantage of using principal components in QTL analyses is that the magnitudes of the estimated effects are difficult to interpret directly in terms of the traits. A transformation that produces traits that are either phenotypically or genetically uncorrelated does not ensure that the QTL only influences a single canonical trait. This is because different QTL affecting a trait may have different patterns of pleiotropy, for example some QTL affect only one trait whereas others affect two or more traits (Knott and Haley, 2000). In this case, it is not possible to find a canonical transform that ensures all QTL only influence one canonical trait. Consequently, it cannot be assumed that QTL found to be affecting two different canonical variables in the same location are actually different QTL, as stated by Weller et al. (1996). One could only conclude that QTL affecting different canonical traits are indeed different if the genetic correlations between traits are the same as the phenotypic correlations and all individual QTL follow the same pattern, a situation that is likely to be rare.

A number of methods to analyze the traits simultaneously have been developed (Jiang and Zeng, 1995; Knott and Haley, 2000; Korol et al., 2001). However, currently, multiple trait approaches suffer in their implementation. These methods have not been widely adopted which is probably a reflection of their relative statistical complexity. In addition, it is not clear how to proceed with the analysis of data containing many traits (e.g., does one start with single trait analysis or with one multitrait analysis that assumes that there are QTL affecting all traits (Haley, 1999). Additionally, in practice, results are observed that seem intuitively incorrect. For example, Knott (2005) stated that single-trait analyses give evidence for all traits in one region of a linkage group, but when the traits are analyzed together, the best location can move some distance away to where there was no evidence for QTL from the individual trait analyses. The aim of this study was to develop a method that can combine genetic information across both correlated traits and half-sib families in livestock species to increase the power of gene mapping experiments.

Materials and methods

Whole genome marker analysis

Rather than testing an individual marker independent of all other markers, it is preferable to model all potential QTL at once. A number of methods are available that allow the inclusion of all of the markers of the entire genome in QTL analysis (Xu, 2003; Gilmour, 2007; Verbyla et al., 2007). Gilmour's (2007) method allows separate variance components for each chromosome. Markers

are fitted as independent random effects and in the absence of any QTL represent error contrasts with expected variance component of zero. A significant variance component indicates that there is additional variation, attributed to QTL, and the best linear unbiased predictions (BLUPs) for the marker effects are then interpreted assuming the correlation between markers is $e^{-2\delta}$, where δ is the distance between markers in Morgans. In this study, a modification in the proposed method by Gilmour (2007) was made to extend the method for multivariate situations. A single genome variance component was estimated rather than chromosomal variances. Based on the magnitude and prediction error of the markers BLUPs given the data, they were converted to a probability of being zero (Verbyla et al., 2003). These probabilities were converted to log scale (-2ln(P)) which is equivalent to a LOD score for each marker and is distributed as a χ^2_2 (Fisher, 1954). Since the method thus far is a marker selection method, the most significant marker is added to the fixed effects to estimate size and significance. The model is then refit to identify whether significant genome variance remains and if so, the process of identifying most probable markers and re-fitting as fixed effects continues. Markers are selected conservatively compared to individual marker regression.

Mixed model for multivariate whole genome markers

Suppose that *p* traits are measured in *n* individuals. The model for combined vector of data across traits is given by:

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\tau}_j + \mathbf{Z}_j\mathbf{u}_j + \mathbf{e}_j \quad (1)$$

where there are q_j fixed effects associated with trait *j* so that $\mathbf{X}_j$ and $\boldsymbol{\tau}_j$ have, respectively, dimensionality $n\times q_j$ and $q_j \times 1$ for each character. $\mathbf{u}_j^{(b_j \times 1)}$ is vector of random effects associated with trait *j*.

It is assumed that the joint distribution of ($\mathbf{u}_j$, $\mathbf{e}_j$) is Gaussian with zero mean and variance matrix

$$\theta_j \begin{pmatrix} \mathbf{G}_j(\gamma) & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_j(\phi) \end{pmatrix} \quad (2)$$

where, θ_j is a scale parameter which in the case of multiple trait or multi-trials is fixed to one. However, in mixed effects models with a single residual variance, θ_j is equal to the residual variance (σ^2). γ and ϕ are vectors of variance parameters. The distribution of the data vector $\mathbf{y}_j$ is thus,

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\tau}, \mathbf{H}) \quad (3)$$

where $\mathbf{H}= \mathbf{ZGZ}' + \mathbf{R}$.

$\mathbf{R}$ is $(np)\times(np)$ co-variance matrix associated with the total vector $\mathbf{e}^T = (\mathbf{e}_1^T,...,\mathbf{e}_p^T)$ of residual errors. $\mathbf{G}$ is co-variance matrix associated with the total vector $\mathbf{u}^T = (\mathbf{u}_1^T,...,\mathbf{u}_p^T)$ of random effects.

The co-variance matrix $\mathbf{G}$, for the *i*th random term, has many possible forms. In the most general case, $\mathbf{G}$ could be completely unstructured, comprising $b_i(b_i+1)/2$ parameters. As stated by Smith *et al.* (2001), interaction terms may be regarded as a vector representation of a t_i dimensional array of effects, where t_i is the number of factors in the interaction. The variance structure for the *j*th dimension is $\mathbf{G}_{ij}$. As with a single dimension random effect, $\mathbf{G}_{ij}$ may take a range of forms. In the model herein, the interaction term is *trait by marker*. Let $\mathbf{u}_m$ be the $mp \times 1$ vector of the effects of *m* markers for *p* traits. A general form for the variance structure of the interaction term is

$$\mathrm{var}(\mathbf{u}_\mathbf{g}) = \mathbf{G}_\mathbf{p} \otimes \mathbf{G}_\mathbf{m} \quad (4)$$

where, $\mathbf{G}_\mathbf{p}$ and $\mathbf{G}_\mathbf{m}$ are positive definite symmetric matrices of dimension $p\times p$ and $m\times m$, respectively. The matrix $\mathbf{G}_\mathbf{p}$ = (σ_{Mjj}) is the marker variance matrix. The diagonal elements are the marker variances for traits and the off-diagonal elements are the marker co-variances between pairs of traits. $\mathbf{G}_\mathbf{m} = \mathbf{I}_\mathbf{m}$ and in the Kronecker product notation, variance of trait by marker interaction term, var($\mathbf{u}_g$), is

$$\mathrm{var}(\mathbf{u}_g) = \mathbf{G}_\mathbf{p} \otimes \mathbf{I}_\mathbf{m} \quad (5)$$

The model (5) implies that *trait* × *marker* effects are correlated between traits. Separating the marked genetic effects from other random terms, including non-marked polygenic effects, the mixed model (1) can then be written as

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\tau}_j + \mathbf{Z}_{0j}\mathbf{u}_{0j} + \mathbf{Z}_{gj}\mathbf{u}_{gj} + \mathbf{e}_j \quad (6)$$

where $\mathbf{u}_{gj}$ are the marker effects for trait *j* with associated design matrix $\mathbf{Z}_{gj}^{(n\times mp)}$ and variance matrix as in (5). $\mathbf{u}_{0j}$ comprise any additional random effects (including non-marked polygenic effects) with associated design matrix $\mathbf{Z}_{0j}$ and variance matrix $\mathbf{G}_0$.

A simple structure for $\mathbf{G}_p$ is a diagonal model (DIAG), assuming the marker effects for different traits are reg-

arded as independent so that $\mathbf{G}_p$=diag(σ_{Mjj}), *j=1...p*. The most general form for $\mathbf{G}_p$ is the unstructured variance model which contains *p(p+1)/2* parameters (i.e. the number of parameters to be estimated increases quadratically with the number of traits). This model will provide the best fit (in a likelihood sense) to the data. However, in cases with a large number of traits and markers, it is difficult to ensure that REML estimates of the variance parameters for such a complex variance model remain within the parameter space. Also estimation of such a structure may be inefficient for a large number of traits and markers so a more parsimonious structure is desirable. This can be achieved using a factor analytic model for the marker effects across traits. Even for small number of traits, a factor analytic structure is preferred for $\mathbf{G}_p$ as given the purpose that is finding pleiotropic QTL and also trait-specific QTL.

