


SOFTWARE

Open Access



# MVQTL-CIM: composite interval mapping of multivariate traits in a hybrid F<sub>1</sub> population of outbred species

Fenxiang Liu<sup>1,2</sup>, Chunfa Tong<sup>1\*</sup> , Shentong Tao<sup>1</sup>, Jiyan Wu<sup>1</sup>, Yuhua Chen<sup>1</sup>, Dan Yao<sup>1</sup>, Huogen Li<sup>1</sup> and Jisen Shi<sup>1</sup>

## Abstract

**Background:** With the plummeting cost of the next-generation sequencing technologies, high-density genetic linkage maps could be constructed in a forest hybrid F<sub>1</sub> population. However, based on such genetic maps, quantitative trait loci (QTL) mapping cannot be directly conducted with traditional statistical methods or tools because the linkage phase and segregation pattern of molecular markers are not always fixed as in inbred lines.

**Results:** We implemented the traditional composite interval mapping (CIM) method to multivariate trait data in forest trees and developed the corresponding software, mvqtlcim. Our method not only incorporated the various segregations and linkage phases of molecular markers, but also applied Takeuchi's information criterion (TIC) to discriminate the QTL segregation type among several possible alternatives. QTL mapping was performed in a hybrid F<sub>1</sub> population of *Populus deltoides* and *P. simonii*, and 12 QTLs were detected for tree height over 6 time points. The software package allowed many options for parameters as well as parallel computing for permutation tests. The features of the software were demonstrated with the real data analysis and a large number of Monte Carlo simulations.

**Conclusions:** We provided a powerful tool for QTL mapping of multiple or longitudinal traits in an outbred F<sub>1</sub> population, in which the traditional software for QTL mapping cannot be used. This tool will facilitate studying of QTL mapping and thus will accelerate molecular breeding programs especially in forest trees. The tool package is freely available from <https://github.com/tongchf/mvqtlcim>.

**Keywords:** Quantitative trait locus, Composite interval mapping, Multivariate linear model, Multivariate traits, *Populus*

## Background

Most forest trees are outbred species and have the characteristics of high heterozygosity and long generation times [1]. These properties make it very difficult to generate inbred lines in forest trees for linkage mapping and then for quantitative trait loci (QTL) mapping with traditional statistical methods. However, with the continuously reducing cost of next-generation sequencing (NGS) technologies and the development of new genetic mapping strategies, thousands of genetic markers could be obtained across many individuals and thus could be used to construct high-density genetic linkage

maps in a forest hybrid F<sub>1</sub> population [2, 3]. Such dense linkage maps would greatly facilitate QTL mapping as well as comparative genomics in forest trees. Yet, the statistical methods of QTL mapping used for populations derived from inbred lines cannot be directly applied to the outcrossing populations because the linkage phase and segregation pattern of markers on genetic maps may vary from locus to locus and are not always fixed as in inbred lines [4–6].

Over the past three decades, many statistical models developed for QTL mapping were mainly based on experimental populations, such as the backcross and F<sub>2</sub>, crossed from two inbred lines. These models initiated with the seminal approach of interval mapping (IM) proposed by Lander and Botstein [7]. To overcome the problem of possibly generating so-called 'ghost QTL'

\* Correspondence: [tongchf@njfu.edu.cn](mailto:tongchf@njfu.edu.cn)

<sup>1</sup>The Southern Modern Forestry Collaborative Innovation Center, College of Forestry, Nanjing Forestry University, Nanjing 210037, China  
Full list of author information is available at the end of the article



with IM [8], Zeng [9, 10] proposed the composite interval mapping (CIM) method by adding a proper number of background markers into the model to absorb effects of other QTLs outside the detected region. Since then, QTL mapping methods were extended to multiple interval mapping (MIM) [11, 12], and also to mapping binary and categorical traits [13, 14]. Moreover, Bayesian models [15–18] and the least absolute shrinkage and selection operator (LASSO) methods [19–23] were applied to mapping single or multiple QTLs. In addition, approaches were inherently established by extending from mapping single trait to multiple traits or longitudinal trait data [11, 24–26]. Specifically, Wu and colleagues proposed a so-called functional mapping method in order to identify QTLs that affect a particular biological process with trait values over multiple stages [27–30].

Meanwhile, several great efforts have been made to develop statistical models used for QTL mapping in outbred species. Haley et al. [31] proposed a method for identifying QTLs in an outcrossing population of pigs, but it has the limitations that it did not consider the possible changes in marker segregation pattern and the linkage phase of the parents. Besides the QTL location and effects, Lin et al. [32] subsequently established an approach that can estimate the linkage phase between the linked QTL and a marker in an outcrossed population. Tong et al. [6] proposed a model selection method to discriminate the most likely QTL segregation pattern within several possible QTL segregation patterns in a full-sib family generated from two outcrossing parents. This method actually was implemented in the context of IM, but it capitalized on the complex genetic architecture of an outcrossing population, such as the various marker segregation patterns and non-fixed linkage phases. Recently, Gazaffi et al. [33] presented a CIM method with a series of hypothesis tests to infer significant QTLs and their segregation types in a full-sib progeny. However, their procedure of testing significant QTLs is similar to the method in the software MapQTL [34], which could lose the power to detect QTLs segregating in the test cross or  $F_2$  pattern in real examples [6].

Although these significant advances have been achieved, there is still a lot of room to improve for QTL mapping in a hybrid  $F_1$  population in outbred species, because of the complex genetic characteristics of such a mapping population. In this study, we developed a model selection method to implement CIM for mapping tree height with values at different growth time points in a hybrid  $F_1$  population of *Populus deltoides*  $\times$  *P. simonii*. The two *Populus* species display substantially different performance in growth rate, resistance to diseases and bad conditions, and rooting ability [2, 35]. Their hybrid progeny provide a permanent material for constructing genetic linkage maps and identifying QTLs in *Populus*. We showed that our

QTL mapping approach can detect 12 QTLs that affect tree height, based on the parent-specific high-density linkage maps constructed in our previous study [2]. Furthermore, the new method can find more QTLs with higher significance compared with the interval mapping method. The software developed for implementing the algorithm can be downloaded from <https://github.com/tongchf/mvqtlcim> as package *mvqtlcim*.

## Methods

### Mapping population

The mapping population was an interspecific  $F_1$  hybrid population between *P. deltoides* ( $P_1$ ) and *P. simonii* ( $P_2$ ), which was established in 2011 [2]. The parental genetic linkage maps were constructed by using 1601 and 940 SNPs and covered 4249.12 and 3816.24 cM of the whole genomes of  $P_1$  and  $P_2$ , respectively [2]. A total of 177 individuals were selected for QTL mapping. The tree height of each individual was measured at six different time points during the growth period in 2014. The phenotype data showed large variation at different development stages in the  $F_1$  hybrid population.

