Theor Appl Genet (2011) 122:33–48
DOI 10.1007/s00122-010-1420-8

ORIGINAL PAPER

Mapping of epistatic quantitative trait loci in four-way crosses

Xiao-Hong He · Hongde Qin · Zhongli Hu · Tianzhen Zhang · Yuan-Ming Zhang

Received: 21 August 2009 / Accepted: 24 July 2010 / Published online: 9 September 2010

Abstract Four-way crosses (4WC) involving four different inbred lines often appear in plant and animal commercial breeding programs. Direct mapping of quantitative trait loci (QTL) in these commercial populations is both economical and practical. However, the existing statistical methods for mapping QTL in a 4WC population are built on the single-QTL genetic model. This simple genetic model fails to take into account QTL interactions, which play an important role in the genetic architecture of complex traits. In this paper, therefore, we attempted to develop a statistical method to detect epistatic QTL in 4WC population. Conditional probabilities of QTL genotypes, computed by the multi-point single locus method, were used to sample the genotypes of all putative QTL in the entire genome. The sampled genotypes were used to construct the design matrix for QTL effects. All QTL effects, including main and epistatic effects, were simultaneously estimated by the penalized maximum likelihood method. The proposed method was confirmed by a series of Monte Carlo simulation studies and real data analysis of cotton. The new method will provide novel tools for the genetic dissection of complex traits, construction of QTL networks, and analysis of heterosis.

Communicated by F. van Eeuwijk.

X.-H. He · Y.-M. Zhang (✉)
Section on Statistical Genomics, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, People's Republic of China
e-mail: soyzhang@njau.edu.cn

X.-H. He · Y.-M. Zhang
Chinese National Center for Soybean Improvement, Nanjing Agricultural University, Nanjing 210095, People's Republic of China

H. Qin · T. Zhang
Cotton Research Institute, State Key Laboratory of Crop Genetics and Germplasm Enhancement, College of Agriculture, Nanjing Agricultural University, Nanjing 210095, People's Republic of China

Z. Hu
Key Lab of the Ministry of Education for Plant Developmental Biology, College of Life Science, Wuhan University, Wuhan 430072, People's Republic of China

Introduction

Since the 1980s, quantitative trait locus (QTL) mapping has been increasingly used to understand the genetic architecture of quantitative traits. In plants and laboratory animals, most QTL mapping strategies focus on segregating populations derived from an initial cross between two inbred lines. However, this biparental-based population mapping strategy has several shortcomings. First, this simple cross is rarely used alone in commercial breeding, and therefore the results from these single-cross experiments have limited roles in breeding practice (Liu and Zeng 2000; Blanc et al. 2006; Verhoeven et al. 2006). Second, the statistical inference space is limited to the two inbred lines, and the results from the line cross cannot be generalized to other line crosses. If the two lines being crossed do not segregate at particular QTL, no matter how many offspring are sampled in the mapping population, the QTL cannot be detected and type II error is occurred (Xu 1996). Finally, for most outbred species, such as most trees and livestock, inbred lines could not be developed, making the biparental-based mapping strategy impractical. To overcome these issues, four-way cross (4WC) strategy was developed. It is often used in the commercial breeding of animals and plants. Direct mapping of QTL in a 4WC

population is both economical and practical because the population being mapped is readily available and the identified QTL are directly applicable (Rao and Xu 1998; Harmegnies et al. 2006; Qin et al. 2008). Due to its multi-parent cross characteristic, more QTL can be detected with 4WC than with simple crosses and the type II error, caused by random sampling of parents, can be reduced (Xu 1996). Therefore, the 4WC is more valuable than the biparental cross for analyzing the inheritance of complex traits. Moreover, methods of QTL mapping in a 4WC population can be easily adopted in outbred populations, where inbred lines are not available (Groover et al. 1994; Haley et al. 1994; Knott et al. 1997, 1998; Rao and Xu 1998; Xu 1998). The advantages of detecting QTL from 4WC justify seeking improvements to the QTL mapping methodology.

Statistical methods for QTL mapping are well established. These methods can be classified based on parameter estimation into linear regression (Haley and Knott 1992; Martínez and Curnow 1992), maximum likelihood (Lander and Botstein 1989; Jansen and Stam 1994; Zeng 1994; Kao et al. 1999; Zhang and Xu 2005), Bayesian (Hoeschele and Vanranden 1993a, b; Satagopan et al. 1996; Sillanpää and Arjas 1998; Wang et al. 2005), empirical Bayesian (Xu 2007; Xu and Jia 2007), Bayesian LASSO (Yi and Xu 2008; Park and Casella 2008), the hierarchical generalized linear model (Yi and Banerjee 2009), and nonparametric methods (Kruglyak and Lander 1995). However, all of these approaches focus on segregating populations derived from an initial cross between two inbred lines rather than on the 4WC.

To date, several theoretical studies on the mapping of 4WC have been carried out. In the early 1990s, the analysis of variance technique was first applied to identify QTL influencing wood specific gravity in an outbred pedigree of loblolly pine (Groover et al. 1994). Haley et al. (1994) and Xu (1996) extended the simple linear regression (Haley and Knott 1992) to the 4WC. However, estimation of the residual variance with the simple regression method of Haley and Knott (1992) is confounded by part of the QTL variance (Xu 1995). To overcome the deficiency, Xu (1998) proposed the use of iteratively reweighted least squares (IRWLS) under a heterogeneous variance model for QTL mapping in 4WC. Thereafter, dominant and missing markers were incorporated into this model (Xie and Xu 1999). Apart from the ordinary quantitative traits, Rao and Xu (1998) developed a generalized linear model approach (GLM) to map QTL for ordered categorical traits in 4WC, and Luo and Xu (2003) suggested a maximum likelihood (ML) method for viability loci mapping in 4WC. However, these methods were all based on the single-QTL genetic model. Estimates of the effects and the positions of QTL will be biased in the presence of more than one QTL in any given linkage group or when epistatic QTL are present (Lander and Botstein 1989; Haley and Knott 1992; Jansen and Stam 1994; Zeng 1994; Kao et al. 1999; Zhang 2006).