Factor analytic model

Factor analytic variance structures have been proposed for genotype by environment effects in mixed model analyses of data from multi-environment trials (Cullis et al., 1998; Smith et al., 2001; Thompson et al., 2003; Smith et al., 2005). Smith et al. (2001; 2005) use a factor analytic structure to model variety by environment interactions, whilst simultaneously estimating a separate spatial correlation structure for the errors for each trial. Herein, the same formulation as Smith et al. (2001; 2005) was used to explain the application of factor analysis in modelling trait by marker effects.

When applied to the marker effects for each trait, the factor analytic model for marker effects, $\mathbf{u}_g$ will be

$$\mathbf{u}_g = (\boldsymbol{\lambda}_1 \otimes \mathbf{I}_m)\xi_1 + ... + (\boldsymbol{\lambda}_k \otimes \mathbf{I}_m)\xi_k + \boldsymbol{\delta} \quad (7)$$

where $\xi_r^{(m\times 1)}$ are a few, random quantities called factors (*r=1...k<p*), the coefficients $\boldsymbol{\lambda}_r^{(p\times 1)}$ are known as loadings, and $\boldsymbol{\delta}^{(mp\times 1)}$ is the vector of residuals or lack of fit for the model.

Equation (7) has the form of a random regression on *k* trait co-variates $\boldsymbol{\lambda}_1,...,\boldsymbol{\lambda}_k$. However, the difference between this equation and standard random regression problems is that in this formulation both the co-variates and the regression coefficients are unknown and therefore, must be estimated from the data (Smith et al., 2001) .

Matrix notation allows the entire system of equations to be written quite compactly as

$$\mathbf{u}_g = (\boldsymbol{\Lambda} \otimes \mathbf{I}_m)\xi + \boldsymbol{\delta} \quad (8)$$

Considering the full model (6) the distribution of ($\boldsymbol{\xi}$, $\boldsymbol{\delta}$, $\mathbf{e}$) is assumed to be multivariate normal, with mean the zero vector and variance matrix

$$\begin{pmatrix} \mathbf{I}_K \otimes \mathbf{I}_m & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \otimes \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{pmatrix} \quad (9)$$

where $\boldsymbol{\Psi}$ is a diagonal matrix with elements $(\psi_1, \psi_2, ..., \psi_p)$ and ψ_i is known as the specific variance for the *i*th trait. The variance matrix for the marker effects for each trait, $\mathbf{var}(\mathbf{u}_g)$, is then given by

$$\mathrm{var}(\mathbf{u}_g) = (\boldsymbol{\Lambda} \otimes \mathbf{I}_m)\,\mathrm{var}(\boldsymbol{\xi})\,(\boldsymbol{\Lambda}' \otimes \mathbf{I}_m) + \mathrm{var}(\boldsymbol{\delta}) = (\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}) \otimes \mathbf{I}_m \quad (10)$$

Researchers may be interested only in $\boldsymbol{\xi}$ (e.g., in modelling variety by environment interaction in plants). However, herein both $\boldsymbol{\xi}$ and $\boldsymbol{\delta}$ are of interest.

Estimation of the parameters

Estimates of the fixed and random effects in equation (6) are obtained as solutions to the mixed-model equations (Smith et al., 2001), which are given by

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_g \\ \mathbf{Z}_g'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_g'\mathbf{R}^{-1}\mathbf{Z}_g + \mathbf{G}_p^{-1} \otimes \mathbf{I}_m \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\tau}} \\ \tilde{\mathbf{u}}_\mathbf{g} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_g'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \quad (11)$$

This leads to best linear unbiased estimates (BLUEs) of the fixed effects,

$$\hat{\boldsymbol{\tau}} = (\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}^{-1}\mathbf{y} \quad (12)$$

and best linear unbiased predictors (BLUPs) of the random effects ,

$$\tilde{\mathbf{u}}_\mathbf{g} = (\mathbf{G}_p \otimes \mathbf{I}_\mathbf{m})\mathbf{Z}_g'\mathbf{P}\mathbf{y} \quad (13)$$

where $\mathbf{H} = \mathrm{var}(y) = \mathbf{Z}_p(\mathbf{G}_p \otimes \mathbf{I}_m)\mathbf{Z}_p' + \mathbf{R}$ (14)

and $\mathbf{P} = \mathbf{H}^{-1} - \mathbf{H}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}^{-1}$ (15)

In practice, BLUEs and BLUPs and the variance components are obtained through an iterative scheme. However, extra calculations in a factor analytic model are parameters in $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$. The parameters in $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ are usually unknown and require to be estimated from the experimental data. The number of parameters in the factor analytic model with *k* terms is given by

*pk+p-k (k-1)/*2. Estimation in factor analysis is a two-stage procedure. First, the parameters in the model are estimated, and then these are used to provide estimates of individual factor scores.

The use of model (10) for marker effects can lead to models with variance structures of less than full rank, which may occur when estimates of one or more specific variances tend to zero. In the literature on factor analysis, this is known as the Heywood case (Lawley and Maxwell, 1971; Johnson and Wichern, 1998). In this situation, REML estimation using the average information algorithm (Gilmour et al., 1995) or other standard algorithms is no longer possible.

Thompson et al. (2003) presented a sparse implementation of the average information algorithm for REML estimation of the factor analytic variance parameters. The algorithm is computationally efficient as exploits the regression underpinning the factor analytic model thereby facilitating substantial time savings. Additionally, the (commonly occurring) case of factor analytic variance structures with less than full rank (reduced rank variance models) has been accommodated in the algorithm, which is useful in the multivariate analysis. The algorithm has been implemented in ASReml (Glimour et al., 2006) and can be accessed via the "XFA" variance model.

Testing for pleiotropic QTL

The first step is to fit model (6) where markers of the entire genome are fitted simultaneously as random regression genetic effects, considering two co-variance models for $\mathbf{G_p}$ (The diagonal co-variance model (DIAG) and the factor analytic model with one factor (FA1)). The DIAG model implies no marker co-variance between traits (that is, the traits are independent with heterogeneous variances). Since DIAG and FA1 are nested models, the REMLRT statistic, $2\Delta\ell$ (twice the log likelihood difference), can be approximated by the χ^2 distribution with the degree of freedom equal to the difference in the number of free parameters in the two nested models (Stuart et al., 1999).

Rejection of the null hypothesis would provide supporting evidence for the existence of either QTL that cause pleiotropic effects or multiple linked QTL. Under the null hypothesis of no pleiotropic QTL, since the markers are neutral, there should be no co-variance associated with markers. The alternative is that one or more QTL affecting two or more than two traits occur on the genome. Then, all of the marker co-variables in general, and in particular, those closest to the QTL, will take up some of the co-variation caused by the QTL, thereby inflating the co-variance component of the random regression term. Therefore, if a likelihood ratio test is significant for a FA1 model for the trait by random regression marker effects, there is evidence for at least one pleiotropic QTL in the genome, this will be fitted as a fixed co-variate and the process continues. If there is no significant FA1 model for this term, then there is no evidence of pleiotropic QTL, and the process is terminated.

