BMC Plant Biology

METHODOLOGY ARTICLE   **Open Access**

Multiallelic epistatic model for an out-bred cross and mapping algorithm of interactive quantitative trait loci

Chunfa Tong[1,2], Bo Zhang[1], Zhong Wang[2,3], Meng Xu[1], Xiaoming Pang[3], Jingna Si[3], Minren Huang[1] and Rongling Wu[3,2*]

Abstract

Background: Genetic mapping has proven to be powerful for studying the genetic architecture of complex traits by characterizing a network of the underlying interacting quantitative trait loci (QTLs). Current statistical models for genetic mapping were mostly founded on the biallelic epistasis of QTLs, incapable of analyzing multiallelic QTLs and their interactions that are widespread in an outcrossing population.

Results: Here we have formulated a general framework to model and define the epistasis between multiallelic QTLs. Based on this framework, we have derived a statistical algorithm for the estimation and test of multiallelic epistasis between different QTLs in a full-sib family of outcrossing species. We used this algorithm to genomewide scan for the distribution of mul-tiallelic epistasis for a rooting ability trait in an outbred cross derived from two heterozygous poplar trees. The results from simulation studies indicate that the positions and effects of multiallelic QTLs can well be estimated with a modest sample and heritability.

Conclusions: The model and algorithm developed provide a useful tool for better characterizing the genetic control of complex traits in a heterozygous family derived from outcrossing species, such as forest trees, and thus fill a gap that occurs in genetic mapping of this group of important but underrepresented species.

Background

Approaches for quantitative trait locus (QTL) mapping were developed originally for experimental crosses, such as the backcross, double haploid, RILs or F_2, derived from inbred lines [1-3]. Because of the homozygosity of inbred lines, the Mendelian (co)segregation of all markers each with two alternative alleles in such crosses can be observed directly. In practice, there is also a group of species of great economical and environmental importance - out-crossing species, such as forest trees, in which traditional QTL mapping approaches cannot be appropriately used. For these species, it is difficult or impossible to generate inbred lines due to long generation intervals and high heterozygosity [4], although

experimental hybrids have been commercially used in practical breeding programs.

For a given outbred line, some markers may be heterozygous, whereas others may be homozygous over the genome. All markers may, or may not, have the same allele system between any two outbred lines used for a cross. Also, for a pair of heterozygous loci, their allelic configuration along two homologous chromosomes (i.e., linkage phase) cannot be observed from the segregation pattern of genotypes in the cross [5,6]. Unfortunately, a consistent number of alleles across different markers and their known linkage phases are the prerequisites for statistical mapping approaches described for the backcross or F_2. Grattapaglia and Sederoff [7] proposed a so-called pseudo-test backcross strategy for linkage mapping in a controlled cross between two outbred parents. This strategy is powerful for the linkage analysis of those testcross markers that are heterozygous in one parent and null in the other, although it fails to consider many other marker cross types, such as intercross

* Correspondence: rwu@hes.hmc.psu.edu
[3]Center for Computational Biology, National Engineering Laboratory for Tree Breeding, Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants, Beijing Forestry University, Beijing 100083, China
Full list of author information is available at the end of the article

markers and dominant markers, that occur for an outbred cross. Maliepaard et al. [8] derived numerous formulas for estimating the linkage between different types of markers by correctly determining the linkage phase of markers. A general model has been developed for simultaneous estimation of the linkage and linkage phase for any marker cross type in outcrossing populations [9,10]. Stam [11] wrote powerful software for integrating genetic linkage maps using different types of markers.

Statistical methods for QTL mapping in a full-sib family of outcrossing species have not received adequate attention. Lin et al. [12] developed a model that takes into account uncertainties about the number of alleles across the genome. Wu et al. [13] used this model to reanalyze a full-sib family data for poplar trees [14], leading to the detection of new QTLs for biomass traits which were not discovered by traditional approaches. With increasing recognition of the role of epistasis in controlling and maintaining quantitative variation [15], it is crucial to extend Lin et al.'s model to map the epistatic of QTLs by which to elucidate a detailed and comprehensive perspective on the genetic architecture of a quantitative trait. However, the well-established theory and model for epistasis are mostly based on biallelic genes [16] and their estimation and test are made for a pedigree derived from inbred lines [17]. Until now, no models and algorithms have been available for characterizing the epistasis of multiallelic QTLs in an outcrossing population.

In this article, we will extend the theory for biallelic epistasis to model the epistasis between different QTLs each with multiple alleles. The multiallelic epistatic theory is then implemented into a statistical model for QTL mapping based on a mixture model. We have derived a closed form for the estimation of the main and interactive effects of multiallelic QTLs within the EM framework. Our model allows geneticists to test the effects of individual genetic components on trait variation. The estimating model has been investigated through simulation studies and validated by an example of QTL mapping for poplar trees [18]. The algorithm has been packed to a newly developed package of software, 3FunMap, derived to map QTLs in a full-sib family [19].

Quantitative Genetic Model
Additive-dominance Model
Randomly select two heterozygous lines as parents P_1 and P_2 to produce a full-sib family, in which a QTL will form four genotypes if the two lines have completely different allele systems. Let μ_{uv} be the value of a QTL genotype inheriting allele u ($u = 1,2$) from parent P_1 and allele

v ($v = 3, 4$) from parent P_2. Based on quantitative genetic theory, this genotypic value can be partitioned into the additive and dominant effects as follows:

$$\mu_{uv} = \mu + \alpha_u + \beta_v + \gamma_{uv}, \tag{1}$$

where μ is the overall mean, α_u and β_v are the allelic (additive) effects of allele u and v, respectively, and γ_{uv} is the interaction (dominant) effect at the QTL. Considering all possible alleles and allele combinations between the two parent, there are a total of four additive effects (α_1 and α_2 from parent P_1 and β_3 and β_4 from parent P_2 and four dominant effects (γ_{13}, γ_{14}, γ_{23} and γ_{34}). But these additive and dominant effects are not independent and, therefore, are not estimable. After parameterization, there are two independent additive effects, $\alpha = \alpha_1 = -\alpha_2$ and $\beta_3 = \beta_3 = -\beta_4$, and one dominant effect, $\gamma = \gamma_{13} = -\gamma_{14} = -\gamma_{23} = \gamma_{24}$, to be estimated.

