Genet. Res., Camb. (2010), **92**, *pp.* 239–250.
doi:10.1017/S0016672310000236

Bayesian mixture structural equation modelling in multiple-trait QTL mapping

XIAOJUAN MI[1], KENT ESKRIDGE[1]*, DONG WANG[1], P. STEPHEN BAENZIGER[2], B. TODD CAMPBELL[3], KULVINDER S. GILL[4] AND ISMAIL DWEIKAT[2]
[1] *Department of Statistics, University of Nebraska, Lincoln, NE 68583-0963, USA*
[2] *Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583, USA*
[3] *USDA-ARS-Coastal Plains Research Ctr., Florence, SC 29501, USA*
[4] *Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164, USA*

(*Received 18 January 2010 and in revised form 28 April 2010*)

Summary

Quantitative trait loci (QTLs) mapping often results in data on a number of traits that have well-established causal relationships. Many multi-trait QTL mapping methods that account for correlation among the multiple traits have been developed to improve the statistical power and the precision of QTL parameter estimation. However, none of these methods are capable of incorporating the causal structure among the traits. Consequently, genetic functions of the QTL may not be fully understood. In this paper, we developed a Bayesian multiple QTL mapping method for causally related traits using a mixture structural equation model (SEM), which allows researchers to decompose QTL effects into direct, indirect and total effects. Parameters are estimated based on their marginal posterior distribution. The posterior distributions of parameters are estimated using Markov Chain Monte Carlo methods such as the Gibbs sampler and the Metropolis–Hasting algorithm. The number of QTLs affecting traits is determined by the Bayes factor. The performance of the proposed method is evaluated by simulation study and applied to data from a wheat experiment. Compared with single trait Bayesian analysis, our proposed method not only improved the statistical power of QTL detection, accuracy and precision of parameter estimates but also provided important insight into how genes regulate traits directly and indirectly by fitting a more biologically sensible model.

1. Introduction

Research on quantitative trait loci (QTLs) often provides information on multiple complex traits that have well-established causal relationships. For example, in wheat genetics, it is common to collect data on grain yield (GRYL) and yield components such as thousand kernel weight (TKW), spikes per square metre (SPSM) and kernels per spike (KPS), where the causal relationships among these traits are well-established (Fig. 1), because yield components develop sequentially with later-developing components under the control of earlier-developing ones (Dofing & Knight, 1992). The primary goal of QTL mapping is to locate regions or genes that are associated with quantitative traits. The commonly used procedures capture only total QTL effects while providing no understanding of direct and indirect effects. However, these direct and indirect effects can help answer important questions that are not addressed by examining the total effect alone. For instance, a pleiotropic QTL can have a positive direct effect on GRYL, but a negative effect on a yield component. Without knowing the full pathway of the causal relationship, a breeder might select against the yield component QTL thinking it only affects the yield component detrimentally, not knowing that it is actually beneficial to the important trait of GRYL. Thus, the total effect can provide a misleading impression. To understand

* Corresponding author. Department of Statistics, University of Nebraska, Lincoln, NE 68583-0963, USA. Tel: (402) 472-7213. Fax: (402) 472-5179 ong. e-mail: keskridg@unlserve.unl.edu

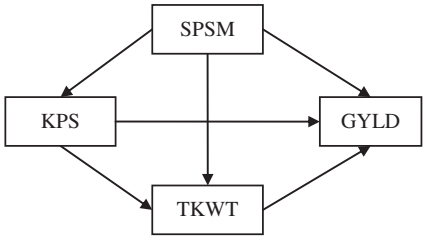

Fig. 1. The path diagram of the causal relationship among GRYL and yield components.

the genetic effects of a QTL thoroughly, it is necessary to understand not only the total QTL effect but also the direct and indirect effects of a QTL through other traits by incorporating the causal structure among the traits. Such a strategy of QTL mapping can provide additional insight into how QTLs regulate traits directly and indirectly through other traits. It should also improve the power of the QTL detection and the precision of the parameter estimate.

QTL-mapping studies are usually conducted for each trait separately using single trait analyses (Lander & Botstein, 1989; Haley & Knott, 1992; Jansen & Stam, 1994; Zeng, 1994). However, such a single trait analysis may result in biased estimates and lower statistical power of QTL detection when data observations are collected on multiple causally related or genetically correlated traits. To combat these problems, several multiple trait QTL analysis (joint analysis) methods have been developed to take into account the correlation among multiple traits. These methods may be classified into multi-trait maximum-likelihood (ML) (Jiang & Zeng, 1995; Williams *et al*., 1999; Xu *et al*., 2005), multi-trait least squares (LS) (Calinski *et al*., 2000; Knott & Haley, 2000; Hackett *et al*., 2001), principal component analysis (PCA) (Weller *et al*., 1996; Mangin *et al*., 1998) and discriminant analysis (DA) (Gilbert & Le Rol, 2003). Multi-trait ML, implemented with the expectation conditional maximization (ECM) algorithm, provides a powerful tool to multi-trait QTL mapping. However, there are problems with this method when the number of QTLs and traits increase. The likelihood is a finite mixture of densities and becomes very difficult to evaluate (Satagopan *et al*., 1996). The gains in power from joint analysis may compensate for the critical value for the test, due to the increase in the number of unknown parameters to be estimated (Mangin *et al*., 1998). Multi-trait LS, which regresses the quantitative trait value on the conditional expected genotypic value, produces results very similar to ML and simplifies computation (Haley & Knott, 1992). The PCA and DA dimension reduction techniques, decompose the traits into a number of linear combinations that can be analysed separately. However, the approaches of PCA and DA may cause spurious linkages and difficulties in the biological interpretation of study results (Mähler *et al*., 2002; Gilbert & Le Roy, 2003).

With new and powerful computational techniques available, Bayesian QTL mapping provides an extremely flexible way to search for multiple QTLs simultaneously. There are many practical advantages of using a Bayesian approach over frequentist approaches, such as the ability to fit more complex and biologically sensible models, the ability to incorporate prior information into the specification of the model, and the ability to obtain estimates of the posterior distributions of any function of the model parameters (Dunson, 2001; Yi & Shriner, 2008). The Bayesian approach has been extensively applied to QTL mapping for a single trait (Satagopan *et al*., 1996; Sillanpaa & Arjas, 1998, 1999; Stephens & Fisch, 1998; Yi & Xu, 2002; Yi *et al*., 2003, 2005, 2007; Narita & Sasaki, 2004; Yi, 2004; Wang *et al*., 2005). Recently, several Bayesian methods implemented via the Markov chain Monte Carlo (MCMC) algorithm have been developed for mapping multiple trait QTLs taking into account the correlation among traits. Meuwissen & Goddard (2004) combined linkage and linkage disequilibrium (LD) information to improve the power and precision of QTL mapping. Liu *et al*. (2007) developed a variance component method to model multiple complex traits in outbred populations using a Bayesian approach. In this approach, the number of QTLs is determined by reversible-jump MCMC (Green, 1995; Sillanpaa & Arjas, 1998). The problem with the reversible-jump MCMC for model selection is that it is usually subject to slow mixing of the Markov Chains and high computational demand associated with the algorithm (Wang *et al*., 2005). Yang & Xu (2007) extended the Bayesian shrinkage analysis (Wang *et al*., 2005) to dynamic complex traits by fitting the growth trajectory using Legendre polynomials. The advantage of this method is that it fits any trend in time but parameters of polynomials have no biological interpretation. Banerjee *et al*. (2008) introduced the seemingly unrelated regression (SUR) model, which allows different genetic models for different traits. However, none are capable of dealing with the causal relationships among traits, resulting in the omission of the direct and indirect QTL effects.