Epistasis, the interaction between genes, is a common, key concept for understanding the adaptation and evolution of natural populations, response to selection in breeding programs, and determination of complex disease (Carlborg and Haley 2004). Epistasis is related to canalization and stabilizing selection and may act as a genetic buffer, creating interdependency of genes in the network (Moore 2005). Epistasis can lead to heterosis (Yu et al. 1997; Lynch and Walsh 1998; Lippman and Zamir 2006; Melchinger et al. 2007), which is very important in the hybrid breeding of rice, corn and rapeseed. In addition, epistatic variance may reduce the resemblance of offspring to their parents (Lynch and Walsh 1998). Although Hanlon and Lorenz (2005) proposed a random walk-based method to detect multi-locus QTL interaction and Hanlon et al. (2006) adopted this method to identify the three- and four-locus interactions responsible for mouse insulin-like growth factor-I in one 4WC population, this method did not fully explore the space of epistasis and could not guarantee that the corresponding estimates be the global optimum values. To overcome this issue, several methods are available, i.e., penalized maximum likelihood (PML) (Zhang and Xu 2005). The PML method assumes that there is one putative QTL residing on each marker in the entire genome, considers all main and epistatic effects in one model, adopts a penalty that depends on the values of the parameters, and allows spurious QTL effects to be shrunk towards zero and large QTL effects to be estimated without shrinkage. In the PML marker genotypes are assumed to be known unambiguously. However, in real data analysis marker genotype information is sometimes incomplete. In this situation, multiple imputations for incomplete marker genotypes are generally adopted (Sen and Churchill 2001). In this study, we calculated multi-point single locus conditional QTL probabilities for additive and dominant effects and then took the products of the single locus conditional QTL probabilities as joint conditional QTL probabilities that form the basis for constructing regressors for pairwise epistatic interactions. Therefore, a current focal point is to incorporate epistatic QTL mapping into the 4WC under the framework of full genetic model using the PML approach.

The purpose of this research was to develop a statistical method for systematic and genome-wide mapping of epistatic QTL in 4WC. Detailed genetic and statistical models of QTL mapping using all markers simultaneously have been described. All parameters were estimated by the PML method of Zhang and Xu (2005). Our proposed method

was tested via systemic Monte Carlo simulations and real data analysis in cotton.

Theory and method

Genetic design and data collection

In animal and plant breeding, four inbred lines (GP_1, GP_2, GP_3, and GP_4) are often selected to serve as grandparents. The four lines are used to construct two crosses. One is derived from GP_1 and GP_2 ($F_1 = GP_1 \times GP_2$) and another is from GP_3 and GP_4 ($F_1' = GP_3 \times GP_4$). Then, the F_1 and F_1' are crossed to form a 4WC population (GP_1/GP_2)//(GP_3/GP_4). Phenotypic values of the quantitative trait of interest and information on polymorphic markers, based on pre-designed primers from four grandparents and/or two parents using polymerase chain reaction (PCR) or other techniques, were measured for each individual in the 4WC population.

Genetic model

If a quantitative trait is controlled by a single QTL and the genotypes for the four grandparents are denoted by Q_1Q_1, Q_2Q_2, Q_3Q_3 and Q_4Q_4, respectively, the 4WC population produces four possible genotypes, i.e., Q_1Q_3, Q_1Q_4, Q_2Q_3, and Q_2Q_4. The phenotypic value of individual i in the 4WC can be described as

$$y_i = \mu + x_{i1}a_1 + x_{i2}a_2 + z_i d + e_i \quad (i = 1, 2, \ldots, n) \qquad (1)$$

where μ is the population mean; a_1 (a_2) is the first (second) additive effect, measuring the average allele substitution effect of Q_2 by Q_1 (Q_4 by Q_3); d is the dominance effect; $e_i \sim N(0, \sigma_e^2)$ is random error; and x_{i1}, x_{i2} and z_i are indicator variables, depending on the genotype of individual i (G_i), which can be defined as

$$x_{i1} = \begin{cases} +1 & \text{if } G_i = Q_1Q_3 \\ +1 & \text{if } G_i = Q_1Q_4 \\ -1 & \text{if } G_i = Q_2Q_3 \\ -1 & \text{if } G_i = Q_2Q_4 \end{cases},$$

$$x_{i2} = \begin{cases} +1 & \text{if } G_i = Q_1Q_3 \\ -1 & \text{if } G_i = Q_1Q_4 \\ +1 & \text{if } G_i = Q_2Q_3 \\ -1 & \text{if } G_i = Q_2Q_4 \end{cases} \quad \text{and}$$

$$z_i = \begin{cases} +1 & \text{if } G_i = Q_1Q_3 \\ -1 & \text{if } G_i = Q_1Q_4 \\ -1 & \text{if } G_i = Q_2Q_3 \\ +1 & \text{if } G_i = Q_2Q_4 \end{cases}$$

If we assume that there is one QTL residing on each marker in the entire genome and all pair-wise epistatic QTL are considered, model (1) should be extended as

$$\begin{aligned} y_i = \mu + & \sum_{j=1}^{q} (x_{ij1}a_{j1} + x_{ij2}a_{j2} + z_{ij}d_j) \\ & + \sum_{r=1}^{q-1} \sum_{s=r+1}^{q} [(x_{ir1}x_{is1})i_{a_{r1}a_{s1}} + (x_{ir1}x_{is2})i_{a_{r1}a_{s2}} + (x_{ir1}z_{is})i_{a_{r1}d_s} \\ & + (x_{ir2}x_{is1})i_{a_{r2}a_{s1}} + (x_{ir2}x_{is2})i_{a_{r2}a_{s2}} + (x_{ir2}z_{is})i_{a_{r2}d_s} \\ & + (z_{ir}x_{is1})i_{d_ra_{s1}} + (z_{ir}x_{is2})i_{d_ra_{s2}} \\ & + (z_{ir}z_{is})i_{d_rd_s}] + e_i \end{aligned} \qquad (2)$$

where q equals the number of markers on the genome; $i_{a_{r1}a_{s1}}$, $i_{a_{r1}a_{s2}}$, $i_{a_{r1}d_s}$, $i_{a_{r2}a_{s1}}$, $i_{a_{r2}a_{s2}}$, $i_{a_{r2}d_s}$, $i_{d_ra_{s1}}$, $i_{d_ra_{s2}}$, and $i_{d_rd_s}$ are epistatic effects between the rth and sth QTL ($r = 1, \ldots, q-1$; $s = r+1, \ldots, q$); and the other parameters are the same as those in model (1).

Because both main and epistatic effects are treated in the same way under the PML, for the sake of clarification, model (2) can be rewritten as

$$y_i = b_0 + \sum_{k=1}^{p} b_k x_{ik} + e_i \qquad (3)$$

where $p = \frac{3}{2}q(3q-1)$ is the total number of genetic effects, $\mathbf{b} = \{b_1, b_2, \ldots, b_p\}^{\mathrm{T}} = \{a_{11}, a_{12}, d_1, \ldots, a_{q1}, a_{q2}, d_q, i_{a_{11}a_{21}}, i_{a_{11}a_{22}}, i_{a_{11}d_2}, i_{a_{12}a_{21}}, i_{a_{12}a_{22}}, i_{a_{12}d_2}, i_{d_1a_{21}}, i_{d_1a_{22}}, i_{d_1d_2}, \ldots, i_{a_{(q-1)1}a_{q1}}, i_{a_{(q-1)1}a_{q2}}, i_{a_{(q-1)1}d_q}, i_{a_{(q-1)2}a_{q1}}, i_{a_{(q-1)2}a_{q2}}, i_{a_{(q-1)2}d_q}, i_{d_{(q-1)}a_{q1}}, i_{d_{(q-1)}a_{q2}}, i_{d_{(q-1)}d_q}\}^{\mathrm{T}}$, $\mathbf{x}'_k = \{x'_{1k}, \ldots, x'_{nk}\}^{\mathrm{T}}$ is an $n \times 1$ incidence vector, corresponding to the effect b_k ($k = 1, \ldots, p$), $b_0 = \mu$, and e_i are defined the same way as in model (1).