Let $\mathbf{u} = (\mu_{uv})_{4 \times 1}$ and $\mathbf{a} = (\mu, \alpha, \beta, \gamma)^{\mathrm{T}}$, which can be connected by a design matrix $\mathbf{D}$. We have

$$\mathbf{u} = \mathbf{Da},$$

where

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

The expression of a can be obtained from the expression of $\mathbf{u}$ by

$$\mathbf{a} = \mathbf{D}^{-1}\mathbf{u}. \tag{2}$$

Additive-dominance-epistatic Model
If there are two segregating QTL in the full-sib family, the epistatic effects due to their nonallelic interactions should be considered. The theory for epistasis in an inbred family [16] can be readily extended to specify different epistatic components for outbred crosses. Consider two epistatic multiallelic QTL, each of which has four different genotypes, 13, 14, 23, and 24, in the outbred progeny. Let $\mu_{u_1v_1/u_2v_2}$ be the genotypic value for QTL genotype u_1v_1/u_2v_2 for $u_1,u_2 = 1,2$ and $v_1,v_2 = 3,4$ and $\mathbf{u} = (\mu_{u_1v_1/u_2v_2})$ be the corresponding mean vector. The two-QTL genotypic value is partitioned into different components as follows:

$$\begin{aligned} \mu_{u_1v_1/u_2v_2} = {} & \mu + \alpha_1 + \beta_1 + \gamma_1 + \alpha_2 + \beta_2 + \gamma_2 \\ & + I_{\alpha\alpha} + I_{\alpha\beta} + I_{\beta\alpha} + I_{\beta\beta} + J_{\alpha\gamma} + J_{\beta\gamma} + K_{\gamma\alpha} + K_{\gamma\beta} + L_{\gamma\gamma} \end{aligned} \tag{3}$$

where

(1) μ is the overall mean;

(2) α_1 is the additive effect due to the substitution from allele 1 to 2 at the first QTL;

(3) β_1 is the additive effect due to the substitution from allele 3 to 4 at the first QTL;

(4) γ_1 is the dominant effect due to the interaction between alleles from different parents;

(5) α_2 is the additive effect due to the substitution from allele 1 to 2 at the second QTL;

(6) β_2 is the additive effect due to the substitution from allele 3 to 4 at the second QTL;

(7) γ_2 is the dominant effect due to the interaction between alleles from different parents;

(8) $I_{\alpha\alpha}$ is the additive × additive epistatic effect due to the interaction between the substitutions from allele 1 to 2 at the first and second QTLs;

(9) $I_{\alpha\beta}$ is the additive × additive epistatic effect due to the interaction between the substitutions from allele 1 to 2 at the first QTL and from allele 3 to 4 at the second QTL;

(10) $I_{\beta\alpha}$ is the additive × additive epistatic effect due to the interaction between the sub-stitutions from allele 3 to 4 at the first QTL and from allele 1 to 2 at the second QTL;

(11) $I_{\alpha\beta}$ is the additive × additive epistatic effect due to the interaction between the sub-stitutions from allele 3 to 4 at the first and second QTLs;

(12) $J_{\alpha\gamma}$ is the additive × dominant epistatic effect due to the interaction between the substitutions from allele 1 to 2 at the first QTL and the dominant effect at the second QTL;

(13) $J_{\beta\gamma}$ is the additive × dominant epistatic effect due to the interaction between the substitutions from allele 3 to 4 at the first QTL and the dominant effect at the second QTL;

(14) $K_{\gamma\alpha}$ is the dominant × additive epistatic effect due to the interaction between the dominant effect at the first QTL and the substitutions from allele 1 to 2 at the second QTL;

(15) $K_{\gamma\beta}$ is the dominant × additive epistatic effect due to the interaction between the dominant effect at the first QTL and the substitutions from allele 3 to 4 at the second QTL;

(16) $L_{\gamma\gamma}$ is the dominant × dominant epistatic effect due to the interaction between the dominant effects at the first and second QTLs.

Genetic effect parameters for two interacting QTL are arrayed in $\mathbf{a} = (\mu, \alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2, I_{\alpha\alpha}, I_{\alpha\beta}, I_{\beta\alpha}, I_{\beta\beta}, J_{\alpha\gamma}, J_{\beta\gamma}, K_{\gamma\alpha}, K_{\gamma\beta}, L_{\gamma\gamma})^{\mathrm{T}}$. We relate the genotypic value vector and genetic effect vector by

$$\mathbf{u} = \mathbf{Da},$$

where design matrix

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

Thus, the genetic effect vector can be expressed, in terms of the genotypic value vector, as

$$\mathbf{a} = \mathbf{D}^{-1}\mathbf{1u}. \tag{4}$$

If we have alleles 1 = 3 and 2 = 4 for an outbred family, Equations 1 and 3 will be reduced to traditional biallelic additive-dominant and biallelic additive-dominant-epistatic genetic models, respectively [20].

Statistical Model
Likelihood
Suppose there is a full-sib family of size n derived from two outbred lines. Consider two interacting QTLs for a quantitative trait. Let u_1v_1 and u_2v_2 denote a general genotype at QTL 1 and 2, respectively, where u_1 and u_2 ($u_1,u_2 = 1,2$) are the alleles inherited from parent P_1 and v_1 and v_2 ($v_1,v_2 = 3,4$) are the alleles inherited from parent P_2. The linear model of the trait value (y_i) for individual i affected by the two QTLs is written as

$$y_i = \sum_{u_1=1}^{2}\sum_{v_1=3}^{4}\sum_{u_2=1}^{2}\sum_{v_2=3}^{4} \xi_{iu_1v_1/u_2v_2}\mu_{u_1v_1/u_2v_2} + e_i, \tag{5}$$

where $\xi_{iu_1v_1/u_2v_2}$ is the indicator variable for QTL genotypes defined as 1 if a particular genotype u_1v_1/u_2v_2 is considered for individual i and 0 otherwise, and e_i is the residual error normally distributed with mean 0 and variance σ^2. The probability with which individual i carries QTL genotype u_1v_1/u_2v_2 can be inferred from its marker genotype, with this conditional probability expressed as $\omega_{u_1v_1/u_2v_2|i}$[20].