The structural equation model (SEM) is a generalization of simultaneous equation procedures originating from path analysis (Wright, 1921) and initially popularized in econometics and genetics. It is a useful method for estimating and evaluating simultaneous causal relationships among variables which allows variables to be both dependents and predictors. It is best explained by considering a path diagram. In particular, SEM allows researchers to decompose the effects of one variable on another into direct, indirect

and total effects. The direct effect is the path coefficient between an independent variable and the dependent variable that are not causally explained by any other intermediary variable. The indirect effects of a variable are mediated by at least one other intervening variable. The indirect effects are calculated by multiplying the path coefficients for each path of the associated variable to the dependent variable. The total effect is the sum of direct and all indirect effects. By explicitly accounting for the complex multicomponent causal structure among traits, SEM can provide a better understanding of multiple trait QTL analysis by allowing researchers to decompose the effects of one variable on another into direct, indirect and total effects within a QTL framework.

Recently, SEM has been applied to functionally related traits in genetic research with the goal of characterizing genetic architecture precisely and intuitively. Zhu & Zhang (2009) conducted simulation studies to compare the performance of multiple trait analysis and single trait analysis in family-based association studies. They found that multiple trait analysis improved the power of association tests and precision of parameter estimates when there are causal relations among the traits themselves. Nadeau *et al.* (2003) used the Bayesian network analysis to infer a functional/causal trait network of the cardiovascular system from a RIL population. Li *et al.* (2006) analysed data from mouse inbred crosses to identify the causal networks including subphenotypes and QTL related to obesity and bone geometry. However, their approaches were limited to testing and quantifying the relationships among identified QTLs and phenotypes without QTL detection.

In this paper, we propose SEM in the context of QTL detection. The goal is to develop a Bayesian SEM approach mapping multiple traits QTL using recombinant inbred line (RIL) populations. The RIL populations are commonly used in QTL mapping experiments and are usually derived from a cross between two inbred parents followed by self-pollination and single-seed descent to reach homozygosity. The performance of the proposed method is evaluated by a simulation study and applied to data from a wheat experiment.

2. Statistical method

(i) *Mixture SEM*

Consider m QTLs located at positions $\lambda_1, \lambda_2, \ldots, \lambda_m$ in m different marker intervals $I_1, I_2, \ldots, I_m$ in a linkage group on the same chromosome. Let the value of each marker and putative QTL be coded as 2 for one homozygous parent type and 0 with the other homozygous parent type, since the RILs are homozygous at every locus. Assume that p causally related quantitative traits $y_1, y_2, \ldots,$ and y_p are affected by these m QTLs additively. The SEM in matrix form is

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 0 & \beta_{12} & \cdots & \beta_{1p} \\ 0 & 0 & \cdots & \beta_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}}_{\mathbf{y}} + \underbrace{\begin{bmatrix} \alpha_{11} \\ \alpha_{12} \\ \vdots \\ \alpha_{1p} \end{bmatrix}}_{\boldsymbol{\alpha}_1} Q_1 + \ldots + \underbrace{\begin{bmatrix} \alpha_{21} \\ \alpha_{22} \\ \vdots \\ \alpha_{2p} \end{bmatrix}}_{\boldsymbol{\alpha}_m} Q_m + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix}}_{\boldsymbol{\zeta}}, \quad (1)$$

where y_k is the phenotypic value for trait k ($k = 1, 2, \ldots, p$), β_{kh} is the regression coefficient of trait h on trait k ($h = 1, 2, \ldots, m$), α_{lk} is the direct effect of the lth putative QTL on trait k ($l = 1, 2, \ldots, m$), Q_l is the lth putative QTL genotype, taking the value of 2 for genotype QQ and 0 for genotype qq and e_k, the residual effect on trait k, is assumed to be multivariate normal distributed with means zero and covariance matrix

$$\boldsymbol{\Psi} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_p^2 \end{pmatrix}.$$

More compactly, the model (1) for ith individual ($i = 1, 2, \ldots, n$) can be rewritten as

$$\mathbf{y}_i = \mathbf{B}\mathbf{y}_i + \sum_{l=1}^{m} \boldsymbol{\alpha}_l Q_{il} + \boldsymbol{\zeta}_i \quad (2)$$

and the reduced model

$$\mathbf{y}_i = (\mathbf{I} - \mathbf{B})^{-1} \left(\sum_{l=1}^{m} \boldsymbol{\alpha}_l Q_{il} + \boldsymbol{\zeta}_i \right), \quad (3)$$

where $\mathbf{y}_i$ is a $p \times 1$ vector of y_k for the ith individual, $\mathbf{B}$ is the $p \times p$ coefficient matrix (contains βs) that describes causal relationship among p traits, where $(\mathbf{I} - \mathbf{B})^{-1}$ exists; $\boldsymbol{\alpha}_l$ is a $p \times 1$ vector of α_{lk}, Q_{il} is the putative QTL genotype for the lth QTL and ith individual; $\boldsymbol{\zeta}_i$, a $p \times 1$ vector of errors in the equation, is assumed to be multivariate and normally distributed with mean zero and diagonal covariance matrix $\boldsymbol{\Psi}$. Elements in $\mathbf{B}$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_m)$ and $\boldsymbol{\Psi}$ are parameters that need to be estimated.

In practice, we observe the marker genotypes and trait values, but not the putative QTL genotypes. However, given the lth QTL positions λ_l and observed flanking marker genotypes, the distribution of the lth QTL genotype Q_l in interval I_l can be inferred in terms

https://doi.org/10.1017/S0016672310000236 Published online by Cambridge University Press

of the recombination frequency between them. We assume that there is no crossing-over interference. The conditional distributions of the individual putative QTL genotypes are independent given the flanking marker genotypes (Kao *et al.*, 1999). The joint conditional probability of the genotype of the m putative QTL for individual i can be expressed as

$$f(Q_i|\lambda, I_i) = f(Q_{i1}, Q_{i2}, \ldots, Q_{im}|\lambda, I_i) = \prod_{l=1}^{m} f(Q_{il}|\lambda_l, I_{il}),$$
$$i = 1, 2, \ldots, n, \tag{4}$$

where Q_i is the joint QTL genotype for individual i; $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ are the locations of m QTLs; $I_i = (I_{i1}, I_{i2}, \ldots, I_{im})$ are the intervals of flanking markers for individual i; Q_{il} is the putative QTL genotype of the lth QTL for the ith individual. The joint conditional probability of the m QTL is the product of the marginal conditional probabilities of individual QTL. There are 2^m possible different QTL genotypes in the population. We denote q_{ij} ($j = 1, 2, \ldots, 2^m$) as the 2^m possible QTL genotypes with the conditional probabilities of p_{ij} respectively for the ith individual, where p_{ij} containing information on QTL positions is non-negative and $\sum_{j=1}^{2^m} p_{ij} = 1$. That is, $f(Q_i = q_{ij}) = p_{ij}$.