### Stepwise regression model

In order to apply CIM into an outbred full-sib family for multivariate phenotype data, the first step is to choose background markers to control other QTL effects when scanning a putative QTL at a specific position on genome. We used stepwise regression method to choose the background markers among the whole available molecular marker data. Considering  $n$  individuals with genotype data of  $M$  markers and the phenotypic values of a trait at  $T$  time points, the linear regression model can be described as

$$y_{it} = \mu_t + \sum_{j=1}^M \sum_{k=1}^{K_j} x_{ijk} B_{jkt} + e_{it}, \quad i = 1, \dots, n; \quad t = 1, \dots, T \quad (1)$$

where  $y_{it}$  is the trait value at the  $t$ th time point of the  $i$ th individual;  $\mu_t$  is the overall mean of the trait value at the  $t$ th time point;  $x_{ijk}$  is an indicator variable for the  $k$ th genotype of the  $j$ th marker for the  $i$ th individual, taking the value of 1 or 0;  $B_{jkt}$  is the effect of the  $k$ th genotype of the  $j$ th marker at the  $t$ th time point, with the restriction

$$\sum_{k=1}^{K_j} B_{jkt} = 0 \quad (2)$$

where  $K_j$  is the number of genotypes of the  $j$ th marker;  $e_{it}$  is a random error at the  $t$ th time point for the  $i$ th individual. Because a molecular marker could segregate in the ratio of 1:1, 1:2:1, or 1:1:1:1 in an outbred full-sib family [4, 5], the number of genotypes of the  $j$ th marker possibly takes the value of 2,

3, or 4. For the random errors of individual  $i$ , let  $e_i = (e_{i1}, e_{i2}, \dots, e_{iT})$  and assume that  $e_i \sim N(0, \Sigma)$ .

The stepwise regression involved starting with no markers in the model, adding a marker to the model with the most significance at a specified entry level, removing a candidate marker from the model if its significance is reduced below a specified staying level, and repeating this process until no markers can be added or deleted. Since model (1) actually belongs to a multivariate multiple regression model, the significance for a candidate marker in the model can be tested with Wilks' lambda statistic

$$\Lambda = |\hat{\Sigma}|/|\hat{\Sigma}_0| \tag{3}$$

where  $\hat{\Sigma}$  is the maximum likelihood estimate of  $\Sigma$  for the full model and  $\hat{\Sigma}_0$  for the reduced model under a null hypothesis. The lambda statistic can be approximated by an  $F$  or chi-square distribution in some cases for calculating  $p$ -value in testing the significance of a marker in our regression model [36, 37].

### Composite interval mapping model

Unlike in inbred lines, not only molecular markers but also QTLs may segregate in any patterns in an  $F_1$  outcrossing population. In our CIM model, we focused on the markers segregating in the types of  $aa \times ab$ ,  $ab \times aa$ ,  $ab \times ab$  and  $ab \times cd$ , and QTLs segregating in the types of test cross (i.e.  $Qq \times Qq$  or  $Qq \times QQ$ ),  $F_2$  cross (i.e.  $Qq \times Qq$  or  $Qq \times qQ$ ) and full cross (i.e.  $Q_1Q_2 \times Q_3Q_4$ ) [4, 6]. Assuming that there exists a QTL in an interval of markers  $M_s$  and  $M_{s+1}$  on a chromosome, our CIM model for multivariate phenotype data can be described by incorporating the QTL genotype effects into model (1) as

$$y_{it} = \sum_{j=1}^J x_{ij}^* \mu_{jt} + \sum_{j=1}^M \sum_{k=1}^{K_j} x_{ijk} B_{jkt} + e_{it}, \tag{4}$$

$j \neq s, s+1$

$i = 1, \dots, n; t = 1, \dots, T$

where  $\mu_{jt}$  is the value of the  $j$ th QTL genotype at the  $t$ th time point;  $J$  is the number of QTL genotypes, determined by the QTL segregation type, possibly taking the value of 2, 3 or 4;  $M$  is the number of markers chosen as background markers in CIM;  $x_{ij}^*$  is an indicator variable for the  $j$ th QTL genotype for the  $i$ th individual, taking the value of 1 or 0; The other variables are defined as in model (1). Let  $B$  denote the matrix composed of non-redundant parameters of  $B_{jkt}$  with the  $t$ th column corresponding to the  $t$ th time point, and  $X_i$  the row vector for individual  $i$

corresponding to the coefficients of parameter  $B_{jkt}$  in any column of matrix  $B$ . The likelihood of the unknown parameters in model (2) can be written as

$$L(\Theta) = \prod_{i=1}^n L_i(\Theta) = \prod_{i=1}^n \sum_{j=1}^J p_{ij} f(y_i; \mu_j + B'X_i, \Sigma) \tag{5}$$

where  $\Theta = (\mu_1, \dots, \mu_J, B, \Sigma)$  is the unknown parameters,

$$y_i = (y_{i1} \ y_{i2} \ \dots \ y_{iT})', f(y_i; \mu_j + B'X_i, \Sigma) = \frac{T}{2|\Sigma|} \frac{1}{2e^{-\frac{1}{2}(y_i - \mu_j - B'X_i)' \Sigma^{-1}(y_i - \mu_j - B'X_i)}}$$

and  $p_{ij}$  is the conditional probability of the  $j$ th QTL genotype on the flanking marker genotype. Although there are many cases for the combination of any two markers due to several different marker segregation types, the conditional probability can be calculated in a uniform procedure [6].

Differentiating eq. (5) with respect to the unknown parameters of  $\mu_j$ s,  $B$  and  $\Sigma$ , and setting these partial derivatives to zero, we obtained the following equations as

$$B = (X'X)^{-1}X' \left( Y - \sum_{j=1}^J P_j \otimes \mu_j \right) \tag{6}$$

$$\mu_j = \frac{\sum_{i=1}^n P_{ij}(y_i - B'X_i)}{\sum_{i=1}^n P_{ij}} \quad (j = 1, \dots, J) \tag{7}$$

$$\Sigma = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^J P_{ij} (y_i - \mu_j - B'X_i)(y_i - \mu_j - B'X_i)' \tag{8}$$

Where

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}, P_j = \begin{pmatrix} P_{1j} \\ P_{2j} \\ \vdots \\ P_{nj} \end{pmatrix} \quad (j = 1, \dots, J)$$

and

$$P_{ij} = \frac{p_{ij} f(y_i; \mu_j + B'X_i, \Sigma)}{\sum_{j=1}^J p_{ij} f(y_i; \mu_j + B'X_i, \Sigma)} \quad (j = 1, \dots, J) \tag{9}$$

To obtain the maximum likelihood estimates (MLEs) of the unknown parameters, we performed the expectation-maximization (EM) algorithm [38]. In the E-step, the posterior probability of the  $j$ th QTL genotype for individual  $i$  was calculated by eq. (7) with initial values of the unknown parameters. In the M-step, the

estimates of parameters  $B$ ,  $\mu_j$ s and  $\Sigma$  were calculated by eqs. (4–6), respectively. The two steps were repeated until all the parameters converged.