The log-likelihood of the putative QTLs given the trait value (y) and marker information (M) is given by

$$L(\Theta|y, M) = \prod_{i=1}^{n}\sum_{u_1=1}^{2}\sum_{v_1=3}^{4}\sum_{u_2=1}^{2}\sum_{v_2=3}^{4}\omega_{u_1v_1/u_2v_2|i}f_{u_1v_1/u_2v_2}(y_i), \tag{6}$$

where Θ is the vector for unknown parameters that include the QTL position expressed by the conditional probabilities $(\omega_{u_1v_1/u_2v_2|i})$, QTL genotypic values $(\mu_{u_1v_1/u_2v_2})$ and the residual variance (σ^2). The first parameters, denoted by Θ_p, are contained in the mixture proportions of the above model, whereas the second two, denoted by Θ_q, are quantitative genetic parameters. Normal distribution density $fu_1v_1/u_2v_2(\gamma_i)$ has mean $\mu_{u_1v_1/u_2v_2}$ and variance σ^2.

EM Algorithm

The standard EM algorithm is developed to obtain the estimates of the unknown vector. By differentiating the log-likelihood of equation (6) with respect to two groups of unknown parameters (Θ_p, Θ_q), we have

$$\frac{\partial}{\partial\Theta}\log L(\Theta|\gamma, M)$$

$$= \sum_{i=1}^{n}\sum_{u_1=1}^{2}\sum_{v_1=3}^{4}\sum_{u_2=1}^{2}\sum_{v_2=3}^{4}\frac{f_{u_1v_1/u_2v_2}(\gamma i)\frac{\partial}{\partial\Theta_p}\omega_{u_1v_1/u_2v_2|i} + \omega_{u_1v_1/u_2v_2|i}\frac{\partial}{\partial\Theta_p}f_{u_1v_1u_2v_2}(\gamma i)}{\sum_{u'_1=1}^{2}\sum_{v'_1=3}^{4}\sum_{u'_2=1}^{2}\sum_{v'_2=3}^{4}\omega u'_1v'_1/u'_2v'_2|if u'_1v'_1/u'_2v'_2(\gamma_i)}$$

$$= \sum_{i=1}^{n}\sum_{u_1=1}^{2}\sum_{v_1=3}^{4}\sum_{u_2=1}^{2}\sum_{v2=3}^{4}\left[\frac{\omega u_1v_1/u_2v_2|if u_1v_1/u_2v_2(\gamma i)\frac{1}{\omega u_1v_1/u_2v_2|i}\frac{\partial}{\partial\Theta_p}\omega_{uv|i}}{\sum_{u'_1=1}^{2}\sum_{v'_1=3}^{4}\sum_{u'_2=1}^{2}\sum_{v'_2=3}^{4}\omega_{u'_1v'_1/u'_2v'_2|i}f_{u'_1v'_1/u'_2v'_2}(\gamma i)}\right.$$

$$\left. + \frac{\omega_{u_1v_1/u_2v_2|i}f_{u_1v_1/u_2v_2}(\gamma i)\frac{\partial}{\partial\Theta_q}\log f_{u_1v_1/u_2v_2}(\gamma i)}{\sum_{u'_1=1}^{2}\sum_{v'_1=3}^{4}\sum_{u'_2=1}^{2}\sum_{v'_2=3}^{4}f_{u'_1v'_1/u'_2v'_2|i}f_{u_1v' 1/u'_2v'_2}(\gamma i)}\right]$$

$$= \sum_{i=1}^{n}\sum_{u_1=1}^{2}\sum_{v_1=3}^{4}\sum_{u_2=1}^{2}\sum_{v_2=3}^{4}\prod_{u_1v_1/u_2v_2|i}\left[\frac{1}{\omega_{u_1v_1/u_2v_2|i}}\frac{\partial}{\partial\Theta_p}\omega_{u_1v_1/u_2v_2|i} + \frac{\partial}{\partial\Theta_q}\log f_{u_1v_1/u_2v_2}(\gamma i)\right],$$

where we define

$$\prod{}_{u_1v_1/u_2v_2|i} = \frac{\omega_{u_1v_1/u_2v_2|i}f_{u_1v_1/u_2v_2}(\gamma i)}{\sum_{u'_1=1}^{2}\sum_{v'_1=3}^{4}\sum_{u'_2=1}^{2}\sum_{v'_2=3}^{4}\omega_{u'_1v'_1/u'_2v'_2|i}f_{u'_1v'_1/u'_2v'_2}(\gamma_i)} \quad (7)$$

which could be thought of as a posterior probability that individual i has a QTL genotype u_1v_1/u_2v_2.

In the E step, calculate the posterior probabilities of QTL genotypes given the marker genotype of individual i by equation (7). In the M step, estimate the maximum likelihood estimates (MLEs) of the unknown parameters by solving $\frac{\partial}{\partial\Theta}\log L(\Theta|\gamma, M) = 0$. The closed forms for estimating the genotypic values and residual variance are derived as

$$\hat{\mu}_{u_1v_1/u_2v_2} = \frac{\sum_{u_1=1}^{2}\sum_{v_1=3}^{4}\sum_{u_2=1}^{2}\sum_{v_2=3}^{4}\prod_{u_1v_1/u_2v_2|i}\gamma_i}{\sum_{u_1=1}^{2}\sum_{v_1=3}^{4}\sum_{u_2=1}^{2}\sum_{v_2=3}^{4}\prod_{u_1v_1/u_2v_2|i}}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{u_1=1}^{2}\sum_{v_1=3}^{4}\sum_{u_2=1}^{2}\sum_{v_2=3}^{4}\left(\gamma_i\hat{\mu}_{u_1v_1/u_2v_2}\right)^2\prod{}_{u_1v_1/u_2v_2|i\cdot} \quad (8)$$

By giving initial values for the parameters, the E and M steps are iterated until the estimates are stable. The stable values are the MLEs of the unknown parameters. Note that the QTL position within a marker interval

can be estimated by treating the position is fixed. Using a grid search, we can obtain the MLE of the QTL position from the peak of the profile of the log-likelihood ratio test statistics across a chromosome.

