

A model for quantitative trait loci mapping, linkage phase, and segregation pattern estimation for a full-sib progeny

Rodrigo Gazaffi · Gabriel R. A. Margarido ·
Maria Marta Pastina · Marcelo Mollinari ·
Antonio Augusto F. Garcia

Received: 20 November 2012 / Revised: 4 September 2013 / Accepted: 14 September 2013 / Published online: 23 April 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract Quantitative trait loci (QTL) mapping is an important approach for the study of the genetic architecture of quantitative traits. For perennial species, inbred lines cannot be obtained due to inbreed depression and a long juvenile period. Instead, linkage mapping can be performed by using a full-sib progeny. This creates a complex scenario because both markers and QTL alleles can have different segregation patterns as well as different linkage phases between them. We present a two-step method for QTL mapping using full-sib progeny based on composite interval mapping (i.e., interval mapping with cofactors), considering an integrated genetic map with markers with different segregation patterns and conditional probabilities obtained by a multipoint approach. The model is based on three orthogonal contrasts to estimate the additive effect (one in each parent) and dominance effect. These estimatives are obtained using the EM algorithm. In the first step, the genome is scanned to detect QTL. After, segregation pattern and linkage phases between QTL and markers are estimated. A simulated example is presented to validate the methodology. In general, the new model is more effective than existing approaches, because it can reveal QTL present in a full-sib progeny that segregates in any pattern present and

can also identify dominance effects. Also, the inclusion of cofactors provided more statistical power for QTL mapping.

Keywords Genetic architecture · Full-sib progeny · Composite interval mapping · Mixture models · F_1 population · Outbred species

Introduction

Quantitative trait loci (QTL) mapping enables a better understanding of the genetic architecture of quantitative traits. One of the most important applications of this type of study is the possibility to incorporate architecture-derived information into breeding programs to make them more effective. It also allows a better understanding of the genetic correlation among traits (Jiang and Zeng 1995; Mackay 2001), the interaction between genotypes and environments (Malosetti et al 2004, 2008; van Eeuwijk et al 2005, 2007, 2009; Boer et al 2007; Mathews et al 2008; Messmer et al 2009; Pastina et al 2012), and the determination of the breeding value of individuals for marker-assisted selection (Kao et al 1999; Zeng et al 1999; Dekkers and Hospital 2002; Hospital 2009).

Several statistical models are available for QTL mapping in populations based on inbred lines (e.g., F_2 , backcross and recombinant inbred lines), including interval mapping (IM) (Lander and Botstein 1989), composite interval mapping (CIM), (Zeng 1993, 1994; Jansen and Stam 1994) and multiple interval mapping (Kao and Zeng 1997; Kao et al 1999). IM proposed modeling QTL genotypes as latent variables by using mixture models for the analyses. In CIM, cofactors are included in the model to remove the effect of QTL located outside the mapping region, resulting in a significant increase in statistical power. Models based on CIM have been used for QTL mapping in several economically important plant species, e.g., maize and soybeans (Sabadin et al 2008; Li et al

Communicated by D. Grattapaglia

R. Gazaffi
Centre for Agricultural Sciences, Federal University of São Carlos,
Road Anhanguera, km 174, 13600-970 Araras, SP, Brazil

G. R. A. Margarido · M. Mollinari · A. A. F. Garcia (✉)
Departamento de Genética, Escola Superior de Agricultura Luiz de
Queiroz (ESALQ), Universidade de São Paulo (USP), CP 83,
13400-970 Piracicaba, São Paulo, Brazil
e-mail: augusto.garcia@usp.br

M. M. Pastina
Nucleus of Applied Biology, Embrapa Maize and Sorghum, Road
MG424, km 65, 35701-970 Sete Lagoas, MG, Brazil

2008; García-Lara et al 2009, 2010; Tang et al 2010; Tucker et al 2010; Warburton et al 2011; Xu et al 2011).

For perennial species (citrus, eucalyptus, loblolly pine, rubber tree, and others), inbred lines are unavailable so mapping populations can be generated by a biparental cross between non-inbred individuals, resulting in full-sib progeny. In diploids, either molecular markers or QTL may have 1:1:1:1, 1:2:1, 3:1, or 1:1 segregation patterns, depending on the number and configuration of the alleles of the parents. In this situation, statistical analyses are frequently carried out using an approach named *double pseudo-testcross* (Grattapaglia and Sederoff 1994). For this analysis, only markers with 1:1 segregation patterns are considered, which allows obtaining separate linkage maps for each parent, and the usage of QTL models developed for backcross populations on each individual map. However, this approach cannot be directly employed in integrated maps using new markers with distinct segregation patterns (1:1:1:1, 1:2:1, and 3:1), which have become common recently (e.g., single nucleotide polymorphism, microsatellite, etc.).

Several authors proposed the construction of integrated linkage maps using markers exhibiting different segregation patterns. Ritter et al (1990), Ritter and Salamini (1996), and Maliepaard et al (1997) have developed methods to determine recombination fractions using two-point estimates. Ridout et al (1998) have proposed the estimation of recombination fractions based on three-point tests. Wu et al (2002a) and Lu et al (2004) have developed approaches based on maximum likelihood to simultaneously estimate the recombination fraction and linkage phase between markers. Ling (2000), Wu et al (2002b), and Tong et al (2010) proposed methods based on multipoint maximum likelihood using hidden Markov models (HMM). HMMs have been incorporated into software, such as OneMap (Margarido et al 2007, 2011). The main advantage of these methods is the ability to obtain linkage maps with higher saturation and good representation of the genetic polymorphism generated by the cross, because markers with all segregation patterns can be used in the statistical analysis.

Several alternatives are available for QTL mapping in outbred populations, in which two common situations can be considered: a complex pedigree or a large progeny. In the former, QTL mapping is done based on multiple families, and both fixed or random models can be used (Knott and Haley 1992; Kruglyak and Lander 1995; Xu and Atchley 1995; Xu and Gessler 1998; Gessler and Xu 1999; Yi and Xu 1999). The latter considers a single biparental cross to obtain tens or hundreds of offspring, usually modeling QTL genotypes as fixed effects (Haley et al 1994; Knott et al 1997; Schäfer-Pregl et al 1996; Sillanpää and Arjas 1999; Lin et al 2003; Wu et al 2007; Hu and Xu 2009; Xiong 2010; Payne et al 2010). Considering the approaches developed for large progeny, Haley et al (1994) proposed a model for F_2 population (two segregating alleles), and then applied it to a full-sib progeny, under some assumptions. When more than two alleles are

considered, Knott et al (1997) expanded the previous approach, but still requires pedigree information. Schäfer-Pregl et al (1996) also presented models for a full-sib progeny considering either one scorable allele common to both parents or considering four alleles per marker locus under a non-linear approach. Sillanpää and Arjas (1999) proposed a Bayesian QTL mapping method for outbred species, which was initially proposed for inbred-based populations (Sillanpää and Arjas 1998). Lin et al (2003) developed an IM model using a maximum likelihood approach, considering QTL effects and linkage phases between markers with different patterns of segregation. However, the conditional probabilities for QTL genotypes are not based on multipoint estimates, and, in this approach, there are difficulties in estimating the linkage phase between QTL and markers using the expectation–maximization (EM) algorithm. As observed, none of these models incorporated the advantages of CIM, which is widely used for inbred-based populations, including the incorporation of cofactors and a high statistical power. Other software-based approaches were also suggested (Xiong 2010; Payne et al 2010; Hu and Xu 2009), but the segregation patterns of QTL and their linkage phases with markers were not fully addressed. Moreover, the estimation of QTL probabilities based on HMMs (multipoint) is of core importance in this context, because with outcrossing, it is common to have genomic regions with different marker types.