Locating the pleiotropic QTL

Once the significance of co-variance for the marker random regression is established by the factor analytic model, the next stage is to detect the most likely marker linked to the pleiotropic QTL. In order to locate pleiotropic QTL, individual marker effects for individual traits and the factor are converted to LOD scores. The marker with highest LOD value for the factor is considered and based on the map information, a QTL co-variate is calculated and added to the fixed effect part of the model (6) nested within traits. If there is only one pleiotropic QTL and its location is identified, the FA1 model for the marker co-variance will become non-significant in the presence of the QTL co-variate, confirming the location. However, the QTL co-variate may not remove all the marker co-variance leading to the need for further investigation. If there is another pleiotropic QTL, then the QTL co-variate may have explained a substantial amount of the marker co-variance, but the remaining marker co-variance will indicate the location of the second pleiotropic QTL. A co-variate is added for the second pleiotropic QTL and the process is repeated until the random marker co-variance becomes effectively zero (that is, non-significant FA1 model compared to the DIAG model). In this stage, only the QTL affecting individual traits (that is, trait specific QTL) remain. In order to locate the trait specific QTL, the analysis is continued using the univariate whole genome marker analysis framework explained above, fitting the detected pleiotropic QTL as fixed co-variates and testing the marker variance for that specific trait.

To investigate the behavior of the approach, extensive simulation studies were conducted. Four normally distributed quantitative traits were considered, each with a residual standard deviation of unity, with individuals being assigned a random value from this distribution. The simulation design was based on: sample sizes of 125, 250, 500 and 750, a total chromosome number

of 8 with 6 markers for each chromosome, and an average marker distance of 20 cM and 1000 replications. Inheritance of all loci was determined assuming random assortment and that recombination events occurred independently, allowing use of Haldane's mapping function (Haldane,1919). A total of 10 QTL were set (Table 1) for the whole genome. Among these QTL, there are two QTL with pleiotropic effects. Three sets of simulations for each population were generated, QTL with small effects, QTL with medium effects and QTL with large effects, in which each QTL (on average) explains 7, 10 and 13 %, respectively, of the phenotypic variance in the backcross. Two QTL were set in repulsion phase on chromosome 6 affected trait 2. Two QTL on chromosome 5 in coupling phase affected trait 1. Two QTL on chromosome 8 in coupling phase affected trait 4. One QTL in the centromeric position (first marker) of chromosome 2 and another QTL in the telomeric end of chromosome 7 (last marker) both affected trait 2.

Results

Factor loadings and specific variances

In terms of the trait by markers effects, a DIAG model and a factor analytic model with k=1 factor (denoted FA1) was fitted sequentially. In the simulation study, the DIAG model had 4 parameters and the FA1 model had 8 variance parameters (4 loadings and 4 specific variances for four traits). As the QTL size of effect increased the marker variance estimated from DIAG model for all four traits increased (Table 2). In all cases, the marker variances for traits 1 and 2 were higher than those of traits 3 and 4, a result which one would expect as the heritability for these later traits was lower than that of the former. Since two pleiotropic QTL were simulated (QTL1 and QTL3) for traits 1 and 3 and only one of them was considered to affect one of the traits 2 or 4 (QTL1 for trait 4 and QTL3 for trait 2), the factor loadings on trait 1 and 3 were higher than those for traits 2 and 4.

Three specific QTL (QTL2, 6 and 7) were simulated for trait 2 and only one specific QTL was simulated for trait 3 (QTL8). This fact was reflected in their specific variances so that, in general, traits 2 and 3 had highest and lowest, respectively, specific variances among four traits (Table 2). It should be noted that in a few of the replicates, the estimated marker variance using DIAG model for small populations and small QTL size was on the boundary for one trait; that is, it was estimated as zero. Also in some replicates, using the FA1 model led to zero estimation of specific variances for one or two traits when the sample size and QTL effect were small.

Power of pleiotropy test

Since the DIAG model is nested within the FA1 model, a direct comparison can be made using a REML likelihood ratio test. Herein, the power of the pleiotropic test is the chance of detecting a common factor (pleiotropic QTL) if that factor really exists. Thus, the power of the pleiotropic test was defined as the number of analyses (out of 1000 replicates) resulting in a significant FA1 model compared with the DIAG model. The power of the test depended on sample size and QTL size of effect (Figure 1). For a given QTL effect, as the population size increased, the power of the test increased and there was low power when both QTL effect and sample size were small. A population size of 500 seems to be a critical limit, in which for the small QTL considered in the simulation study, the test can reach up to 80% power.

Power of QTL detection and false positives

The power of the experiment was calculated as the average number of QTL detected divided by the number of QTL present. The results showed that the ability to detect QTL using both univariate and multivariate analyses was strongly influenced by the QTL size of effect, and sample size so that the power to detect QTL improved significantly with increasing sample size and QTL effect (Figure 2 and Tables 3-5).

The overall power of detecting a QTL using the FA1 model was generally higher than that obtained in the

Table 1. Simulated pleiotropic and trait specific QTL

	QTL1	QTL2	QTL3	QTL4	QTL5	QTL6	QTL7	QTL8	QTL9	QTL10
	Chr1	Chr2	Chr4	Chr5	Chr5	Chr6	Chr6	Chr7	Chr8	Chr8
Position (cM)	20	0	60	20	80	20	60	100	40	80
Trait 1	PLTC		PLTC	COUP#	COUP#					
Trait 2		#	PLTC			REPL#	REPL#			
Trait 3	PLTC		PLTC					#		
Trait 4	PLTC								COUP#	COUP#

PLTC: Pleiotropic QTL. #: Trait specific QTL. COUP: Two linked QTL in coupling phase, REPL: Two linked QTL in repulsion phase

Table 2. Mean loadings (×10), and marker variances (×100) estimated for four traits and averaged over 1000 replicates

QTL Effect	Sample Size	Trait	FA1 model		DIAG model
			Loadings	Specific variance	Marker variance
Small[a]	125	Trait1	0.61	0.35	0.78
		Trait2	0.37	0.67	0.92
		Trait3	0.59	0.18	0.59
		Trait4	0.23	0.45	0.55
	250	Trait1	0.59	0.35	0.73
		Trait2	0.35	0.80	0.97
		Trait3	0.61	0.18	0.58
		Trait4	0.25	0.43	0.51
	500	Trait1	0.59	0.34	0.71
		Trait2	0.34	0.89	1.03
		Trait3	0.61	0.18	0.57
		Trait4	0.24	0.40	0.47
	750	Trait1	0.60	0.34	0.71
		Trait2	0.34	0.91	1.04
		Trait3	0.61	0.18	0.56
		Trait4	0.24	0.39	0.45
Medium[b]	125	Trait1	0.79	0.57	1.24
		Trait2	0.46	1.27	1.60
		Trait3	0.75	0.29	0.92
		Trait4	0.29	0.70	0.82
	250	Trait1	0.76	0.59	1.18
		Trait2	0.45	1.43	1.68
		Trait3	0.78	0.27	0.91
		Trait4	0.30	0.65	0.76
	500	Trait1	0.76	0.57	1.17
		Trait2	0.44	1.53	1.74
		Trait3	0.77	0.28	0.89
		Trait4	0.31	0.61	0.71
	750	Trait1	0.77	0.56	1.18
		Trait2	0.44	1.56	1.76
		Trait3	0.78	0.28	0.89
		Trait4	0.31	0.59	0.69
Large[c]	125	Trait1	0.99	0.91	1.95
		Trait2	0.58	2.24	2.70
		Trait3	0.92	0.44	1.36
		Trait4	0.35	1.02	1.17
	250	Trait1	0.98	0.92	1.89
		Trait2	0.57	2.43	2.80
		Trait3	0.95	0.42	1.35
		Trait4	0.37	0.94	1.09
	500	Trait1	0.97	0.93	1.91
		Trait2	0.56	2.53	2.86
		Trait3	0.95	0.41	1.33
		Trait4	0.38	0.89	1.03
	750	Trait1	0.98	0.95	1.93
		Trait2	0.55	2.56	2.88
		Trait3	0.96	0.41	1.32
		Trait4	0.38	0.86	1.01

[a] Each QTL accounted for 7% of phenotypic variation, [b] Each QTL accounted for 10% of phenotypic variation,[c] Each QTL accounted for 13% of phenotypic variation

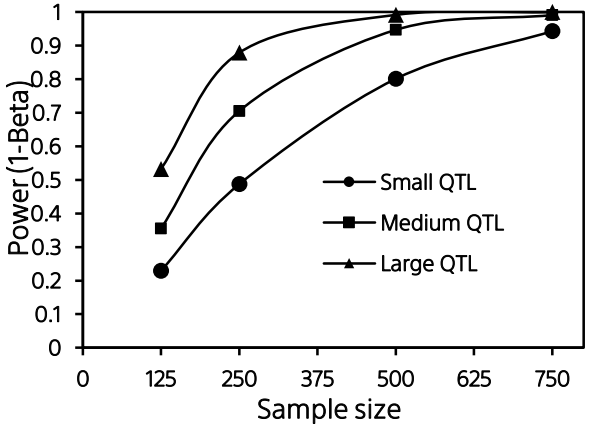

Figure 1. Observed statistical power (proportion of replicates with significant (P<0.05) FA1 model compared with DIAG model) for the pleiotropy model test.