Hypothesis Tests

After the parameters are estimated, a number of hypothesis tests can be made. The existence of a QTL can be tested by formulating the null hypothesis expressed as

$$H_0 : \quad \mu_{u_1v_1/u_2v_2} \equiv \mu, \text{ for } u_1, v_1 = 1, 2 \text{ and } u_2, v_2 = 3, 4$$
$$H_1 : \quad \text{at least one of the equalities above does not hold.} \quad (9)$$

The likelihoods under the reduced (H_0) and full model (H_1) are calculated and their log-likelihood ratio (LR) is then estimated by

$$LR = -2\ln\left[\frac{L_0(\tilde{\Theta}_p, \tilde{\Theta}_q|\gamma)}{L_0(\hat{\Theta}_p, \hat{\Theta}_q|\gamma, M)}\right], \quad (10)$$

where the tildes and hats are the MLEs under the H_0 and H_1, respectively. The critical threshold for declaring the existence of a QTL can be empirically determined from permutation tests [21].

Hypothesis tests for different genetic effects including the additive $(\alpha_1, \beta_1, \alpha_2, \beta_2)$, dominant (γ_1, γ_2) and additive × additive $(I_{\alpha\alpha}, I_{\alpha\beta}, I_{\beta\alpha}, I_{\beta\beta})$, additive × dominant $(J_{\alpha\gamma}, J_{\beta\gamma})$, dominant × additive $(K_{\gamma\alpha}, K_{\gamma\beta})$ and dominant × dominant $(L_{\gamma\gamma})$ epistatic effects can be formulated, with the respective null hypotheses:

Under each null hypothesis, the genotypic values should be constrained by

$$\mu_{13/13} + \mu_{13/14} + \mu_{13/23} + \mu_{13/24} + \mu_{14/13} + \mu_{14/14} + \mu_{14/23} + \mu_{14/24}$$
$$= \mu_{23/13} + \mu_{23/14} + \mu_{23/23} + \mu_{23/24} + \mu_{24/13} + \mu_{24/14} + \mu_{24/23} + \mu_{24/24} \quad (11)$$

for $H_0 : \alpha_1 = 0$,

$$\mu_{13/13} + \mu_{13/14} + \mu_{13/23} + \mu_{13/24} + \mu_{23/13} + \mu_{23/14} + \mu_{23/23} + \mu_{23/24}$$
$$= \mu_{14/13} + \mu_{14/14} + \mu_{14/23} + \mu_{14/24} + \mu_{24/13} + \mu_{24/14} + \mu_{24/23} + \mu_{24/24} \quad (12)$$

for $H_0 : \beta_1 = 0$,

$$\mu_{13/13} + \mu_{13/14} + \mu_{14/13} + \mu_{14/14} + \mu_{23/13} + \mu_{23/14} + \mu_{24/13} + \mu_{24/14}$$
$$= \mu_{13/23} + \mu_{13/24} + \mu_{14/23} + \mu_{14/24} + \mu_{23/23} + \mu_{23/24} + \mu_{24/23} + \mu_{24/24} \quad (13)$$

for $H_0 : \alpha_2 = 0$,

$$\mu_{13/13} + \mu_{13/23} + \mu_{14/13} + \mu_{14/23} + \mu_{23/13} + \mu_{23/23} + \mu_{24/13} + \mu_{24/23}$$
$$= \mu_{13/14} + \mu_{13/24} + \mu_{14/14} + \mu_{14/24} + \mu_{23/14} + \mu_{23/24} + \mu_{24/14} + \mu_{24/24}, \quad (14)$$

for $H_0 : \beta_2 = 0$,

$$\mu_{13/13} + \mu_{13/14} + \mu_{13/23} + \mu_{13/24} + \mu_{24/13} + \mu_{24/14} + \mu_{24/23} + \mu_{24/24}$$
$$= \mu_{14/13} + \mu_{14/14} + \mu_{14/23} + \mu_{14/24} + \mu_{23/13} + \mu_{23/14} + \mu_{23/23} + \mu_{23/24}, \quad (15)$$

for $H_0 : \gamma_1 = 0$,

$$\mu_{13/13} + \mu_{13/24} + \mu_{14/13} + \mu_{14/24} + \mu_{23/13} + \mu_{23/24} + \mu_{24/13} + \mu_{24/24}$$
$$= \mu_{13/14} + \mu_{13/23} + \mu_{14/14} + \mu_{14/23} + \mu_{23/14} + \mu_{23/23} + \mu_{24/14} + \mu_{24/23}, \quad (16)$$

for $H_0 : \gamma_2 = 0$,

$$\mu_{13/13} + \mu_{13/14} + \mu_{14/13} + \mu_{14/14} + \mu_{23/23} + \mu_{23/24} + \mu_{24/23} + \mu_{24/24}$$
$$= \mu_{13/23} + \mu_{13/24} + \mu_{14/23} + \mu_{14/24} + \mu_{23/13} + \mu_{23/14} + \mu_{24/13} + \mu_{24/14}, \quad (17)$$

for $H_0 : I_{\alpha\alpha} = 0$,

$$\mu_{13/13} + \mu_{13/23} + \mu_{14/13} + \mu_{14/23} + \mu_{23/14} + \mu_{23/24} + \mu_{24/14} + \mu_{24/24}$$
$$= \mu_{13/14} + \mu_{13/24} + \mu_{14/14} + \mu_{14/24} + \mu_{23/13} + \mu_{23/23} + \mu_{24/13} + \mu_{24/23}, \quad (18)$$

for $H_0 : I_{\alpha\beta} = 0$,

$$\mu_{13/13} + \mu_{13/14} + \mu_{14/23} + \mu_{14/24} + \mu_{23/13} + \mu_{23/14} + \mu_{24/23} + \mu_{24/24}$$
$$= \mu_{13/23} + \mu_{13/24} + \mu_{14/13} + \mu_{14/14} + \mu_{23/23} + \mu_{23/24} + \mu_{24/13} + \mu_{24/14}, \quad (19)$$

for $H_0 : I_{\beta\alpha} = 0$,

$$\mu_{13/13} + \mu_{13/23} + \mu_{14/14} + \mu_{14/24} + \mu_{23/13} + \mu_{23/23} + \mu_{24/14} + \mu_{24/24}$$
$$= \mu_{13/14} + \mu_{13/24} + \mu_{14/13} + \mu_{14/23} + \mu_{23/14} + \mu_{23/24} + \mu_{24/13} + \mu_{24/23}, \quad (20)$$