Given eqn (4), model (3) is defined as a finite mixture SEM. We assume the multivariate normal distribution of residual errors, the likelihood conditional on all unknowns $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$, $\theta = (\beta$'s, $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_m)$ and $\delta = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_p{}^2)$ is defined as

$$f(\mathbf{y}|\lambda, \theta, \delta) = \prod_{i=1}^{n} \sum_{j=1}^{2^m} f(Q_i = q_{ij}|\lambda) f_j(\mathbf{y}_i|Q_i = q_{ij}, \theta, \delta)$$
$$= \prod_{i=1}^{n} \sum_{j=1}^{2^m} p_{ij} f_j(\mathbf{y}_i|\mathbf{u}_j, \boldsymbol{\Sigma}), \tag{5}$$

this is a 2^m component mixture SEM, $f_j(\mathbf{y}_i|\mathbf{u}_j, \boldsymbol{\Sigma})$ is a multivariate normal density for the jth QTL genotype ($j = 1, 2, \ldots, 2^m$) with probability p_{ij}, mean vector $\mathbf{u}_j$ and a covariance matrix $\boldsymbol{\Sigma} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})^{-1'}$, assuming the same covariance matrix across all components. The mean vector $\mathbf{u}_j$s are derived from eqn (3) corresponding to the genotypic values of the 2^m different QTL genotypes. For instance, in the two QTLs model, there are four multivariate normal densities with mean vectors $\boldsymbol{\mu}_1 = (\mathbf{I} - \mathbf{B})^{-1}\sum_{j=1}^{2} 2\boldsymbol{\alpha}_j$, $\boldsymbol{\mu}_2 = (\mathbf{I} - \mathbf{B})^{-1}2\boldsymbol{\alpha}_1$, $\boldsymbol{\mu}_3 = (\mathbf{I} - \mathbf{B})^{-1}2\boldsymbol{\alpha}_2$ and $\boldsymbol{\mu}_4 = 0$, respectively.

The mean vectors and covariance matrix are functions of unknown parameters, which make the likelihood very difficult to evaluate by applying maximum likelihood procedures when the number of QTLs and traits increase. In response to this problem, we apply a Bayesian approach using an MCMC algorithm, which provides a powerful tool for solving complex mixtures. Inferences are based on the joint posterior distribution of all unknowns given the prior distribution of all unknowns and the observed data. We can also make use of posterior probability to obtain estimates of the posterior distributions of any function of the parameters, such as indirect and total QTL effects based on our proposed multi-trait SEM. The number of QTLs affecting traits is determined by the Bayes factor (BF).

Joint posterior distribution: in the Bayesian framework, the joint posterior distribution of all unknowns $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$, $\theta = (\beta$'s, $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_m)$ and $\delta = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_p{}^2)$, given the trait values $\mathbf{y}$, the marker genotypes ($\mathbf{M}$) and prior information, can be expressed as

$$f(\lambda, \theta, \delta, Q|\mathbf{y}, \mathbf{M}) \propto f(\mathbf{y}|Q, \theta, \delta) f(Q|\lambda) f(\lambda, \theta, \delta)$$
$$= \prod_{i=1}^{n} f(\mathbf{y}_i|Q_i, \theta, \delta) f(Q_i|\lambda) f(\lambda, \theta, \delta). \tag{6}$$

The terms $f(\mathbf{y}|Q, \theta, \delta)f(Q|\lambda)$ on the right-side of eqn (6) is the likelihood conditional on all unknowns, which is defined in eqn (5). The last term $f(\lambda, \theta, \delta)$ is the joint prior distribution of all parameters (λ, θ, δ). We assume independence of prior distributions

$$f(\lambda, \theta, \delta) = \prod_{l=1}^{m} f(\lambda_l) f(\theta) \prod_{k=1}^{p} f(\sigma_k^2). \tag{7}$$

The prior distribution of the QTL location λ is assumed to be uniform on a predefined interval. When no information regarding the locations is available, a prior uniform distribution with an interval equivalent to the length of the chromosome can be used. For convenience in elicitation and computation, we chose conjugate priors for the remaining parameters. The prior distribution of θ is assumed to be multivariate normal. The prior distributions of the variance component parameters are assumed to be independent inverse-gamma distributions.

Full conditional posterior distributions: from eqn (6) we can derive the full conditional posterior distributions.

The conditional distribution of a QTL genotype is

$$f(Q_i = q_{ij}|\mathbf{y}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda) = \frac{f(Q_i = q_{ij}|\lambda) f_j(\mathbf{y}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma})}{\sum_{j=1}^{2^m} f(Q_i = q_{ij}|\lambda) f_j(\mathbf{y}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma})}$$
$$= \frac{p_{ij} f_j(\mathbf{y}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma})}{\sum_{j=1}^{2^m} p_{ij} f_j(\mathbf{y}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma})}, \tag{8}$$

where $f_j(\mathbf{y}_i|\mathbf{u}_j, \boldsymbol{\Sigma})$ is a multivariate normal density function for the jth QTL genotype ($j = 1, 2, \ldots, 2^m$) with mean vector $\mathbf{u}_j$ and a covariance matrix $\boldsymbol{\Sigma} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})^{-1'}$. Let $\boldsymbol{\theta}$ be a vector of the path coefficients (elements in $\boldsymbol{\alpha}$ and $\mathbf{B}$). The model (2) $\mathbf{y}_i = \mathbf{B}\mathbf{y}_i + \sum_{l=1}^{m} \boldsymbol{\alpha}_l Q_{il} + \boldsymbol{\zeta}_i$ can be rewritten as $\mathbf{y}_i = \mathbf{A}_i\boldsymbol{\theta} + \boldsymbol{\zeta}_i$,

https://doi.org/10.1017/S0016672310000236 Published online by Cambridge University Press

$V(\boldsymbol{\zeta}_i)=\boldsymbol{\Psi}$, where $\mathbf{y}_i$ is a $p\times 1$ vector of trait values for observation i ($i=1, 2, \ldots, n$) with a known $p\times q$ matrix $\mathbf{A}_i$ containing $\mathbf{y}_i$ and Q_i; $\boldsymbol{\theta}$ is a $q\times 1$ vector of coefficients to be estimated; $\boldsymbol{\zeta}_i$ is a $p\times 1$ vector of random residuals, residual ζ_{ik} is assumed to be uncorrelated and have variance

$$V(\zeta_{ik})=\sigma_k^2;\quad k=1, \ldots, p$$

$\boldsymbol{\theta}$ and σ_k^2 are assumed to follow the conjugate priors

$$\boldsymbol{\theta}\sim N(\gamma^*, \Omega_\gamma),$$
$$\sigma_k^2\sim \text{InvGamma}(e,f).$$

The expected value γ^*, covariance matrix Ω_γ, e and f are chosen by the researcher. Arminger & Muthén (1998) suggested that one may set γ^* to vector $\{1, \ldots, 1\}$, Ω_γ^{-1} to a diagonal matrix with small values, such as 0·01, set e to $\frac{1}{2}$, and set f^{-1} to a small value, for instance 0·01. The posterior distribution of $\boldsymbol{\theta}$ and σ_k^2 is given as the following (Arminger & Muthén, 1998):

$$\begin{aligned}\boldsymbol{\theta}|\mathbf{y}, Q, \boldsymbol{\Psi}\sim \text{MVN}\Bigg(&\left(\left[\sum_{i=1}^{n}\mathbf{A}_i'\boldsymbol{\Psi}^{-1}\mathbf{A}_i\right]+\Omega_\gamma^{-1}\right)^{-1}\\ &\times\left(\left[\sum_{i=1}^{n}\mathbf{A}_i'\boldsymbol{\Psi}^{-1}\mathbf{y}_i\right]+\Omega_\gamma^{-1}\gamma^*\right),\\ &\left(\left[\sum_{i=1}^{n}\mathbf{A}_i'\boldsymbol{\Psi}^{-1}\mathbf{A}_i\right]+\Omega_\gamma^{-1}\right)^{-1}\Bigg),\end{aligned}\quad (9)$$

$$\begin{aligned}\sigma_k^2|\mathbf{y}, Q, \boldsymbol{\theta}\sim&\text{InvGamma}\Bigg(\frac{n}{2}+e,\\ &\left[\frac{1}{2}\sum_{i=1}^{n}(y_{ik}-\mathbf{A}_i\boldsymbol{\theta})'(y_{ik}-\mathbf{A}_i\boldsymbol{\theta})+f^{-1}\right]^{-1}\Bigg).\end{aligned}\quad (10)$$