To test if there is a significant QTL at a specified position of the genome, a null hypothesis was claimed as

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_j \tag{10}$$

The log-likelihood ratio (LR) statistic can be used for the test as

$$LR = 2 \log \left[ \frac{L(\hat{\theta})}{L(\hat{\theta}_0)} \right] \tag{11}$$

where  $\hat{\theta}$  and  $\hat{\theta}_0$  are the two MLE sets of all the parameters under the full model and the reduced model, respectively. Generally, the critical threshold for asserting a QTL existence can be determined by performing permutation tests [39, 40].

### QTL model selection

As described above, the QTL at a fixed position of the genome may segregate in several different patterns, but the true segregation pattern is unknown a priori. Therefore, a model selection method will be helpful to infer the QTL segregation. Here, we applied Akaike’s information criterion (AIC) [41], Bayesian information criterion (BIC) [42] and Takeuchi’s information criterion (TIC) [43] to infer the best QTL segregation pattern among the five alternatives. These criteria are defined as

$$AIC = -2 \log L(\hat{\theta}) + 2d \tag{12}$$

$$BIC = -2 \log L(\hat{\theta}) + \log(n)d \tag{13}$$

$$TIC = -2 \log L(\hat{\theta}) + 2 \text{tr} \left( \hat{J}(\hat{\theta}) \hat{I}^{-1}(\hat{\theta}) \right) \tag{14}$$

where  $L(\hat{\theta})$  is the maximum likelihood of the model,  $d$  is the number of parameters to be estimated in the model, and  $J(\hat{\theta})$  and  $I(\hat{\theta})$  can be calculated as

$$\hat{J}(\hat{\theta}) = \sum_{i=1}^n \left( \frac{\partial \log L_i(\hat{\theta})}{\partial \theta} \right) \left( \frac{\partial \log L_i(\hat{\theta})}{\partial \theta} \right)' \tag{15}$$

$$\hat{I}(\hat{\theta}) = - \frac{\partial^2 \log L(\hat{\theta})}{\partial \theta^2} \tag{16}$$

The first and second derivatives involved in Eqs. (15) and (16) can be derived as in Additional file 1: Appendixes S1 and S2. We chose a proper index for discriminating QTL patterns by assessing the power of each criterion through computer simulations.

### Monte Carlo simulation

In order to validate the accuracy of parameter estimates and to evaluate the power of each model selection index, we performed a large number of computer simulations. Five chromosomes were considered in our simulations, each with 100 cM long and six markers evenly distributed. These simulated markers have the segregation types of  $aa \times ab$ ,  $ab \times aa$ ,  $ab \times ab$  and  $ab \times cd$ , and the linkage phases between any two adjacent markers are not fixed. Five QTLs with segregation patterns of  $QQ \times Qq$ ,  $Qq \times QQ$ ,  $Qq \times Qq$ ,  $Qq \times qQ$  and  $Q_1Q_2 \times Q_3Q_4$  were supposed to control tree height in a growth period, whose positions on genome and genotype effects at eight sequential time points were set as shown in Tables 1 and Additional file 2: Tables S1-S5. Here, for consistency, a QTL genotype effect is defined as the deviation from the mean of the genotype values. The phenotype values of the  $i$ th individual were sampled from the multivariate normal distribution as  $N(v_i, \Sigma)$ , where the mean vector  $v_i$  is the sum of the overall mean vector

$$v = (34.46, 69.94, 83.93, 103.54, 114.73, 120.80, 124.01, 125.69)$$

and the combination genotype effects of all the five QTLs involved over the eight time points. The covariance matrix  $\Sigma$  was determined by setting the sum of the heritabilities of all the five QTLs to 0.9 at each time point and the correlation coefficient of trait values between the  $i$ th and  $j$ th time points equal to  $0.9^{|i-j|}$ , which can be calculated as

**Table 1** The assumed QTL segregation patterns, positions on genome and the power of detecting the true QTL pattern with different model selection criteria under different sample sizes

Size	Pattern	Chromosome	Interval	Position	AIC	BIC	TIC
300	$QQ \times Qq$	1	1	5	972	1000	975
	$Qq \times Qq$	2	6	5	942	1000	947
	$Q_1Q_2 \times Q_3Q_4$	4	2	3	1000	920	1000
	$Qq \times QQ$	4	5	8	984	1000	988
	$Qq \times qQ$	5	3	7	918	1000	943
200	$QQ \times Qq$	1	1	5	957	1000	960
	$Qq \times Qq$	2	6	5	909	969	923
	$Q_1Q_2 \times Q_3Q_4$	4	2	3	998	634	991
	$Qq \times QQ$	4	5	8	967	1000	972
	$Qq \times qQ$	5	3	7	917	972	950
150	$QQ \times Qq$	1	1	5	957	1000	950
	$Qq \times Qq$	2	6	5	873	849	904
	$Q_1Q_2 \times Q_3Q_4$	4	2	3	986	426	925
	$Qq \times QQ$	4	5	8	953	1000	928
	$Qq \times qQ$	5	3	7	895	839	940

$$\Sigma = \begin{pmatrix} 13.02 & 18.38 & 19.03 & 20.97 & 21.43 & 20.81 & 19.58 & 18.08 \\ 18.38 & 32.04 & 33.17 & 36.55 & 37.34 & 36.26 & 34.12 & 31.51 \\ 19.03 & 33.17 & 42.40 & 46.72 & 47.73 & 46.36 & 43.62 & 40.28 \\ 20.97 & 36.55 & 46.72 & 63.55 & 64.92 & 63.05 & 59.33 & 54.79 \\ 21.43 & 37.34 & 47.73 & 64.92 & 81.89 & 79.53 & 74.83 & 69.10 \\ 20.81 & 36.26 & 46.36 & 63.05 & 79.53 & 95.35 & 89.73 & 82.85 \\ 19.58 & 34.12 & 43.62 & 59.33 & 74.83 & 89.73 & 104.24 & 96.25 \\ 18.08 & 31.51 & 40.28 & 54.79 & 69.10 & 82.85 & 96.25 & 109.73 \end{pmatrix} \tag{17}$$

We considered sample sizes of 300, 200 and 150 each with 1000 replicates. For each case, the average parameter estimates and their standard deviations were calculated. In addition, under the three different model selection criteria described above, the power of detecting a specific QTL segregation pattern for each QTL model was obtained by counting the number of runs out of the 1000 repeats in which the correct pattern was chosen.