In this work, we developed a QTL mapping model based on the CIM approach considering a full-sib progeny and multipoint genetic mapping using molecular markers with several segregation patterns. The proposed method enables the determination of QTL, the estimation of their position, effects and segregation patterns, and the inference of their linkage phase with markers. A simulation study showed the advantages of the proposed model.

Methodology

Statistical model

We considered a full-sib progeny from the cross between two non-inbred, diploid parents P and Q with a known genetic map (Fig. 1). For an interval defined by two adjacent markers m and $m+1$, with alleles 1 or 2 for each parent, the genotypes of these loci may be generically represented by $P_m^{\{1,2\}}, P_{m+1}^{\{1,2\}}, Q_m^{\{1,2\}}, Q_{m+1}^{\{1,2\}}$, where $\{1,2\}$ indicates the allele possibilities for each locus in each parent. Assuming that there is a QTL in the interval, these alleles are represented as P^1 and P^2 for parent P and as Q^1 and Q^2 for parent Q . It is also assumed that the alleles P^1 and Q^1 have a positive effect on the phenotype. The crossing is then represented as $P_m^1 P_{m+1}^1 / P_m^2 P_{m+1}^2 \times Q_m^1 Q_{m+1}^1 / Q_m^2 Q_{m+1}^2$ for the three loci considered.

The segregation of the QTL in the progeny results in four genotypic classes ($P^1 Q^1, P^1 Q^2, P^2 Q^1,$ and $P^2 Q^2$) in 1:1:1:1

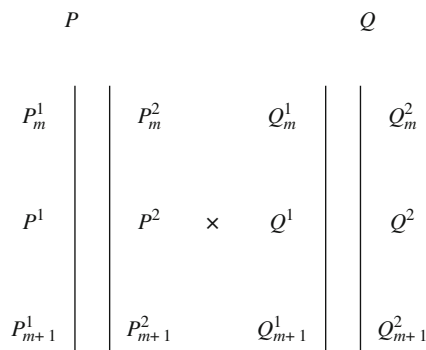


Fig. 1 Schematic representation of the type of cross considered in the model. The parents are represented by P and Q , and the alleles for the markers at the m and $m+1$ loci are represented by $P_m^{1,2}$, $Q_m^{1,2}$, $P_{m+1}^{1,2}$ and $Q_{m+1}^{1,2}$. The QTL alleles are represented as P^1 , P^2 , Q^1 and Q^2

proportion. Therefore, it is possible to define three orthogonal contrasts between the means of these classes, similar to those suggested by Knott et al (1997), Lin et al (2003) and Payne et al (2010):

$$\begin{aligned} &P^1Q^1 + P^1Q^2 - P^2Q^1 - P^2Q^2 \\ &P^1Q^1 - P^1Q^2 + P^2Q^1 - P^2Q^2 \\ &P^1Q^1 - P^1Q^2 - P^2Q^1 + P^2Q^2 \end{aligned}$$

The first two contrasts represent the additive effects of the QTL alleles in parents P and Q , respectively, and the third contrast is the intra-locus interaction (dominance) between additive effects on each parent. The contrast coefficients can be represented in the columns of the matrix \mathbf{D} (genetic design matrix), similar to the notation of Kao and Zeng (1997):

$$\mathbf{D} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix} \tag{1}$$

It is important to note that QTL genotypes are not directly observed, so they need to be inferred based on the genotype of their flanking markers. The conditional probabilities for QTL genotypes can be obtained by either two-point (Lynch and Walsh 1998) or a multipoint approach (using hidden Markov models) (Jiang and Zeng 1997; Wu et al 2002b). Although two-point analysis could be used, multipoint methods are strongly recommended because they allow the inclusion of all individuals, including the ones with missing markers, and also because of the partial information provided by non-fully informative markers on the genetic map (Jiang and Zeng 1997; Wu et al 2002b). For this reason, in the present work, conditional probabilities for QTL genotypes were obtained using OneMap software, which implements a multipoint approach using hidden Markov models (Margarido et al 2007, 2011).

From the contrasts, it is possible to define a statistical model for QTL mapping:

$$y_j = \mathbf{Z}_j\boldsymbol{\gamma} + \alpha_p^*x_{pj}^* + \alpha_q^*x_{qj}^* + \delta_{pq}^*x_{pj}^*x_{qj}^* + \varepsilon_j \tag{2}$$

where y_j : the phenotype of the j^{th} individual ($j=1, \dots, n$); \mathbf{Z}_j : j^{th} line of the indicator matrix \mathbf{Z} , with dimensions $n \times (1+3c)$, and a column of 1's and variables related to the genotypes of c cofactors, according to the contrasts represented in \mathbf{D} matrix (Eq. 1) and similar to x_{pj}^* and x_{qj}^* ; parameter vector $\boldsymbol{\gamma}$: vector $(1+3c) \times 1$ containing the intercept (μ) and the coefficients of the multiple linear regression parameters (α_{pc} , α_{qc} and δ_{pqc}) for each cofactor. Cofactors are selected in a previous step using for an example stepwise regression (Basten et al 1999) and they are fixed for each genomic position, given the window size; α_p^* and α_q^* : additive effects of the QTL for parents P and Q , respectively; δ_{pq}^* : effect of the intra-locus interaction (dominance) between additive effects; and ε_j : error. It is assumed that $\varepsilon_j \sim \mathcal{N}(0, \sigma^2)$. The variables x_{pj}^* and x_{qj}^* indicate the contrasts for QTL genotypes:

$$x_{pj}^* = \begin{cases} 1 & \text{if } P^1Q^1 \\ 1 & \text{if } P^1Q^2 \\ -1 & \text{if } P^2Q^1 \\ -1 & \text{if } P^2Q^2 \end{cases}; \quad x_{qj}^* = \begin{cases} 1 & \text{if } P^1Q^1 \\ -1 & \text{if } P^1Q^2 \\ 1 & \text{if } P^2Q^1 \\ -1 & \text{if } P^2Q^2 \end{cases}$$

The elements of \mathbf{Z} are defined in a similar way; however, they refer to markers proposed for CIM models (Zeng 1993, 1994). To select cofactors, procedures described for CIM for inbred-based populations were used. We found satisfactory results using multiple regression methods between markers and phenotypes using the Bayesian information criterion, or BIC (Schwarz 1978) to select the final model with a maximum of $2\sqrt{n}$ parameters to avoid super-parameterization (Sakamoto and Kitagawa 1987; Wang et al 2007). Because three effects (α_{pc} , α_{qc} and δ_{pqc}) may be included for each marker that is added as a cofactor, the non-significant markers may be removed to reduce the number of parameters to be estimated.

Likelihood and estimation

Considering that QTL genotypes could not be observed, the model 2 was considered as a mixture model, with QTL genotypes as latent variables. The likelihood function for the model is:

$$L(\boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma) = \prod_{j=1}^n \left[\sum_{u=1}^2 \sum_{v=1}^2 p_{uvj} \phi\left(\frac{y_j - \mu_{uvj}}{\sigma}\right) \right] \tag{3}$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ are vectors of QTL and cofactor parameters, respectively; p_{uvj} is the conditional multipoint probability for genotype P^uQ^v and the j^{th} individual, in a given position on the genome. The procedure to obtain these probabilities are detailed by Wu et al (2002b); $\phi(\cdot)$ is the standard normal probability with mean $\mu_{11j} = \mathbf{Z}_j\boldsymbol{\gamma} + \alpha_p^* + \alpha_q^*$

δ_{pq}^* ; $\mu_{12j} = \mathbf{Z}_j \boldsymbol{\gamma} + \alpha_p^* - \alpha_q^* - \delta_{pq}^*$; $\mu_{21j} = \mathbf{Z}_j \boldsymbol{\gamma} - \alpha_p^* + \alpha_q^* - \delta_{pq}^*$; $\mu_{22j} = \mathbf{Z}_j \boldsymbol{\gamma} - \alpha_p^* - \alpha_q^* + \delta_{pq}^*$, and variance σ^2 .