Based on fitting of the multi-trait SEM and posterior distribution of the path coefficients θ, we can obtain estimates of the posterior distributions of the indirect and total QTL effects, which are functions of θ. Unlike the above parameters θ and σ_k^2, there is no explicit expression for the full conditional posterior distributions of the parameters λ. The Metropolis–Hastings sampler can be used to draw the samples from the joint posterior distribution (see detailed description in the following section).

(ii) *Parameter estimate*

Once all the full conditional posteriors are specified, the following MCMC algorithm can implemented.

Step 1. Initialization: Set the initial values of ($\lambda^{(0)}$, $\boldsymbol{\theta}^{(0)}$ and $\delta^{(0)}$) in the space of each parameter.

Step 2. Update the QTL position (λ): There is no closed form for the conditional posterior probability density of a QTL position. Therefore, we take the Metropolis–Hastings (Metropolis *et al.*, 1953; Hastings, 1970) approach for sampling the position of a QTL. Elements of λ are updated one at a time sequentially. Specifically, for the lth QTL, a proposal position λ_l^* is sampled from a uniform distribution with symmetric interval (max (λ_{l-1}, λ_l-d), min(λ_l+d, λ_{l+1})) around the previous position λ_l, where d is the predefined tuning parameter, usually taking a value of 2 cM. The proposed position λ_l^* is accepted with probability.

$$\alpha(\lambda_l, \lambda_l^*)=\min\left\{1, \frac{f(\lambda_l^*|\lambda_{-l}, \theta, \delta, \mathbf{y})}{f(\lambda_l|\lambda_{-l}, \theta, \delta, \mathbf{y})}\right\},\quad (11)$$

where λ_{-l} represents all elements of λ except λ_l. If the new position is accepted, the joint conditional probabilities ($p_{ij}=f(Q_i=q_{ij}|\lambda)$) of the m QTL genotypes (see eqn (4)) is also updated simultaneously. Otherwise, the state remains unchanged, and the algorithm proceeds to update the next QTL position. The QTL position can be updated more than once between updates of other parameters if there is evidence that the chain is mixing slowly (Satagopan *et al.*, 1996).

Step 3. Update QTL genotype (Q): The genotype of joint m putative QTLs (Q_i) is updated one individual at a time. It is sampled from its full conditional probability distribution (8)

$$f(Q_i=q_{ij}|\mathbf{y}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda)=\frac{p_{ij}f_j(\mathbf{y}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma})}{\sum_{j=1}^{2^m}p_{ij}f_j(\mathbf{y}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma})}.\quad (12)$$

Step 4. Update the path coefficients (θ): Elements of θ are simultaneously sampled from their full conditional distribution $\pi(\theta|\lambda, \delta, Q, \mathbf{y})$ given in eqn (9), which is a multivariate normal distribution.

Step 5. Update the residual variances (δ): Elements in δ are updated individually based on their full conditional distribution $f(\sigma_k^2|\lambda, \theta, \sigma_{-k}^2, Q, \mathbf{y})$ given in eqn (10), which is an inverted-gamma distribution.

Step 6. Update the indirect and total QTL effects: The indirect and total QTL effects are functions of the path coefficients θ. They are updated based on the current values of θ. The indirect QTL effect for a particular indirect path from the QTL to the trait is calculated by multiplying all the coefficients in the path. The total indirect QTL effect on the trait is the sum of all the indirect effects from all indirect paths.

Continued sampling by repeating steps 2–6 is known as the MCMC method, because the previous sample values are used as parameters to sample the next values, generating a Markov chain. When the chain is long enough, it converges to the stationary distribution; the sampled parameters actually follow the joint posterior distribution. Likewise, a sample of any single parameter is drawn from its marginal posterior density. Parameters can be estimated based on the samples from the corresponding posterior distributions. Here, we use the posterior mean as the Bayesian estimate.

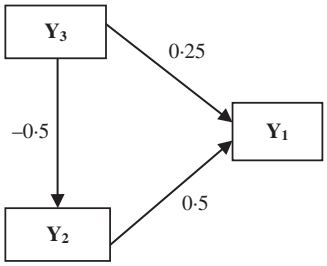

Fig. 2. Causal relationships among three traits in the simulation. Numbers by the arrow lines represent the true path coefficients.

(iii) *Number of QTLs*

We use the BF criterion (Jeffreys, 1961; Kass & Raftery, 1995) as a test statistic to detect QTLs. Specifically, this is the comparison of two different models with different number of QTLs affecting the traits using BFs. Let model_1 and model_0 be two competing models for a given data set. From Bayes theorem, the BF is defined as

$$B_{1-0}=\frac{f(y|\text{model}_1)}{f(y|\text{model}_0)}. \tag{13}$$

The marginal probability of the data under model_j $f(y|\text{model}_j)$ can be estimated using its harmonic mean estimator (Newton & Raftery, 1994).

$$\hat{f}(y|\text{model}_j)=\frac{N}{\sum_{t=1}^{N}\frac{1}{f(\mathbf{y}|\lambda^t,\theta^t,\delta^t,Q^t)}}, \quad j=0,1, \tag{14}$$

where N is the number of total iterations in the MCMC process; λ^t, θ^t, δ^t, Q^t are samples of all unknowns drawn from the tth iteration. Unlike the significance test approach that is based on P-values, this comparison does not depend on the assumption that either model is 'true', and can be applied to non-nested models. It is useful to consider the natural logarithm of the BF and interpret the resulting statistic based on the following criterion given by Kass & Raftery (1995): a negative log B_{1-0} is taken as support for model_0, while a value between 1 and 3 indicates support for model_1 and a value in excess of 3 points to strong support for model_1; a value between 0 and 1 does not allow any conclusion to be drawn.

