Journal of Genetics, Vol. 97, No. 1, March 2018, pp. 267–274
https://doi.org/10.1007/s12041-018-0913-0

RESEARCH ARTICLE

Simultaneous estimation of QTL parameters for mapping multiple traits

LIANG TONG[1,2], XIAOXIA SUN[2] and YING ZHOU[1]*

[1] *School of Mathematical Sciences, Heilongjiang University, Harbin 150080, People's Republic of China*
[2] *School of Information Engineering, Suihua University, Suihua 152061, People's Republic of China*
*For correspondence. E-mail: yzhou@aliyun.com.

Received 11 January 2017; revised 6 June 2017; accepted 21 August 2017; published online 13 March 2018

Abstract. The analysis of quantitative trait loci (QTLs) aims at mapping and estimating the positions and effects of the genes that may affect the quantitative trait, and evaluating the relationship between the gene variation and the phenotype. In existing studies, most methods mainly focus on the association/linkage between multiple gene loci and one trait, in which some useful joint information of multiple traits may be ignored. In this paper, we proposed a method of simultaneously estimating all QTL parameters in the framework of multiple-trait multiple-interval mapping. Simulation results show that in accuracy aspect, the proposed method outperforms an existing method for mapping multiple traits. A real example is also provided to validate the performance of the new method.

Keywords. EM algorithm; estimation; multiple-interval mapping; recombination rate.

Introduction

Currently, statistical methods of gene mapping can be divided into two kinds, i.e. association analysis and linkage analysis. Although many researchers prefer to employ the former in their gene mapping, the latter still has its attractive advantages. Among the methods of linkage analysis, the interval mapping (IM) provides a good beginning, which uses multiple markers on the genome to infer the latent quantitative trait loci (QTLs) (Lander and Botstein 1989), and the method can well estimate the additive and dominant effects of QTLs. By combining multivariate regressions with interval mapping, Zeng (1994) proposed a composite interval mapping method (CIM), which can control the background effects through fitting QTLs located outside a tested interval in the statistical model. The method further improved the precision of QTL mapping. To take advantage of information of more markers, a multiple-interval mapping (MIM) method was proposed by Kao *et al.* (1999), and this method further improved upon the CIM by simultaneously fitting multiple putative QTLs in multiple marker intervals via an EM algorithm (Dempster *et al.* 1977).

The above methods mainly focus on mapping one trait controlled by multiple QTLs, however, in practical investigations, multiple traits are often recorded. It is worth mentioning that some research results show that testing for multiple traits together is more powerful than testing for a single trait at a time in the gene-mapping studies (Zhu and Zhang 2009; Joehanes 2009; Da Costa E Silva *et al.* 2012). Currently, the idea of simultaneously considering multiple traits is widely used in gene-mapping test. For instance, Jiang and Zeng (1995) proposed a multiple-trait CIM method, which exploited the correlation structure of the considered traits to improve the accuracy of QTL detection. Hereafter, several MIM methods based on multivariate regressions were proposed (Korol *et al.* 1998; Calinski *et al.* 2000; Knott and Haley 2000; Hackett *et al.* 2001). Xu *et al.* (2005) considered the joint mapping of multiple binary traits. Guo and Nelson (2008) discussed multiple-trait QTL mapping with incomplete phenotypic data. Besides, Bayesian mapping of QTL for multiple traits was also considered by some researchers (Liu *et al.* 2007; Banerjee *et al.* 2008). Based on the MIM method for single trait, Joehanes (2009) similarly developed the multiple-trait method via the EM algorithm. Da Costa E Silva *et al.* (2012) also provided a theoretical version of multiple trait multiple-interval mapping (MTMIM) of QTLs for inbred line crosses, and developed a new method for estimating genomewide

significance level of putative QTL effects for the MTMIM model.

In the QTL mapping for multiple traits, each putative QTL may exhibit additive and/or dominant QTL effects to each trait. Comparing with the single-trait methods, multiple-trait methods can improve the detecting power and mapping precision. Under the condition when multiple traits exist, the single-trait methods may ignore the correlation relationship among multiple traits and the common function of one QTL to multiple traits, and therefore, the corresponding detection power of QTLs for multiple traits will lose in some certain degree (Evans 2002; Neuschl *et al.* 2007; Thomasen *et al.* 2008; Malosetti *et al.* 2008). On the other hand, in the current existing MTMIM methods, the parameters of QTL positions cannot be estimated simultaneously by a composite estimating algorithm.

To overcome this difficulty in estimating all QTL parameters and improve the detection power over that of single-trait methods, in this paper, we extend the MTMIM developed in Joehanes (2009) and realize the simultaneous estimation of all QTL parameters under the considered statistical model. In the proposed algorithm, the QTL positions have closed estimating formulae, and therefore, the estimating precision shows apparent improvement. Our simulation results present the advantages of the new method in estimating parameters over an existing method. We also apply the new method to real data set.