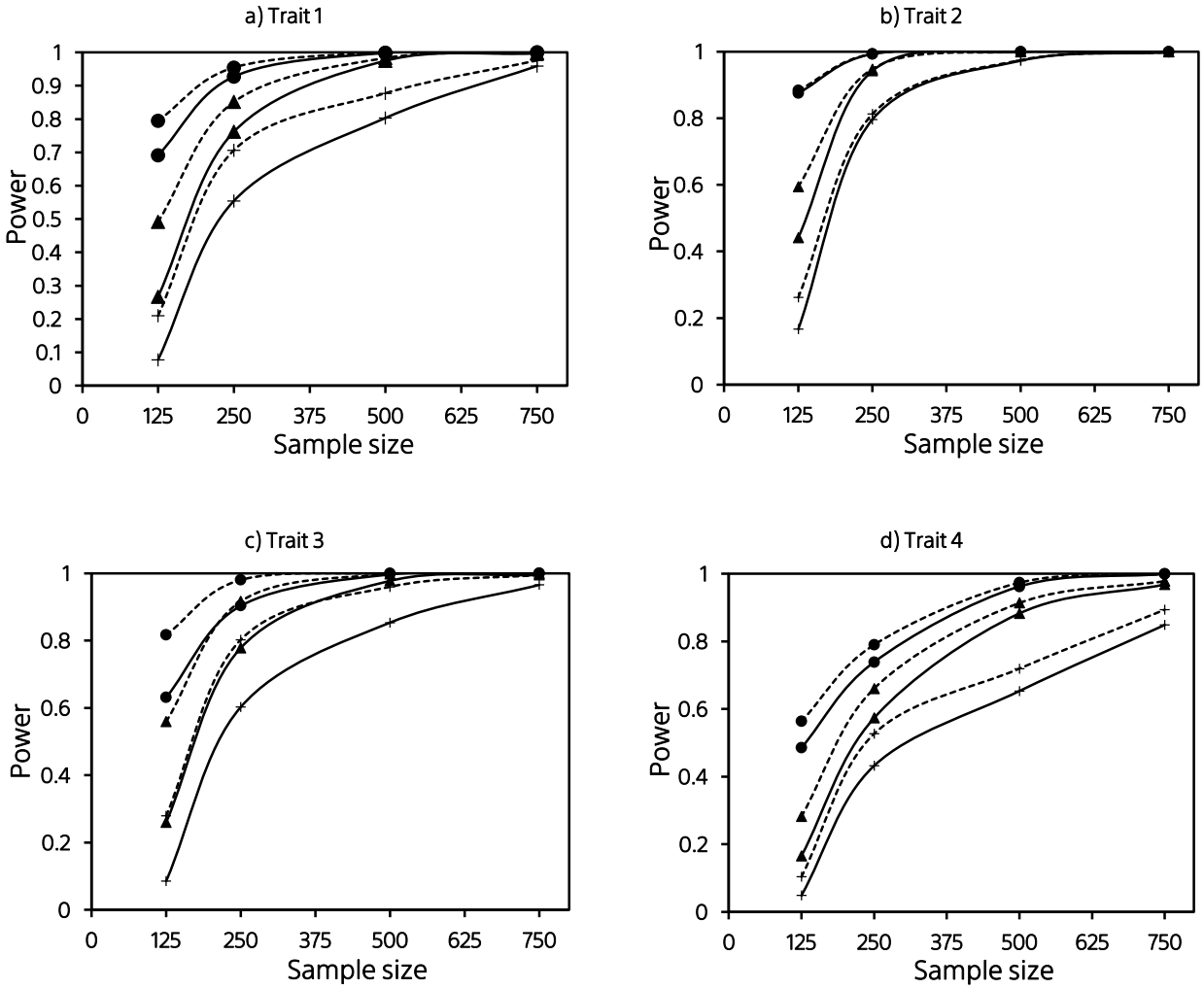

Figure 2. Comparison of the power of univariate and multivariate (FA1) for QTL by the number of QTL present. Legend: Small QTL +, Medium QTL ▲, Large QTL ●; Univariate analysis solid lines and multivariate dashed lines.

Table 3. The power of QTL detection (proportion of significant replicates over all 1000 replicates) and the probability of false QTL detected under univariate and multivariate analysis (Trait 1)

QTL Effect	Sample Size	Model	QTL												
									Pleiotropic [a]		Linked [b]		False QTL [c]		
			1	3	4	5	One	All	One	Two	One	Two	Lk	Uk	Total
Small	125	UNI [d]	5.6	6.4	9.7	9.2	24.6	0.1	11.1	0.9	16.9	2.0	2.1	0.8	2.9
		FA1 [e]	28.4	32.6	11.6	11.4	58.2	0.7	48.0	13.0	20.7	2.3	4.2	3.0	7.2
	250	UNI	53.4	50.5	60.2	57.7	95.2	11.0	76.4	27.5	83.6	34.3	8.8	2.2	11.0
		FA1	83.2	85.3	58.2	55.8	99.3	25.3	96.6	71.9	81.0	33.0	10.7	6.0	16.7
	500	UNI	78.7	79.5	82.0	81.2	99.8	42.2	95.2	63.0	97.2	66.0	2.2	0.1	2.3
		FA1	95.4	96.3	78.9	80.3	57.8	100	99.8	91.9	95.5	63.7	1.7	1.0	2.7
	750	UNI	95.5	95.8	96.5	96.0	100	84.8	100	91.3	99.8	92.7	0.6	0.1	0.7
		FA1	99.5	99.7	94.8	96.2	100	90.3	100	99.2	99.9	91.1	0.8	0.2	1.0
Medium	125	UNI	24.6	24.0	29.0	29.3	60.3	2.7	38.2	10.4	47.3	11.0	4.8	1.5	6.3
		FA1	59.3	65.4	35.1	37.0	92.4	7.4	82.8	41.9	57.2	14.9	6.5	5.2	11.7
	250	UNI	75.3	73.3	79.1	77.1	99.5	35.7	92.8	55.8	95.7	60.5	4.1	0.8	4.9
		FA1	94.3	93.9	77.0	75.4	100	52.2	99.7	88.5	57.9	94.5	4.4	2.0	6.4
	500	UNI	96.8	97.4	98.2	97.6	100	90.2	100	94.2	100	95.8	1.2	0.1	1.3
		FA1	99.6	99.7	96.6	97.3	100	93.3	100	99.3	99.9	94.0	1.0	0.4	1.4
	750	UNI	99.7	99.8	99.8	99.6	100	98.9	100	99.5	100	99.4	0.0	0.0	0.0
		FA1	100	100	99.9	99.4	100	99.3	100	99.3	100	99.3	0.1	0.0	0.1
Large	125	UNI	66.1	65.4	72.3	72.7	98.3	26.4	85.0	45.9	92.8	52.2	7.3	2.0	9.5
		FA1	88.6	89.8	69.4	70.2	100	41.1	97.7	80.7	90.3	49.3	8.4	5.7	14.1
	250	UNI	92.6	91.7	93.6	93.1	100	74.2	99.5	84.8	99.5	87.2	1.7	0.4	2.1
		FA1	99.1	98.8	92.1	91.9	100	82.4	99.9	98.0	99.6	84.4	1.6	1.0	2.6
	500	Diag	99.8	99.7	99.9	99.9	100	99.3	100	99.5	100	99.8	0.2	0.0	0.2
		FA1	100	100	99.9	99.5	100	99.6	100	100	100	99.7	0.4	0.2	0.6
	750	UNI	100	100	100	100	100	100	100	100	100	100	0.1	0.0	0.1
		FA1	100	100	100	100	100	100	100	100	100	100	0.3	0.1	0.4