Using the notation presented by Kao and Zeng (1997), and expanding the ideas of Zeng (1994) for a full-sib cross, the maximum likelihood estimates are obtained using the EM algorithm (Dempster et al 1977), in two steps:

Step E: *a posteriori* probabilities ($\pi_{uvj}^{(t+1)}$) for QTL genotypes, which can be obtained applying the Bayes theorem:

$$\pi_{uvj}^{(t+1)} = \frac{p_{uvj} \phi \left(\frac{y_j - \mu_{uvj}^{(t)}}{\sigma^{(t)}} \right)}{\sum_{u=1}^2 \sum_{v=1}^2 p_{uvj} \phi \left(\frac{y_j - \mu_{uvj}^{(t)}}{\sigma^{(t)}} \right)}$$

Step M: maximum likelihood estimates:

$$\begin{aligned} \hat{\alpha}_p^{*(t+1)} &= \frac{1}{\mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_1 \circ \mathbf{D}_1)} \left[\left(\mathbf{y} - \mathbf{Z} \hat{\boldsymbol{\gamma}}^{(t)} \right)' \boldsymbol{\Pi}^{(t+1)} \mathbf{D}_1 - \mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_1 \circ \mathbf{D}_2) \hat{\alpha}_q^{*(t)} - \mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_1 \circ \mathbf{D}_3) \hat{\delta}_{pq}^{*(t)} \right] \\ \hat{\alpha}_q^{*(t+1)} &= \frac{1}{\mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_2 \circ \mathbf{D}_2)} \left[\left(\mathbf{y} - \mathbf{Z} \hat{\boldsymbol{\gamma}}^{(t)} \right)' \boldsymbol{\Pi}^{(t+1)} \mathbf{D}_2 - \mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_2 \circ \mathbf{D}_1) \hat{\alpha}_p^{*(t)} - \mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_2 \circ \mathbf{D}_3) \hat{\delta}_{pq}^{*(t)} \right] \\ \hat{\delta}_{pq}^{*(t+1)} &= \frac{1}{\mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_3 \circ \mathbf{D}_3)} \left[\left(\mathbf{y} - \mathbf{Z} \hat{\boldsymbol{\gamma}}^{(t)} \right)' \boldsymbol{\Pi}^{(t+1)} \mathbf{D}_3 - \mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_3 \circ \mathbf{D}_1) \hat{\alpha}_p^{*(t)} - \mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_3 \circ \mathbf{D}_2) \hat{\alpha}_q^{*(t)} \right] \\ \hat{\boldsymbol{\gamma}}^{(t+1)} &= (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \left(\mathbf{y} - \boldsymbol{\Pi}^{(t+1)} \mathbf{D} \hat{\boldsymbol{\theta}}^{(t+1)} \right) \\ \hat{\sigma}^2(t+1) &= (1/n) \left[\left(\mathbf{y} - \mathbf{Z} \hat{\boldsymbol{\gamma}}^{(t+1)} \right)' \left(\mathbf{y} - \mathbf{Z} \hat{\boldsymbol{\gamma}}^{(t+1)} \right) - 2 \left(\mathbf{y} - \mathbf{Z} \hat{\boldsymbol{\gamma}}^{(t+1)} \right)' \boldsymbol{\Pi}^{(t+1)} \mathbf{D} \hat{\boldsymbol{\theta}}^{(t+1)} + \hat{\boldsymbol{\theta}}^{(t+1)' \prime} \mathbf{V}^{(t+1)} \hat{\boldsymbol{\theta}}^{(t+1)} \right] \end{aligned}$$

where:

$$\mathbf{V}^{(t+1)} = \begin{bmatrix} \mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_1 \circ \mathbf{D}_1) \mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_1 \circ \mathbf{D}_2) \mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_1 \circ \mathbf{D}_3) \\ \mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_2 \circ \mathbf{D}_1) \mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_2 \circ \mathbf{D}_2) \mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_2 \circ \mathbf{D}_3) \\ \mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_3 \circ \mathbf{D}_1) \mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_3 \circ \mathbf{D}_2) \mathbf{1}' \boldsymbol{\Pi}^{(t+1)} (\mathbf{D}_3 \circ \mathbf{D}_3) \end{bmatrix}$$

\mathbf{D}_1 , \mathbf{D}_2 , and \mathbf{D}_3 are the columns of the matrix \mathbf{D} , $\boldsymbol{\Pi}^{(t+1)} = \{\pi_{uvj}\}_{(n \times 4)}$ is the *a posteriori* probability matrix of QTL genotypes, \circ represents the Hadamard product, and primes indicate a matrix (vector) transposed.

The algorithm is initiated by arbitrarily attributing values in iteration t to the parameters contained in $\boldsymbol{\theta}$, which allows the calculation of the *a posteriori* probabilities in step E ($t+1$); the new probability estimates are then employed to update the model parameters according to the estimators obtained in step M. The procedure is repeated until convergence is obtained.

QTL mapping

The procedure to test for QTL evidence is carried out by comparing the likelihood of the models, considering the presence of a QTL (H_a) versus a model without QTL (H_0):

$$\begin{aligned} H_0 : \alpha_p^* &= \alpha_q^* = \delta_{pq}^* = 0 \\ H_a : &\text{at least one is different from zero} \end{aligned}$$

These hypotheses can be tested in all genome positions, using the LOD Score or likelihood ratio test (LRT) in a way similar to that presented by Zeng (1994). It is necessary to account for problems that occur with multiple tests, which can be done using strategies already available for inbred-based

populations, such as permutation tests (Churchill and Doerge 1994). In short, this is a non parametric resampling method, allowing to obtain the empirical distribution under a null hypothesis of test statistic used for QTL mapping. The method starts shuffling the phenotypic values a number of times in order to break any correlation between QTL and phenotypes, and then performing the QTL mapping for these new data sets. The maximum test value obtained along the genome is recorded for each analysis, and the 95th percentile indicates the genome-wide threshold value.

Linkage phase and QTL segregation pattern

After finding evidence for QTL, it is possible to infer their linkage phase with their flanking markers simply based on the signals of the estimates of α_p^* and α_q^* . Because the configuration $P_m^1 P^1 P_{m+1}^1 / P_m^2 P^2 P_{m+1}^2 \times Q_m^1 Q^1 Q_{m+1}^1 / Q_m^2 Q^2 Q_{m+1}^2$ was used, with alleles P^1 and Q^1 having a positive effect on the phenotype both the estimates for α_p^* and α_q^* are positive (Fig. 1). If distinct configurations occur, the signal of the estimates will be negative accordingly. Therefore, the linkage phase can be inferred simply by identifying the alleles that have a positive or negative effect on the phenotype, and it is not necessary to include the linkage phases in the model, as seen in Lin et al (2003). This approach sensibly reduces the complexity and numerical problems of the EM algorithm.

The QTL segregation pattern depends on the relations between alleles. To infer these relations, several statistical hypotheses have been defined, and they need to be tested in the positions with evidence for QTL presence in one or two

steps (Table 1). In step 1, H_{01} , H_{02} , and H_{03} are tested, one at a time. Depending on what hypotheses are rejected and on the signal of the significant estimates of the QTL effects, another hypothesis may be necessary (step 2).