3. A simulation study

The Bayesian analysis for multi-trait QTL mapping described above was investigated using simulation experiments. The data were simulated for 100 replicates of 250 lines from an RIL population. On a single chromosome segment of length 100 cM, 11 evenly spaced markers were simulated. Two QTLs (Q_1 and Q_2) were placed at 42 and 78 cM to affect three traits, which are causally related as in Fig. 2. The phenotypic values for each individual are determined by eqn (15), the causal relationship among the traits, the effects of two QTLs sampled (where QTL takes values of 2 and 0 for genotype QQ and qq, respectively) and the random residual effects were sampled from the multivariate normal distribution with mean zero and covariance matrix (16).

$$\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}=\begin{pmatrix}0 & 0{\cdot}5 & 0{\cdot}25\\0 & 0 & -0{\cdot}5\\0 & 0 & 0\end{pmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}+\begin{bmatrix}-0{\cdot}125\\0{\cdot}5\\0{\cdot}25\end{bmatrix}Q_1+\begin{bmatrix}0{\cdot}25\\-0{\cdot}5\\-0{\cdot}125\end{bmatrix}Q_2+\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\varepsilon_3\end{bmatrix}, \tag{15}$$

$$\boldsymbol{\Psi}=\begin{pmatrix}1{\cdot}6 & 0 & 0\\0 & 1{\cdot}8 & 0\\0 & 0 & 2{\cdot}5\end{pmatrix}. \tag{16}$$

A total of 100 replicates were analysed to study the variation across different generated samples and to estimate the power of QTL detection. Rather than using subsampling which is inefficient in comparison with full sampling (MacEachern & Berliner, 1994), we used the full Gibbs sample. For each of the MCMC analyses, the first 500 samples (burn-in) were discarded and an additional 2000 Gibbs samples from which parameters of the posterior distribution were estimated. There was evidence that the M-H chain for QTL position chain was mixing slowly. We tried a different number of M-H cycles to check the convergence (Arminger & Muthén, 1998), and 25 cycles yielded satisfactory results in all simulations performed.

Single trait Bayesian analysis was also applied to the simulated data to compare the precision and efficiency with our proposed method. However, the single-trait method only estimates the total QTL effect on each trait, which is the sum of direct and indirect QTL effects. With the multi-trait SEM, the estimates of direct, indirect and total QTL effects are provided.

Models with different numbers of QTLs are compared, and the best one is selected based on the commonly used selection criterion BF. We compared the following models: (1) there is no QTL (model_0) versus there is one QTL (model_1), (2) there is one QTL (model_1) versus there are two QTLs (model_2) and (3) there are two QTLs (model_2) versus there are three QTLs (model_3). For all MCMC analyses, the same initial values and priors were used. The initial values for the QTL locations were set as 50 cM for model_1, as 49 and 74 cM for model_2, as 45, 74 and 89 cM for model_3. The prior distribution for QTL locations was uniform over the chromosome. The tuning parameter of the proposal distribution for QTL locations was chosen to be 2·0 cM. The starting values were set as

https://doi.org/10.1017/S0016672310000236 Published online by Cambridge University Press

Table 1. *BFs (using harmonic mean estimator) for multi-trait QTL-mapping model selection*

Models	Log(BF)
$Model_0$–$Model_1$	−5·79
$Model_1$–$Model_2$	−23·56
$Model_2$–$Model_3$	4·37

$Model_0$, $Model_1$, $Model_2$ and $Model_3$ are the models with zero, one, two and three QTLs, respectively. Estimates are average over 100 replicates.

Table 2. *Observed powers (%) of QTL detection of two methods obtained from 100 replicates in the simulation study*

QTL	Multi-trait SEM	Single-trait analysis			
		Y_1	Y_2	Y_3	Overall
1	99	51	87	67	93
2	100	0	92	38	96
1 and 2	99	0	80	32	91

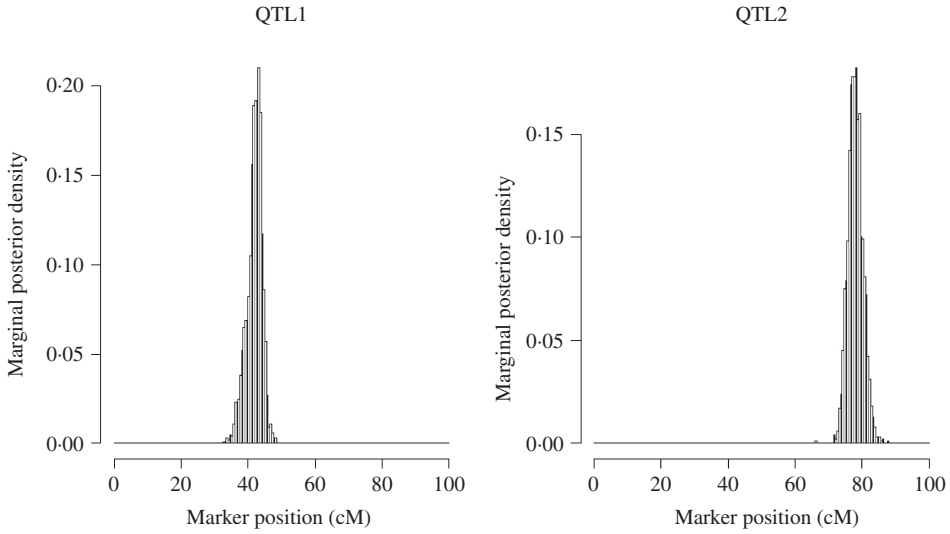

Fig. 3. Approximate posterior distribution of the QTL position in the simulation. The true number of QTL is two, located at position 42 and 78 cM.

0·1 for all regression parameters. The priors for the regression parameters were normally distributed with mean 1 and variance 100. The priors for the residual variances were inverse gamma (InvGamma(0.5,100)). All of the QTL models were fit for each of the 100 simulated data using the same chain length and burn-in.

4. Results

The estimated natural logarithm of the BFs (log(BF)s) for comparing different QTL models, averaged over 100 replicates, are given in Table 1. The average log(BF) comparing $model_0$ versus $model_1$ was −5·79 (against no QTL model 97 times out of 100 simulated datasets). The average log(BF) comparing $model_1$ versus $model_2$ was −23·56 with evidence in favour of two QTL model 99% of the times. An average log(BF) of 4·37 comparing $model_2$ versus $model_3$ favoured two QTL model 87% of the times. Therefore, it is concluded that the two QTL model was selected as the best fitting model. This conclusion is consistent with the simulated number of QTLs.

Now we restrict our attention to the two QTL model ($model_2$). The approximate posterior distributions for the QTL locations are presented in Fig. 3. The graphs are symmetric and concentrated around the true simulated values.

The statistical power was determined by the proportion of the number of replicates in which the QTL was 'detected' over the total number of replicates. A QTL was claimed as detected if there was an obvious peak around the true simulated position. The overall power of single-trait analysis was calculated as the proportion of times the QTL was detected for at least one of the three traits. The power of detecting both simulated QTLs was calculated as the proportion of the number of replicates in which both QTLs were detected over the total number of replicates. The estimated QTL detection powers over 100 replicates are given in Table 2 by multi-trait SEM and single-trait Bayesian mapping methods. The QTL detection powers of the multi-trait SEM analysis were higher than those of the single-trait analysis for both QTLs. This result likely happened, because the single-trait method only estimates the total QTL effects which

https://doi.org/10.1017/S0016672310000236 Published online by Cambridge University Press

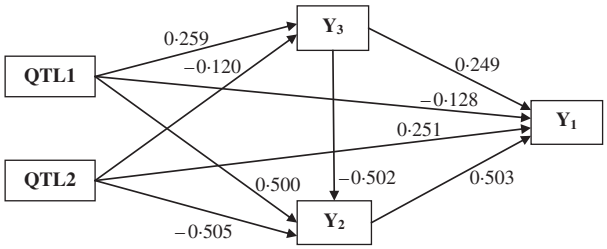

Fig. 4. Path model of multi-trait SEM in the simulation. Single arrows indicate causal relationships. Numbers by the arrow lines represent the Bayesian estimates of the path coefficients.

may be reduced, due to the compensating direct and indirect QTL effects. For instance, the QTL1 had a larger positive direct effect on Y_2 but a negative indirect effect, which in turn reduces the total QTL effect on Y_2. The relatively small total QTL effects associated with Y_1, Y_2 and Y_3 may not be detected using single trait analysis. Based on the power analysis, the multi-trait SEM improved power of detecting individual QTL and both QTLs over the single trait analysis. However, the power of QTL detection may not necessarily always be higher for the multi-trait SEM than that for the single-trait analysis. If the direct and indirect QTL effects are in the same direction for all traits, the total QTL effect will be larger than the direct effects tested with the multi-trait SEM approach. In this case, the power of the multi-trait analysis may be less than the overall power of single-trait analysis since the single-trait method tests the total QTL effect. However, SEM can estimate indirect effects of any path which is not possible with the single-trait approach.