Multi-point single locus approach for determining the conditional probabilities of QTL genotypes

QTL genotypes were assumed to be known unambiguously in model (2). In real data analysis, however, QTL genotypes were unobservable if complete marker information was not available. Therefore, we first used marker data from an entire linkage group to calculate the conditional probability of each QTL (marker) genotype using a multi-point single locus approach (Lander and Green 1987; Jiang and Zeng 1997; Rao and Xu 1998); and we then imputed complete marker information using these conditional probabilities (Sen and Churchill 2001). The sampled genotypes were used to assign values for indicator variables of QTL additive and dominant effects and these values were further cross multiplied to assign coefficients for pairwise epistatic effects in model (2). Thus, the design matrix for all QTL effects was constructed.

Consider m ordered marker loci on a chromosome of interest with a known linkage relationship. Let $\{Q_{t1}Q_{t3}\ Q_{t1}Q_{t4}\ Q_{t2}Q_{t3}\ Q_{t2}Q_{t4}\}$ be the four possible genotypes of the tth locus (marker or QTL) and $r_{t(t+1)}$ be the

recombination fractions between the tth and $(t+1)$th loci. The conditional probability of the tth QTL genotype for individual i, $\text{Prob}(G_{it} = G_h|\mathbf{M})$ $(h = 1, \ldots, 4)$, would be computed by

$$\text{Prob}(G_{it} = G_h|\mathbf{M}) = \frac{\mathbf{1}'\mathbf{D}_1\mathbf{T}_{12}\mathbf{D}_2\mathbf{T}_{23}\ldots\mathbf{D}_t\mathbf{T}_{tQ_t}\mathbf{D}_{G_{it}=G_h}\mathbf{T}_{Q_t(t+1)}\mathbf{D}_{(t+1)}\mathbf{T}_{(t+1)(t+2)}\ldots\mathbf{D}_{m-1}\mathbf{T}_{(m-1)m}\mathbf{D}_m\mathbf{1}}{\sum_{o=1}^{4}\mathbf{1}'\mathbf{D}_1\mathbf{T}_{12}\mathbf{D}_2\mathbf{T}_{23}\ldots\mathbf{D}_t\mathbf{T}_{tQ_t}\mathbf{D}_{G_{it}=G_o}\mathbf{T}_{Q_t(t+1)}\mathbf{D}_{(t+1)}\mathbf{T}_{(t+1)(t+2)}\ldots\mathbf{D}_{m-1}\mathbf{T}_{(m-1)m}\mathbf{D}_m\mathbf{1}} \tag{4}$$

where $\mathbf{1}' = [1\ 1\ 1\ 1]$, $\mathbf{D}_t = \text{diag}[\text{Prob}(Q_{t1}Q_{t3})\ \text{Prob}(Q_{t1}Q_{t4})\ \text{Prob}(Q_{t2}Q_{t3})\ \text{Prob}(Q_{t2}Q_{t4})]$ is determined by the phenotype (or genotype) of the tth marker (or QTL). For the tth marker, $\mathbf{D}_t$ values are shown in Table 1, as the JoinMap formatted marker data (Van Ooijen and Voorrips 2001); for the tth QTL, $\mathbf{D}_{G_{it}=Q_{t1}Q_{t3}} = \text{diag}[1\ 0\ 0\ 0]$ for $h = 1$, $\mathbf{D}_{G_{it}=Q_{t1}Q_{t4}} = \text{diag}[0\ 1\ 0\ 0]$ for $h = 2$, $\mathbf{D}_{G_{it}=Q_{t2}Q_{t3}} = \text{diag}[0\ 0\ 1\ 0]$ for $h = 3$, and $\mathbf{D}_{G_{it}=Q_{t2}Q_{t4}} = \text{diag}[0\ 0\ 0\ 1]$ for $h = 4$; $\mathbf{M}$ represents marker information; and

$$\mathbf{T}_{t(t+1)} = \begin{bmatrix} (1-r_{t(t+1)})^2 & r_{t(t+1)}(1-r_{t(t+1)}) & r_{t(t+1)}(1-r_{t(t+1)}) & r_{t(t+1)}^2 \\ r_{t(t+1)}(1-r_{t(t+1)}) & (1-r_{t(t+1)})^2 & r_{t(t+1)}^2 & r_{t(t+1)}(1-r_{t(t+1)}) \\ (1-r_{t(t+1)})r_{t(t+1)} & r_{t(t+1)}^2 & (1-r_{t(t+1)})^2 & (1-r_{t(t+1)})r_{t(t+1)} \\ r_{t(t+1)}^2 & r_{t(t+1)}(1-r_{t(t+1)}) & (1-r_{t(t+1)})r_{t(t+1)} & (1-r_{t(t+1)})^2 \end{bmatrix} \tag{5}$$

is the transition probability matrix from the tth locus to the $(t+1)$th locus.

Parameter estimation

Several methods exist to estimate the parameters in model (3), e.g., PML (Zhang and Xu 2005), Bayesian LASSO (Yi and Xu 2008; Park and Casella 2008) and hierarchical generalized linear model (Yi and Banerjee 2009). Here, we adopt the PML.

Let $\boldsymbol{\theta} = (b_0, b_1, b_2, \ldots, b_p, \sigma^2)$ be the vector of parameters of interest. The likelihood function is

$$L(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{M}) = \prod_{i=1}^{n}\varphi(y_i; m_i, \sigma^2) \tag{6}$$

where $\mathbf{Y} = (y_1, y_2, \ldots, y_n)^{\mathrm{T}}$; and $\varphi(y_i; m_i, \sigma^2)$ is a normal probability density function with mean $m_i = b_0 + \sum_{k=1}^{p} b_k x_{ik}$ and variance σ^2.