**Implementation**

We developed a command-based software, namely *mvqtlcim*, to implement the computing for our CIM mapping method in an outbred full-sib family. *Mvqtlcim* was written in C++ with Boost C++ 1.62 (<http://www.boost.org>) and can run on Windows, Linux and Mac OS operating systems. The software utilizes a genetic linkage map constructed with different segregation molecular markers such as 1:1, 1:2:1 and 1:1:1:1, and assumes that QTL may segregate in the five different segregation patterns on a specific position of the genetic map. It allows users to select the best QTL segregation pattern with AIC, BIC and TIC for a significant QTL. It also provides command line parameters to be chosen for alternative analyses, including the number of background markers, window size [44], QTL segregation type, genetic map function and number of permutations. Specifically, when performing permutations to determine the empirical threshold of significant QTLs, *mvqtlcim* permits to use multithreads to accelerate computing speed. When an analysis completes, the software will generate two files for each QTL model, of which one contains the parameter estimates and the corresponding statistic values at every 1 cM on the genome, and the other saves the maximum LR value of each permutation. With these result files, we wrote an R script, *lrPlot.r*, to summarize the significant QTL information and generate scatter plots of LR against genome position. These plots can be optionally saved in pdf, jpg, png, tif or bmp format. The software and R script with the manuals are available from <https://github.com/tongchf/mvqtlcim>.

**Results**

**Monte Carlo simulation**

A large number of computer simulations were performed under different scenarios of sample sizes to assess the power of selecting the optimal QTL segregation pattern and the accuracy and precision of parameter estimates, using our multivariate CIM method with the background marker number of 5 and the window size of 15.0 cM. Table 1 shows the power of our statistical model to select the correct QTL segregation pattern among the five alternatives with AIC, BIC and TIC criteria under three different sample sizes. It is observed that all the powers for distinguishing the five QTL patterns are very high ( $\geq 93\%$ ) when the sample size is 300. Although the powers of BIC for the QTL segregation pattern of  $Q_1Q_2 \times Q_3Q_4$  are significantly lower (63.4% and 42.6%) under the sample sizes of 200 and 150, the powers of all the criteria for the other cases are still high ( $\geq 83.9\%$ ). It is interesting to note that the powers of AIC and TIC consistently keep high levels whatever the sample size is large or small, but the powers of TIC are more stable than those of AIC and keep at high levels of  $>90\%$ . Therefore, the TIC criterion is highly recommended to use for selecting the best QTL segregation pattern with the CIM method developed here for an outbred  $F_1$  population.

Additional file 2: Tables S1-S5 list the parameter estimates in detail of the QTL position and genotype effects at each time point under the three cases of sample sizes. Overall, the estimated QTL positions tend towards the setting locations. But for the three QTL segregation patterns of  $Q_1Q_2 \times Q_3Q_4$ ,  $Qq \times QQ$  and  $Qq \times qQ$ , which were set in non-central locations, the position estimates are a little biased to the interval center. The average estimates of QTL genotype effects at the different time points for each case of the QTL segregation pattern are well close to the true values, but the standard deviations expand as expected when the sample size decreases from 300 to 150. Therefore, on average, the heritability of each QTL at each time point closes to the previously set value, and the sum of all the five QTL heritabilities at each time point is around the set value of 90% (Additional file 2: Table S6). In contrast, the estimate of the residual covariance matrix  $\Sigma$  for each QTL segregation pattern under each sample size expands averagely 2–3 times compared with the sum of the variances over the eight time points set in eq. (15) (Additional file 2: Table S7).

**QTL mapping in *Populus***

We performed QTL mapping for the tree heights over 6 time points in the  $F_1$  hybrid population of *P. deltoides*  $\times$  *P. simonii* with the new developed tool *mvqtlcim*. The linkage maps used for QTL mapping were two parental specific maps; All the markers on the maternal map

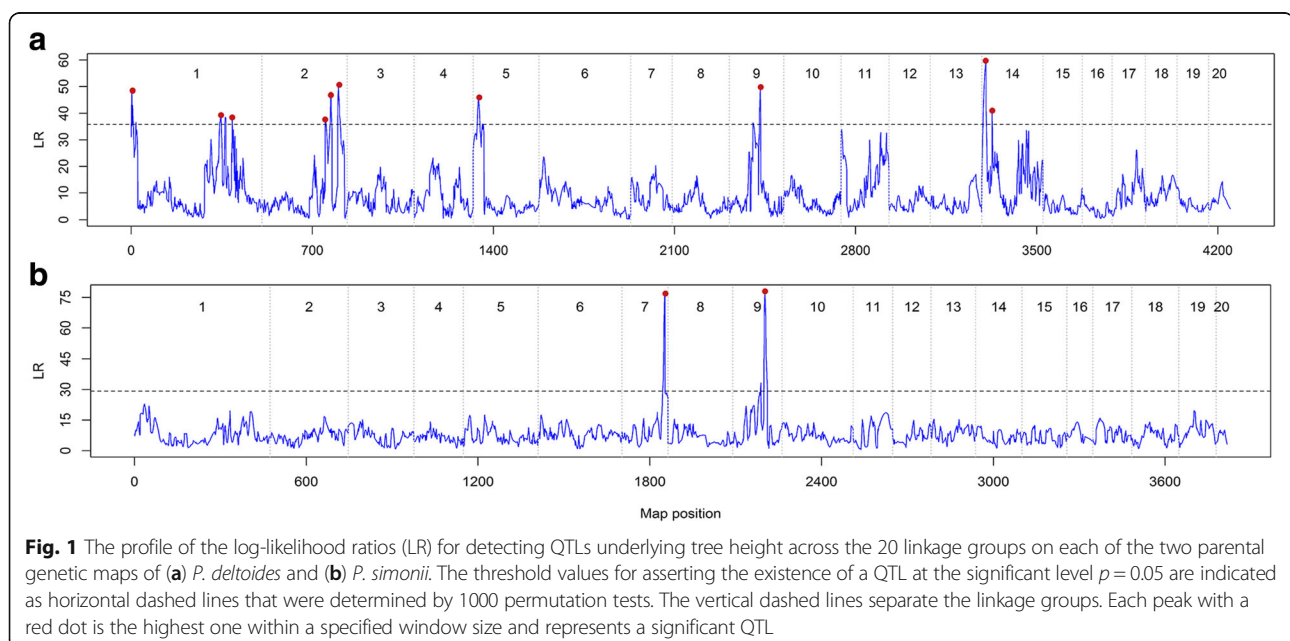
segregate in the type of  $ab \times aa$ , while the markers on the paternal map in the type of  $aa \times ab$  [2]. Therefore, the QTL segregation patterns were assumed to be  $Qq \times Qq$  for the maternal map and  $QQ \times Qq$  for the paternal map when scanning QTLs. In order to obtain the optimal mapping result, we ran *mvqtlcim* with different number of background markers and different window sizes, leaving the other optional parameters as defaults. The number of background markers was iterated from 3 to 39 with a step length of 2 and the window size from 5.0 to 30.0 cM with a step length of 5.0 cM. The optimal mapping result was defined as the one that all the significant QTLs account for the maximum proportion of the phenotypic variance in the population.