Figure 4 shows the Bayesian estimates of path coefficients with multi-trait SEM at the positions where the QTLs were detected. The estimated path coefficients are very close to the true simulated values. Table 3 presents the summary results of Bayesian parameter estimates in the simulation using our proposed method and the single-trait analysis. The means and standard deviations of the posterior distributions of the individual parameters are the averages over the 100 Monte Carlo replications. The estimates of QTL positions and effects by the multi-trait SEM method are very close to the true simulated values with small standard errors. However, the single-trait analysis provided estimates that appear to be biased with much higher standard deviations when compared with multi-trait SEM. The differences between two methods are especially large for QTLs with small total effects. Thus, multi-trait SEM method is more accurate and precise than single-trait analysis on the estimates of QTL positions and effects. Another obvious difference between our proposed method and the single-trait analysis is that multi-trait SEM allows one to fit a more complex and biologically sensible model, which provides the estimation of direct and indirect QTL effects, and therefore important insight giving a richer understanding of the nature of QTLs affecting traits compared with the single-trait analysis.

5. Application to recombinant inbred chromosome line (RICL) wheat experiment

We illustrate our Bayesian multi-trait SEM approach with an analysis of data from a RICLs wheat experiment, which contains a population of 98 RICLs-3A derived from a cross between 'Cheyenne' (CNN) and CNN with a 'Wichita' 3A chromosome substitution (CNN(WI3A)) and thus, the lines differed only in the which portion of 'Wichita' chromosome 3A was contained in each line. This population was evaluated in multi-environment field trials from 1999 to 2001 to identify QTL and QTL-by-environment interactions for GRYL and other agronomic traits in seven environments. Details of the experiment and results of the data analysis performed by univariate QTL detection techniques have been described by Campbell *et al*. (2003). These data were also analysed for genotype-by-environment interaction using a least squares SEM (Dhungana *et al*., 2007).

In this study, we focused on GYLD and yield component traits (TKWT, SPSM and KPS), since the causal relationships among these traits (Fig. 1) are well established (Dofing & Knight, 1992). To illustrate MCMC, we only considered 10 molecular markers covering 71·7 cM of the chromosome 3A in which two QTL regions were detected by Campbell *et al*. (2003). Prior to the analysis, analysis of variance (ANOVA) was performed for each trait to remove the main effects of environments and blocks. Residuals of the four traits were standardized to mean 0 and variance 1, and then were used as observed trait values.

We evaluated the one QTL model against the two QTL model and the two QTL model against the three QTL model using the BF. The prior distributions for the locations and additive effects of QTL, and other path coefficients were set to be the same as those in the simulation. The length of the Markov chain was also set to be the same as that in the simulation. The estimate of log(BF) of comparing one versus two QTL models was −39·15. Comparing the two versus three QTL models gave a log(BF) of 11·44, providing decisive evidence in favour of the two QTL model. Thus, the two QTL model was selected as the best-fitting model.

Figure 5 shows the marginal posterior probability distributions of the QTL locations obtained from the MCMC of the two QTL model. The first QTL is estimated at 5·31 cM (between *Xbcd907* and *Xtam055*), affecting TKWT, KPS and SPSM directly. The second one is estimated at 56·1 cM (between markers

https://doi.org/10.1017/S0016672310000236 Published online by Cambridge University Press

Table 3. *Bayesian estimates of QTL positions and additive effects in the simulation by multi-trait SEM and single-trait analysis*

Methods	QTL	Trait	Position (cM)	Putative QTL effect		
				Total	Direct	Indirect
Parameters	1	Y_1	42	0·125	−0·125	0·25
		Y_2		0·375	0·5	−0·125
		Y_3		0·25	0·25	0
	2	Y_1	78	0	0·250	−0·250
		Y_2		−0·435	−0·50	0·065
		Y_3		−0·125	−0·125	0
Multi-trait SEM	1	Y_1	42·09 (3·48)	0·122 (0·094)	−0·128 (0·086)	0·250 (0·059)
		Y_2		0·369 (0·096)	0·500 (0·088)	−0·131 (0·051)
		Y_3		0·259 (0·090)	0·259 (0·090)	0
	2	Y_1	78·03 (2·23)	−0·003 (0·085)	0·251 (0·092)	−0·254 (0·059)
		Y_2		−0·443 (0·101)	−0·505 (0·088)	0·062 (0·052)
		Y_3		−0·120 (0·101)	−0·120 (0·101)	0
Single-trait	1	Y_1	35·40 (10·30)	0·072 (0·134)		
		Y_2	42·25 (6·53)	0·354 (0·158)		
		Y_3	35·39 (9·32)	0·129 (0·219)		
	2	Y_1	59·11 (9·95)	0·046 (0·126)		
		Y_2	77·59 (5·10)	−0·432 (0·146)		
		Y_3	61·36 (10·93)	0·070 (0·234)		

Estimates are means over 100 replications with standard deviation in parentheses. Sample size = 250, Gibbs: Cycles = 2000, Burn in = 500 Metropolis: Cycles = 25.

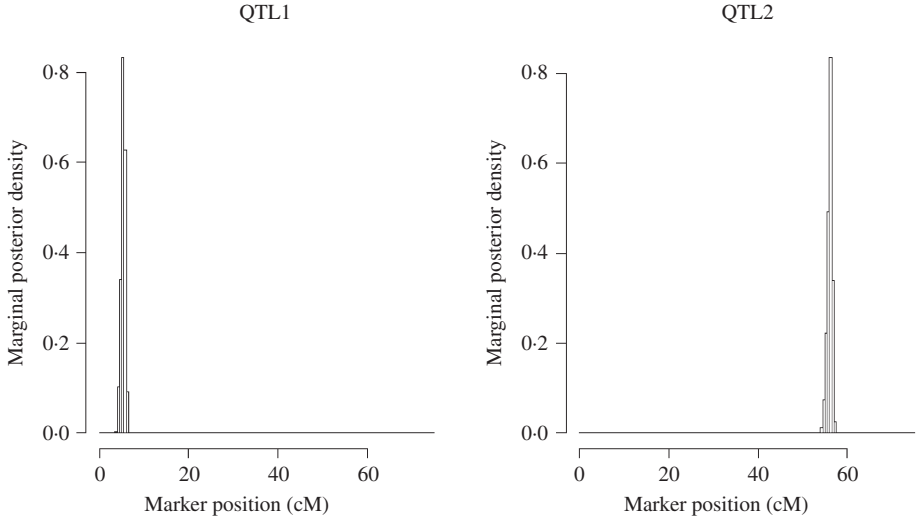

Fig. 5. Two QTL model. Approximate posterior distribution of two QTL locations based on joint analysis (multi-trait SEM) for GRYL and yield components on chromosome 3A of wheat.