[a] One: percentage of runs in which at least one of the four simulated QTL was identified, All: percentage of runs in which all four simulated QTL were identified, One pleiotropic: percentage of replicates in which at least one of the two simulated pleiotropic QTL was identified, Two pleiotropic: percentage of replicates in which both of the two simulated pleiotropic QTL were identified. [b] One linked: percentage of replicates in which at least one of the two simulated linked QTL was identified, Two linked: percentage of replicates in which both of the two simulated linked QTL were identified. [c] Lk: proportion of falsely chosen markers linked to QTL, Uk: proportion of falsely chosen markers unlinked to QTL, Total false QTL: Lk plus Uk. [d]: univariate. [e]: Factor analytic model.

univariate analysis. The power of QTL detection using the FA1 was 100% or was almost 100% when the relative QTL effect was large or a relatively large sample size was considered. The increasing power using the FA1 model was more evident when two pleiotropic QTL were affecting the trait (Traits 1 and 3, Figure 2). For the large sample size and large QTL effect, the two methods had relatively similar power to identify QTL.

The main feature to be noticed (Tables 3-6) is the higher ability of the FA1 model compared to univariate analysis to detect QTL1 and QTL3, which were simulated to have common effect on the traits. Multivariate and univariate analyses were equally efficient in detecting trait specific QTL with large effects. However, trait specific QTL with small effect could only be detected with very low efficiency using both multivariate and univariate analyses (Tables 3-6). In the case of the probability for false QTL detection, in general, both methods gave small likelihoods of finding false QTL. The highest likelihood of detecting false QTL is for small sample size (Tables 3-6).

In the situation where the two linked QTL were in coupling phase, for a few of the replicates, both methods tended to choose the marker between two correct markers, particularly when the sample size and QTL effect was small (Table 6). However, in the case of the two linked QTL in repulsion phase, declaring the middle marker as the correct marker rarely happened (Table 4).

Both univariate and multivariate techniques chose a rather low portion of unlinked loci to a QTL. This effect was not evident with large sample sizes. Both approaches seem quite conservative, delivering only about 0-4.7% (univariate) and 0-7.1 % (FA model) of false positive unlinked loci.

Discussion

The most serious problem in multiple-QTL analysis

Table 4. The power of QTL detection (proportion of significant replicates over all 1000 replicates) and the probability of false QTL detected under univariate and multivariate analysis (Trait 2)

QTL Effect	Sample Size	Model	QTL											
									Repulsion[a]			False QTL [b]		
			2	3	6	7	One	All	One	Two	Mdl	Lk	Uk	Total
Small	125	UNI [c]	27.8	16.9	11.3	10.8	41.3	0.8	18.3	3.8	0.0	1.8	0.8	2.6
		FA1 [d]	34.6	41.2	14.3	14.8	64.7	1.4	24.3	4.8	0.0	3.0	3.7	6.7
	250	UNI	93.8	80.6	70.8	73.2	99.5	43.6	89.0	55.0	0.5	8.1	4.7	12.8
		FA1	93.2	90.3	69.9	71.7	99.9	46.6	87.5	54.1	0.6	8.2	7.1	15.3
	500	UNI	99.7	96.5	96.2	96.6	100	89.4	93.1	99.7	0.0	0.8	0.7	1.5
		FA1	99.5	98.6	95.8	95.9	100	90.3	99.7	92.1	0.0	0.9	1.1	2.0
	750	UNI	100	99.6	99.8	99.8	100	99.2	100	99.6	0.1	0.3	0.1	0.4
		FA1	100	99.9	99.8	99.7	100	99.4	100	99.5	0.0	0.4	0.3	0.7
Medium	125	UNI	60.6	45.0	35.9	34.9	71.9	12.9	50.3	20.5	0.0	3.4	1.0	4.4
		FA1	73.8	74.6	45.4	44.1	95.3	17.7	62.4	27.1	0.0	3.5	2.6	6.1
	250	UNI	99.1	94.5	91.6	91.7	100	78.7	98.8	84.5	0.3	3.1	2.3	5.4
		FA1	99.0	97.3	91.2	91.6	100	80.8	98.7	84.1	0.2	2.8	3.8	6.6
	500	UNI	100	99.9	99.9	100	100	99.8	100	99.9	0.0	0.4	0.3	0.7
		FA1	100	100	99.9	99.9	100	99.8	100	99.8	0.0	0.3	0.5	0.8
	750	UNI	100	100	100	100	100	100	100	100	0.0	0.1	0.0	0.1
		FA1	100	100	100	100	100	100	100	100	0.0	0.1	0.0	0.1
Large	125	UNI	96.5	88.5	83.4	81.6	99.9	62.0	94.3	70.7	0.3	6.3	3.0	9.3
		FA1	95.6	94.0	82.7	81.3	99.9	63.0	94.6	69.4	0.2	6.8	3.9	10.7
	250	UNI	100	99.5	98.8	98.9	100	97.2	100	97.7	0.2	1.2	1.4	2.6
		FA1	100	99.7	98.7	98.8	100	97.3	100	97.5	0.3	1.1	2.0	3.1
	500	UNI	100	100	100	100	100	100	100	100	0.0	0.1	0.0	0.1
		FA1	100	100	100	100	100	100	100	100	0.0	0.1	0.1	0.2
	750	UNI	100	100	100	100	100	100	100	100	0.0	0.1	0.0	0.1
		FA1	100	100	100	100	100	100	100	100	0.0	0.1	0.0	0.1

[a] One: percentage or runs in which at least one of the four simulated QTL was identified, All: percentage of runs in which all four simulated QTL were identified, One repulsion: percentage of replicates in which at least one of the two simulated linked QTL in repulsion phase was identified, Two repulsion: percentage of replicates in which both of the two simulated linked QTL in repulsion phase were identified, Mdl: percentage of replicates in which the marker between two linked QTL was chosen. [b] Lk: proportion of falsely chosen markers linked to QTL, Uk: proportion of falsely chosen markers unlinked to QTL, Total false QTL: Lk plus Uk. [c]: Univariate. [d] Factor analytic model.

comes from the model selection, which has been the focus of many QTL-mapping studies (Kao et al., 1999; Ball, 2001; Piepho and Gauch, 2001; Sen and Churchill, 2001; Broman and Speed, 2002; Xu, 2003; Glimour, 2007; Verbyla et al., 2007). The whole genome marker analysis approach, developed and used herein, is clearly a model selection strategy fitting relatively few models when compared with other methods. The method focuses first on the null hypothesis of no QTL in the genome (the variance component for the distribution of size of QTL is zero) and only after rejecting the null hypothesis, it is concluded that there is an evidence for QTL and then the QTL are located. The REML likelihood ratio test statistic (REMLRT) allows a genome-wide assessment of significance of the QTL.