If only one of these three hypotheses is rejected, an examination of the signal of the significant effect estimates allows the conclusion of the segregation pattern and linkage phase (no step 2 required). For example, if only H_{01} is rejected and the signal of α_p^* is positive, the inferred segregation is 1:1 and the linkage phase is $P_m^1 P^1 P_{m+1}^1 / P_m^2 P^2 P_{m+1}^2$. If α_p^* is negative, the linkage phase is $P_m^1 P^2 P_{m+1}^1 / P_m^2 P^1 P_{m+1}^2$.

If more than one hypotheses is rejected at step 1, new tests are performed (step 2) that also consider the signals of the estimates. If two hypotheses are rejected, they need to be identified and the signals of the estimates of the significant effects need to be examined to check what hypotheses need to be examined in step 2. If both additive effects are significant, H_{04} needs to be tested; if α_p^* and δ_{pq}^* are not 0, H_{05} needs to be tested; otherwise, H_{06} needs to be tested. Depending on the result of the test of H_{04} , H_{05} , and H_{06} (conditional on the signs of the estimates), the segregation patterns and linkage phases are inferred. For example, if H_{01} and H_{02} are both rejected and α_p^* and α_q^* are positive and negative, respectively, H_{04} $\alpha_p^* = -\alpha_q^*$ will be tested (bilateral); if rejected, the segregation is 1:1:1:1 and the linkage phase is $P_m^1 P^1 P_{m+1}^1 / P_m^2 P^2 P_{m+1}^2$ and $Q_m^1 Q^2 Q_{m+1}^1 / Q_m^2 Q^1 Q_{m+1}^2$ for parents P and Q , respectively. When H_{01} , H_{02} , and H_{03} are all rejected, it is necessary to test H_{04} , H_{05} , and H_{06} in step 2. In a similar way, conclusions are reached based on the signs of the estimates and on what hypotheses were rejected in step 2.

The hypothesis tests of step 2 are implemented by obtaining new estimates for the parameters under the constrained new hypothesis. Thus, a **T** matrix is defined, to impose constraints by multiplying the **D** matrix. The **DT** matrix substitutes **D** in the steps of the EM algorithm and then new likelihood estimates are easily obtained (Appendix A).

Simulation

To exemplify and validate the proposed model, a simulation study was conducted in a similar way to those presented by Zeng (1994), Kao and Zeng (1997) and Lin et al (2003). A full-sib progeny consisting of 300 individuals was considered, with a genetic map composed of four chromosomes. Each chromosome had 15 molecular markers equally spaced at 10 centiMorgans (cM), employing the Kosambi function (Kosambi 1944). Markers exhibiting distinct segregation patterns were considered using the notation proposed by Wu et al (2002a). Briefly, the markers are classified into four types according to their segregation type as follows: *A* (1:1:1:1), *B* (1:2:1, separated into B_1 , B_2 , or B_3 depending on the presence of the null allele in the parent P , Q or neither of them, respectively), *C* (3:1), and *D* (1:1, labeled D_1 when the heterozygous parent is

P and D_2 when it is Q). From the total simulation, 15 markers were fully informative (type *A*); 15 markers were of the *B* type (equally distributed among the B_1 , B_2 , and B_3 types); 10 markers were of the *C* type; and 20 markers belonged to the *D* type of markers, with half of them being D_1 and the other half being D_2 . The markers were randomly distributed along the chromosomes, resulting in the following distribution: chromosome one, 5 *A*, 1 B_1 , 0 B_2 , 2 B_3 , 4 *C*, 2 D_1 , and 1 D_2 ; chromosome two, 3 *A*, 2 B_1 , 1 B_2 , 1 B_3 , 1 *C*, 2 D_1 , and 5 D_2 ; chromosome three, 4 *A*, 0 B_1 , 1 B_2 , 2 B_3 , 2 *C*, 4 D_1 , and 2 D_2 ; and chromosome four, 3 *A*, 2 B_1 , 3 B_2 , 0 B_3 , 3 *C*, 2 D_1 , and 2 D_2 . The order of the markers is indicated in Fig. 2.

The simulated trait had a heritability of 0.70 and was controlled by eight QTL located along the four chromosomes whose genetic effects were distributed such that they represented distinct linkage phases and segregation patterns. The effects were simulated as deviations of the mean, which was zero.

The conditional probabilities p_{uvj} were calculated at every 1 cM along each chromosome using the multipoint approach implemented in the OneMap software (Margarido et al 2007, 2011). Subsequently, composite interval mapping was performed, using the new model. Additionally, QTL mapping was carried out in the absence of cofactors to determine if the properties of the new proposed model were similar to those described by Zeng (1994). Cofactor selection was performed by stepwise multiple linear regression using the BIC. For each included marker, three parameters (α_p , α_q and δ_{pq}), were added and their significance was tested. The parameters exhibiting non-significant (5 %) effects were removed from the model. As proposed by Zeng (1994), from all the selected cofactors, only markers located at a distance greater than 10 cM (*window size*) from the markers flanking the interval to be mapped were considered.

To search for QTL along the genome, we used the likelihood ratio test, with three degrees of freedom. To declare a QTL, the threshold value used was $LRT=16.89$ (LOD Score 3.66) obtained by employing 1,000 permutations, with 95 % significance level (Churchill and Doerge 1994). The remaining tests carried out for step 1 (H_{01} , H_{02} , and H_{03}) and step 2 (H_{04} , H_{05} , and H_{06}) were performed using one degree of freedom. These tests are performed only at positions with a putative QTL, thus, the problems derived from the use of multiple tests are not present (Jiang and Zeng 1995).

Results

Interval mapping

As expected, interval mapping did not perform well for QTL detection (Fig. 2). For chromosome 1, two QTL were simulated (15 and 115 cM), but only one was mapped, at 5 cM. On chromosome 2, two of the three simulated QTL were detected. For chromosome 3, a large region of approximately 80 cM was

Table 1 Statistical hypotheses for testing segregation patterns and linkage phase of QTL