In the PML, many spurious QTL effects in model (3) should be minimized to zero, whereas QTL with large effects are estimated with virtually no reduction. Therefore, the objective function that should be maximized is penalized likelihood, which is the product of likelihood function and penalty function. As suggested by Zhang and Xu (2005), the penalty function is

$$P(\boldsymbol{\theta}, \boldsymbol{\xi}) = \prod_{l=1}^{p}\left[\varphi(b_l; \mu_l, \sigma_l^2)\varphi(\mu_l; 0, \sigma_l^2/\eta)\right], \tag{7}$$

where $\boldsymbol{\xi} = (\mu_1, \mu_2, \ldots, \mu_p, \sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2)$ is the vector of hyperparameters and $\eta > 0$ is the prior sample size for accessing μ_k. Therefore, the penalized likelihood is defined as

$$\psi(\boldsymbol{\theta}, \boldsymbol{\xi}) = L(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{M})P(\boldsymbol{\theta}, \boldsymbol{\xi}). \tag{8}$$

The penalized maximum likelihood estimates (PMLE) of model parameter $\boldsymbol{\theta}$ and the nuisance parameter $\boldsymbol{\xi}$ can be obtained by maximizing $\ln\psi(\boldsymbol{\theta}, \boldsymbol{\xi})$. Thus, the PMLE for the above parameters are

$$\begin{aligned} b_0 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \sum_{l=1}^{p}x_{il}b_l) \\ b_l &= \left[\sum_{i=1}^{n}x_{il}^2 + \sigma^2/\sigma_l^2\right]^{-1} \\ &\quad \times \left[\sum_{i=1}^{n}x_{il}(y_i - b_0 - \sum_{t\neq l}^{p}x_{it}b_t) + \mu_l\sigma^2/\sigma_l^2\right] \\ \sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - b_0 - \sum_{l=1}^{p}x_{il}b_l)^2 \\ \mu_l &= b_l/(\eta+1) \end{aligned} \tag{9}$$

Table 1 Diag($\mathbf{D}_t$) values, determined by segregation type, phase type and offspring marker phenotypes (JoinMap data format)

Segregation type	Offspring marker phenotype	Phase			
		{00}	{01}	{10}	{11}
ab × cd	ac	[1 0 0 0]	[0 1 0 0]	[0 0 1 0]	[0 0 0 1]
	ad	[0 1 0 0]	[1 0 0 0]	[0 0 0 1]	[0 0 1 0]
	bc	[0 0 1 0]	[0 0 0 1]	[1 0 0 0]	[0 1 0 0]
	bd	[0 0 0 1]	[0 0 1 0]	[0 1 0 0]	[1 0 0 0]
ef × eg	ee	[1 0 0 0]	[0 1 0 0]	[0 0 1 0]	[0 0 0 1]
	ef	[0 0 1 0]	[0 0 0 1]	[1 0 0 0]	[0 1 0 0]
	eg	[0 1 0 0]	[1 0 0 0]	[0 0 0 1]	[0 0 1 0]
	fg	[0 0 0 1]	[0 0 1 0]	[0 1 0 0]	[1 0 0 0]
hk×hk	hh	[1 0 0 0]	[0 1 0 0]	[0 0 1 0]	[0 0 0 1]
	hk	[0 1/2 1/2 0]	[1/2 0 0 1/2]	[1/2 0 0 1/2]	[0 1/2 1/2 0]
	kk	[0 0 0 1]	[0 0 1 0]	[0 1 0 0]	[1 0 0 0]
	h-	[1/3 1/3 1/3 0]	[1/3 1/3 0 1/3]	[1/3 0 1/3 1/3]	[0 1/3 1/3 1/3]
	k-	[0 1/3 1/3 1/3]	[1/3 0 1/3 1/3]	[1/3 1/3 0 1/3]	[1/3 1/3 1/3 0]
		{0-}	{1-}		
lm × ll	ll	[1/2 1/2 0 0]	[0 0 1/2 1/2]		
	lm	[0 0 1/2 1/2]	[1/2 1/2 0 0]		
		{-0}	{-1}		
nn × np	nn	[1/2 0 1/2 0]	[0 1/2 0 1/2]		
	np	[0 1/2 0 1/2]	[1/2 0 1/2 0]		
		all			
all	–	[1/4 1/4 1/4 1/4]			

Diag($\mathbf{D}_t$)=[$P_{t1}(M_{t1}M_{t3})$ $P_{t2}(M_{t1}M_{t4})$ $P_{t3}(M_{t2}M_{t3})$ $P_{t4}(M_{t2}M_{t4})$]. Phase {0} denotes that the grandparental origin is consistent with the segregation type. Phase {1} denotes that the grandparental origin is switched. In two-digit phase type code, the first is related to F_1 derived from P_1 and P_2, whereas the second is related to F_1 derived from P_3 and P_4. h- and k- are dominant phenotypes; and '–' denotes missing phenotype

$$\sigma_l^2 = \frac{1}{2}[(b_l - \mu_l)^2 + \eta\mu_l^2],$$

where $l = 1, 2, \ldots, p$, and $\eta = 5$ (Zhang and Xu 2005). The prior sample size η does not have a major influence on the result as long as $\eta < \infty$ (Zhang and Xu 2005). The initial values $\boldsymbol{\theta}^{(0)} = \{b_0^{(0)}, b_1^{(0)}, b_2^{(0)}, \ldots, b_p^{(0)}, \sigma^{2(0)}\} = \{\bar{y}, 0, \ldots, 0, s_y^2\}$. are suggested for $\boldsymbol{\theta}$, where $\bar{y}$ and s_y^2 are the sample mean and variance of the phenotypic values. The initial values $\boldsymbol{\xi}^{(0)} = \left\{\mu_1^{(0)}, \mu_2^{(0)}, \ldots, \mu_p^{(0)}, \sigma_1^{2(0)}, \sigma_2^{2(0)}, \ldots, \sigma_p^{2(0)}\right\} = \{0,0,\ldots,0,0.5,0.5,\ldots,0.5\}$ for $\boldsymbol{\xi}$. Iteration is then performed using Eq. (9) and terminated when a predetermined convergence criterion, i.e., 10^{-15}, is satisfied.

Statistical test

As described by Zhang and Xu (2005), the usual likelihood ratio test (LRT) cannot be carried out with the PML method owing to oversaturated epistatic genetic model. We proposed the following two-stage selection process to screen the markers (Zhang and Xu 2005). In the first stage, all markers with $\left|\hat{b}_k/\hat{\sigma}\right| > 10^{-6}$ are picked up. In the second stage, the epistatic genetic model is modified so that only effects past the first round of selection are included in the model. Owing to the smaller dimensionality of the reduced model, we can use the maximum likelihood method to reanalyze the data and perform the LRT. The procedure for the LRT is as follows.

The overall null hypothesis is no effect of the QTL at the locus of interest, denoted by $H_0 : a_1 = a_2 = d = 0$ or $H_0 : \mathbf{Lu} = 0$, where $\mathbf{L} = \{1\ 0\ 0; 0\ 1\ 0; 0\ 0\ 1\}$ and $\mathbf{u} = \{a_1\ a_2\ d\}^{\mathrm{T}}$. If we solve the maximum likelihood estimates of the parameters under the restriction of $\mathbf{Lu} = 0$ and calculate the log-likelihood value using the solutions with this restriction, we obtain $L(\hat{\boldsymbol{\theta}}|\mathbf{Lu} = 0)$. Meanwhile, we can also evaluate the log-likelihood value of the solutions without restriction and obtain $L(\hat{\boldsymbol{\theta}})$. Therefore, the LR test statistic is

$$\mathrm{LR} = -2\left[L(\hat{\boldsymbol{\theta}}|\mathbf{Lu} = 0) - L(\hat{\boldsymbol{\theta}})\right]. \quad (10)$$

Various other test statistics can be used by redefining the $\mathbf{L}$ matrix. To test the hypothesis of $H_1 : a_1 = 0$, for example, we define $\mathbf{L}_1 = \{1\ 0\ 0\}$. The LR test statistic is $\mathrm{LR}_1 = -2\left[L(\hat{\boldsymbol{\theta}}|\mathbf{L}_1\mathbf{u} = 0) - L(\hat{\boldsymbol{\theta}})\right]$.