With the maternal linkage map, we found that the optimal mapping result corresponding to the run with 29 background markers and the window size of 20.0 cM, leading to 10 significant QTLs detected. The threshold determined by 1000 permutations was 35.84 for asserting the existence of a QTL at the significant level of 0.05. Fig. 1(a) displays the scatter plot of the LR against the position of the linkage map of *P. deltoides* with the dashed threshold line. A significant QTL corresponds to a peak which is above the threshold. If more than one significant peaks are within the specified window size, we chose the highest one as a significant QTL and ignored the others. It can be seen that the identified QTLs are distributed on the linkage groups of 1, 2, 5, 9 and 14. In the same way, we detected two significant QTLs on the paternal linkage map of *P. simonii* under the experimental threshold value of 29.23 with 3 background markers and the window size of 10.0 cM in running *mvqtlcim* (Fig. 1(b)). Table 2 summarizes the position,

effects at each time point and the average heritability over the six time points for each significant QTL. These QTL IDs were named after the linkage group number, the order within a linkage group and either of the two parental linkage maps, where D stands for *P. deltoides* and S for *P. simonii* (e.g. Q2D1 indicates the second QTL located in group 1 on the linkage map of *P. deltoides*). It is observed that, on average, Q1D14 explains the maximum proportion (27.43%) of the phenotypic variance, while Q1D1 accounts for the minimum (only 1.11%).

### Candidate gene investigations

In order to investigate the candidate genes of these QTLs, we searched for the coding genes within the physical interval of each QTL in the gene annotation database of *Populus trichocarpa* v3.0 at Phytozome (<https://phytozome.jgi.doe.gov>). Because of the limited information in the annotation database [45], the coding sequences (CDS) of those genes related to each QTL were re-annotated by first blasting and then mapping on Gene Ontology (GO) terms with Blast2GO (<https://www.blast2go.com>). Consequently, the genomic region covering a QTL has an average length of 801 kb (Table 2) and contained 7–247 genes, of which 79% have 19.7 blast hits and 5.0 GO terms received on average (Additional file 3: Excel Sheets Q1D1-QS9). Additional file 4: Figures S1–12 showed the biological process GO category for the genes within the local region of each QTL. Interestingly, we found that the biological processes (BP) of three genes (Potri.014G029100, Potri.014G031100, Potri.014G031300) in the interval of Q1D14 and one gene (Potri.014G041600) in the interval of Q2D14 involved in brassinosteroids, which have great effects on plant height [46]. Another



**Fig. 1** The profile of the log-likelihood ratios (LR) for detecting QTLs underlying tree height across the 20 linkage groups on each of the two parental genetic maps of (a) *P. deltoides* and (b) *P. simonii*. The threshold values for asserting the existence of a QTL at the significant level  $p = 0.05$  are indicated as horizontal dashed lines that were determined by 1000 permutation tests. The vertical dashed lines separate the linkage groups. Each peak with a red dot is the highest one within a specified window size and represents a significant QTL

**Table 2** Summary of the identified QTL position, LR, effect of QQ at each time point and the average heritability over the time points on the two parental linkage maps of *P. deltoides* and *P. simonii*

Linkage Map	QTL ID	Chr/LG <sup>a</sup>	Marker Interval	Map Position (cM)	Genome Position <sup>b</sup> (Mb)	Region Length <sup>b</sup> (kb)	LR	T1 <sup>c</sup>	T2	T3	T4	T5	T6	Average Heritability (%)
<i>P. deltoides</i>	Q1D1	1	2	2.12	1.24	701	48.24	-5.83	-4.39	-5.34	-7.32	-8.89	-11.59	1.11
	Q2D1	1	145	346.67	31.30	332	39.11	27.01	30.86	35.04	34.92	36.87	38.24	22.13
	Q3D1	1	159	390.19	35.01	57	38.11	-10.24	-13.18	-15.49	-18.51	-20.04	-24.91	6.00
	Q1D2	2	99	244.74	20.86	739	38.14	-26.04	-24.07	-22.74	-22.00	-21.80	-21.26	10.14
	Q2D2	2	102	265.38	19.82	1799	46.40	-29.44	-28.21	-27.24	-26.36	-26.57	-26.78	14.40
	Q3D2	2	111	296.05	23.72	222	49.57	23.82	25.13	23.93	22.51	23.07	23.25	10.67
	QD5	5	3	20.13	2.24	844	46.32	-17.64	-19.79	-21.67	-22.95	-22.22	-23.74	8.78
	QD9	9	44	119.31	8.18	344	49.43	21.82	26.66	24.76	27.05	25.34	21.66	11.60
	Q1D14	14	1	16.01	1.37	1734	59.85	-33.74	-37.42	-38.52	-40.54	-38.65	-38.13	27.43
	Q2D14	14	12	39.12	3.10	529	41.03	22.96	26.29	29.36	30.96	28.58	29.43	15.03
<i>P. simonii</i>	QS7	7	28	149.75	15.28	1115	77.47	7.16	7.90	12.36	11.03	19.11	29.92	5.25
	QS9	9	31	112.18	7.87	1200	77.59	11.91	14.76	19.68	19.57	25.60	35.70	9.74

<sup>a</sup>Chr, chromosome; LG, linkage group

<sup>b</sup>Estimated by the flanking SNPs on the reference sequence of *Populus trichocarpa* v3.0

<sup>c</sup>T1-T6 are the QTL effects of genotype QQ over six time point

interesting finding was that two candidate genes (Potri.014G016500, Potri.014G027200) in the interval of Q1D14 and one gene in the interval of QD5 (Potri.005G021800) were related to shoot formation or development. Moreover, candidate genes for embryo or root development can be found in the flanking regions of Q1D14, QD5, QD9, QS7 and QS9, and for response to stress such as salt and heat in the regions of Q1D1, Q2D1, Q1D2, Q2D2, Q3D2, Q1D14 and QS9. Additionally, we also searched the candidate genes associated with photosynthesis, which plays the most important role in tree growth and development [47]. As a result, the candidate genes related to photosynthesis were located in the regions of Q1D2, Q2D2, Q1D14, QD5 and QS9. Other interesting candidate genes could be searched out in the Blast2GO annotation results presented in Additional file 3: Excel Sheets Q1D1-QS9 and Additional file 4: Figures S1–12.

## Discussion

Statistical methods for QTL mapping have been greatly developed for the past three decades, from the seminal work of interval mapping by Lander and Botstein [7] to the recent more popular Bayesian LASSO approaches [19–23]. However, there were few successful examples of identifying QTLs in outbred forest trees. One of the reasons may be due to the fact that most QTL mapping tools are not for the outbred species in which inbred lines are difficultly or even impossibility derived. Here, we implemented the traditional CIM method into an F<sub>1</sub> population generated by hybridizing two outbred parents for mapping multiple or longitudinal traits. It is essentially useful for forest trees because such species have

the characteristic of long generation times and high heterozygosity so that phenotypic data over long time can be easily observed but the genetic structures are more complicated. With the model selection criterion of TIC, our method could discriminate a QTL segregation type among five alternatives with a higher power (see section 3.2). In contrast to our previous work [6], the IM method with the LEC criterion for mapping a single trait could select an appropriate QTL segregation type by considering only three alternative patterns. Compared with the recent work of Gazaffi et al. [33], our work has a great advantage in the aspect of inferring a QTL segregation pattern (as described in introduction). We also provided the software *mvqtlcim* to put our method into practice. The software permits to use multithreads for performing a large number of permutations to determine the experimental threshold of LR for a significant QTL.