| Step 1 | Rejected hypotheses | Signal of the estimates ^a | Step 2 | | Conclusions | | Case |
|------------------------------------|--|--|--|--|--|-------------------|------|
| | | | Hypotheses | Segregation | Phase ^b | | |
| H_{01} , H_{02} , and H_{03} | H_{01} : $\alpha_p^* = 0$ | $+\ \emptyset\ \emptyset$ or $-\ \emptyset\ \emptyset$ | – | 1:1 | $P^1 P^2$ (+ $\emptyset\emptyset$) or $P^2 P^1$ (- $\emptyset\emptyset$) | (a) | |
| | H_{02} : $\alpha_q^* = 0$ | $\emptyset + \emptyset$ or $\emptyset - \emptyset$ | – | 1:1 | $Q^1 Q^2$ ($\emptyset + \emptyset$) or $Q^2 Q^1$ ($\emptyset - \emptyset$) | (b) | |
| | H_{03} : $\delta_{pq}^* = 0$ | $\emptyset\emptyset +$ or $\emptyset\emptyset -$ | – | 1:1 | Undetermined | (c) | |
| | H_{01} and H_{02} | $+\ +\ \emptyset$ | H_{04} : $\alpha_p^* = \alpha_q^*$ rejected H_{04} : $\alpha_p^* = -\alpha_q^*$ rejected H_{04} : $-\alpha_p^* = \alpha_q^*$ rejected H_{04} : $-\alpha_p^* = -\alpha_q^*$ rejected | 1:1:1:1 1:1:1:1 1:1:1:1 1:1:1:1 | $P^1 P^2 \times Q^1 Q^2$ $P^1 P^2 \times Q^2 Q^1$ $P^2 P^1 \times Q^1 Q^2$ $P^2 P^1 \times Q^2 Q^1$ | (d) (e) (f) | |
| H_{01} and H_{03} | $+\ +\ \emptyset$ or $+\ -\ \emptyset$ or $-\ +\ \emptyset$ or $-\ -\ \emptyset$ | H_{04} not rejected | 1:2:1 | (c), (d), (e), or (f) | | | |
| | $+\ \emptyset +$ or $-\ \emptyset -$ | H_{05} : $\alpha_p^* = \delta_{pq}^*$ or $-\alpha_p^* = -\delta_{pq}^*$ rejected H_{05} : $\alpha_p^* = -\delta_{pq}^*$ or $-\alpha_p^* = \delta_{pq}^*$ rejected | 1:1:1:1 1:1:1:1 | (a) (a) | | | |
| H_{02} and H_{03} | $+\ \emptyset -$ or $-\ \emptyset +$ | H_{05} not rejected | 1:2:1 | (a) | | | |
| | $+\ \emptyset \pm$ or $-\ \emptyset \pm$ | H_{06} : $\alpha_q^* = \delta_{pq}^*$ or $-\alpha_q^* = -\delta_{pq}^*$ rejected H_{06} : $\alpha_q^* = -\delta_{pq}^*$ or $-\alpha_q^* = \delta_{pq}^*$ rejected | 1:1:1:1 1:1:1:1 | (b) (b) | | | |
| | $\emptyset + +$ or $\emptyset - -$ | H_{06} not rejected | 1:2:1 | (b) | | | |
| | $\emptyset + -$ or $\emptyset - +$ | One or three hypotheses from H_{04} , H_{05} , and H_{06} rejected | 1:1:1:1 | (c), (d), (e), or (f) | | | |
| H_{01} , H_{02} , and H_{03} | $+\ +\ \emptyset$ or $+\ -\ \emptyset$ or $-\ +\ \emptyset$ or $-\ -\ \emptyset$ | Two hypotheses from H_{04} , H_{05} , and H_{06} rejected | 1:2:1 | (c), (d), (e), or (f) | | | |
| H_{01} , H_{02} , and H_{03} | $+\ +\ \emptyset$ or $+\ -\ \emptyset$ or $-\ +\ \emptyset$ or $-\ -\ \emptyset$ | No hypothesis from H_{04} , H_{05} , and H_{06} rejected | 3:1 | (c), (d), (e), or (f) | | | |

^aThe signals +, -, and \emptyset indicate, respectively, QTL with positive, negative, and null (non-significant) effects on the phenotype. The \pm signal indicates a significant effect whose signal has no interference on Step 2

^bTo simplify, only the QTL alleles are represented; $P^1 P^2$ represents the configuration $P_m^1 P_{m+1}^1 / P_m^2 P_{m+1}^2$, and so on

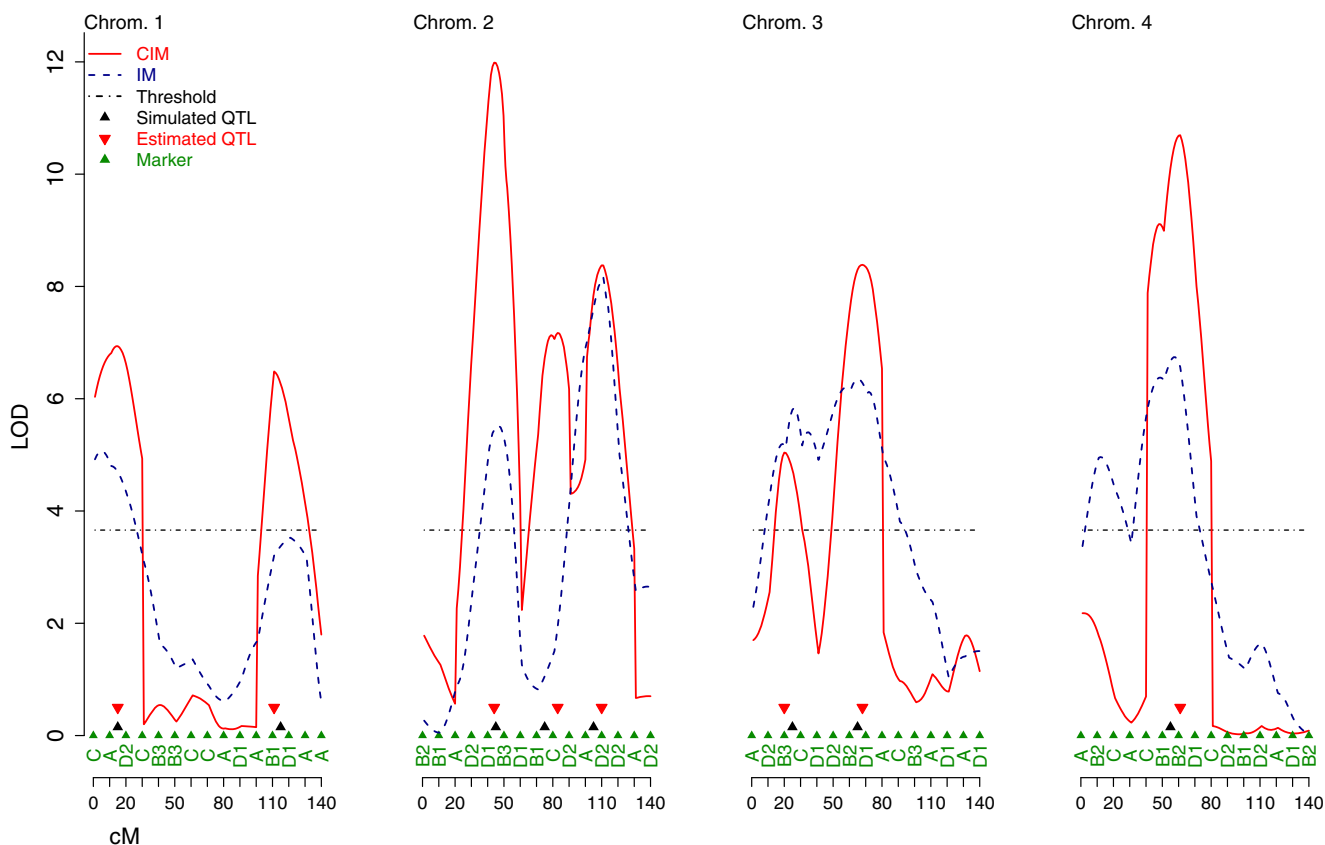


Fig. 2 QTL mapping for the simulated population without inclusion of cofactors (IM) and with inclusion of cofactors (CIM). Genetic markers are indicated by green triangles; the corresponding types (A, B1, B2, B3, C, D1, and D2) follow the notation of Wu et al (2002a)

found to display an LOD Score superior to the threshold. However, the mapping results did not show conclusively if there exists two QTL located at 25 and 65 cM, as simulated, or if there is only a single QTL with a residual effect on the adjacent intervals. On chromosome 4, a QTL located at 60 cM was detected but there was also a possible false positive at 12 cM.

Cofactor selection

Eight cofactors were selected and all of them flanked the regions spanning the simulated QTL. No super-parameterization occurred in the CIM model because the actual sample size would accommodate a total of $2\sqrt{300} = 34$ parameters. Fifteen genetic effects were included in the model, with eight markers used as cofactors (Table 2). Although some of the selected cofactors are informative, only in one parent (D_1 or D_2), dominance effects were retained in some cases because when the multipoint approach is employed to obtain the probabilities, the genotype information is recovered, even for markers that are not fully informative.