Genome-wide searches for loci influencing quantitative traits are often plagued by low power and interpretive difficulties. Attempts to remedy these difficulties have typically relied on, and have promoted the use of, larger sample sizes, a greater density of molecular markers, and more-sophisticated statistical modelling. Many of these remedies can be costly to implement. In addition, as pointed out by Broman and Speed (2002), more sophisticated methods may not necessary lead to improved estimates. There have been numerous publications that address the power issue in QTL mapping. For example, it has been reported that multivariate approaches can increase the power of the test and the precision of parameter estimates (Jiang and Zeng, 1995; Korol et al., 1995; Gilbert and Le Roy, 2004; Meuwissen and Goddard, 2004). Meta-analysis of results from different studies (Allison and Heo, 1998; Wood et al., 2006) or joint analysis of the original data (Walling et al., 2000) are other strategies that have been proposed to improve QTL mapping resolution. By exploiting a multiplicative mixed model approach, the present study has ad-

Table 5. The power of QTL detection (proportion of significant replicates over all 1000 replicates) and the probability of false QTL detected under univariate and multivariate analysis (Trait 3)

QTL Effect	Sample Size	Model	QTL									
								Pleiotropic [a]		False QTL [b]		
			1	3	8	One	All	One	Two	Lk	Uk	Total
Small	125	UNI [c]	5.9	6.4	13.3	21.8	0.1	11.3	1.0	1.2	0.6	1.8
		FA1 [d]	32.6	36.6	14.5	61.1	2.7	55.2	14.0	2.9	3.1	6.0
	250	UNI	53.2	54.2	73.3	91.2	23.7	76.8	30.6	4.7	2.0	6.7
		FA1	84.9	88.0	67.8	99.0	51.8	97.0	75.9	6.1	4.0	10.1
	500	UNI	80.7	79.9	95.3	99.6	63.1	94.8	65.8	0.9	0.1	1.0
		FA1	98.1	97.6	92.1	100	88.7	95.7	92.1	1.1	1.0	2.1
	750	UNI	96.0	94.2	99.3	100	90.0	99.7	90.5	0.4	0.0	0.4
		FA1	99.9	99.4	99.1	100	98.5	100	99.3	0.2	0.8	1.0
Medium	125	UNI	22.8	18.9	36.4	51.6	4.4	7.3	34.4	2.7	1.1	3.8
		FA1	61.7	65.6	40.3	90.6	16.9	85.4	41.9	3.0	2.2	5.2
	250	UNI	71.1	72.3	90.1	98.1	49.0	90.4	53.0	2.7	1.2	3.9
		FA1	94.1	96.3	84.5	100	76.5	99.7	90.7	2.3	2.6	4.9
	500	UNI	97.1	96.4	99.7	100	93.5	99.7	93.8	0.5	0.1	0.6
		FA1	100	99.9	99.3	100	99.2	100	99.9	0.3	0.5	0.8
	750	UNI	99.8	98.7	100	100	98.5	100	98.5	0.0	0.0	0.0
		FA1	100	100	100	100	100	100	100	0.0	0.4	0.4
Large	125	UNI	56.5	55.1	77.7	91.2	29.4	77.0	34.6	5.3	2.4	7.7
		FA1	86.7	88.6	69.6	99.0	54.6	97.7	77.6	6.0	2.0	8.0
	250	UNI	84.9	87.7	98.4	99.9	73.9	98.0	74.6	1.4	0.3	1.7
		FA1	98.9	98.9	96.2	100	94.2	100	97.8	1.1	0.9	2.0
	500	Diag	99.5	99.6	99.9	100	99.0	100	99.1	0.2	0.0	0.2
		FA1	100	100	99.9	100	99.9	100	100	0.1	0.1	0.2
	750	UNI	100	100	100	100	100	100	100	0.0	0.0	0.0
		FA1	100	100	100	100	100	100	100	0.0	0.1	0.1

[a] One: percentage or runs in which at least one of the three simulated QTL was identified, All: percentage of runs in which all three simulated QTL were identified, One pleiotropic: percentage of replicates in which at least one of the two simulated pleiotropic QTL was identified, Two pleiotropic: percentage of replicates in which both of the two simulated pleiotropic QTL were identified. [b] Lk: proportion of false chosen markers linked to QTL, Uk: proportion of false chosen markers unlinked to QTL, Total false QTL: Lk plus Uk. [c]: Univariate. [d]: Factor analytic model.

dressed these two views to improve the power of QTL identification.

When modelling multiple trait or multiple trials, it is important to avoid over-parameterization, especially in small experimental populations used for QTL detection (Sillanpaa and Corander, 2002). As Piepho (2000) suggested, to avoid over- parameterization, a certain variance-co-variance structure was imposed. The specific multiplicative model considered herein was the factor analytic model. This provided a parsimonious model specification to limit the number of parameters to be estimated. The proposed approach considered an unstructured co-variance model for the residuals for traits and fitted a FA1 and a DIAG model sequentially for the interaction terms (trait by marker term, family by marker interaction or trait by family by marker interaction). Considering the correlations among QTL effects at a single gene are either +1 or –1 as suggested by Goddard (2001), a factor analytic model is more appropriate structure for a pleiotropic QTL. The aim of fitting FA1 structure for trait by marker effects was to account for the genetic marker co-variances among *p* traits in terms of an unknown factor. Because the model was fitted within a mixed-model framework, the importance of the co-variance due to the markers could be formally tested using a comparison of a model assuming no marker correlation (the DIAG model) and a model assuming marker correlation (the FA1 model). The DIAG model was nested within the FA1 model. Therefore, residual maximum likelihood ratio tests could be used to compare these models. This provided a formal test for common QTL (across families) or pleiotropic QTL effects. An extensive simulation study was undertaken to investigate the power of the pleiotropy test. The results indicated that the test was robust when the QTL size or sample size were high. However, there was a relatively low power to detect a QTL with small pleiotropic effects or in small populations.

Multiplicative models have been popularized (Cullis et al., 1998; Smith et al., 2001; Thompson et al., 2003;

Table 6. The power of QTL detection (proportion of significant replicates over all 1000 replicates) and the probability of false QTL detected using univariate and multivariate analysis (Trait 4)

QTL Effect	Sample Size	Model	QTL										
								Coupling[a]			False QTL[b]		
			1	9	10	One[c]	All[c]	One	Two	Mdl	Lk	Uk	Total
Small	125	UNI e	3.0	5.4	6.1	12.8	0.0	10.8	0.7	0.6	1.4	0.4	1.8
		FA1[f]	14.8	8.0	8.3	26.7	0.2	15.1	1.2	1.3	2.2	2.9	5.1
	250	UNI	32.0	50.0	47.5	79.5	8.7	73.1	24.4	4.7	7.6	1.1	8.7
		FA1	61.1	48.9	47.9	88.8	13.7	73.7	23.1	5.4	8.7	2.1	10.8
	500	UNI	57.5	69.4	68.9	95.8	29.6	91.4	46.9	0.5	0.7	0.3	1.0
		FA1	78.2	69.3	68.2	97.9	38.1	90.4	47.1	0.7	0.8	1.0	1.8
	750	UNI	81.2	86.7	86.5	99.8	60.4	99.2	74.0	0.4	0.7	0.0	0.7
		FA1	93.2	87.3	87.4	100	70.7	99.4	75.3	0.5	0.8	0.5	1.3
Medium	125	UNI	11.0	19.8	18.8	39.0	0.9	34.1	4.5	1.2	2.2	1.1	3.3
		FA1	34.8	25.6	24.2	62.9	2.7	43.7	6.1	2.2	3.6	3.2	6.8
	250	UNI	48.4	62.3	61.3	90.9	20.2	85.7	37.9	2.6	4.0	0.6	4.6
		FA1	74.2	61.5	62.4	95.8	28.9	85.7	38.2	2.9	4.1	2.3	6.4
	500	UNI	84.3	90.3	90.2	100	69.0	99.4	81.1	0.5	0.5	0.1	0.6
		FA1	93.9	89.8	90.3	100	76.2	99.2	80.9	0.6	0.6	0.5	1.1
	750	UNI	95.6	97.0	97.8	100	90.8	100	94.8	0.3	0.4	0.0	0.4
		FA1	98.8	96.8	97.8	100	93.5	100	94.6	0.3	0.4	0.1	0.5
Large	125	UNI	37.2	54.3	54.0	86.1	11.7	80.0	28.3	4.0	7.6	2.4	10.0
		FA1	61.5	53.8	53.8	91.1	19.2	77.9	29.7	4.5	8.7	2.3	12.0
	250	UNI	67.6	77.4	76.5	99.0	39.8	96.1	57.8	1.2	1.3	0.1	1.4
		FA1	86.1	75.2	75.6	99.4	48.3	95.1	55.7	1.7	1.7	1.5	3.2
	500	Diag	94.6	97.6	96.3	100	88.9	100	93.9	0.2	0.2	0.0	0.2
		FA1	98.6	97.4	95.9	100	92.0	100	93.3	0.2	0.2	0.1	0.3
	750	UNI	100	99.8	99.9	100	99.7	100	99.7	0.0	0.1	0.0	0.1
		FA1	100	99.8	99.9	100	99.7	100	99.7	0.1	0.2	0.0	0.3