With the multivariate linear model method and EM algorithm, our CIM approach for mapping multivariate data traits has the advantage that the MLEs of unknown parameters can be globally obtained with limited iterative steps at each position on genome. However, in most functional QTL mapping cases [27, 28, 48, 49], owing to the nonlinear growth curves involved in the statistical models, the parameter MLEs could not be always obtained globally. This may decrease the power of identifying QTLs or even possibly generate pseudo QTLs. To overcome the problem in functional QTL mapping, we could first use the multivariate CIM method proposed here to identify QTLs and then to find the growth curves of these QTL genotypes. One strategy is to derive the nonlinear growth curve using the function mapping

method within a small region flanking a QTL, which allows to obtain the optimal solution by iterating over different initiative points with intensive computing. Another way is to directly fit the growth curve with the QTL genotype values over time estimated from our multivariate CIM method. The latter method was illustrated by fitting the estimated genotype values for the 12 QTLs identified in section 3.3 with the Richards' growth curves [50] (Additional file 5: Figure S13).

The results of Monte Carlo simulations indicated that our QTL mapping approach can provide accurate estimates of genetic parameters and a high power of inferring the actual QTL segregation type, but the estimate of the residual covariance matrix  $\Sigma$  expanded several times (Additional file 2: Table S7) compared with the setting values. It is noted that the estimate of residual variance was not assessed and ignored in the pioneer work of CIM approach [10]. This inconsistency between the estimates and the setting values in the residual covariance matrix could be explained by the fact that the setting model for simulations contains all the five QTL effects while the CIM model focuses only one QTL effect at a specific position on genome. The other QTL effects cannot be fully absorbed by the background markers in the CIM model, thus leading to the expanded estimates of the residual errors.

The application of mapping QTLs in *Populus* illustrated that our new multivariate CIM method could detected more number of QTLs underlying tree height in this study than in a previous study (12 vs. 8), in which a modified CIM was applied for tree height measured at a single time point [2]. These included some small-effect QTLs, such as Q1D1, QS9 and Q3D1 that averagely accounted for 1.11%, 5.25% and 6.00% of the phenotypic variances over the 6 time points, respectively (Table 2). This may be the main reason that our QTL mapping approach allows more QTLs to be detected. We also noted that the QTL effect size was not consistent with the LR statistic in our multivariate CIM mapping. For example, Q1D1 has a bigger LR value than Q2D2 (48.24 vs. 39.11), but its heritability is much lower than the later (1.11% vs. 22.13%) (Table 2). The reason is that the CIM statistical model may be different for different positions on genome because the background markers and their number in the model vary with the detected position. Therefore, the LR values of QTLs cannot be compared with each other to determine if one would more significant than the other. However, the LR threshold for significant QTLs is strictly valid in statistics because it was determined by the LR values each with the largest value chosen from a different permutation over the whole genome positions.

Further compared with other previous studies in mapping *Populus* height, our method may not only find

more number of QTLs but also increase the genetic variance explained by them. In the early 1990s, Bradshaw and Stettler [51] found one QTL underlying 2-year height on linkage group D, which accounted for 25.9% of the phenotypic variance in an  $F_2$  population derived by crossing *P. trichocarpa* and *P. deltoides*. Later on, Wu (1998) [52] detected two QTLs for 3-year height on linkage groups D and M with the same materials, totally explaining 27.3% of the phenotypic variance. Because the relationship between *Populus* linkage groups and chromosomes was not so clear in the early two studies, we could not match the QTL positions to our present results. Recently, Monclus et al. (2012) [45] identified 5 QTLs distributed on chromosomes 1, 5, 6, 10 and 14 for the first-year height (Height1) and 7 QTLs on chromosomes 4, 6, 10, 12, 13, 16 and 17 for the second-year height (Height2) using 330  $F_1 P. deltoides \times P. trichocarpa$  progeny. These QTLs could explain 20~30% of the phenotypic variance for Height1 or Height2, but only two QTLs were located consistently on the same chromosomes (6 and 10) for the heights of the 2 years even with the same mapping materials. Among these QTLs, three for Height1 estimated in the confidence intervals of 23.12–54.23 Mb on chromosome 1, 7.11–25.80 Mb on chromosome 5, and 4.87–12.49 Mb on chromosome 14 seem to be in agreement with the QTLs identified in this study that were located in the positions of 31.30/35.01 Mb (Q2D1/Q3D1), 2.24 Mb (QD5), and 1.37 Mb (Q1D14) on the corresponding chromosomes. More recently, Du et al. [53] identified three QTLs affecting tree height in an  $F_1$  population of *Populus*, which were located in linkage groups 8 (Chr01), 12 and 16 (Chr13), and accounted for 3.4%, 8.0% and 6.4% of the phenotypic variance, respectively. One QTL was estimated in the interval of 18.37–21.00 Mb on the same chromosome (Chr01) as the QTLs of Q1D1, Q2D1 and Q3D1 detected in this study, but it was over 10 Mb away from any one of the three QTLs (Table 2). These comparisons between the previous and current studies displayed a large difference in identifying QTLs for *Populus* height, though a few consistent cases existed. The reason may be due to many factors such as mapping materials, genetic data structures, measures of phenotypic traits, and statistical methods [54].

Finally, we also conducted QTL analysis for our *Populus* real datasets each from one parental linkage map using the popular LASSO method with the glmnet/R package (v2.0–10, <http://www.stanford.edu/~hastie/Papers/glmnet.pdf>). In order to select a stable optimal value of the tuning parameter, the leave-one-out cross-validation was performed for each dataset (Additional file 6: Figure S14). As a result, a total of 12 SNPs were identified to be associated with the tree height, exactly half of which come from each SNP dataset. Among these associated SNPs, three were detected



consistently by both CIM and LASSO (Additional file 6: Table S8). The high level of inconsistency between the results of CIM and LASSO was also observed in the most recent work of Xu and his colleagues [54], where they identified 28 and 29 QTLs for eight yield traits in maize by CIM and LASSO, respectively, but only half were consistent with both methods. The reason may be due to the difference in the way to utilize marker information in the two methods. The CIM method takes use of not only the marker segregation information but also the information of marker linkage as well as linkage phase, thus capable of detecting a QTL in an interval of two adjacent markers. However, although LASSO can handle a whole marker dataset simultaneously, it only uses the marker genotype information and provides the associated information between markers and a phenotypic trait. Perhaps, each of the two methods has its own advantages in such a hard task of QTL identification.