Composite interval mapping

The results from CIM (Fig. 2) were more consistent in comparison to those obtained from IM, because all simulated QTL

were mapped. False positives, detected using interval mapping, were eliminated. It is also noteworthy that virtually all QTL mapped using CIM exhibited higher LOD Scores than those mapped using IM, which indicates a higher statistical power of CIM. Therefore, our results are in agreement with those from Zeng (1994) that presented the advantages of CIM.

For chromosome 1, both simulated QTL were detected by CIM at 15 and 111 cM. The first QTL was mapped to the exact location of the simulated one, and the second one was located within the same interval.

The QTL at 15 cM, which was detected by both methods, had a higher LOD Score for CIM analysis, indicating the greater statistical power of this model.

The CIM model displayed good results for chromosome 2 as well because it could identify all three simulated QTL, which were not obtained using IM. The QTL positioned at 44 cM was also mapped with higher resolution by the CIM approach. In the case of chromosome 3, the results were also substantially improved using CIM because IM showed one large region spanning 80 cM, providing an imprecise location of the QTL, while CIM analysis correctly pointed out two distinct peaks at 20 and 68 cM.

The use of CIM for chromosome 4 removed the false QTL at 10 cM detected by IM and also detected a QTL at 61 cM with a higher LOD Score. The simulated QTL was located at

Table 2 Location of the cofactors in relation to the simulated QTL

| QTL | | | Cofactors | | |
|--------|-----|---------|-----------|-----------------------|---|
| Chrom. | cM | Segreg. | cM | Segreg. | Significant effects |
| 1 | 15 | 1:1 | 10 | <i>A</i> | α_p |
| 1 | 115 | 1:1 | 120 | <i>D</i> ₁ | α_q |
| 2 | 45 | 1:2:1 | 40 | <i>D</i> ₁ | α_p and δ_{pq} |
| 2 | 75 | 3:1 | 80 | <i>C</i> | α_p , α_q and δ_{pq} |
| 2 | 105 | 1:2:1 | 110 | <i>D</i> ₂ | α_p and δ_{pq} |
| 3 | 25 | 1:1:1:1 | 20 | <i>B</i> ₃ | α_p and α_q |
| 3 | 65 | 1:1 | 60 | <i>D</i> ₁ | α_p and δ_{pq} |
| 4 | 55 | 1:2:1 | 60 | <i>B</i> ₂ | α_p and α_q |

55 cM, which is outside, but adjacent, to the range of the simulated interval. Eye inspection of Fig. 2 allowed us to infer that the confidence interval for the QTL spans the simulated position, and that during an actual mapping situation, this would not compromise practical applications of the results.

QTL segregation pattern and linkage phases

The QTL segregation patterns result was satisfactory for all tested situations (Table 3). All QTL that segregated in a 1:1 fashion (i, ii, vii) were correctly characterized, with only one hypothesis rejected at step 1. The estimated QTL were very close to the simulated ones. QTL with a 1:2:1 segregation pattern (iii, v, viii) were also correctly inferred. For QTL with a 3:1 fashion (iv), the segregations were well estimated, with three hypothesis rejected at step 1, and three not rejected at step 2.

QTL vi, which was simulated as 1:1:1:1 with additive effects larger than dominance effects, was mapped with a 1:2:1 segregation pattern and with two significant effects $\alpha_p^* = -\alpha_q^*$ and without dominance effect δ_{pq}^* . In this case, the inferred segregation was distinct from the simulated one, most likely due to the small magnitude of the effects, which may have impaired the correct identification.

In general, the CIM model was very effective at estimating the linkage phases between QTL and markers. In all situations where QTL effects were significant, the linkage phases were always correctly estimated.

Discussion

In this work, we have presented a model for QTL mapping using full-sib progeny obtained from two diploid, non-inbred individuals. The model takes into account the distinct segregation patterns that the molecular markers and QTL may assume in the investigated context. The approach is based on the composite interval mapping model (Zeng 1993, 1994) that was first developed for inbred-based populations (BC, *F*₂,

Table 3 QTL mapping and characterization using the CIM model

| QTL | Chrom. | cM ^a | Step 1 | | | Step 2 | | | Conclusion | | | | |
|------|--------|----------------------|---------|--------------------|--------------------|---------------|-----------------------|---------------|------------|-------|------------------------|------------------------|---|
| | | | Effects | | | p-values | | | Segreg. | Phase | | | |
| | | | LOD | $\hat{\alpha}_p^*$ | $\hat{\alpha}_q^*$ | LOD | $\hat{\delta}_{pq}^*$ | LOD | | | <i>H</i> ₀₄ | <i>H</i> ₀₅ | <i>H</i> ₀₆ |
| i | 1 | 15 (15) ^b | 6.94 | 1.53 (1.70) | 6.40 | -0.22 (0.00) | 0.17 | -0.40 (0.00) | 0.40 | - | - | 1:1 (1:1) | <i>P</i> ¹ <i>P</i> ² (<i>P</i> ¹ <i>P</i> ²) |
| ii | 1 | 111 (115) | 6.49 | -0.03 (0.00) | 0.00 | -1.50 (-1.70) | 5.88 | 0.28 (0.00) | 0.74 | - | - | 1:1 (1:1) | <i>Q</i> ¹ <i>Q</i> ¹ (<i>Q</i> ² <i>Q</i> ¹) |
| iii | 2 | 44 (45) | 11.99 | 1.26 (2.00) | 5.84 | 0.45 (1.10) | 0.71 | -1.63 (-2.00) | 5.35 | - | 1 | 1:2:1 (1:2:1) | <i>P</i> ¹ <i>P</i> ² (<i>P</i> ¹ <i>P</i> ² × <i>Q</i> ¹ <i>Q</i> ²) |
| iv | 2 | 82 (75) | 7.13 | -0.47 (-1.60) | 0.95 | -1.11 (-1.60) | 2.95 | 0.98 (1.60) | 2.45 | 1 | 0.46 | 3:1 (3:1) | <i>P</i> ² <i>P</i> ¹ × <i>Q</i> ² <i>Q</i> ¹ (<i>P</i> ² <i>P</i> ¹ × <i>Q</i> ² <i>Q</i> ¹) |
| v | 2 | 110 (105) | 8.38 | 1.26 (1.80) | 4.78 | 0.16 (0.00) | 0.04 | -1.27 (-1.80) | 3.72 | - | 1 | 1:2:1 (1:2:1) | <i>P</i> ¹ <i>P</i> ² (<i>P</i> ¹ <i>P</i> ²) |
| vi | 3 | 20 (25) | 5.04 | 0.73 (1.10) | 1.78 | -1.06 (-1.50) | 3.15 | -0.48 (-0.60) | 0.76 | 0.61 | - | 1:2:1 (1:1:1:1) | <i>P</i> ¹ <i>P</i> ² × <i>Q</i> ¹ <i>Q</i> ² (<i>P</i> ¹ <i>P</i> ² × <i>Q</i> ¹ <i>Q</i> ²) |
| vii | 3 | 68 (65) | 8.39 | 1.60 (1.60) | 7.94 | 0.13 (0.00) | 0.03 | 0.40 (0.00) | 0.38 | - | - | 1:1 (1:1) | <i>P</i> ¹ <i>P</i> ² (<i>P</i> ¹ <i>P</i> ²) |
| viii | 4 | 61 (55) | 10.69 | 1.26 (1.50) | 4.53 | 1.49 (1.50) | 6.68 | 0.31 (0.00) | 0.64 | 0.49 | - | 1:2:1 (1:2:1) | <i>P</i> ¹ <i>P</i> ² × <i>Q</i> ¹ <i>Q</i> ² (<i>P</i> ¹ <i>P</i> ² × <i>Q</i> ¹ <i>Q</i> ²) |

^aFrom the first marker on the chromosome

^bThe actual values are indicated in parentheses

RILs). To validate the model, we have simulated a quantitative trait with a 0.70 heritability controlled by eight QTL exhibiting distinct effects, segregation, and linkage phases.