For epistatic QTL, we may define $\mathbf{L} = \text{diag}\ \{1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\}_{15\times 15}$ and $\mathbf{u} = \{a_{11}\ a_{12}\ \ d_1\ a_{21}\ a_{22}\ d_2\ i_{a_{11}a_{21}}\ i_{a_{11}a_{22}}\ i_{a_{11}d_2}\ \ i_{a_{12}a_{21}}\ i_{a_{12}a_{22}}\ i_{a_{12}d_2}\ i_{d_1a_{21}}\ i_{d_1a_{22}}\ i_{d_1d_2}\}^{\mathrm{T}}$. In the same way, the significance of epistatic effects can be tested. The significance threshold of the LOD score is set at 2.0 for our simulated data and at 3.0 for our real data, where LOD = LR/4.605.

Monte Carlo simulation studies

Simulation design

We conducted four simulation experiments to evaluate the performance of the proposed method. In the first simulation experiment, three chromosomes, each with eight equally spaced markers (15 cM apart), were simulated. Three QTL were located 60.0, 45.0 and 60.0 cM, respectively, from the left-hand end of each chromosome. Genetic effects for the three QTL were: $a_{11} = 2.0$ (marginal variance component: 4.00, see Appendix 1), $a_{12} = 1.5$ (2.25) and $d_1 = 1.0$ (1.00) for the first QTL (partial dominance); $a_{21} = -1.5$ (2.25), $a_{22} = -1.0$ (1.00) and $d_2 = 2.0$ (4.00) for the second QTL (overdominance); and $a_{31} = a_{32} = 0.00$ (0.00) and $d_3 = -1.5$ (2.25) for the third QTL (complete dominance). In addition, there was an additive-by-additive interaction between the second and third QTL, and its epistatic effect $i_{a_{21}a_{32}}$ was assumed to be 1.50 (2.25). The marginal genetic variances explained by the three main-effect QTL were 7.25, 7.25 and 2.25, respectively. The total genetic variance of the trait (σ_g^2) was 19.00. The environmental variance was calculated by $\sigma_e^2 = (1-h^2)\sigma_g^2 \big/ h^2$, where h^2 was the heritability. The sample size was set at 300. Each case was replicated 100 times. For each simulated QTL, we counted the samples in which the LOD statistic was greater than 2.0 and the identified QTL was within 30 cM of the simulated QTL. The estimate for QTL parameter was the average of the corresponding estimates in the counted samples. The ratio of the number of such samples to the total number of replicates represented the empirical power of this QTL. The false-positive rate (FPR) was the ratio of the number of false-positive effects to the total number of zero effects considered in model (3). The other three simulation experiments were carried out similarly and detailed simulation parameters are presented in Table 2.

Effect of QTL heritability on QTL mapping

The first simulation experiment was designed to investigate the effect of heritability on mapping of QTL in a 4WC population with a sample size of 300. Changing the value of the residual variance resulted in the total heritability of the

Table 2 Simulated experimental parameters

Simulation experiment	Three main-effect QTL				Epistatic QTL		Sample size	Heritability	Marker information	No. of replications
	Chromosome: position (cM)	a_1	a_2	d	Position 1 × Position 2	$i_{a_{21}a_{32}}$				
1	1:60.0, 2:45.0, 3:60.0	2.0, −1.5, 0.0	1.5, −1.0, 0.0	1.0, 2.0, −1.5	45.0 × 60.0	1.5	300	0.20, 0.40, 0.60, 0.80	full	100
2	1:60.0, 2:45.0, 3:60.0	2.0, −1.5, 0.0	1.5, −1.0, 0.0	1.0, 2.0, −1.5	45.0 × 60.0	1.5	250, 300, 400, 500, 600, 1000	0.50	Full	100
3	1:60.0, 2:45.0, 3:60.0	2.0, −1.5, 0.0	1.5, −1.0, 0.0	1.0, 2.0, −1.5	45.0 × 60.0	1.5	300	0.50	Missing: 5.0, 10.0, 20.0 Dominant: 12.5, 25.0, 37.5	200
4	1:35.0, 2:52.5, 3:70.0	2.0, −1.5, 0.0	1.5, −1.0, 0.0	1.0, 2.0, −1.5	52.5 × 70.0	1.5	300	0.60	Full	100

three main-effect QTL and one epistatic QTL for a quantitative trait to be set at four levels: 0.20, 0.40, 0.60 and 0.80. Table 3 shows the effect of QTL heritability on QTL mapping. The effects and positions for main and epistatic QTL could be estimated without bias in all simulated heritability situations. As expected, the accuracy of the estimates of the effects and positions of the QTL, as well as the empirical power, increased with heritability. Most coefficients of variance (CV) were less than 30%, and the CV fell below $\sim$10% when the marginal variance of a genetic effect accounted for >10% of the total phenotypic variance. For an effect that accounted for <2.5% of the phenotypic variation, e.g., d_1 and a_{22} in the case of the second heritability level and a_{21}, d_3 and $i_{a_{21}a_{32}}$ in the first heritability level, the empirical power was approximately 60%. When the marginal heritability of a QTL was more than 3%, the empirical power was generally greater than 90%. The FPR in the detection of QTL was relatively low, demonstrating a trend that decreased as heritability increased.

Effect of sample size on QTL mapping

In the second simulation experiment, we evaluated the effect of sample size on mapping of QTL by setting the number of individuals in the 4WC population to 250, 300, 400, 500, 600 and 1,000. The results from the simulated experiment in Table 4 show the general behavior of QTL mapping, i.e., as sample size increases the result becomes better (judged by the decrease in the standard deviation and FPR, and by the increase in empirical power). For an effect that accounted for about 5.0% of the phenotypic variation (e.g., a_{12}, a_{21}, d_3 and $i_{a_{21}a_{32}}$), the power was more than 95% and the CV was less than 30.0% for a sample size of 250, the power increased to 100% and the CV decreased to about 15% when the sample size was increased to 500. For our general purposes, a sample size of 500 was used.