Theory and method

We consider data collected from n subjects of F_2 intercross on t quantitative traits and q marker intervals divided by $q + 1$ genetic marker loci. Let

$$\mathbf{Y}_j = (Y_{j1}, \ldots, Y_{jt}),$$
$$\mathbf{X}_j = (X_{j1}, \ldots, X_{j(q+1)}), \mathbf{X}_j^* = (X_{j1}^*, \ldots, X_{jq}^*),$$

where $Y_{ji} (j = 1, \ldots, n, i = 1, \ldots, t)$ denotes the ith trait value of the jth subject, X_{ji} $(j = 1, \ldots, n, i = 1, \ldots, q + 1)$ denotes the ith marker genotype of the jth subject, and $X_{ji}^* (j = 1, \ldots, n, i = 1, \ldots, q)$ is the genotype of the latent QTL in the ith marker interval of the jth subject. Let γ_i and γ_{i1}, respectively, denote recombination rate of the ith marker interval, which is known and the recombination rate between the ith marker and the latent QTL in the interval. In the following analysis, at most, one QTL in a marker interval is assumed. Let $p(X_{ji}^*|X_{ji}^M)$ denotes the conditional probability of the QTL genotype $X_{ji}^* (Q_iQ_i, Q_iq_i$ or $q_iq_i)$, given the genotype combination X_{ji}^M of the ith marker interval of the jth individual. For intercross families, X_{ji}^M has nine possible values, which are coded as 1, 2, 3, …, 9, respectively. The conditional probabilities $p(X_{ji}^*|X_{ji}^M)$ are presented in table 1.

Table 1. The conditional probabilities of QTL genotype given the marker genotypes.

Code	Marker genotype	QTL genotype		
		Q_iQ_i	Q_iq_i	q_iq_i
1	$M_iM_iM_{i+1}M_{i+1}$	1	0	0
2	$M_iM_iM_{i+1}m_{i+1}$	$1 - r_i$	r_i	0
3	$M_iM_im_{i+1}m_{i+1}$	$(1 - r_i)^2$	$2r_i(1 - r_i)$	r_i^2
4	$M_im_iM_{i+1}M_{i+1}$	r_i	$1-r_i$	0
5	$M_im_iM_{i+1}m_{i+1}$	0	1	0
6	$M_im_im_{i+1}m_{i+1}$	0	$1 - r_i$	r_i
7	$m_im_iM_{i+1}M_{i+1}$	r_i^2	$2r_i(1 - r_i)$	$(1 - r_i)^2$
8	$m_im_iM_{i+1}m_{i+1}$	0	r_i	$1-r_i$
9	$m_im_im_{i+1}m_{i+1}$	0	0	1

$r_i = \gamma_{i1}/\gamma_i$.

Statistical model

In the framework of MIM, the following statistical model for multiple traits is considered (Joehanes 2009)

$$\underset{n\times t}{\mathbf{Y}} = \sum_{i=1}^{q} \left[\underset{n\times 11\times t}{\boldsymbol{\xi}_\mathbf{i}\mathbf{a}_\mathbf{i}} + \underset{n\times 11\times t}{\boldsymbol{\eta}_\mathbf{i}\mathbf{d}_\mathbf{i}} \right] + \underset{n\times(p+1)}{\mathbf{X}} \underset{(p+1)\times t}{\mathbf{B}} + \underset{n\times t}{\mathbf{E}}, \tag{1}$$

where $\underset{n\times t}{\mathbf{Y}}$ represents the matrix of trait values, $\underset{n\times(p+1)}{\mathbf{X}}$ represents the genotype matrix, which may contain non-genetic factors. $\underset{n\times 1}{\boldsymbol{\xi}_\mathbf{i}}$ and $\underset{n\times 1}{\boldsymbol{\eta}_\mathbf{i}}$, respectively, denote the genotype indicator vector for the ith QTL genotype of n subjects, i.e.,

$$\xi_{ji} = 1, \quad \eta_{ji} = -\frac{1}{2}, \text{ if } X_{ji}^* = Q_iQ_i;$$
$$\xi_{ji} = 0, \quad \eta_{ji} = \frac{1}{2}, \text{ if } X_{ji}^* = Q_iq_i;$$
$$\xi_{ji} = -1, \quad \eta_{ji} = -\frac{1}{2}, \text{ otherwise}.$$

$\mathbf{a_i} = (a_{i1}, a_{i2}\ldots a_{it})$ and $\mathbf{d_i} = (d_{i1}, d_{i2}\ldots d_{it})$, respectively, represent the additive effect vector and dominant effect vector of the ith QTL to t traits, and $\underset{(p+1)\times t}{\mathbf{B}}$ is the genotype effect matrix of markers. Further, let the QTL effect matrix

$$\mathbf{C} = \begin{pmatrix} a_{11}, & a_{12}, \ldots, a_{1t} \\ d_{11}, & d_{12}, \ldots, d_{1t} \\ \cdots & \cdots \quad \cdots \\ a_{q1}, & a_{q2}, \ldots, a_{qt} \\ d_{q1}, & d_{q2}, \ldots, d_{qt} \end{pmatrix}.$$

$\mathbf{E} = \{e_{ji}\}_{n\times t}$ represents the residual matrix, where e_{ji} is the random error of the ith trait value of the jth subject, with mean zero and $\text{cov}(e_{ji}, e_{jl}) = \sigma_{il} = \rho\sigma_i\sigma_l, i, l = 1, \ldots, t$.