## Conclusion

The traditional CIM method was implemented for mapping multiple or longitudinal traits in a full-sib family derived by crossing two outbred parents. Our method not only incorporated various marker segregation ratios, such as 1:1, 1:2:1 and 1:1:1:1, but also utilized the model selection index of TIC to discriminate the actual QTL segregation pattern among several possible alternatives. We provided a powerful tool package to implement the algorithms of our method, which is freely available at the website: <https://github.com/tongchf/mvqtlcim>. The software package will facilitate studying of QTL mapping and thus will accelerate molecular breeding programs especially in forest trees.

## Additional files

**Additional file 1: Appendices S1 and S2.** (DOCX 93 kb)

**Additional file 2: Table S1.** Average of parameter estimates with the standard deviation in bracket under different sample sizes when the QTL segregation type is  $QQ \times Qq$ , based on 1000 simulation replicates. **Table S2.** Average of parameter estimates with the standard deviation in bracket under different sample sizes when the QTL segregation type is  $Qq \times Qq$ , based on 1000 simulation replicates. **Table S3.** Average of parameter estimates with the standard deviation in bracket under different sample sizes when the QTL segregation type is  $Q1Q2 \times Q3Q4$ , based on 1000 simulation replicates. **Table S4.** Average of parameter estimates with the standard deviation in bracket under different sample sizes when the QTL segregation type is  $Qq \times QQ$ , based on 1000 simulation replicates. **Table S5.** Average of parameter estimates with the standard deviation in bracket under different sample sizes when the QTL segregation type is  $Qq \times qQ$ , based on 1000 simulation replicates. **Table S6.** Summary on average estimates of QTL heritabilities (%) with the standard deviation in brackets under different time points (T1-T8) and different sample sizes based on 1000 simulation replicates. **Table S7.** The average estimate of the residual covariance matrix with standard deviations in brackets when the sample size is 300 and the QTL segregation type is  $Qq \times qQ$ , based on 1000 simulation replicates. (DOCX 42 kb)

**Additional file 3:** Excel Sheets from Q1D1 to Q59. (XLSX 97 kb)

**Additional file 4: Figure S1.** Biological process GO category for the genes within the region of QTL Q1D1. **Figure S2.** Biological process GO category for the genes within the region of QTL Q2D1. **Figure S3.** Biological process GO category for the genes within the region of QTL Q3D1. **Figure S4.** Biological process GO category for the genes within the region of QTL Q1D2. **Figure S5.** Biological process GO category for the genes within the region of QTL Q2D2. **Figure S6.** Biological process GO category for the genes within the region of QTL Q3D2. **Figure S7.** Biological process GO category for the genes within the region of QTL QD5. **Figure S8.** Biological process GO category for the genes within the region of QTL QD9. **Figure S9.** Biological process GO category for the genes within the region of QTL Q1D14. **Figure S10.** Biological process GO category for the genes within the region of QTL Q2D14. **Figure S11.** Biological process GO category for the genes within the region of QTL Q57. **Figure S12.** Biological process GO category for the genes within the region of QTL Q59. (DOCX 642 kb)

**Additional file 5: Figure S13.** Richards' growth curves of the 12 QTLs underlying the tree height of *Populus*, fitted with their genotype values (dot) over time estimated from the multivariate CIM method. The red is for the genotype QQ and the blue for Qq. (PDF 375 kb)

**Additional file 6: Figure S14.** Plots of the mean cross-validated error against the log of parameter lambda for the female (a) and male (b) SNP datasets. **Table S8.** SNPs identified to be associated with *Populus* height by the LASSO method using the two SNP datasets from each parental linkage map. (DOCX 61 kb)

## Abbreviations

AIC: Akaike's information criterion; BIC: Bayesian information criterion; CIM: Composite interval mapping; EM: Expectation-maximization; IM: Interval mapping; LASSO: Least absolute shrinkage and selection operator; LR: Log-likelihood ratio; MLE: Maximum likelihood estimate; NGS: Next-generation sequencing; QTL: Quantitative trait locus; TIC: Takeuchi's information criterion

## Acknowledgements

Not applicable

## Availability of data and material

The software *mvqtlcim* on different operating systems and an example input file are available at <https://github.com/tongchf/mvqtlcim>. Additional file 1 contains Appendices S1 and S2, Additional file 2: Tables S1-S7, Additional file 3: Supplementary Excel Sheets Q1D1-Q59, Additional file 4: Figures S1-S12, Additional file 5: Figure S13, and Additional file 6: Figure S14 and Table S8.

## Funding

This work has been supported by the National Natural Science Foundation of China (No. 3127076) and the Priority Academic Program Development of the Jiangsu Higher Education Institutions (PAPD). Neither organization contributed to the design or conclusions of this study.

## Authors' contributions

CT, HL, and JS conceived of the study. FL and CT developed the software and wrote the manuscript. FL performed QTL mapping in *Populus*. ST, JW, YC, and DY measured the tree height. All authors read and approved the final version of this manuscript.

## Ethics approval and consent to participate

Not applicable

## Consent for publication

Not applicable

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>The Southern Modern Forestry Collaborative Innovation Center, College of Forestry, Nanjing Forestry University, Nanjing 210037, China. <sup>2</sup>College of Department of Computer Science and Engineering, Sanjiang University, Nanjing 210012, China.