Effect of missing and dominant markers on QTL mapping

This simulation experiment was intended to evaluate the effect of incomplete marker information, i.e., missing and dominant markers, on mapping QTL in a 4WC population. Three levels of marker information content were simulated and compared: (1) all markers were codominant with no missing marker data; (2) all markers were codominant with 5, 10 and 20% random missing marker data; and (3) markers were codominant and dominant, with dominant proportions of 12.5% (the 5th marker on chromosomes 1 and 3, and the 4th marker on chromosome 2), 25.0% (the 4th and 5th markers on chromosomes 1 and 3, and the 3rd and 4th markers on chromosome 2) and 37.5% (the 4th to 6th markers on chromosomes 1 and 3, and the 3rd to 5th

Table 3 Effect of QTL heritability on the results obtained for epistatic QTL mapping in four-way crosses (100 replicates)

Heritability (%)	Statistic	b_0	σ^2	QTL_1				QTL_2				QTL_3		$QTL_2 \times QTL_3$			False-positive rate (%)
				a_{11}	a_{12}	d_1	Posi.	a_{21}	a_{22}	d_2	Posi.	d_3	Posi.	$i_{a_{21}a_{32}}$	Posi.	Posi.	
True value		100.00	–	2.00	1.50	1.00	60.00	−1.50	−1.00	2.00	45.00	−1.50	60.00	1.50	45.00	60.00	
0.20	Mean	100.046	62.522	1.853	1.508	1.432	57.81	−1.548	−1.300	1.861	46.21	−1.597	59.24	1.529	43.09	56.71	0.27
	SD	0.758	9.859	0.665	0.393	0.393	7.57	0.371	0.431	0.522	12.95	0.611	9.41	0.521	21.45	18.72	
	Power			0.90	0.49	0.26		0.59	0.27	0.76		0.59		0.68			
0.40	Mean	100.059	23.566	1.909	1.404	1.032	59.45	−1.376	−0.986	1.901	44.01	−1.295	59.78	1.435	43.471	59.21	0.24
	SD	0.438	3.462	0.423	0.331	0.337	4.41	0.377	0.267	0.403	6.99	0.349	2.71	0.405	8.458	8.36	
	Power			0.98	0.93	0.65		0.96	0.65	0.99		0.97		0.97			
0.60	Mean	99.992	10.510	1.970	1.446	0.924	59.89	−1.467	−0.892	1.932	44.39	−1.480	59.96	1.489	44.901	59.14	0.19
	SD	0.292	1.349	0.213	0.242	0.260	2.38	0.259	0.221	0.274	3.14	0.234	0.28	0.227	1.638	3.33	
	Power			1.00	1.00	0.95		1.00	0.96	1.00		1.00		1.00			
0.80	Mean	99.993	4.159	1.981	1.492	0.987	60.00	−1.483	−0.961	2.004	44.68	−1.488	59.98	1.474	44.987	59.99	0.12
	SD	0.162	0.434	0.127	0.146	0.145	0.20	0.157	0.155	0.149	1.74	0.136	0.25	0.145	0.166	0.26	
	Power			1.00	1.00	1.00		1.00	1.00	1.00		1.00		1.00			

Table 4 Effect of sample size on the results obtained for epistatic QTL mapping in four-way crosses (100 replicates)

Sample size	Statistic	b_0	σ^2	QTL_1				QTL_2				QTL_3		$QTL_2 \times QTL_3$			False-positive rate (%)
				a_{11}	a_{12}	d_1	Posi.	a_{21}	a_{22}	d_2	Posi.	d_3	Posi.	$i_{a_{21}a_{32}}$	Posi.	Posi.	
True value		100.00	–	2.00	1.50	1.00	60.00	−1.50	−1.00	2.00	45.00	−1.50	60.00	1.50	45.00	60.00	
250	Mean	100.054	14.694	1.920	1.402	1.008	58.20	−1.340	−0.954	1.926	44.33	−1.402	59.98	1.463	42.041	59.76	0.25
	SD	0.388	2.490	0.345	0.343	0.319	4.05	0.339	0.270	0.349	4.23	0.361	3.55	0.431	8.147	9.23	
	Power			1.00	0.97	0.67		0.96	0.75	0.99		0.97		0.99			
300	Mean	100.000	15.859	1.922	1.374	0.902	59.61	−1.416	−0.973	1.939	44.92	−1.389	60.13	1.384	44.365	58.71	0.21
	SD	0.329	2.072	0.308	0.295	0.263	2.73	0.351	0.282	0.282	4.90	0.326	1.70	0.348	7.681	6.46	
	Power			1.00	0.99	0.80		1.00	0.770	0.99		0.98		0.98			
400	Mean	100.030	16.539	1.923	1.448	0.951	59.74	−1.456	−0.918	1.962	44.82	−1.426	60.00	1.495	44.450	59.93	0.20
	SD	0.355	1.604	0.295	0.240	0.237	1.93	0.250	0.241	0.252	3.45	0.225	0.61	0.294	5.035	5.40	
	Power			1.00	1.00	0.92		1.00	0.93	1.00		1.00		1.00			
500	Mean	99.976	17.108	1.972	1.441	0.922	59.90	−1.443	−0.943	1.969	44.62	−1.455	60.00	1.452	44.654	59.53	0.17
	SD	0.273	1.181	0.207	0.223	0.227	1.68	0.202	0.236	0.221	3.39	0.253	0.00	0.247	2.865	3.31	
	Power			1.00	1.00	0.95		1.00	0.95	1.00		1.00		1.00			
600	Mean	100.017	17.476	1.939	1.475	0.966	59.93	−1.479	−0.947	1.938	44.96	−1.464	59.94	1.486	44.945	59.88	0.16
	SD	0.269	0.896	0.229	0.188	0.214	0.78	0.179	0.210	0.193	1.48	0.185	0.62	0.213	0.949	1.09	
	Power			1.00	1.00	0.99		1.00	0.99	1.00		1.00		1.00			
1,000	Mean	99.997	18.141	1.979	1.470	0.964	60.01	−1.450	−0.978	1.991	44.87	−1.475	59.99	1.466	44.901	60.08	0.16
	SD	0.204	0.918	0.144	0.153	0.150	0.32	0.151	0.160	0.151	0.78	0.160	0.11	0.149	0.515	0.70	
	Power			1.00	1.00	1.00		1.00	1.000	1.00		1.00		1.00			

Table 5 Effect of marker information on the results obtained for epistatic QTL mapping in four-way crosses (200 replicates)