For convenience, let the covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12}\cdots\sigma_{1t} \\ \sigma_{21} & \sigma_2^2\cdots\sigma_{2t} \\ \vdots & \vdots \quad \ddots \vdots \\ \sigma_{t1} & \sigma_{t2}\cdots\sigma_t^2 \end{pmatrix}.$$

Based on the famous EM algorithm (Dempster *et al.* 1977), the details of our proposed method to infer the parameter matrix $\boldsymbol{\Omega} = (\mathbf{B}, \mathrm{C}, \Sigma, \gamma)$ in the simultaneous MIM for multiple traits can be described as follows.

Likelihood function

The complete likelihood function about $\boldsymbol{\Omega}$ for the considered data is given by

$$\begin{aligned} L = L(\boldsymbol{\Omega}|\mathbf{Y}, \mathbf{X}, \mathbf{X}^*) &= \prod_{j=1}^{n} P(Y_j, X_j, X_j^*|\Omega) \\ &= \prod_{j=1}^{n} P(Y_j|X_j^*, \Omega) \cdot P(X_j^*|X_j, \Omega), \end{aligned}$$

so, the complete log-likelihood function

$$\begin{aligned} l(\Omega) &= \log \prod_{j=1}^{n} P(Y_j|X_j^*, \Omega) \cdot P(X_j^*|X_j, \Omega) \\ &= \sum_{j=1}^{n} [\log P(Y_j|X_j^*, \Omega) + \log P(X_j^*|X_j, \Omega)]. \end{aligned} \tag{2}$$

Full EM algorithm for estimating all the parameters

E-step: Given $\mathbf{X}$, $\mathbf{Y}$, $\boldsymbol{\Omega}^{(s)}$, compute the conditional expectation $Q(\boldsymbol{\Omega}|\mathbf{X}, \mathbf{Y}, \boldsymbol{\Omega}^{(s)})$ of $l(\Omega)$.

$$\begin{aligned} Q(\Omega|\mathbf{X}, \mathbf{Y}, \boldsymbol{\Omega}^{(s)}) &= \sum_{j=1}^{n} E[\log P(Y_j|X_j^*, \boldsymbol{\Omega}) \\ &\quad + \log P(X_j^*|X_j, \boldsymbol{\Omega})|\mathbf{X}, \mathbf{Y}, \boldsymbol{\Omega}^{(s)}] \\ &= \sum_{j=1}^{n} E_{X^*}[\log P(Y_j|X_j^*, \boldsymbol{\Omega}) \\ &\quad + \log P(X_j^*|X_j, \boldsymbol{\Omega})|\mathbf{X}, \mathbf{Y}, \boldsymbol{\Omega}^{(s)}] \\ &= \sum_{j=1}^{n} \sum_{X_j^*} \log P(Y_j|X_j^*, \boldsymbol{\Omega}) \cdot \omega_{X_j^*}^{(s)} \\ &\quad + \sum_{j=1}^{n} \sum_{X_j^*} \log P(X_j^*|X_j, \boldsymbol{\Omega}) \cdot \omega_{X_j^*}^{(s)}, \end{aligned}$$

where $\omega_{X_j^*}^{(s)} = P(X_j^* = x_j^*|X_j, Y_j, \boldsymbol{\Omega}^{(s)})$, and further, let

$$\begin{aligned} \omega_{X_{jk}^*}^{(s)} &= P(X_j^* = x_j^{*k}|X_j, Y_j, \boldsymbol{\Omega}^{(s)}) \\ &= \frac{P(Y_j|x_j^{*k}, \boldsymbol{\Omega}^{(s)}) \cdot P(x_j^{*k}|X_j, \boldsymbol{\Omega}^{(s)})}{\sum_{k=1}^{3^q} P(Y_j|x_j^{*k}, \boldsymbol{\Omega}^{(s)}) \cdot P(x_j^{*k}|X_j, \boldsymbol{\Omega}^{(s)})}, \\ & k \in \{1, \ldots, 3^q\}, \end{aligned}$$

denote the conditional probability of the kth value of X_j^* given X_j, Y_j, $\boldsymbol{\Omega}^{(s)}$. For the multiple quantitative traits that we considered, $P(Y_j|x_j^{*k}, \boldsymbol{\Omega}^{(s)}) = f(Y_j; \mu_{jk}, \boldsymbol{\Sigma})$, i.e., a multivariate normal density function with mean μ_{jk} and covariance matrix $\boldsymbol{\Sigma}$. $P(x_j^{*k}|X_j, \boldsymbol{\Omega}^{(s)})$ can be exactly expressed as the product of q conditional probabilities that listed in table 1, because the genotype of each QTL is conditionally independent when the marker genotypes are given, and thus, $P(x_j^{*k}|X_j, \boldsymbol{\Omega}^{(s)})$ is the function of recombination rates $\gamma_{11}, \gamma_{21}, \ldots, \gamma_{q1}$.

M-step: Maximize the Q-function $Q(\boldsymbol{\Omega}|\mathbf{X}, \mathbf{Y}, \boldsymbol{\Omega}^{(s)})$ to obtain $\boldsymbol{\Omega}^{(s+1)}$.