Received: 11 April 2017 Accepted: 1 November 2017

Published online: 23 November 2017

**References**

- 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.
- Tong CF, Li HG, Wang Y, Li XR, Ou JJ, Wang DY, et al. Construction of high-density linkage maps of *Populus deltoides* × *P. simonii* using restriction-site associated DNA sequencing. *PLoS One.* 2016;11(3):e0150692.
- Mousavi M, Tong C, Liu F, Tao S, Wu J, Li H, et al. De novo SNP discovery and genetic linkage mapping in poplar using restriction site associated DNA and whole-genome sequencing technologies. *BMC Genomics.* 2016;17:656.
- Maliepaard C, Jansen J, Van Ooijen JW. Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. *Genet Res.* 1997;70:237–50.
- Wu RL, Ma CX, Painter I, Zeng ZB. Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. *Theor Popul Biol.* 2002;61(3):349–63.
- Tong CF, Zhang B, Li HG, Shi JS. Model selection for quantitative trait loci mapping in a full-sib family. *Genet Mol Biol.* 2012;35(3):622–31.
- Lander ES, Botstein D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics.* 1989;121:185–99.
- Doerge RW. Mapping and analysis of quantitative trait loci in experimental populations. *Nat Rev Genet.* 2002;3(1):43–52.
- Zeng Z-B. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc Natl Acad Sci U S A.* 1993;90(23):10972–6.
- Zeng Z-B. Precision mapping of quantitative trait loci. *Genetics.* 1994;136:1457–68.
- Kao C-H, Zeng Z-B, Teasdale RD. Multiple interval mapping for quantitative trait loci. *Genetics.* 1999;152(3):1203–16.
- Zeng Z-B, Kao C-H, Basten CJ. Estimating the genetics architecture of quantitative traits. *Genet Res.* 1999;74:279–89.
- Xu S, Atchley WR. Mapping quantitative trait loci for complex binary diseases using line crosses. *Genetics.* 1996;143:1417–24.
- Xu C, Li Z, Xu S. Joint mapping of quantitative trait loci for multiple binary characters. *Genetics.* 2005;169:1045–59.
- Satagopan JM, Yandell BS, Newton MA, Osborn TC. A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics.* 1996;144:805–16.
- Yi N, Yandell BS, Churchill GA, Allison DB, Eisen EJ, Pomp D. Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics.* 2005;170(3):1333–44.
- Xu S. An empirical Bayes method for estimating epistatic effects of quantitative trait loci. *Biometrics.* 2007;63(2):513–21.
- Huang A, Xu S, Cai X. Empirical Bayesian elastic net for multiple quantitative trait locus mapping. *Heredity (Edinb).* 2015;114(1):107–15.
- Yi N, Xu S. Bayesian LASSO for quantitative trait loci mapping. *Genetics.* 2008;179(2):1045–55.
- Mutshinda CM, Sillanpää MJ. Extended Bayesian LASSO for multiple quantitative trait loci mapping and unobserved phenotype prediction. *Genetics.* 2010;186:1067–75.
- Cai X, Huang A, Xu S. Fast empirical Bayesian LASSO for multiple quantitative trait locus mapping. *BMC Bioinformatics.* 2011;12:211.
- Fang M, Jiang D, Li D, Yang R, Fu W, Pu L, et al. Improved LASSO priors for shrinkage quantitative trait loci mapping. *Theor Appl Genet.* 2012;124(7):1315–24.
- Li J, Wang Z, Li R, Wu R. Bayesian group lasso for nonparametric varying-coefficient models with application to functional genome-wide association studies. *Ann Appl Stat.* 2015;9(2):640–64.
- Jiang C, Zeng Z-B. Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics.* 1995;140(3):1111–27.
- Da Costa ESL, Wang S, Zeng Z-B. Multiple trait multiple interval mapping of quantitative trait loci from inbred line crosses. *BMC Genet.* 2012;13:67.
- Macgregor S, Knott SA, White I, Visscher PM. Quantitative trait locus analysis of longitudinal quantitative trait data in complex pedigrees. *Genetics.* 2005;171(3):1365–76.
- Wu RL, Lin M. Functional mapping – how to map and study the genetic architecture of complex dynamic traits. *Nat Rev Genet.* 2006;7:229–37.
- Li Y, Wu R. Functional mapping of growth and development. *Biol Rev Camb Philos Soc.* 2010;85(2):207–16.
- Wang Z, Wang Y, Wang N, Wang J, Wang Z, Vallejos CE, et al. Towards a comprehensive picture of the genetic landscape of complex traits. *Brief Bioinform.* 2014;15(1):30–42.
- Cao J, Wang L, Huang Z, Gai J, Wu R. Functional mapping of multiple dynamic traits. *J Agric Biol Environ Stat.* 2016;22(1):60–75.
- Haley CS, Knott SA, Elsen JM. Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics.* 1994;136:1195–207.
- Lin M, Lou XY, Chang M, Wu R. A general statistical framework for mapping quantitative trait loci in nonmodel systems: issue for characterizing linkage phases. *Genetics.* 2003;165(2):901–13.
- Gazaffi R, Margarido GRA, Pastina MM, Mollinari M, Garcia AAF. A model for quantitative trait loci mapping, linkage phase, and segregation pattern estimation for a full-sib progeny. *Tree Genet Genomes.* 2014;10(4):791–801.
- Van Ooijen JW. MapQTL 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species. *Kyazma BV.* Wageningen, Netherlands. <https://www.kyazma.nl/index.php/MapQTL>. Accessed 1 Jan 2009.
- Zhang B, Tong CF, Yin T, Zhang X, Zhuge Q, Huang M, et al. Detection of quantitative trait loci influencing growth trajectories of adventitious roots in *Populus* using functional mapping. *Tree Genet Genomes.* 2009;5:539–52.
- Rao CR. *Linear statistical inference and its applications.* 2nd ed. New York: Wiley; 1973.
- Johnson RA, Wichern DW. *Applied multivariate statistical analysis.* 6th ed. Beijing: Tsinghua University Press; 2008.
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via EM algorithm. *J R Stat Soc Ser B (Methodological).* 1977;39:1–38.
- Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. *Genetics.* 1994;138:963–71.
- Doerge RW, Churchill GA. Permutation tests for multiple loci affecting a quantitative character. *Genetics.* 1996;142(1):285–94.
- Akaike H. A new look at the statistical model identification. *IEEE Trans Automatic Control AC.* 1974;19:716–23.
- Schwarz G. Estimating the dimension of a model. *Ann Stat.* 1978;6:461–4.
- Takeuchi K. Distribution of informational statistics and a criterion of model fitting. *Suri-Kagaku (Mathematic Sciences, In Japanese).* 1976;153:12–8.
- Wang S, Basten CJ, Zeng Z-B. Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. <http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>. Accessed 1 July 2006.
- Monclus R, Leple JC, Bastien C, Bert PF, Villar M, Marron N, et al. Integrating genome annotation and QTL position to identify candidate genes for productivity, architecture and water-use efficiency in *Populus* spp. *BMC Plant Biol.* 2012;12:173.
- Dubouzet JG, Strabala TJ, Wagner A. Potential transgenic routes to increase tree biomass. *Plant Sci.* 2013;212:72–101.
- Wang L, Wang B, Du Q, Chen J, Tian J, Yang X, et al. Allelic variation in PtoPsbW associated with photosynthesis, growth, and wood properties in *Populus tomentosa*. *Mol Gen Genomics.* 2016;292(1):77–91.
- Huang ZW, Tong CF, Bo WH, Pang XM, Wang Z, Xu JC, et al. An allometric model for mapping seed development in plants. *Brief Bioinform.* 2014;15(4):562–70.
- Tong CF, Shen LY, Lv YF, Wang Z, Wang XL, Feng SS, et al. Structural mapping: how to study the genetic architecture of a phenotypic trait through its formation mechanism. *Brief Bioinform.* 2014;15(1):43–53.
- Kshirsagar AM, Smith WB. *Growth curves.* New York: Marcel Dekker; 1995.
- 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;139:963–73.
- Wu RL. Genetic mapping of QTLs affecting tree growth and architecture in *Populus*: implication for ideotype breeding. *Theor Appl Genet.* 1998;96:447–57.

53. Du Q, Gong C, Wang Q, Zhou D, Yang H, Pan W, et al. Genetic architecture of growth traits in *Populus* revealed by integrated quantitative trait locus (QTL) analysis and association studies. *New Phytol.* 2016;209(3):1067–82.
54. Su C, Wang W, Gong S, Zuo J, Li S, Xu S. High density linkage map construction and mapping of yield trait QTLs in maize (*Zea mays*) using the genotyping-by-sequencing (GBS) technology. *Front Plant Sci.* 2017;8:706.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