Marker information		Statistic	b_0	σ^2	QTL_1				QTL_2				QTL_3		$QTL_2 \times QTL_3$			False-positive rate (%)
					a_{11}	a_{12}	d_1	Posi.	a_{21}	a_{22}	d_2	Posi.	d_3	Posi.	$i_{a_{21}a_{32}}$	Posi.	Posi.	
True value			100.00		2.00	1.50	1.00	60.00	−1.50	−1.00	2.00	45.00	−1.50	60.00	1.50	45.00	60.00	
Full information		Mean	99.979	15.601	1.960	1.398	0.942	59.78	−1.391	−0.929	1.941	43.83	−1.425	59.78	1.490	44.582	58.93	0.22
		SD	0.386	2.127	0.302	0.328	0.266	2.96	0.287	0.263	0.316	4.59	0.305	2.87	0.346	6.539	7.21	
		Power			1.000	0.990	0.820		0.980	0.845	0.99		0.99		0.985			
Missing markers	0.05	Mean	99.980	16.232	1.924	1.373	0.932	59.60	−1.393	−0.900	1.862	44.18	−1.376	59.88	1.384	42.911	58.99	0.22
		SD	0.349	2.285	0.289	0.296	0.260	2.86	0.325	0.265	0.342	4.59	0.316	3.82	0.355	8.259	8.00	
		Power			1.000	0.985	0.805		0.995	0.805	1.000		0.990		0.995			
	0.10	Mean	99.961	16.866	1.879	1.322	0.970	59.11	−1.353	−0.955	1.786	43.90	−1.285	59.54	1.348	43.026	59.89	0.23
		SD	0.400	2.375	0.339	0.316	0.319	4.53	0.321	0.277	0.339	5.19	0.347	4.43	0.344	10.703	11.03	
		Power			1.000	0.975	0.750		0.990	0.725	0.99		0.955		1.000			
	0.20	Mean	100.019	17.905	1.835	1.314	0.908	59.08	−1.294	−0.915	1.704	43.62	−1.310	59.78	1.278	42.292	57.85	0.24
		SD	0.407	2.961	0.384	0.349	0.273	5.33	0.332	0.280	0.390	5.45	0.340	4.89	0.379	12.895	12.72	
		Power			1.000	0.945	0.705		0.975	0.760	0.945		0.910		0.960			
Dominant markers	0.125	Mean	100.011	19.701	1.595	1.167	0.784	59.48	−1.423	−0.872	1.399	43.17	−1.110	57.02	1.272	43.889	57.73	0.29
		SD	0.440	3.011	0.377	0.365	0.337	10.38	0.358	0.231	0.419	7.57	0.346	11.29	0.452	13.168	16.56	
		Power			0.975	0.910	0.335		0.965	0.540	0.735		0.780		0.945			
	0.250	Mean	100.002	20.279	1.602	1.225	0.778	60.02	−1.427	−0.945	1.316	43.80	−1.058	59.88	1.199	43.841	58.11	0.29
		SD	0.440	3.125	0.433	0.357	0.255	10.09	0.327	0.295	0.452	8.81	0.356	10.81	0.413	12.640	14.62	
		Power			0.985	0.875	0.235		0.970	0.555	0.735		0.795		0.920			
	0.375	Mean	100.069	20.715	1.398	1.175	0.500	56.27	−1.381	−0.916	1.337	43.58	−1.040	58.64	1.244	42.766	56.37	0.30
		SD	0.503	3.576	0.459	0.347	0.659	11.80	0.379	0.297	0.428	9.14	0.323	15.94	0.442	15.864	19.57	
		Power			0.980	0.910	0.190		0.960	0.555	0.730		0.695		0.905			

markers on chromosome 2). The results in Table 5 show that the accuracy of the estimates of QTL positions, as well as the empirical power, decreased as the rate of the missing marker and the dominant marker increased. The accuracy of the estimates of QTL effects showed little sensitivity to dominant and missing markers. The QTL effect was estimated without bias in the missing marker information situation and was underestimated in dominant marker situations. The effect of the dominant marker seemed greater than that of missing marker. The possible explanation is as below. Although missing marker provides no information at all and dominant marker may provide some information, the proportion of missing marker on the simulated QTL position is small and the proportion of dominant marker on the simulated QTL position is 100%.

Effect of QTL position on mapping of epistatic QTL in 4WC

In the above simulation experiments, all QTL overlapped with markers. In reality, this is not practical. Therefore, the purpose of this simulation experiment was to investigate the effect of the QTL position on mapping of QTL in 4WC. Three simulated QTL were placed to the left, middle and right of one marker interval, respectively. The results are shown in Table 6. The accuracy of the effects and positions of QTL, as well as the empirical power, were satisfied, but were lower than those presented in Table 3; the estimate of QTL effects incorporated a smaller bias when the QTL were not situated exactly on the markers. The FPR of QTL detection was equivalent to previous results (data not shown). It should be noted that the bias in the QTL effect increased as the distance between the QTL and marker increased, and the bias was larger for the dominant effect than for additive effects. These results are explained in Appendix 2.

Real data analysis in cotton

Qin et al. (2008) published results of interval mapping for eight yield traits and five fiber quality traits in a 4WC population from (Simian3/Sumian12)//(Zhong4133/8891) crosses in cotton. The authors genotyped 286 molecular markers from 280 samples in 2004. These markers covered 2,113.3 cM of the entire genome, with an average marker interval of 9.2 cM. A total of four QTL were identified from four linkages groups [D2(1), D2(2), D4(2) and D9] over a 2.5% fiber span length (FL, mm). The sizes of the identified QTL ranged from 2.83 to 11.07% of the phenotypic variance (Qin et al. 2008). The overall contribution of the four QTL to the phenotypic variance was 27.40%.

This dataset for the FL in 2004 was reanalyzed in this study. Due to missing quantitative trait phenotypes, only

Table 6 Effect of QTL position on the results obtained for epistatic QTL mapping in four-way crosses (100 replicates)

Statistic	b_0	σ^2	QTL_1				QTL_2				QTL_3		$QTL_2 \times QTL_3$		
			a_{11}	a_{12}	d_1	Posi.	a_{21}	a_{22}	d_2	Posi.	d_3	Posi.	$i_{a_{21}a_{32}}$	Posi.	Posi.
True value	100.00	–	2.00	1.50	1.00	35.00	−1.50	−1.00	2.00	52.50	−1.50	70.00	1.50	52.50	70.00
Mean	99.976	13.621	1.799	1.334	0.859	33.045	−1.265	−0.833	1.600	52.294	−1.167	71.124	1.319	49.705	68.440
SD	0.465	2.077	0.278	0.320	0.231	5.877	0.337	0.238	0.395	6.708	0.344	6.552	0.374	10.825	12.941
Power(T)			1.00	0.97	0.63		0.99	0.78	0.94		0.98		0.98		
Power(L)			0.85	0.74	0.38		0.58	0.35	0.65		0.27			0.61	0.33
Power(R)			0.30	0.16	0.18		0.42	0.30	0.64		0.78			0.39	0.60
Power(L & R)			0.15	0.02	0.00		0.05	0.00	0.36		0.08			0.10	0.09

Power(T) the total detection power for the assumed QTL effect, *Power(L)* the detection power for the assumed QTL effect when the QTL was mapped on the left marker of the interval, *Power(R)* the detection power for the assumed QTL effect when the QTL was mapped on the right marker of the interval, *Power(L & R)* the detection power for the assumed QTL effect when the QTL was mapped on both the left and right marker of the interval