Updating B, C and Σ

After necessary derivation, the $(s+1)$th step iterative formula of QTL effect matrix C is given by

$$\mathrm{C}^{(s+1)} = R^{(s)} - M^{(s)}\mathrm{C}^{(s)},$$

where the formulae of $R^{(s)}$ and $M^{(s)}$ can be expressed as

$$R^{(s)} = \begin{bmatrix} \frac{D_1' W'^{(s)}(\mathbf{Y} - \mathbf{X}\mathbf{B}^{(s)})}{1' W^{(s)}(D_1 \# D_1)} \\ \vdots \\ \frac{D_{cq}' W'^{(s)}(\mathbf{Y} - \mathbf{X}\mathbf{B}^{(s)})}{1' W^{(s)}(D_{cq} \# D_{cq})} \end{bmatrix}_{cq \times t}, \quad M^{(s)} = \begin{bmatrix} 0 & \frac{1' W^{(s)}(D_1 \# D_2)}{1' W^{(s)}(D_1 \# D_1)} & \cdots & \frac{1' W^{(s)}(D_1 \# D_{cq})}{1' W^{(s)}(D_1 \# D_1)} \\ \frac{1' W^{(s)}(D_2 \# D_1)}{1' W^{(s)}(D_2 \# D_2)} & 0 & \cdots & \frac{1' W^{(s)}(D_2 \# D_{cq})}{1' W^{(s)}(D_2 \# D_2)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1' W^{(s)}(D_{cq} \# D_1)}{1' W^{(s)}(D_{cq} \# D_{cq})} & \frac{1' W^{(s)}(D_{cq} \# D_2)}{1' W^{(s)}(D_{cq} \# D_{cq})} & \cdots & 0 \end{bmatrix}_{cq \times cq},$$

where # denotes the Hadamard (componentwise) product of two vectors. The iterative formula of $\mathbf{B}$ and $\boldsymbol{\Sigma}$ are respectively,

$$\begin{aligned} \mathbf{B}^{(s+1)} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{Y} - W^{(s)} D \mathrm{C}^{(s+1)}], \\ \boldsymbol{\Sigma}^{(s+1)} &= \frac{1}{n}(\mathbf{Y} - \mathbf{X}\mathbf{B}^{(s+1)} - W^{(s)} D \mathrm{C}^{(s+1)})' \\ &\quad \cdot (\mathbf{Y} - \mathbf{X}\mathbf{B}^{(s+1)} - W^{(s)} D \mathrm{C}^{(s+1)}), \end{aligned}$$

where $W = [\omega_{X^*_{ji}}]$ is a $n \times 3^q$ matrix, and $D = (D_1, D_2, D_3, \ldots, D_{cq})$. The constant $c = 2$ when both additive and dominant effects are considered, and

$$D = \begin{bmatrix} \overbrace{\begin{matrix} 1 & -1/2 & 1 & -1/2 \cdots & 1 & -1/2 & 1 & -1/2 \\ 1 & -1/2 & 1 & -1/2 \cdots & 1 & -1/2 & 0 & 1/2 \end{matrix}}^{2q} \\ \vdots \\ \begin{matrix} -1 & -1/2 & -1 & -1/2 \cdots & -1 & -1/2 & 0 & 1/2 \\ -1 & -1/2 & -1 & -1/2 \cdots & -1 & -1/2 & -1 & -1/2 \end{matrix} \end{bmatrix}$$

$$= \begin{bmatrix} G_{\overbrace{1\,1\ldots1\,1}^{q}} \\ G_{1\,1\ldots1\,0} \\ \vdots \\ G_{-1-1\ldots-1\,0} \\ G_{-1-1\ldots-1-1} \end{bmatrix}.$$

Remark: The iterative formulae $R^{(s)}$ and $M^{(s)}$ given in Joehanes (2009) are not completely correct, and here, we revised them by the ones that we have given above in this paper. In Joehanes (2009), the author only gave the iterative formulae $R^{(s)}$ and $M^{(s)}$ for one QTL's case, although the author considered q QTLs in the method. At the same time, the numerator of each element of $R^{(s)}$ all lacked transposition, which also lead to the mistake of the dimension of $R^{(s)}$. So the algorithm in Joehanes (2009) cannot be conducted if the mistaken iterative formulae $R^{(s)}$ and $M^{(s)}$ were used.

Updating γ

Let the indicator function $I^{ji}_{(st)}$ about the QTL genotype and genotype of marker interval has the following expression (Tong *et al.* 2015)

$$I^{ji}_{(st)} = \begin{cases} 1, & X^*_{ji}{=}s,\ X^M_{ji} = t;\ s = 0, 1, 2;\ t = 1, \ldots, 9 \\ 0, & \text{else} \end{cases},$$