XkusA6 and *Xbcd366*) affecting GYLD, KPS and SPSM directly.

Figure 6 shows the standardized path coefficients with multi-trait SEM located at two QTL locations. Table 4 shows the estimated posterior means and posterior standard deviations for the direct, indirect and total QTL effects of the identified QTLs. The QTL detected at position 5·3086 cM (close to *Xbarc12*) has a small positive direct effect (not significant) and a significant positive indirect effect on trait GYLD resulting in a large significant total QTL effect on trait GYLD. The direct and indirect QTL effects on TKWT are both negative ($P<0·001$) resulting in a large negative total effect ($P<0·001$), which significantly decreases TKWT. The QTL has a large positive direct effect on KPS ($P<0·001$) and a negative indirect effect ($P<0·001$) resulting in a smaller absolute total effect. The second QTL detected at position 56·01 cM (close to *Xbarc67*) affects all traits directly or indirectly. The direct and indirect QTL effects on GYLD are both positive ($P<0·001$) leading to a large total effect ($P<0·001$), which significantly increases

https://doi.org/10.1017/S0016672310000236 Published online by Cambridge University Press

Table 4. *Bayesian estimates of the chromosome 3A QTL locations and effects using multi-trait SEM*

Trait	QTL at position (cM)	Putative QTL effect		
		Total	Direct	Indirect
GYLD	5·31 (0·4481)	0·0715 (0·0195)	0·0135 (0·0101)	0·0580 (0·0170)
TKWT		−0·1144 (0·0192)	−0·0789 (0·0177)	−0·0355 (0·0083)
KPS		0·0821 (0·0200)	0·1121 (0·0155)	−0·0300 (0·0126)
SPSM		0·0475 (0·0199)	0·0475 (0·0199)	0·0000
GYLD	56·01 (0·5429)	0·1649 (0·0196)	0·0526 (0·0098)	0·1123 (0·0171)
TKWT		−0·0247 (0·0202)	0·0245 (0·0187)	−0·0492 (0·0082)
KPS		0·0052 (0·0200)	0·0700 (0·0157)	−0·0648 (0·0127)
SPSM		0·1024 (0·0198)	0·1024 (0·0198)	0·0000

Values in parentheses are respective SD values.

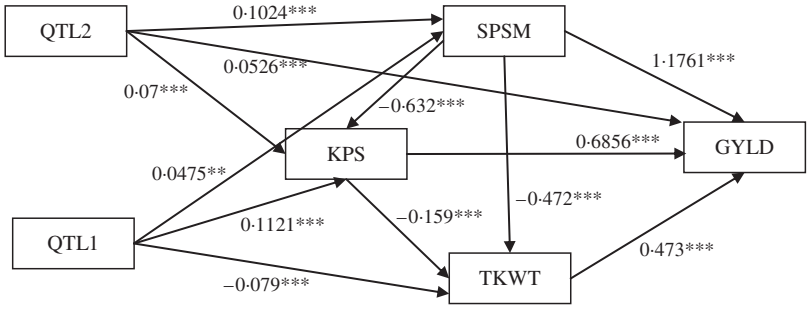

Fig. 6. Path estimates of multi-trait SEM at positions 5·3086 cM (QTL1, close to *Xbarc12*) and 56·005 cM (QTL2, close to *Xbarc67*) on chromosome 3A of wheat. Single arrows indicate causal relationships. Numbers by the arrow lines represent the estimated standardized coefficients with significance level: ***$P<0·001$, **$P<0·01$ and *$P<0·05$.

GYLD. The QTL has significant negative indirect effects on both TKWT and KPS, but positive direct effects leading to non-significant total effects, since the direct and indirect effects are cancelled out. In contrast, in their combined analysis Campbell *et al.* (2003) were not able to detect QTLs for SPSM in either region using univariate QTL detection techniques which only captured the total QTL effects. In addition, they detected a minor QTL for GYLD in region one, while the corresponding QTL that we detected had a greater effect on GYLD. Thus, by considering the GRYL and yield components together in a multi-trait analysis, we not only improved the ability to detect QTLs for GYLD and SPSM but also gained insight into the processes of how the QTLs on chromosome 3A affected agronomic performance directly and indirectly. Understanding the genetic control of GRYL using a biologically relevant yield component framework provides useful information for plant breeders interested in breaking unfavourable indirect QTL effects and for better understanding complex traits.

6. Discussion and conclusion

Research on QTL studies often provides information on multiple complex traits that have well-established causal relationships. However, there has been a lack of a comprehensive multivariate multiple QTL-mapping technique which is capable of incorporating the causal structure among multiple traits. Consequently the genetic functions may not be fully understood. In this study, we have presented a Bayesian approach to multiple traits QTL mapping using an SEM, taking into account the causal relationships among multiple traits. We have explored some aspects of multiple traits QTL mapping that have not been done in previous studies. In particular, it allows one to fit a more complex and biologically sensible model; it provides the estimates of the total, direct and indirect QTL effects and ultimately allows for important insights into how QTLs regulate multiple complex traits. Knowledge of the direct and indirect QTL effects can be very important for plant breeders interested in finding modifier genes to overcome the pleiotropism.

Using our proposed multi-trait SEM, we were able to improve the power of QTL detection and precision of parameter estimates compared with the single-trait analysis. However, the power of the QTL detection may not necessarily always be higher for the multi-trait SEM. It depends on the magnitude of QTL effects and the directions of causal relationship among multiple traits. The power of the multi-trait SEM can

https://doi.org/10.1017/S0016672310000236 Published online by Cambridge University Press

increase significantly if the direct and indirect QTL effects on the trait are relatively large and in the opposite direction, which reduces the total QTL effect. In this case, the reduced total QTL may not be detected by the single-trait analysis since the single-trait method tests only the total QTL effect. If QTL influences traits that are not causally explained by any other intermediary trait, the power of multi-trait SEM is very close to that of the single-trait analysis since the direct and total QTL effects are the same. If the direct and indirect QTL effects are in the same direction for the trait, the total QTL effect will be larger than the direct effects tested with the multi-trait SEM approach. In this case, the power of the multi-trait SEM analysis may be less than the overall power of single-trait analysis.

We also applied Bayesian multi-trait SEM to the RICLs-3A wheat experiment data. As expected, we detected QTLs for SPSM in either region which have not been reported in Campbell *et al*. (2003), where univariate QTL detection techniques were used. In addition, they detected a minor QTL for GYLD in region one, while the corresponding QTL that we detected had a greater effect on GYLD.

A prerequisite of the proposed method is prior biological knowledge of the causal relationships among the multiple traits, since SEM is generally used as a confirmatory rather than exploratory procedure. Theoretical insight and judgment by the researcher is very important in building a correct model. In practice, one can obtain some basic background about the key structure of the model either from knowledge of the related field or from preliminary data analyses. Other applications likely may require more model development based on procedures described elsewhere (see Bollen, 1989).