[a] One: percentage or runs in which at least one of the three simulated QTL was identified. [b] All: percentage of runs in which all three simulated QTL were identified. [c] One coupling: percentage of replicates in which at least one of the two simulated linked QTL in coupling phase was identified, Two coupling: percentage of replicates in which both of the two simulated linked QTL in coupling phase were identified, Mdl: percentage of replicates in which the marker between two linked QTL was chosen. [d] Lk: proportion of false chosen markers linked to QTL, Uk: proportion of false chosen markers unlinked to QTL, Total false QTL: Lk plus Uk. e: Univariate. f: Factor analytic model.

Smith et al., 2005) in the analysis of plant variety trials to model genotype by environment interactions in the analysis of the data from multi-environment trials. The key aims of a multi-environment trial analysis are to provide accurate and precise estimates of overall variety performance and to aid with the interpretation and understanding of variety by environment interaction (Smith et al., 2001). However, in multiplicative modelling of the trait by marker interaction herein, both common and specific factors are of interest. From a breeding point of view, common factors (loci with pleiotropic effects) are important in order to implement an indirect selection program that is getting a response to selection for a trait by selecting on a correlated trait instead. This is particularly the case when the heritability for the secondary trait is smaller (Falconer and Mackay, 1996) or when the secondary trait is difficult or expensive to measure. In addition, knowledge about such loci can be very important for animal breeders who, for example, would like to dissociate the positive correlation between birth weight and carcass weight. Also knowledge about trait-specific genes is important in the case where a breeding objective is to change one trait without affecting other traits.

Specific variances for individual traits sometimes need to be constrained to zero. If more than one trait must be constrained in this way, the factor analytic variance structure then has less than full rank so that the use of standard information-based estimation techniques (e.g. average information algorithm, Gilmour et al. (1995)) is precluded. In the current research, a sparse implementation of the average information algorithm developed by Thompson et al. (2003) was used. The advantage of the algorithm is that convergence for FA models is fast and estimation of parameters in the reduced ranked models is possible. It should be noted

that, in practice, for a few replicates the occurrence of zero specific variances was observed for one or two traits, which means that the marker effect for the trait was completely determined by the multiplicative part of the model or it was too small to be detected. Also Smith et al. (2001) reported this for multi-environment trial data in Australia.

The simulation study to examine the multitrait multiple QTL approach contained several situations including some cases, which could present difficulties for a QTL analysis. These cases included QTL near the ends of the chromosome (two QTL), QTL in coupling phase (QTL with loose linkage and QTL with relatively tight linkage), QTL in repulsion phase (two QTL), and pleiotropic QTL (two QTL). The simulation results showed the general behavior and performance of the method with respect to these situations.

Multiple trait QTL analysis should increase the power of detection, and hence, increase the significance of a QTL if the QTL is not a false positive result (Jiang and Zeng, 1995; Korol et al., 1995; Mangin et al., 1998; Henshall and Goddard, 1999; Knott and, Haley 2000). In terms of the number of QTL correctly identified, the FA1 model performed better than the univariate analysis, though it was only slightly better than univariate analysis for large QTL sizes and large populations. Apart from the issue of power, it is important to understand the nature of a genetic correlation between traits, which can provide relevant information for selection decisions. In this regard, the key advantage of FA1 over univariate analysis is that it provides a formal test for pleiotropic effects. The superior performance of multivariate analysis was due largely to its ability to detect the QTL with common effect on different traits. If the pleiotropic model is the correct one, it would be expected that fitting this model would give highest power and smallest standard deviations especially for location, as in this case, a number of traits are being used to estimate the same parameter (Jinag and Zeng, 1995; Knott and Haley 2000).

The situation was considered where pleiotropic QTL had the same effect on different traits. In general, multitrait analysis will have a greater benefit when a QTL has small effects on one trait and the same QTL has greater effect on another trait (Sorensen et al., 2003) or when the pleiotropic effects of the QTL differ substantially from the most frequently observed effects of the environment and background genes, which is reflected by the environmental and background genetic correlations (Meuwissen and Goddard, 2004).

The power of the FA1 model compared to the univariate analysis was more evident when the pleiotropic effect was small. In general, a pleiotropic locus, too small to be detected by single-trait analyses, can be detected with the help of a multitrait analysis (Mangin et al., 1998).

With respect to false QTL detection, in general, both the univariate and multivariate methods chose a rather low portion of linked or unlinked loci to a QTL. In the simulation study herein, a relatively sparse marker density was considered. However, a high marker density may increase the likelihood of choosing linked false markers.

The proposed method herein is a one-stage process, which models residuals and genetic effects simultaneously. In addition, it includes all the markers in one analysis. Moreover, the approach utilizes widely available statistical procedures, namely the linear mixed model and restricted maximum likelihood. It can easily accommodate co-variates, extra sources of variation, fixed or random including polygenic effects and it can easily be generalized to experimental and crossing designs commonly used.

Only four traits were considered in the simulation data set. However, the model, as formulated in equation (6) is obviously expandable to an unlimited number of phenotypes. However, the behavior of the approach using very dense marker maps needs to be investigated. The fact that the same variance for the total markers of the entire genome is used is problematic, since the majority of the markers will not be linked to the QTL and they may dominate the estimate of the marker variance/co-variance. Consequently, the estimate of the genome variance/co-variance will be close to zero. It should be noted that the extra markers on a chromosome would not dilute the marker variance associated with the chromosome because they also have co-variances between them. However, adding extra chromosomes may dilute the genome variance, as most of the linkage groups are not linked to the QTL. Therefore, one solution is to allow the markers of the same linkage group have common variance. The extension of this approach to the multivariate analysis would be removing the linkage groups with non-significant variance for the traits of interest, then those traits that had non-zero variances for a chromosome were combined and the factor model fitted across the chromosome instead of the genome.

Correlations among phenotypes can arise from several different causal processes, which may have differ-

ent implications for the power and performance of the multivariate linkage analysis. Obviously, further studies using the approach suggested herein for multitrait QTL mapping that specifically consider different situations should be undertaken. Furthermore, the efficiency of the model to distinguish between a pleiotropic QTL and closely linked QTL affecting different traits is another area that needs more investigation.

References

Allison, D.B., Heo, M., 1998. Meta-analysis of linkage data under worst-case conditions: A demonstration using the human OB region. *Genetics* 148, 859-866.

Ball, R.D., 2001. Bayesian methods for quantitative trait loci mapping based on model selection: Approximate analysis using the Bayesian information criterion. *Genetics* 159, 1351-1364.

Broman, K.W., Speed, T.P., 2002. A model selection approach for the identification of quantitative trait loci in experimental crosses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64, 641-656.