Where $j = 1, \ldots, n$, $i = 1, \ldots, q$. Then, based on the above indicator functions, the second part of the $Q(\mathbf{\Omega}|\mathbf{X}, \mathbf{Y}, \mathbf{\Omega}^{(s)})$ can be transformed into

$$\sum_{j=1}^{n} \sum_{X_j^*} \log P(X_j^*|X_j, \Omega) \cdot \omega^{(s)}_{X_j^*}$$

$$= \sum_{j=1}^{n} \sum_{X_j^*} \omega^{(s)}_{X_j^*} \cdot \sum_{i=1}^{q} \log P(X^*_{ji}|X^M_{ji}, \Omega)$$

$$= \sum_{j=1}^{n} \sum_{X_j^*} \omega^{(s)}_{X_j^*} \cdot \sum_{i=1}^{q} \Bigg[\log \frac{\gamma_i - \gamma_{i1}}{\gamma_i} \cdot I^{ji}_{02} + \log \frac{\gamma_{i1}}{\gamma_i} \cdot I^{ji}_{12} + \log \left(\frac{\gamma_i - \gamma_{i1}}{\gamma_i} \right)^2 \cdot I^{ji}_{03} + \log \left(\frac{\gamma_{i1}}{\gamma_i} \right)^2 \cdot I^{ji}_{23} + \log \frac{\gamma_{i1}}{\gamma_i} \cdot I^{ji}_{04} + \log \frac{\gamma_i - \gamma_{i1}}{\gamma_i} \cdot I^{ji}_{14} + \log \frac{\gamma_i - \gamma_{i1}}{\gamma_i} \cdot I^{ji}_{16} + \log \frac{\gamma_{i1}}{\gamma_i} \cdot I^{ji}_{26} + \log \left(\frac{\gamma_{i1}}{\gamma_i} \right)^2 \cdot I^{ji}_{07} + \log \left(\frac{\gamma_i - \gamma_{i1}}{\gamma_i} \right)^2 \cdot I^{ji}_{27} + \log \frac{\gamma_{i1}}{\gamma_i} \cdot I^{ji}_{18} + + \log \frac{\gamma_i - \gamma_{i1}}{\gamma_i} \cdot I^{ji}_{28} + \left(\log \frac{\gamma_{i1}}{\gamma_i} + \log \frac{\gamma_i - \gamma_{i1}}{\gamma_i} \right) \cdot (I^{ji}_{13} + I^{ji}_{17}) \Bigg]. \tag{3}$$

Through maximizing equation (3), we obtain the closed iterative formula of γ is given by

$$\gamma_{i1}^{(s+1)} = \frac{\gamma_i \cdot \sum_{j=1}^{n} \sum_{X_j^*} \overset{(s)}{\omega}_{X_j^*} \cdot (I^{ji}_{04} + 2I^{ji}_{07} + I^{ji}_{12} + I^{ji}_{18} + 2I^{ji}_{23} + I^{ji}_{26} + I^{ji}_{13} + I^{ji}_{17})}{\sum_{j=1}^{n} \sum_{X_j^*} \overset{(s)}{\omega}_{X_j^*} \cdot (I^{ji}_{02} + 2I^{ji}_{03} + I^{ji}_{04} + 2I^{ji}_{07} + I^{ji}_{12} + I^{ji}_{14} + I^{ji}_{16} + I^{ji}_{18} + I^{ji}_{26} + 2I^{ji}_{23} + I^{ji}_{28} + 2I^{ji}_{27} + 2I^{ji}_{13} + 2I^{ji}_{17})}.$$

The above procedure is iteratively carried out until convergence, and then the estimates of parameter matrix $\mathbf{\Omega}$ can be obtained. For convenience, the proposed method is called MTMIM-NEW.

Simulation studies

In this section, simulation experiments are performed to evaluate the proposed estimating method (MTMIM-NEW) for mapping multiple traits, and compare the performance of the new method with the existing multiple-trait method MTMIM (Joehanes 2009), and a single-trait multipleinterval mapping method (named as ST-MIM) by revising the MTMIM-NEW into single-trait case.

Here, a F_2 population is considered for the analysis. Two correlated quantitative traits are simulated in the simulations, which are controlled by two QTLs. The two QTLs are respectively, located in two equally-spaced maker intervals on a chromosome. To randomly generate maker genotypes, we take 0.0906 as the true value of the recombination rates of two marker intervals. The two true values of the recombination rates between the two QTLs and their flanking

Table 2. The estimates of QTL effects and the MSEs under different heritability.

Heritability	Trait	QTL	Effect parameter		ST-MIM		MTMIM		MTMIM-NEW	
			Additive	Dominant	Additive	Dominant	Additive	Dominant	Additive	Dominant
$h^2 = 0.2$	T1	Q1	−0.1478	0.1182	–[a]	–	−0.1030	0.1244	−0.1233	0.1201
					–	–	(0.0126)	(0.0124)	(0.0110)	(0.0121)
		Q2	0.1288	0.1492	–	–	0.0759	0.0933	0.0925	0.1221
					–	–	(0.0232)	(0.0227)	(0.0161)	(0.0226)
	T2	Q1	−0.1478	0.1135	−0.0785	0.0718	−0.0828	0.0835	−0.0844	0.0981
					(0.0141)	(0.0182)	(0.0121)	(0.0170)	(0.0117)	(0.0155)
		Q2	0.2593	0	0.1350	0.0291	0.1709	0.0166	0.1934	0.0131
					(0.0220)	(0.0180)	(0.0175)	(0.0171)	(0.0163)	(0.0172)
$h^2 = 0.05$	T1	Q1	−0.0678	0.0542	–	–	−0.0159	0.0253	−0.0323	0.0526
					–	–	(0.0172)	(0.0281)	(0.0123)	(0.0223)
		Q2	0.0932	−0.0510	–	–	0.0358	−0.0338	0.0585	−0.0456
					–	–	(0.0277)	(0.0350)	(0.0132)	(0.0219)
	T2	Q1	−0.0678	0.0520	−0.0261	−0.0043	−0.0285	0.0237	−0.0469	0.0325
					(0.0347)	(0.0468)	(0.0334)	(0.0594)	(0.0343)	(0.0241)
		Q2	0.1089	0	0.0485	−0.0112	0.0498	0.0179	0.0667	0.0107
					(0.0256)	(0.0224)	(0.0243)	(0.0184)	(0.0252)	(0.0212)

[a]–Denotes no corresponding result is obtained since only trait 2 is analysed by the ST-MIM; the numbers in the parentheses are MSEs of the estimates.