The model considered in this paper was illustrated using an RIL-simulated population to provide a general idea of the nature of QTLs affecting the traits, and did not include epistatic genetic effects. However, the general approach can be easily applied to different population structures (such as F2 and backcross), and genetic models by setting up the corresponding conditional QTL genotype probability given QTL locations. Here, we assumed a pleiotropic QTL model (each QTL affects all traits). It is important to separate pleiotropic effects against closely linked QTL. We plan to use BF to test pleiotropic effects against closely linked QTL in the future. In this study, we assumed complete phenotypic data. However, we acknowledged that a large amount of missing phenotypic data may reduce the power of QTL detection and precision of QTL location and effect estimation in joint analyses (Fridley & de Andrade, 2008; Guo & Nelson, 2008). Our proposed method can be extended to deal with missing phenotypic data by using multiple imputations. The proposed model here did not account for the experimental design issue, thus ignoring non-genetic sources of variation such as environments, blocks or gene–environment interactions. Methods incorporating these innovations could result in increased statistical power of QTL detection, precision in estimation of QTL effects and position and an improved understanding of how QTL interact with environmental factors. In addition, researchers may collect data of different types for a sample set (e.g. both binary and continuous traits). Methods that are capable of dealing with a mixture of continuous and binary traits could be valuable in a variety of situations.

Programs were written in SAS PROC IML and are available by sending email to xjmixu@yahoo.com

References

Arminger, G. & Muthén, B. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis–Hastings algorithm. *Psychometrika* **6**, 271–300.

Banerjee, S., Yandell, B. S. & Yi, N. (2008). Bayesian quantitative trait loci mapping for multiple traits. *Genetics* **179**, 2275–2289.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley Interscience.

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

Campbell, B. T., Baenziger, P. S., Gill, K. S., Eskridge, K. M., Budak, H., Erayman, M. & Yen, Y. (2003). Identification of QTLs and environmental interactions associated with agronomic traits on chromosome 3A of wheat. *Crop Science* **43**, 1493–1505.

Dhungana, P., Eskridge, K. M., Baenziger, P. S., Campbell, B. T., Gill, K. S. & Dweikat, I. (2007). Analysis of genotype-by-environment interaction in wheat using a structural equation model and chromosome substitution lines. *Crop Science* **47**, 477–484.

Dofing, S. M. & Knight, C. W. (1992). Alternative model for path analysis of small-grain yield. *Crop Science* **32**, 487–489.

Dunson, D. B. (2001). Commentary: practical advantages of Bayesian analysis of epidemiologic data. *American Journal of Epidemiology* **153**, 1222–1226.

Fridley, B. L. & de Andrade, M. (2008). Missing phenotype data imputation in pedigree data analysis. *Genetic Epidemiology* **32**, 52–60.

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

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.

Guo, Z. & Nelson, J. C. (2008). Multiple-trait quantitative trait locus mapping with incomplete phenotypic data. *BMC Genetics* **9**, 82.

Hackett, C. A., Meyer, R. C. & Thomas, W. T. B. (2001). Multi-trait QTL mapping in barley using multivariate regression. *Genetic Research* **77**, 95–106.

Haley, C. S. & Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

Jansen, R. C. & Stam, P. (1994). High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**, 1447–1455.

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

Jeffreys, H. (1961). *Theory of Probability*, 3rd edn. Oxford, UK: Oxford University Press.

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

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795.

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

Lander, E. S. & Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.

Li, R., Tsaih, S. W., Shockley, K., Stylianou, I. M., Wergedahl, J., Paigen, B. & Churchill, G. A. (2006). Structural model analysis of multiple quantitative traits. *PLoS Genetics* **2**, e114.

Liu, J., Liu, Y., Liu, X. & Deng, H.-W. (2007). Bayesian mapping of quantitative trait loci for multiple complex traits with the use of variance components. *American Journal of Human Genetics* **81**, 304–320.

MacEachern, S. N. & Berliner, M. L. (1994). Subsampling the Gibbs sampler. *The American Statistician* **48**, 188–190.

Mähler, M., Most, C., Schmidtke, S., Sundberg, J. P., Li, R., Hedrich, H. J. & Churchill, G. A. (2002). Genetics of colitis susceptibility in IL-10-deficient mice: backcross versus F2 results contrasted by principal component analysis. *Genomics* **80**, 274–282.

Mangin, B., Thoquet, P. & Grimsley, N. (1998). Pleiotropic QTL analysis. *Biometrics* **54**, 88–99.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1092.

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.

Nadeau, J. H., Burrage, L. C., Restivo, J., Pao, Y. H., Churchill, G. & Hoit, B. D. (2003). Pleiotropy, homeostasis, and functional networks based on assays of cardiovascular traits in genetically randomized populations. *Genome Research* **13**, 2082–2091.

Narita, A. & Sasaki, Y. (2004). Detection of multiple QTL with epistatic effects under a mixed inheritance model in an outbred population. *Genetics, Selection, Evolution* **36**, 415–433.

Newton, M. A. & Raftery, A. E. (1994). Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *Journal of the Royal Statistical Society: Series B* **56**, 3–48.

Satagopan, J. M., Yandell, B. S., Newton, M. A. & Osborn, T. C. (1996). A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics* **144**, 805–816.

Sillanpaa, M. J. & Arjas, E. (1998). Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics* **148**, 1373–1388.

Sillanpaa, M. J. & Arjas, E. (1999). Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. *Genetics* **151**, 1605–1619.

Stephens, D. A. & Fisch, R. D. (1998). Bayesian analysis of quantitative trait locus data using reversible jump Markov chain Monte Carlo. *Biometrics* **54**, 1334–1347.

Wang, H., Zhang, Y. M., Li, X., Masinde, G. L., Mohan, S., Baylink, D. J. & Xu, S. (2005). Bayesian shrinkage estimation of quantitative trait loci parameters. *Genetics* **170**, 465–480.

Weller, J. I., Wiggans, G. R., Van Raden, 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.

Williams, J. T., Van Eerdewegh, P., Almasy, L. & Blangero, J. (1999). Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. I. Likelihood formulation and simulation results. *American Journal of Human Genetics* **65**, 1134–1147.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research* **20**, 557–585.

Xu, C., Li, Z. & Xu, S. (2005). Joint mapping of quantitative trait loci for multiple binary characters. *Genetics* **169**, 1045–1059.

Yang, R. & Xu, S. (2007). Bayesian shrinkage analysis of quantitative trait loci for dynamic traits. *Genetics* **176**, 1169–1185.

Yi, N. (2004). A unified markov chain monte carlo framework for mapping multiple quantitative trait loci. *Genetics* **167**, 967–975.

Yi, N. & Shriner, D. (2008). Advantages in Bayesian multiple quantitative trait loci mapping in experimental crosses. *Heredity* **100**, 240–252.

Yi, N. & Xu, S. (2002). Mapping quantitative trait loci with epistatic effects. *Genetical Research, Cambridge* **79**, 185–198.

Yi, N., Xu, S. & Allison, D. B. (2003). Bayesian model choice and search strategies for mapping interacting quantitative trait loci. *Genetics* **165**, 867–883.

Yi, N., Yandell, B. S., Churchill, G. A., Allison, D. B., Eisen, E. J. & Pomp, D. (2005). Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics* **170**, 1333–1344.

Yi, N., Shriner, D., Banerjee, S., Mehta, T., Pomp, D. & Yandell, B. S. (2007). An efficient Bayesian model selection approach for interacting QTL models with many effects. *Genetics* **176**, 1865–1877.

Zeng, Z.-B. (1994). Precision mapping of quantitative trait loci. *Genetics* **136**, 1457–1468.

Zhu, W. S. & Zhang, H. P. (2009). Why do we test multiple traits in genetic association studies? *Journal of the Korean Statistical Society* **38**, 1–10.