Cullis, B.R., Gogel, B. J., Verbyla, A.P., Thompson, R., 1998. Spatial analysis of multi-environment early generation trials. *Biometrics* 54, 1-18.

Esmailizadeh, A., Bottema, C.D.K., Sellick, G.S., Verbyla, A. P., Morris, C.A., Cullen, N. G., Pitchford, W. S., 2008. Effects of the myostatin F94I substitution on beef traits. *Animal Science* 86, 1038–1046.

Esmailizadeh K.A., Mohammadabadi M.R. 2010. A molecular genome scan to map quantitative trait loci affecting bovine carcass weight. *Agricultural Biotechnology Journal* 1, 117-130.

Falconer, D.S., Mackay, T.F.C., 1996. Introduction to Quantitative Genetics. 4th ed., Longmans Green, Harlow, Essex, UK.

Fisher, R.A., 1954. Statistical Methods for Research Workers, 12th ed., Oliver and Boyd, Edinburgh, UK.

Gilbert, H., Le Roy, P., 2003. Comparison of three multitrait methods for QTL detection. *Genetics Selection Evolution* 35, 281-304.

Gilbert, H., Le Roy, P., 2004. Power of three multitrait methods for QTL detection in crossbred populations. *Genetics Selection Evolution* 36, 347-361.

Gilmour, A.R., 2007. Mixed model regression mapping for QTL detection in experimental crosses. *Computational Statistics & Data Analysis* 51, 3749-3764

Gilmour, A.R., Gogel, B.J., Cullis B.R., Thompson, R., 2006. ASReml user guide Release 2.0 VSN International Ltd., Hemel Hempstead, HP1 1ES, UK.

Gilmour, A.R., Thompson, R., Cullis, B.R., 1995. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51, 1440–1450.

Goddard, M.E., 2001. The validity of genetic models underlying quantitative traits. *Livestock Production Science* 72, 117-127.

Haldane, J.B.S., 1919. The combination of linkage values, and the calculation of distance between linked factors. *Genetic*s 8, 299–309.

Haley, C.S., 1999. Advances in quantitative trait locus mapping. Proceedings of From Jay L. Lush to Genomics: Visions for Animal Breeding and Genetics, Iowa State University, Ames, Iowa, USA, pp. 47-59.

Henshall, J.M., Goddard, M.E., 1999. Multiple-trait mapping of quantitative trait loci after selective genotyping using logistic regression. *Genetics* 151, 885-894.

Jiang, C., Zeng, Z.B., 1995. Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* 140, 1111-1127.

Johnson, R.A., Wichern, D.W., 1998. Applied Multivariate Analysis. 4th ed., Prentice Hall, Englewood Cliffs, New Jersey, USA.

Kao, C.H., Zeng, Z.B., Teasdale, R.D., 1999. Multiple interval mapping for quantitative trait loci. *Genetics* 152, 1203-1216.

Knott, S.A., 2005. Regression-based quantitative trait loci mapping: Robust, efficient and effective. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 360, 1435-1442.

Knott, S.A., Haley, C.S., 2000. Multitrait least squares for quantitative trait loci detection. *Genetics* 156, 899-911.

Korol, A.B., Ronin, Y.I., Itskovich, A.M., Peng, J., Nevo, E., 2001. Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits. *Genetics* 157, 1789-1803.

Korol, A. B., Ronin, Y.I., Kirzhner, V. M., 1995. Interval mapping of quantitative trait loci employing correlated trait complexes. *Genetics* 140: 1137-1147.

Lawley, D.N., Maxwell, A.E., 1971. Factor Analysis as a Statistical Method. 2nd ed., Butterworths, London. ***UK***.

Mangin, B., Thoquet, P., Grimsley, N., 1998. Pleiotropic QTL analysis. *Biometrics* 54, 88-99.

Meuwissen, T.H., Goddard, M.E., 2004. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genetics Selection Evolution* 36, 261-279.

Moradian H., Esmailizadeh A.K., Sohrabi S., Mohammadabadi M.R. 2015. Identification of quantitative trait loci associated with weight and percentage of internal organs on chromosome 1 in Japanese quail. *Journal of Agricultural Biotechnology* 6, 143-158.

Moradian H., Esmailizadeh A.K., Sohrabi S.S., Nasirifar E., Askari N., Mohammadabadi M.R., Baghizadeh A. 2014. Genetic analysis of an F_2 intercross between two strains of Japanese quail provided evidence for quantitative trait loci affecting carcass composition and internal. *Molecular Biology Reports* 41, 4455-4462.

Piepho, H.P., 2000. A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. *Genetics* 156, 2043-2050.

Piepho, H.P., Gauch, H.G., JR., 2001. Marker pair selection for mapping quantitative trait loci. *Genetics* 157, 433-444.

Sen, S., Churchill, G.A., 2001. A statistical framework for quantitative trait mapping. *Genetics* 159, 371-387.

Sillanpaa, M.J., Corander, J., 2002. Model choice in gene mapping: What and why. *Trends in Genetics* 18, 301-307.

Smith, A., Cullis, B., Thompson, R., 2001. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* 57, 1138-1147.

Smith, A.B., Cullis, B.R., Thompson, R., 2005. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. *Agricultural Science* 143, 1–14.

Sohrabi S., Esmailizaseh K.A., Mohammadabadi M.R., Moradian H. 2014. Mapping quantitative trait loci underlying Kleiber ratio and identification of their mode of action in an F2 population of Japanese quail (*Coturnix coturnix japonica*). *Agricultural Biotechnology Journal* 6, 111-122.

Sorensen, P., Lund, M.S., Guldbrandtsen, B., Jensen, J., Sorensen, D., 2003. A comparison of bivariate and univariate QTL mapping in livestock populations. *Genetics Selection Evolution* 35, 605-622.

Stearns, T. M., Beever, J. E., Southey, B. R., Ellis, M., F. Mckeith, K., Rodriguez-Zas, S. L., 2005. Evaluation of approaches to detect quantitative trait loci for growth, carcass, and meat quality on swine chromosomes 2, 6, 13, and 18. I. Univariate outbred F2 and sib-pair analyses. *Animal Science* 83, 1481-1493.

Stuart, A., Ord, J.K., Arnold, S., 1999. Kendall's Advanced Theory of Statistics.Classical Inference and the Linear Model, Vol. 2A, 6 ed., Oxford University Press Inc., New York, USA.

Thompson, R., Cullis, B., Smith, A., Gilmour, A., 2003. A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. *Australian & New Zealand Journal of Statistics* 45, 445-459.

Verbyla, A.P., Cullis, B.R., Thompson, R., 2007. The analysis of QTLs by simultaneous use of the full linkage map. *Theoretical and Applied Genetics* 116, 95-111.

Verbyla, A.P., Eckermann, P.J., Thompson, R., Cullis, B.R., 2003. The analysis of quantitative trait loci in multi-environment trials using a multiplicative mixed model. *Australian Journal of Agricultural Research* 54, 1395–1408.

Walling, G.A., Visscher, P.M., Andersson, L., Rothschild, M.F., Wang, L., Moser, G., Groenen, M.A.M., Bidanel, J.P., Cepica, S., Archibald, A.L., Geldermann, H., de Koning, D.J., Milan, D., Haley, C.S., 2000. Combined analyses of data from quantitative trait loci mapping studies: Chromosome 4 effects on porcine growth and fatness. *Genetics* 155, 1369-1378.

Weller, J.I., Wiggans, G.R., Vanraden, P.M., Ron, M., 1996. Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multi-trait experiment. *Theoretical and Applied Genetics* 92, 998-1002.

Wood, I.A., Moser, G., Burrell, D.L., Mengersen, K.L., Hetzel, D.J., 2006. A meta-analytic assessment of a thyroglobulin marker for marbling in beef cattle. *Genetics Selection Evolution* 38, 479-494.

Xu, S., 2003. Estimating polygenic effects using markers of the entire genome. *Genetics* 163, 789-801.