Table 3. The estimates of recombination rates and covariance matrix and the MSEs under different heritability.

Heritability	Parameter	True value	ST-MIM (MSE)	MTMIM (MSE)	MTMIM-NEW (MSE)
$h^2 = 0.2$	Q1	0.03	0.0167 (0.0438)	0.0310 (0.0428)	0.0291 (0.0171)
	Q2	0.04	0.0253 (0.0425)	0.0413 (0.0404)	0.0394 (0.0143)
	ρ	0.9	–	0.8947 (0.0574)	0.8953 (0.0563)
	σ_1^2	1	–	0.9968 (0.0640)	0.9976 (0.0634)
	σ_2^2	1	0.9864 (0.0671)	0.9931 (0.0600)	0.9941(0.0518)
$h^2 = 0.05$	Q1	0.03	0.0137 (0.0533)	0.0381 (0.0568)	0.0287 (0.0214)
	Q2	0.04	0.0223 (0.0583)	0.0429 (0.0504)	0.0390 (0.0232)
	ρ	0.9	–	0.8870 (0.0648)	0.8943 (0.0537)
	σ_1^2	1	–	0.9848 (0.0700)	0.9965 (0.0574)
	σ_2^2	1	0.9867 (0.0673)	0.9934 (0.0670)	0.9987 (0.0536)

markers are respectively, taken as 0.03 and 0.04. The sample size $n = 500$ is assigned, and two simulated scenarios are designed by taking the heritability $h^2 = 0.05$ and $h^2 = 0.2$, respectively. Next, we give the generating process of simulation data for each in detail:

(1) According to the true values of recombination rates of two marker intervals, randomly generate genotypes X_{j1}^M and X_{j2}^M for the two maker intervals for individual j; and based on the genotypes of marker intervals, generate genotypes X_{j1}^* and X_{j2}^* of two latent QTLs according to the conditional probabilities given in table 1.

(2) Generate phenotype vector value of two traits for individual j from the model: $\mathrm{Y} = \sum_{i=1}^{2} [\xi_i a_i + \eta_i d_i] + \mathrm{XB} + E$, where E follows a two-variate normal distribution with mean vector zero and covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$, and the true values of all parameters are listed in tables 2–4.

(3) Repeat steps (1) and (2) for n times, then observed data can be obtained.

In each scenario of the simulation, for each set of parameters, we compute the MLEs of all the parameters for the two-trait data by the MTMIM method and the proposed MTMIM-NEW method, and compute the MLEs of all the parameters for the data $\{(\mathrm{X}_j, \mathrm{Y}_{j2}), j = 1, \ldots, n\}$ by the ST-MIM (i.e., we only consider trait 2 here). Repeating the whole process for $M = 1000$ times, we obtained the averages of MLEs and the estimated MSE for each parameter. The simulation results are provided in tables 2–4.

The estimating results of the additive and dominant effects by the three methods are listed in table 2. It is clear to see that the estimates obtained by the MTMIM-NEW

Table 4. The estimates of marker genotype effects and the MSEs under different heritability.

Heritability	Trait	Marker	Effect value	ST-MIM (MSE)	MTMIM (MSE)	MTMIM-NEW (MSE)
$h^2 = 0.2$	T1	B1	0.18	–	0.1450 (0.0176)	0.1538 (0.0189)
		B2	0.12	–	0.1398 (0.0201)	0.1327 (0.0192)
		B3	0.15	–	0.1466 (0.0122)	0.1358 (0.0125)
	T2	B1	0.13	0.1501 (0.0231)	0.1280 (0.0160)	0.1163 (0.0164)
		B2	0.07	0.1303 (0.0246)	0.1053 (0.0209)	0.0658 (0.0165)
		B3	0.1	0.1788 (0.0172)	0.1373 (0.0143)	0.1123 (0.0120)
$h^2 = 0.05$	T1	B1	0.18	–	0.1565 (0.0216)	0.1751 (0.0143)
		B2	0.12	–	0.1362 (0.0276)	0.1279 (0.0163)
		B3	0.15	–	0.1483 (0.0232)	0.1485 (0.0125)
	T2	B1	0.13	0.1368 (0.0219)	0.1286 (0.0216)	0.1288 (0.0155)
		B2	0.07	0.0640 (0.0250)	0.0863 (0.0230)	0.0842 (0.0152)
		B3	0.10	0.1530 (0.0190)	0.1218 (0.0143)	0.1134 (0.0116)

Table 5. The results of QTLs identified for gallstone data.