In general, the model allowed us to map the simulated QTL with their correct characterizations. The model also provided correct estimates of the linkage phases for all QTL with significant effects, meaning that it was possible to identify the origin of the alleles that increased or reduced the phenotype. The main advantage of this feature is that the mapping results may be useful for marker-assisted selection in plant breeding programs, even if the inferences for segregation and/or the estimates of QTL effects are eventually incorrect.

The proposed model exhibits advantages over the approach devised by Lin et al (2003). In contrast to that previous work, we have not considered the linkage phases as parameters to be estimated by the model, but instead obtained them by interpreting the signals of the estimates, α_p^* and α_q^* . This change sensibly reduced the complexity of the EM algorithm, which allowed the model to be easily expanded to the CIM context. More complex situations found in multiple interval mapping (Kao and Zeng 1997; Kao et al 1999) and multiple trait and environmental mapping (Jiang and Zeng 1997) may also be easily investigated using the proposed model. Future studies may include investigations on epistatic interactions, interactions between QTL and environments and correlation between traits.

Lin et al (2003) noted that the 1:1 segregation is tested by the assumption that one of the additive effects is zero and that the 1:2:1 pattern, similar to that found in F_2 , occurs when marginal effects are statistically equal. However, the present model allows for the identification of the possible segregation patterns via a procedure to identify these situations and a bypass to avoid multiple testing problems.

A great advantage of the proposed model is that it is based on multipoint conditional probabilities, and therefore, the presence of informative markers along the linkage groups allows the detection of QTL exhibiting distinct segregation patterns (1:1:1:1, 1:2:1, 3:1), even for regions with less informative markers. As an example, in the work by Lin et al (2003), the authors did not present effective means to estimate the conditional probabilities among less informative markers, such as those with 1:1 or 3:1 segregation patterns. In the current model, the information on adjacent markers is recovered and the probabilities are estimated in a more precise way. Jiang and Zeng (1997) have proven the effectiveness of the model for inbred lines, but to our knowledge, our work is the first to use multipoint conditional probabilities through HMM for QTL genotypes with a full-sib progeny.

The successful use of the strategy suggested for identifying QTL segregation and linkage phases depends on correctly estimating their location. Thus, models allowing higher control of the residual variance are advantageous. Zeng (1994) notes that the use of multiple linear regressions combined with interval

mapping (Lander and Botstein 1989) provides more reliable estimates for QTL effects. In the present work, the CIM model more precisely positioned the QTL in comparison to the IM approach and also displayed higher statistical power for QTL mapping. In the model proposed by Lin et al (2003), the inclusion of cofactors is complicated because the extension of the EM algorithm is not straightforward in their model. Moreover, by modeling QTL effects based on multipoint QTL probabilities, we were able to easily expand from IM to CIM. This method is also valid for more sophisticated models, such as multiple interval mapping (Kao and Zeng 1997; Kao et al 1999). Thus, the suggested model provided a sound basis for future research. An R package named *fullsibQTL* to implement the models hereby presented is under development and will be released soon.

Acknowledgments This research was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grant 141355/2006-9) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP grants: 2012/13272-6, 2007/02775-9, 2010/00083-5 and 2008/54402-4). AAFG has a research fellowship from CNPq.

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Appendix

Table 4 Matrices to obtain the constrained estimates of parameters, required for Step 2 of the testing procedures

| Hypotheses | T |
|--|--|
| $H_{04} : \alpha_p^* = \alpha_q^* \text{ or } -\alpha_p^* = -\alpha_q^*$ | $\begin{bmatrix} 1/2 & 0 \\ 1/2 & 0 \\ 0 & 1 \end{bmatrix}$ |
| $H_{04} : \alpha_p^* = -\alpha_q^* \text{ or } -\alpha_p^* = \alpha_q^*$ | $\begin{bmatrix} 1/2 & 0 \\ -1/2 & 0 \\ 0 & 1 \end{bmatrix}$ |
| $H_{05} : \alpha_p^* = \delta_{pq}^* \text{ or } -\alpha_p^* = -\delta_{pq}^*$ | $\begin{bmatrix} 1/2 & 0 \\ 0 & 1 \\ 1/2 & 0 \end{bmatrix}$ |
| $H_{05} : \alpha_p^* = -\delta_{pq}^* \text{ or } -\alpha_p^* = \delta_{pq}^*$ | $\begin{bmatrix} 1/2 & 0 \\ 0 & 1 \\ -1/2 & 0 \end{bmatrix}$ |
| $H_{06} : \alpha_q^* = \delta_{pq}^* \text{ or } -\alpha_q^* = -\delta_{pq}^*$ | $\begin{bmatrix} 0 & 1 \\ 1/2 & 0 \\ 1/2 & 0 \end{bmatrix}$ |
| $H_{06} : \alpha_q^* = -\delta_{pq}^* \text{ or } -\alpha_q^* = \delta_{pq}^*$ | $\begin{bmatrix} 0 & 1 \\ 1/2 & 0 \\ -1/2 & 0 \end{bmatrix}$ |