for $H_0 : I_{\beta\beta} = 0$,

$$\mu_{13/13} + \mu_{13/24} + \mu_{14/13} + \mu_{14/24} + \mu_{23/14} + \mu_{23/23} + \mu_{24/14} + \mu_{24/23}$$
$$= \mu_{13/14} + \mu_{13/23} + \mu_{14/14} + \mu_{14/23} + \mu_{23/13} + \mu_{23/24} + \mu_{24/13} + \mu_{24/24}, \quad (21)$$

for $H_0 : J_{\alpha\gamma} = 0$,

$$\mu_{13/13} + \mu_{13/24} + \mu_{14/14} + \mu_{14/23} + \mu_{23/13} + \mu_{23/24} + \mu_{24/14} + \mu_{24/23}$$
$$= \mu_{13/14} + \mu_{13/23} + \mu_{14/13} + \mu_{14/24} + \mu_{23/14} + \mu_{23/13} + \mu_{24/13} + \mu_{24/24}, \quad (22)$$

for $H_0 : J_{\beta\gamma} = 0$,

$$\mu_{13/13} + \mu_{13/14} + \mu_{14/23} + \mu_{14/24} + \mu_{23/23} + \mu_{23/24} + \mu_{24/13} + \mu_{24/14}$$
$$= \mu_{13/23} + \mu_{13/24} + \mu_{14/13} + \mu_{14/14} + \mu_{23/13} + \mu_{23/14} + \mu_{24/23} + \mu_{24/24}, \quad (23)$$

for $H_0 : K_{\gamma\alpha} = 0$,

$$\mu_{13/13} + \mu_{13/23} + \mu_{14/14} + \mu_{14/24} + \mu_{23/14} + \mu_{13/24} + \mu_{24/13} + \mu_{24/23}$$
$$= \mu_{13/14} + \mu_{13/24} + \mu_{14/13} + \mu_{14/23} + \mu_{23/13} + \mu_{23/23} + \mu_{24/14} + \mu_{24/24}, \quad (24)$$

for $H_0 : K_{\gamma\beta} = 0$, and

$$\mu_{13/13} + \mu_{13/24} + \mu_{14/14} + \mu_{14/23} + \mu_{23/14} + \mu_{23/23} + \mu_{24/13} + \mu_{24/24}$$
$$= \mu_{13/14} + \mu_{13/23} + \mu_{14/13} + \mu_{14/24} + \mu_{23/13} + \mu_{23/24} + \mu_{24/14} + \mu_{24/23}, \quad (25)$$

for $H_0 : L_{\gamma\gamma} = 0$, respectively. Each of these constraints is implemented with the EM algorithm as described above, which will lead to the MLEs of the genotypic values that satisfies equations (11) - (25), respectively. The critical thresholds for each of the tests (11) - (25) can be determined from simulation studies.

Results
A Worked Example
We use a real example of a forest tree to illustrate our multiallelic epistatic QTL mapping method in an outbred population. The material was an interspecific F_1 hybrid population between *Populus deltoides* (P_1) and *P. euramericana* (P_2). A total of 86 individuals were selected for QTL mapping. A genetic linkage map was constructed by using 74 SSR markers of segregating genotypes 12 × 34, which covers 822.35 cM of the whole genome and contains 14 linkage groups. The total number of roots per cutting (TNR) was measured and showed large variation in the hybrid population during the later development stage of adventitious rooting in water culture.

Through a systematic search over these linkage groups, the multiallelic espistatic model identifies six significant pairs of QTLs from different groups for TNR at the 5% significance level (Figure 1). The group × group-wide LR threshold for asserting that a pair of interacting QTLs exist was determined from 1000 permutation tests. Linkage group 2 has multiple regions that contain QTLs, which are located between markers L2_G_3592 and L2_O_10, markers L2_P_422 and L2_P_667, markers L2_P_667 and L2_G_876, and markers L2_O_286 and L2_O_222. These QTLs form five epistatic combinations by interacting with each other or with those on linkage groups 4, 7, 12 and 14 (Table 1). The sixth pair comes from linkage groups 6 and 12.

Table 1 gives the estimates of genetic effect parameters for the six pairs of interacting QTLs. At QTLs on linkage group 2, parent *P. euramericana* tends to contribute unfavorable alleles to root number, as seen by many negative β values, although this parent shows a better rooting capacity than parent *P. deltoides*. At these QTLs, parent *P. deltoides* generally contributes a small-effect allele to root number, as seen by small α values. At the QTL on linkage group 6, this parent triggers a large positive additive effect. It is interesting to find that there are pronounced interactions between alleles from these two parents, as seen by large γ values, suggesting the importance of dominance in rooting capacity. In many cases, additive × additive epistatic effects are important, as indicated by many large I values. Our model can further discern which kind of additive × additive epistasis contribute. For example, the additive × additive epistasis between QTLs from linkage group 2 is due to the interaction between alleles from parent *P. euramericana*, while for QTL pair from linkage groups 2 and 14 this is due to the interaction between alleles from parent *P. deltoides*. The pattern of how the QTLs interact with each other in terms of additive × dominant, dominant × additive, and dominant × dominant epistasis can also be identified (Table 1).