					QTL effect			
			QTL position (recombination rate)		MTMIM		MTMIM-NEW	
Phenotype	Chr.	QTL	MTMIM	MTMIM-NEW	Additive	Dominant	Additive	Dominant
GW[a]	4	QTL1	64.1	62.04	0.1247	− 0.0541	0.1252	− 0.0496
	10	QTL2	58.3	56.1	0.0714	− 0.1455	0.0635	− 0.1405
HDL	4	QTL1	64.1	62.04	− 0.0346	− 0.0747	− 0.0324	− 0.0709
	10	QTL2	58.3	56.1	0.0337	− 0.0236	0.0322	− 0.0192

[a]GW, denotes Gallstone weight; the unit of QTL position is cM.

are very close to the corresponding true values of parameters in each scenario of heritability, and in most cases, the new method outperforms the MTMIM for estimating parameters (the estimates obtained by the new method have smaller MSE than the MTMIM). We also find that the estimates by the proposed method become more accurate with the increase of heritability. At the same time, the MTMIM-NEW outperforms the single-trait method, ST-MIM by comparing the corresponding MSEs of each parameter. This is expected because the proposed method simultaneously using the joint information of multiple traits, however, the ST-MIM ignores some useful trait information. In addition, the MTMIM also outperforms the ST-MIM in most cases.

Table 3 presents the estimating results of recombination rates and covariance matrix and their corresponding MSEs. The estimates of recombination rates and other parameters are obtained simultaneously in our algorithm, which is the special character of the proposed method. Whereas, in the MTMIM method, the estimates are obtained by scanning the chromosome region to establish a log-likelihood ratio (LR) graph to find the maximum LR, which is really an indirect strategy. From table 3, we can see that under each case of heritability, the estimating precision of the MTMIM-NEW is higher than those of the MTMIM and the ST-MIM consistently, and the MTMIM outperforms the ST-MIM in most cases.

The estimating results of marker genotype effects and their MSEs by the three methods are listed in table 4. It can be seen that similar results are obtained, i.e., the MTMIM-NEW consistently outperforms the ST-MIM, and in most cases the new method has higher precision than the MTMIM for estimating parameters. Moreover, the MTMIM outperforms the ST-MIM when estimating these parameters.

In general, through comparing the estimating precision of the three methods, we validate that our proposed MTMIM-NEW is effective and powerful. No matter of the QTL parameter, the marker effects, or the other model parameters, all reasonable and accurate estimating results can be obtained by the MTMIM-NEW method.

A worked example

To better demonstrate the performance and practicability of the proposed method, we take advantage of the proposed MTMIM-NEW method and the MTMIM method to analyse a real data set, which is from published literature (Wittenburg *et al.* 2003). The data set comprises of 305 mice of F_2 progeny. Here, we choose two quantitative traits: gallstone weight and high-density lipoprotein (HDL). By computing, we obtain that the mean and the SD of trait gallstone weight are respectively, 5.6×10^{-16} and 1, the mean and the SD of trait HDL are respectively, -5.7×10^{-7} and 1, and the correlation coefficient of the two traits is 0.0903. To detect the QTLs that may control

the two traits, markers D4Mit31 and D4Mit126, respectively, located at 51.3 and 71 cM on chromosome 4, and markers D10Mit66 and D10Mit34, respectively, located at 49 and 62 cM on chromosome 10 are selected to construct two marker intervals.

We use the two multiple-trait methods to estimate all QTL parameters, and the estimated results are listed in table 5. Two latent QTLs are found by the MTMIM-NEW, i.e., QTL1 is located at 62.04 cM on chromosome 4, and QTL2 is located at 56.1 cM on chromosome 10, which are approximately consistent with the reported results in the published paper (Wittenburg *et al.* 2003). Besides, from the estimates of QTL effects, the detected QTL1 shows two additive effects in opposite directions on the two traits, but QTL2 shows two additive effects in the same directions on the two traits. At the same time, the signs of all dominant effects are negative. In a certain sense, these theoretical estimating results can be supplements of the experimental results in the published paper (Wittenburg *et al.* 2003). Similar results were obtained by the MTMIM, but the estimates of QTL positions by the MTMIM-NEW are more close to the ones given in Wittenburg *et al.* (2003).

Discussion

With the rapid development of biotechnology, researchers obtain a huge number of genetic data, among which the large amount of multiple-trait data impel us to find out new statistical methods in the genetic-mapping studies. Since more latent information can be mined from multiple traits, researchers have begun to discuss the gene-mapping problem between multiple traits and multiple genetic loci (Joehanes 2009; Da Costa E Silva *et al.* 2012). Referring to Joehanes'method, we proposed a strategy of simultaneously estimating all QTL parameters in the multiple-trait multiple-interval mapping. We revised two important iterative formulae in Joehanes (2009), and presented a simultaneous estimation of all the parameters including the QTL positions.

The simulation results suggested that the newly proposed method outperforms the existing MTMIM method when multiple correlated traits are analysed jointly, and the MIM (especially the proposed MTMIM-NEW) outperforms the usual single-trait method. The proposed method has higher estimating precision, a main reason for this, is that we treat the recombination rates between the first marker and the putative QTL within each interval as unknown parameters, and embed them into the whole algorithm to get the estimates simultaneously with the other model parameters. This is an important statistical idea by which the overall error from each estimating step can be decreased. The results obtained by this method are more direct. However, in most existing studies, the positions of QTLs are mostly determined by scanning a genomewide region in a linkage group to establish a—LR-graph to find the maximum LR, which may lead to some precision loss. The proposed method was also applied to analyse a real data set, and two QTLs located at 56.1 cM on chromosome 10 and 62.04 cM on chromosome 4 were detected to have significant impact on the two traits gallstone weight and high-density lipoprotein. Likewise, the new method showed more precise results than the MTMIM.