References

- Basten C, Weir B, Zeng Z (1999) QTL Cartographer, version 1.13: program in statistical genetics
- Boer MP, Wright D, Feng L, Podlich DW, Luo L, Cooper M, van Eeuwijk FA (2007) A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. *Genetics* 177:1801–1813
- Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait loci. *Genetics* 138:963–971
- Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. *Nat Rev Genet* 3:22–32
- Dempster AP, Laird NM, Rubin D (1977) Maximum likelihood from incomplete data via EM algorithm. *J R Stat Soc Ser B Methodological* 39:1–38
- García-Lara S, Khairallah MM, Vargas M, Bergvinson DJ (2009) Mapping of QTL associated with maize weevil resistance in tropical maize. *Crop Sci* 49:139–149
- García-Lara S, Burt AJ, Amason JT, Bergvinson DJ (2010) QTL mapping of tropical maize grain components associated with maize weevil resistance. *Crop Sci* 50:815–825
- Gessler DDG, Xu S (1999) Multipoint genetic mapping of quantitative trait loci with dominant markers in outbred populations. *Genetica* 105:281–291
- Grattapaglia D, Sederoff R (1994) Genetic linkage maps of eucalyptus grandis and eucalyptus urophylla using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 137:1121–1137
- Haley CS, Knott SA, Elsen JM (1994) Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* 136:1195–1207
- Hospital F (2009) Challenges for effective marker-assisted selection in plants. *Genetica* 136:303–310
- Hu Z, Xu S (2009) Proc qtl—a SAS procedure for mapping quantitative trait loci. *Int J Plant Genomics*. doi:10.1155/2009/141234
- Jansen RC, Stam P (1994) Resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136:1447–1455
- Jiang C, Zeng ZB (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* 140:1111–1127
- Jiang C, Zeng ZB (1997) Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. *Genetica* 101:47–58
- Kao CH, Zeng ZB (1997) Identification of quantitative trait loci in rice for yield, yield components, and agronomic traits across years and locations. *Biometrics* 73:75–83
- Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. *Genetics* 152:1203–1216
- Knott SA, Haley CS (1992) Maximum likelihood mapping of quantitative trait loci using full-sib families. *Genetics* 132:1211–1222
- Knott SA, Neale DB, Sewell MM, Haley CS (1997) Multiple marker mapping of quantitative trait loci in an outbred pedigree of loblolly pine. *Theor Appl Genet* 94:810–820
- Kosambi DD (1944) The estimation of map distances from recombination values. *Ann Eugene* 12:172–175
- Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 57:439–454
- Lander E, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–199
- Li D, Pfeiffer TW, Cornelius PL (2008) Soybean QTL for yield and yield components associated with glycine soja alleles. *Crop Sci* 48:571–581
- Lin M, Lou XY, Chang M, Wu R (2003) General statistical framework for mapping quantitative trait loci in nonmodel systems: issue for characterizing linkage phases. *Genetics* 165:901–913
- Ling S (2000) Constructing genetic maps for outbred experimental crosses. PhD thesis, University of California
- Lu Q, Cui Y, Wu R (2004) A multilocus likelihood approach to joint modeling of linkage, parental diplotypes and gene order in a full-sib family. *BMC Genet* 5(20). doi:10.1186/1471-2156-5-20
- Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland
- Mackay TFC (2001) The genetic architecture of quantitative traits. *Annu Rev Genet* 35:303–339
- Maliepaard C, Jansen J, van Ooijen JW (1997) Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. *Genet Res* 70:237–250
- Malosetti M, Voltas J, Romagosa I, Ullrich SE, van Eeuwijk FA (2004) Mixed models including environmental covariables for studying QTL by environment interaction. *Euphytica* 137:139–145
- Malosetti M, Ribaut JM, Vargas M, Crossa J, van Eeuwijk FA (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
- Margarido GRA, Souza AP, Garcia AAF (2007) Onemap: software for genetic mapping in outcrossing species. *Hereditas* 144:78–79
- Margarido GRA, Mollinari M, Garcia AAF (2011) OneMap tutorial: software for constructing genetic maps in experimental crosses: full-sib, RILs, F_2 and backcrosses. Available: <http://cran.r-project.org/web/packages/onemap/onemap.pdf>. Accessed 27 Sept 2011
- Mathews KL, Malosetti M, Chapman S, McIntyre L, Reynolds M, Shorter R, van Eeuwijk FA (2008) Multi-environment QTL mixed models for drought stress adaptation in wheat. *Theor Appl Genet* 117:1077–1091
- Messmer R, Fracheboud Y, Bänziger M, Vargas M, Stamp P, Ribaut JM (2009) Drought stress and tropical maize: QTL-by-environment interactions and stability of QTLs across environments for yield components and secondary traits. *Theor Appl Genet* 119:913–930
- Pastina MM, Malosetti M, Gazaffi R, Mollinari M, Margarido GRA, Oliveira KM, Pinto LR, Souza AP, van Eeuwijk FA, Garcia, AAF (2012) A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. *Theor Appl Genet* 124:835–849
- Payne RW, Harding SA, Murray DA, Soutar DM, Baird DB, Glaser AI, Channing IC, Welham SJ, Gilmour AR, Thompson R, Webster R (2010) GenStat Release 13 Reference Manual, Part 3 Procedure Library PL21. VSN International, Hemel Hempstead
- Ridout MS, Tong S, Vowden CJ, Tobutt KR (1998) Three-point linkage analysis in crosses of allogamous plant species. *Genet Res* 72:111–121
- Ritter E, Salamini F (1996) The calculation of recombination frequencies in crosses of allogamous plant species with applications to linkage mapping. *Genet Res* 67:55–65
- Ritter E, Gebhardt C, Salamini F (1990) Estimation of recombination frequencies and construction RFLP linkage maps in plants from crosses between heterozygous parents. *Genetics* 125:645–654
- Sabadin PK, Souza Junior CL, Souza AP, GARCIA AAF (2008) QTL mapping for yield components in a tropical maize population using microsatellite markers. *Hereditas* 145:194–203
- Sakamoto Y, Kitagawa G (1987) Akaike information criterion statistics. Kluwer Academic Publishers, Norwell
- Schäfer-Pregl R, Salamini F, Gebhardt C (1996) Models for mapping quantitative trait loci (QTL) in progeny of non-inbredparents and their behaviour in presence of distorted segregation ratios. *Genet Res* 67:43–54
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464
- Sillanpää M, Arjas E (1998) Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics* 148:1373–1388
- Sillanpää M, Arjas E (1999) Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. *Genetics* 151:1605–1619
- Tang J, Yan J, Ma X, Teng W, Wu W, Dai J, Dhillion BS, Melchinger AE, Li J (2010) Dissection of the genetic basis of heterosis in an elite

- maize hybrid by QTL mapping in an immortalized F_2 population. *Theor Appl Genet* 120:333–340
- Tong C, Zhang B, Shi J (2010) A hidden Markov model approach to multilocus linkage analysis in a full-sib family. *Tree Genet Genomes* 6:651–662
- Tucker DM, Saghai Maroof MA, Mideros S, Skoneczka JA, Nabati DA, Buss GR, Hoeschele I, Tyler BM, Martin SKS, Dorrance AE (2010) Mapping quantitative trait loci for partial resistance to *Phytophthora sojae* in a soybean interspecific cross. *Crop Sci* 50:628–635
- van Eeuwijk FA, Malosetti M, Yin X, Struik PC, Stam P (2005) Statistical models for genotype by environment data: from conventional ANOVA models to eco-physiological QTL models. *Aust J Agric Res* 56:883–894
- van Eeuwijk FA, Malosetti M, Boer MP (2007) Modelling the genetic basis of response curves underlying genotype \times environment interaction. In: Spiertz JHJ, Struik PC, van Laar HH (eds) *Scale and complexity in plant systems research: gene-plant-crop relations*. Springer, Dordrecht, pp 115–126
- van Eeuwijk FA, Boer M, Totir LR, Bink M, Wright D, Winkler CR, Podlich D, Boldman K, Baumgarten A, Smalley M, Arbelbide M, ter Braak CJF, Cooper M (2009) Mixed model approaches for the identification of QTLs within a maize hybrid breeding program. *Theor Appl Genet*. doi:10.1007/s00122-009-1205-0
- Wang S, Basten CJ, Zeng ZB (2007) Windows QTL cartographer 2.5. Available: <http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>. Accessed 03 Dec 2009
- Warburton ML, Brooks TD, Windham GL, Williams WP (2011) Identification of novel QTL contributing resistance to aflatoxin accumulation in maize. *Mol Breeding* 27:491–499
- Wu R, Ma CX, Painter I, Zeng ZB (2002a) Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. *Theor Popul Biol* 61:349–363
- Wu R, Ma CX, Wu SS, Zeng ZB (2002b) Linkage mapping of sex-specific differences. *Genet Res* 79:85–96
- Wu R, Ma CX, Casella G (2007) *Statistical genetics of quantitative traits: linkage, maps, and QTL*. Springer, New York
- Xiong J (2010) Genetic analysis of forking defects in loblolly pine. PhD thesis, North Carolina State University
- Xu S, Atchley WR (1995) A random model approach to interval mapping to quantitative trait loci. *Genetics* 141:1189–1197
- Xu S, Gessler DDG (1998) Multipoint genetic mapping of quantitative trait loci using a variable number of sibs per family. *Genet Res* 71:73–83
- Xu HN, Li Y, Li GJ, Wang X, Cheng LG, Zhang YM (2011) Mapping quantitative trait loci for seed size traits in soybean (*Glycine max* L. Merr.). *Theor Appl Genet* 122:581–594
- Yi N, Xu S (1999) Mapping quantitative trait loci for complex binary traits in outbred populations. *Heredity* 82:668–676
- Zeng ZB (1993) Theoretical basis of precision mapping of quantitative trait loci. *Proc Natl Acad Sci U S A* 90:10,972–10,976
- Zeng ZB (1994) Precision mapping of quantitative trait loci. *Genetics* 136:1457–1468
- Zeng ZB, Kao CH, Basten CJ (1999) Estimating the genetic architecture of quantitative traits. *Genet Res* 74:279–289