Monte Carlo Simulation
We performed simulation studies to investigate the statistical properties of the multiallelic epistatic model. We simulated a full-sib family of sample size 400, 800 and 2000 derived from two outcrossing parents. Two QTLs were assumed at different locations of a 100 cM-long linkage group with 6 even-spaced markers. Phenotypic values of a quantitative trait for each individual were simulated as the genotypic values at these QTLs plus normally distributed errors (scaled to have different heritabilities, 0.1 and 0.4). Genotypic values are expressed
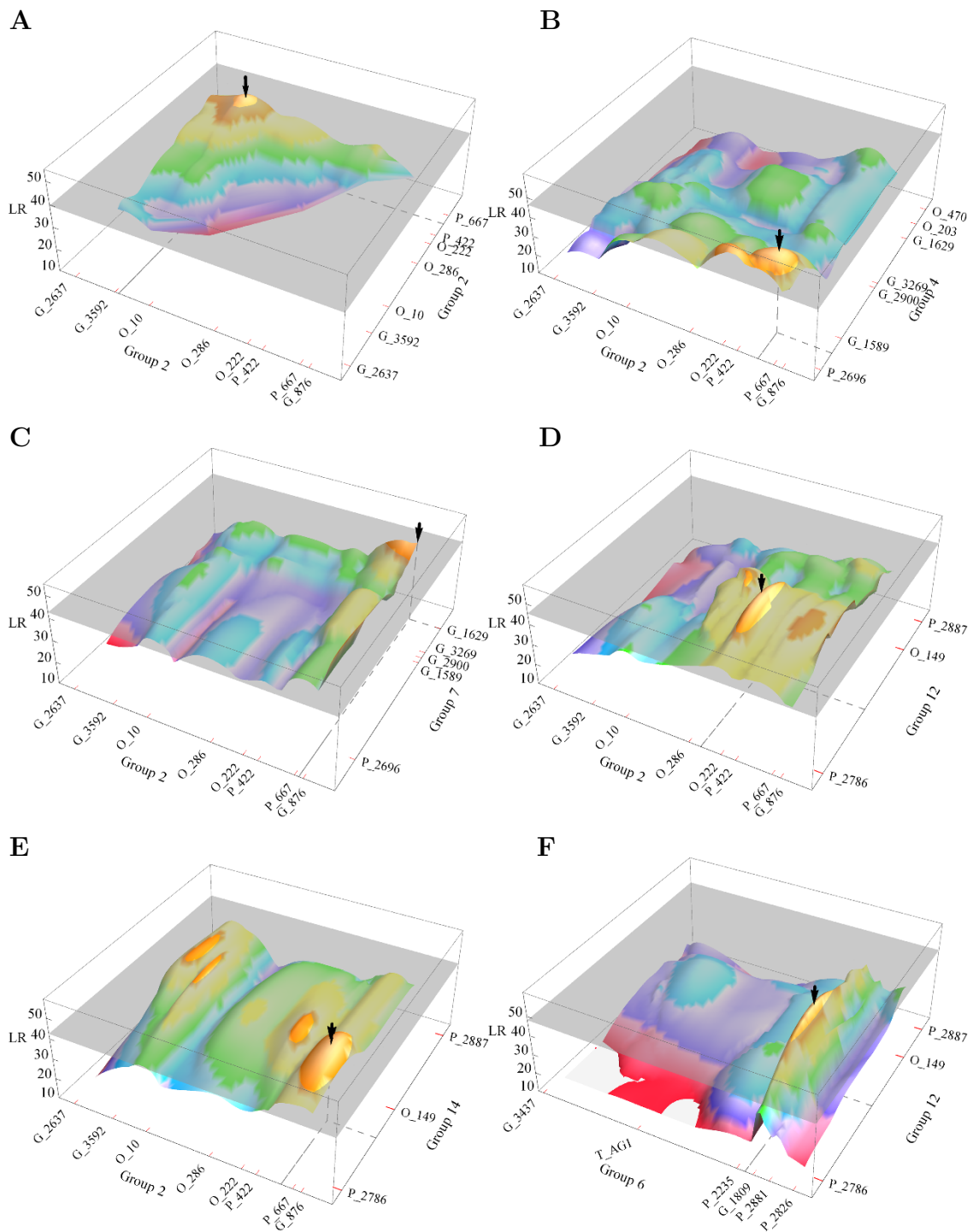
Figure 1 The landscapes of log-likelihood ratio (LR) values testing the existence of two interacting QTLs controlling the total number of roots per cutting over different linkage groups. **A**. one QTL from linkage group 2 interacting with the second QTL from linkage group 2. **B**. one QTL from linkage group 2 interacting with the second QTL from linkage group 4. **C**. one QTL from linkage group 2 interacting with the second QTL from linkage group 7. **D**. one QTL from linkage group 2 interacting with the second QTL from linkage group 12. **E**. one QTL from linkage group 2 interacting with the second QTL from linkage group 14. **F**. one QTL from linkage group 6 interacting with the second QTL from linkage group 12. In each case, the peak of the LR landscape (shown by an arrow) beyond the threshold surface (indicated in grey) shows the positions of two epistatic QTLs. The names and positions of markers at each group are indicated.

Table 1 Parameter estimates of interacting QTLs for root numbers in a full-sib family of poplars

Parameter	Estimate					
	L2 G 3592	L2 P 422	L2 P 667	L2 O 286	L2 P 422	L6 P 2235
QTL 1 Position	|	|	|	|	|	|
	L2 O 10	L2 P 667	L2 G 876	L2 O 222	L2 P 667	L6 G 1809
	L2 P 422	L4 P 2696	L7 G 3269	L12 P 2786	L14 P 2786	L12 P 2786
QTL 2 Position	|	|	|	|	|	|
	L2 P 667	L4 G 1589	L7 G 1629	L12 O 149	L14 O 149	L12 O 149
μ	2.0536	2.0578	2.0306	2.088	2.0438	2.0715
α_1	-0.0081	0.0382	0.0113	-0.0462	0.0457	0.1275
β_1	-0.1126	-0.1637	-0.2318	-0.181	-0.2119	0.0394
γ_1	0.0452	0.1798	0.0907	0.1888	0.1785	0.1662
α_2	0.0584	-0.0607	-0.0344	-0.1208	-0.0281	-0.0667
β_2	-0.1898	-0.0096	-0.1207	0.0073	0.0758	0.0067
γ_2	0.1276	-0.1329	0.1373	0.1029	0.1241	0.0435
$I_{\alpha\alpha}$	-0.0638	-0.1602	-0.0113	-0.1002	0.1761	-0.0693
$I_{\alpha\beta}$	0.0329	-0.1755	0.0945	-0.1029	-0.0044	0.1519
$I_{\beta\alpha}$	0.0535	-0.0064	0.0306	-0.2074	-0.0799	-0.1880
$I_{\beta\beta}$	0.1380	0.0480	-0.1335	-0.3300	0.0738	-0.1594
$J_{\alpha\gamma}$	-0.0498	0.0886	-0.0412	-0.0422	0.0613	0.0721
$J_{\beta\gamma}$	0.0592	0.0309	0.0317	0.1105	-0.1096	0.0716
$K_{\gamma\alpha}$	-0.0006	0.1228	0.054	-0.018	0.0165	0.2394
$K_{\gamma\beta}$	-0.0928	-0.0063	-0.1124	-0.0905	-0.0433	-0.1011
$L_{\gamma\gamma}$	-0.0979	0.0886	-0.04	-0.1476	-0.0011	0.1952
σ^2	0.187	0.1481	0.1809	0.0763	0.1549	0.0808
LR	40.9709	51.6811	46.4261	50.3592	47.6986	52.4637
$LR_{0.05}$	39.6061	46.4006	42.7068	45.3719	42.5698	48.0733

in terms of genetic actions and interactions with true values tabulated in Table 2.