Although the new method was developed based on data of F_2 intercross families, in fact, it can be easily extended to other populations, such as backcross, or double haploid (DH) population, in which we need to adjust the conditional probabilities in table 1 when conducting QTL mapping. The process of estimating parameter matrix $\mathbf{\Omega}$ is completely analogous with F_2 case, but the estimate of γ has new iterative formula.

In addition, after obtaining the estimate of $\mathbf{\Omega}$, we can further discuss whether the QTLs significantly exist in the considered marker intervals, which may control the multiple traits. The two hypotheses about the effect matrix $\mathbf{C}$ are

$$H_0 : \mathbf{C} = 0 \quad \text{vs} \quad H_1 : \mathbf{C} \neq 0,$$

and the test statistic $\text{LOD} = Log_{10}[L(\hat{\mathbf{C}}, \hat{\mathbf{B}}, \hat{\mathbf{S}}, \hat{\gamma})/L(0, \tilde{\mathbf{B}}, \tilde{\mathbf{S}}, \tilde{\gamma})]$ can be employed to test the hypotheses.

Of course, just as usual interval-mapping methods, the proposed method also has some shortcomings. For example, the proposed method may encounter a heavy amount of computation when the number of selected marker intervals is much larger. Now it is clear that MTMIM have advantages over single-trait MIM, since many pleiotropic genes commonly affect multiple traits, therefore, we will make further investigations and develop more effective MIM that are suitable for more markers in our future work.

Acknowledgements

This research was supported by the Natural Science Foundation of Heilongjiang Province (A201207), the Science and Technology Innovation Team in Higher Education Institutions of Heilongjiang Province (no. 2014TD005), the Fundamental Research Funds on basic research project of Heilongjiang Provincial Colleges and Universities (no. 2016-KYYWF-0929), and the Innovation Project for University Students of Heilongjiang Province (no. 201610236006).

References

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

Calinski T., Kaczmarek Z. and Krajewski P. 2000 A multivariate approach to the problem of QTL localization. *Heredity* **84**, 303–310.

Da Costa E Silva L., Wang S. C. and Zeng Z. B. 2012 Multiple trait multiple interval mapping of quantitative trait loci from inbred line crosses. *BMC Genet.* **13**, 67.

Dempster A. P., Laird N. M. and Rubin D. B. 1977 Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc. B* **39**, 1–38.

Evans D. M. 2002 The power of multivariate quantitative-trait loci linkage analysis is influenced by the correlation between variables. *Am. J. Hum. Genet*. **70**, 1599–1602.

Guo Z. and Nelson J. C. 2008 Multiple-trait quantitative trait locus mapping with incomplete phenotypic data. *BMC Genet.* **9**, 82.

Hackett C. A., Meyer R. C. and Thomas W. T. 2001 Multi-trait QTL mapping in barley using multivariate regression. *Genet. Res.* **77**, 95–106.

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

Joehanes R. 2009 *Multiple-trait multiple-interval mapping of quantitative-trait loci*. Kansas State University, Manhattan.

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

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

Korol A. B., Ronin Y. and Nevo E. 1998 Multi-interval mapping of correlated trait complexes. *Heredity* **80**, 273–284.

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

Liu J., Liu Y., Liu X. and Deng H. W. 2007 Bayesian mapping of quantitative trait loci for multiple complex traits with the use of variance components. *Am. J. Hum. Genet.* **81**, 304–320.

Malosetti M., Ribaut J. M., Vargas M., Crossa J. and van Eeuwijk F. A. 2008 A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (Zea mays L.). *Euphytica* **161**, 241–257.

Neuschl C., Brockmann G. A. and Knott S. A. 2007 Multiple-trait QTL mapping for body and organ weights in a cross between NMRI8 and DBA/2 mice. *Genet. Res.* **89**, 47–59.

Thomasen J. R., Guldbrandtsen B. and Sørensen P. 2008 Quantitative trait loci affecting calving traits in Danish Holstein cattle. *J. Dairy Sci.* **91**, 2098–2105.

Tong L., Ma W., Liu H., Yuan C. and Zhou Y. 2015 Simultaneous estimation of QTL effects and positions when using genotype data with errors. *J. Genet.* **94**, 27–34.

Wittenburg H., Lyons M. A., Li R., Churchill G. A., Carey M. C. and Paigen B. 2003 FXR and ABCG5/ABCG8 as determinants of cholesterol gallstone formation from quantitative trait locus mapping in mice. *Gastroenterology* **125**, 868–881.

Xu C., Li Z. and Xu S. 2005 Joint mapping of quantitative trait loci for multiple binary characters. *Genetic*s **169**, 1045–1059.

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

Zhu W. S. and Zhang H. P. 2009 Why do we test multiple traits in genetic association studies? *J. Korean Stat. Soc.* **38**, 1–10.

Corresponding editor: Rajiva Raman