It was found that the QTL positions can well be estimated using our model (Table 2). The additive effects at individual QTLs and additive × additive epistatic effects can be reasonably estimated even when a modest sample size is used for a modest heritability. The other genetic effect parameters, especially dominant × dominant epistatic effects, need a large sample size to be reasonably estimated especially when the heritability is low. Because of a large number of parameters involved, the outcrossing design requires much larger sample sizes than backcross or F$_2$ designs.

Discussion

The past two decades have seen a tremendous interest in developing statistical models for QTL mapping of complex traits inspired by Lander and Botestin's (1989) pioneered interval mapping [2,3,17,22-25]. However, model development for QTL mapping in outbred populations, a group of species of great environmental and economical importance [26], has not received adequate attention. Only a few publications are available to QTL mapping in outcrossing species [12,13]. In this article, we present a quantitative genetic model for studying the

epistasis of multiallelic QTLs and a computational algorithm for estimating and testing epistatic interactions.

The central issue of QTL mapping for outcrossing populations is how to model genetic actions and interactions between multiple alleles at different QTLs. Traditional quantitative genetic models have been developed for biallelic genetic effects [16] and their extension to multiallelic cases have not been clearly explored. This study gives a first attempt to characterize epistatic interactions between multiallelic QTLs that pervade outcrossing populations. We partition additive effects at each QTL into two subcomponents based on different parental origins of alleles. Similarly, we partition the additive × additive epistasis into four different subcomponents, the additive × dominant epistasis into two subcomponents, and the dominant × additive epistasis into two subcomponents based on the interactions of alleles of different parental origins. These subcomponents have unique biological meanings because they are derived from distinct parents. In practice, hybridization is made between two genetically distant parents, thus an understanding of each of these subcomponent helps to study the genetic basis of heterosis.

We tested the new multiallelic epistasis model through simulation studies. In general, because of a number of

Table 2 Parameter estimates and their standard errors of the multiallelic epistatic model for an outbred cross based on 1000 repeat simulations

Parameter	True Value	$H^2 = 0.1$			$H^2 = 0.4$		
		$N = 400$	$N = 800$	$N = 2000$	$N = 400$	$N = 800$	$N = 2000$
QTL 1 Position	30	29.64 (5.66)	29.90 (4.27)	29.86 (2.51)	30.00 (3.27)	30.04 (2.14)	29.99 (1.32)
QTL 2 Position	70	70.26 (5.73)	70.13(4.16)	70.04 (2.37)	70.07 (3.06)	70.03 (2.04)	69.96 (1.28)
μ	50.0	50.12 (3.02)	50.15 (2.08)	49.98 (1.25)	50.15 (1.26)	50.04 (0.82)	50.04 (0.51)
α_1	2.0	2.03 (3.27)	1.95 (2.20)	2.02 (1.37)	2.11 (1.35)	2.07 (0.85)	2.03 (0.55)
β_1	3.0	2.90 (3.33)	2.95 (2.21)	2.99 (1.39)	3.08 (1.33)	2.99 (0.90)	3.01 (0.54)
γ_1	4.0	3.72 (3.61)	4.08 (2.52)	3.95 (1.50)	3.89 (1.47)	3.99 (0.98)	3.98 (0.58)
α_2	-3.0	-3.11 (3.37)	-2.92 (2.11)	-2.99 (1.37)	-3.14 (1.34)	-3.04 (0.89)	-3.02 (0.53)
β_2	1.0	1.06 (3.26)	1.02 (2.13)	0.96 (1.36)	1.01 (1.36)	0.98 (0.88)	1.01 (0.53)
γ_2	-2.5	-2.62 (3.69)	-2.39 (2.44)	-2.57 (1.52)	-2.59 (1.42)	-2.48 (0.99)	-2.53 (0.60)
$I_{\alpha\alpha}$	-2.0	-2.22 (3.62)	-2.30 (2.57)	-2.04 (1.58)	-2.15 (1.49)	-2.11 (0.95)	-2.02 (0.61)
$I_{\alpha\beta}$	2.5	2.67 (3.64)	2.43 (2.44)	2.53 (1.52)	2.61 (1.47)	2.50 (0.94)	2.53 (0.59)
$I_{\beta\alpha}$	-3.0	-2.64 (3.61)	-3.10 (2.44)	-2.95 (1.53)	-2.94 (1.49)	-3.02 (0.95)	-2.98 (0.61)
$I_{\beta\beta}$	3.5	3.11 (3.48)	3.30 (2.47)	3.43 (1.51)	3.32 (1.47)	3.43 (0.95)	3.46 (0.60)
$J_{\alpha\gamma}$	-4.0	-3.82 (4.00)	-3.90 (2.77)	-3.91 (1.70)	-3.99 (1.56)	-3.97 (1.07)	-3.99 (0.67)
$J_{\beta\gamma}$	-4.5	-4.07 (4.03)	-4.36 (2.66)	-4.47 (1.67)	-4.37 (1.50)	-4.41 (1.01)	-4.48 (0.65)
$K_{\gamma\alpha}$	-2.0	-1.93 (4.10)	-2.02 (2.74)	-2.06 (1.67)	-2.12 (1.51)	-2.05 (1.06)	-2.01 (0.67)
$K_{\gamma\beta}$	2.5	2.32 (4.12)	2.41 (2.68)	2.46 (1.68)	2.39 (1.46)	2.40 (1.02)	2.48 (0.64)
$L_{\gamma\gamma}$	-5.0	-4.42 (4.16)	-4.68 (3.07)	-4.88 (1.90)	-4.84 (1.69)	-4.90 (1.11)	-4.98 (0.73)
σ^2	1334.3	1242.64 (99.80)	1286.12 (71.09)	1314.93 (43.50)			
σ^2	222.4				206.83 (17.44)	214.62 (11.88)	219.40 (7.45)

Two QTLs are set at 30 cM and 70 cM in a chromosome of 100cM with 6 markers evenly spaced and the true parameters are set as in the second column.

parameters involved, a larger sample size is required to obtain reasonably precise estimation for QTL mapping in outcrossing populations. According to our experience, the increased heritability of traits by precise phenotyping can improve parameter estimation and model power than augmented experiment scales. We recommend that more efforts are given to field management that can improve the quality of phenotype measurements than experimental size. By analyzing a real data set from a poplar genetic study, the new model has been well validated. It is interesting to find that interactions between alleles from different poplar species contribute substantially to rooting capacity from cuttings, larger than genetic effects of alleles that operate alone. This result may help to understand the role of dominance in mediating heterosis.

Conclusions

We have developed a statistical model for mapping interactive QTLs in a full-sib family of outcrossing species. By capitalizing on traditional quantitative genetic theory, we define epistatic components due to interactions between two outcrossing multiallelic QTLs. An algorithmic procedure was derived to estimate all types of outcrossing epistasis and test their significance in controlling a quantitative trait. Our model provides a

useful tool for studying the genetic architecture of complex traits for outcrossing species, such as forest trees, and fill a gap that occurs in genetic mapping of this group of important but underrepresented species.

Author details
[1]The Key Laboratory of Forest Genetics and Gene Engineering, Nanjing Forestry University, Nanjing, Jiangsu 210037, China. [2]Center for Statistical Genetics, The Pennsylvania State University, Hershey, PA 17033, USA. [3]Center for Computational Biology, National Engineering Laboratory for Tree Breeding, Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants, Beijing Forestry University, Beijing 100083, China.

Authors' contributions
CT derived the model and performed computer simulation and data analysis. BZ and MX collected the data from poplar hybrids. ZW and JS participated in simulation studies. XP participated in model design and result interpretation. MH conceived of the experiment. RW developed the model and algorithm, coordinated simulation and data analysis, and wrote the paper. All authors have read and approved the final manuscript.

References
1. Lander ES, Botstein D: **Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps.** *Genetics* 1989, **121**:185-199.

2. Zeng ZB: **Precision mapping of quantitative trait loci.** *Genetics* 1994, **136**:1457-1468.
3. Lynch M, Walsh B: *Genetics and Analysis of Quantitative Traits* Sinauer Associates, Sunderland, MA; 1998.
4. Wu RL, Zeng ZB, McKend SE, O'Malley DM: **The case for molecular mapping in forest tree breeding.** *Plant Breed Rev* 2000, **19**:41-68.
5. Ritter E, Gebhardt C, Salamini F: **Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozy gous parents.** *Genetics* 1999, **125**:645-654.
6. Ritter E, Salamini F: **The calculation of recombination frequencies in crosses of allogamous plant species with applications to linkage mapping.** *Genet Res* 1996, **67**:55-65.
7. Grattapaglia D, Sederoff RR: **Genetic linkage maps of** *Eucalyptus grandis* **and** *Eucalyptus urophylla* **using a pseudo-testcross: mapping strategy and RAPD markers.** *Genetics* 1994, **137**:1121-1137.
8. Maliepaard C, Jansen J, van Ooijen JW: **Linkage analysis in a fullsib family of an outbreeding plant species: overview and consequences for applications.** *Genet Res* 1997, **70**:237-250.
9. Wu RL, Ma CX, Painter I, Zeng ZB: **Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species.** *Theor Pop Biol* 2002, **61**:349-363.
10. Lu Q, Cui YH, Wu RL: **A multilocus likelihood approach to joint modeling of linkage, parental diplotype and gene order in a full-sib family.** *BMC Genet* 2004, **5**:20.
11. Stam P: **Construction of integrated genetic linkage maps by means of a new computer package: JoinMap.** *Plant J* 1993, **3**:739-744.
12. Lin M, Lou XY, Chang M, Wu RL: **A general statistical framework for mapping quantitative trait loci in non-model systems: Issue for characterizing linkage phases.** *Genetics* 2002, **165**:901-913.
13. Wu S, Yang J, Huang YJ, Li Y, Yin T, Wullschleger SD, Tuskan GA, Wu RL: **An improved approach for mapping quantitative trait loci in a pseudo-testcross design: Revisiting a poplar genome study.** *Bioinformat Biol Insights* 2010, **4**:1-8.
14. Wullschleger SD, Yin TM, DiFazio SP, Tschaplinski TJ, Gunter LE, Davis MF, Tuskan GA: **Phenotypic variation in growth and biomass distribution for two advanced-generation pedigrees of hybrid poplar (Populus spp.).** *Can J For Res* 2005, **5**:1779-1789.
15. Whitlock MC, Phillips PC, Moore FBG, Tonsor SJ: **Multiple fitness peaks and epistasis.** *Ann Rev Ecol Syst* 1995, **26**:601-629.
16. Mather K, Jinks JL: **Biometrical Genetics.** Chapman & Hall London;, 3 1982.
17. Kao CH, Zeng ZB: **Modeling epistasis of quantitative trait loci using Cocker-ham's model.** *Genetics* 2002, **160**:1243-1261.
18. Zhang B, Tong CF, Yin TM, Zhang XY, Zhuge Q, Huang MR, Wang MX, Wu RL: **Detection of quantitative trait loci influencing growth trajectories of adventitious roots in Populus using functional mapping.** *Tree Genet Genom* 2009, **5**:539-552.
19. Tong CF, Wang Z, Zhang B, Shi JS, Wu RL: **3FunMap: Full-sib family functional mapping of dynamic traits.** *Bioinformatics* .
20. Wu RL, Ma CX, Casella G: *Statistical Genetics of Quantitative Traits: Linkage Maps and QTL* Springer-Verlag, New York; 2007.
21. Churchill GA, Doerge RW: **Empirical threshold values for quantitative trait mapping.** *Genetics* 1994, **138**:963-971.
22. Yi NJ, Xu SZ, Allison DB: **Bayesian model choice and search strategies for mapping interacting quantitative trait loci.** *Genetics* 2003, **165**:867-883.
23. Broman KW: **Mapping quantitative trait loci in the case of a spike in the phenotype distribution.** *Genetics* 2003, **163**:1169-1175.
24. Zou F, Nie L, Wright FA, Sen PK: **A robust QTL mapping procedure.** *J Stat Plan Infer* 2009, **139**:978-989.
25. Cheng JY, Tzeng SJ: **Parametric and semiparametric methods for mapping quantitative trait loci.** *Computat Stat Data Analy* 2009, **53**:1843-1849.
26. Bradshaw HD, Stettler RF: **Molecular genetics of growth and development in Populus. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree.** *Genetics* 1995, **139**:963-973.