262 individuals were included in the analysis. The number of markers was reduced to 184 to avoid the high degree of multicollinearity that could be caused by very tightly linked markers. Because of incomplete marker genotypic information, multiple imputations for incomplete marker genotypes (Sen and Churchill 2001) were adopted here to simulate complete information marker genotypes. This requires multiple analyses of the data, each for one imputed data set. Although 10–20 imputed data sets may suffice (Sen and Churchill 2001), we imputed 50 samples in this paper. The critical LOD for declaring a QTL was set at 3.0. The ratio of the number of a QTL detected to the total number of imputed samples represented the relative frequency of this QTL. The phenotype distribution for the FL is depicted in Fig. 1. The FL ranged from 26.57 to 33.27 mm, with a mean of 29.24 mm and a standard deviation of 1.13 mm. The skewness and kurtosis were 0.29 and 0.17, respectively. The results for QTL mapping of the FL are listed in Table 7 for main-effect QTL and in Table 8 for epistatic QTL. In total, 13 main-effect QTL and 5 epistatic QTL were detected. The four main-effect QTL that were originally identified by Qin et al. (2008) using MapQTL 5.0 were all detected by the new method. The relative frequency of the QTL for linkage group D2(2) was very low, in agreement with the low LOD (LOD = 2.1, less than critical LOD of 2.7) described in the study by Qin et al. (2008). The nine additional main-effect QTL mapped in the analysis demonstrated that the new method seemed to be more powerful, even for the detection of main-effect QTL, compared with the interval mapping described by Qin et al. (2008). More importantly, five epistatic QTL were detected using our proposed method. The graphical representation of the phenotypic effects of identified epistatic QTL in Fig. 2 would be helpful to understand the five interactions, because all the paired lines in Fig. 2 were not parallel. It should be noted that all the five epistatic QTL

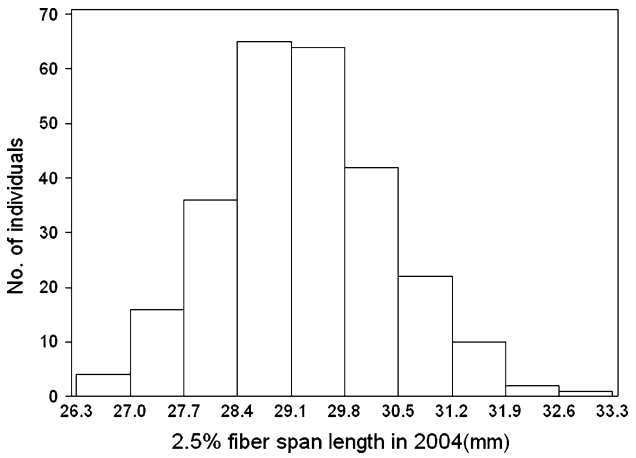

Fig. 1 Phenotypic distribution of 2.5% fiber span length in cotton in 2004. The data were derived from Qin et al. (2008)

Table 7 Main effect QTL controlling 2.5% fiber span length

QTL	Linkage group	Linked marker	Position	a_1			a_2			d			Heritability (%)
				Effect[a]	LOD[a]	Relative frequency[b]	Effect[a]	LOD[a]	Relative frequency[b]	Effect[a]	LOD[a]	Relative frequency[b]	
1	A2(2)	NAU1529	5.891							−0.137 (0.036)	9.769 (5.597)	0.480	1.48
2	A5(2)	CIR253	14.800				−0.146 (0.047)	11.200 (7.195)	0.640				1.68
3	A5(2)	NAU5077	77.773	0.125 (0.045)	9.133 (6.311)	0.140	0.141 (0.045)	9.685 (5.751)	0.340	−0.136 (0.038)	9.752 (5.766)	0.600	4.25
4	A8(2)	JESPR291	4.944	−0.124 (0.031)	8.200(3.999)	0.340	−0.148 (0.063)	9.208 (5.868)	0.160	−0.124 (0.032)	7.519 (3.334)	0.080	4.14
5	D1(3)	NAU3736	30.673				−0.120 (0.036)	7.961 (5.125)	0.540				1.13
6	D2(1)	NAU4024	7.183	0.238 (0.069)	22.152 (9.577)	1.000	−0.173 (0.049)	13.623 (6.443)	0.880				6.81
7	D2(2)	STV030	10.501	0.125 (0.028)	7.333 (3.081)	0.060	−0.141 (0.046)	10.357 (7.587)	0.140	0.108 (0.000)	6.272 (0.000)	0.020	3.71
8	D3	NAU2761	9.412							0.141 (0.045)	10.614 (6.290)	0.500	1.56
9	D4(2)	NAU5236	53.682				0.124 (0.034)	7.982 (4.368)	0.680	0.130 (0.028)	8.568 (2.093)	0.040	2.54
10	D9	BNL1317	85.181	0.136 (0.045)	9.061 (5.466)	0.360	0.164 (0.052)	13.437 (7.185)	0.380				3.57
11	D10	NAU5195	18.752	0.114 (0.025)	6.139 (1.747)	0.180	0.118 (0.030)	7.153 (3.661)	0.400	−0.151 (0.051)	10.802 (6.810)	0.420	3.91
12	D12(2)	NAU3293	1.464	0.122 (0.066)	8.731 (8.794)	0.080	0.141 (0.036)	10.144 (5.073)	0.640				2.73
13	D13(2)	NAU4102	16.546	−0.134 (0.060)	9.618 (7.321)	0.100	0.127 (0.037)	7.981 (3.997)	0.400	−0.110 (0.025)	5.673 (1.426)	0.080	3.63

[a] The numbers in parentheses are the standard deviations for the 50 imputed samples

[b] Relative frequency is the ratio of the number of the detected QTL to the total number of imputed samples. The same is true for Table 8

Table 8 Epistatic QTL controlling 2.5% fiber span length

QTL_1			QTL_2			Type	Effect[a]	LOD[a]	Relative frequency	Heritability (%)
Linkage group	Linked marker	Position	Linkage group	Linked marker	Position					
A10	NAU2532	62.890	D1(3)	NAU3736	34.200	$i_{a_{11}a_{22}}$	−0.134 (0.040)	8.553 (4.390)	0.36	1.41
A10	NAU2542	82.461	D4(2)	JESPR220	29.200	$i_{a_{11}a_{22}}$	−0.106 (0.018)	5.345 (1.727)	0.22	0.88
A13	NAU1522	24.429	D2(1)	NAU4024	1.100	$i_{a_{11}a_{21}}$	0.129 (0.033)	8.293 (4.216)	0.44	1.31
D2(1)	NAU4024	0.000	D13(1)	BNL3558	31.603	$i_{a_{11}a_{22}}$	0.136 (0.035)	8.521 (4.491)	0.20	1.45
D3	NAU2761	9.455	D5(2)	NAU2126	7.807	$i_{a_{11}a_{21}}$	−0.140 (0.040)	9.654 (5.194)	0.20	1.54

[a] The numbers in parentheses are standard deviations for the 50 imputed samples

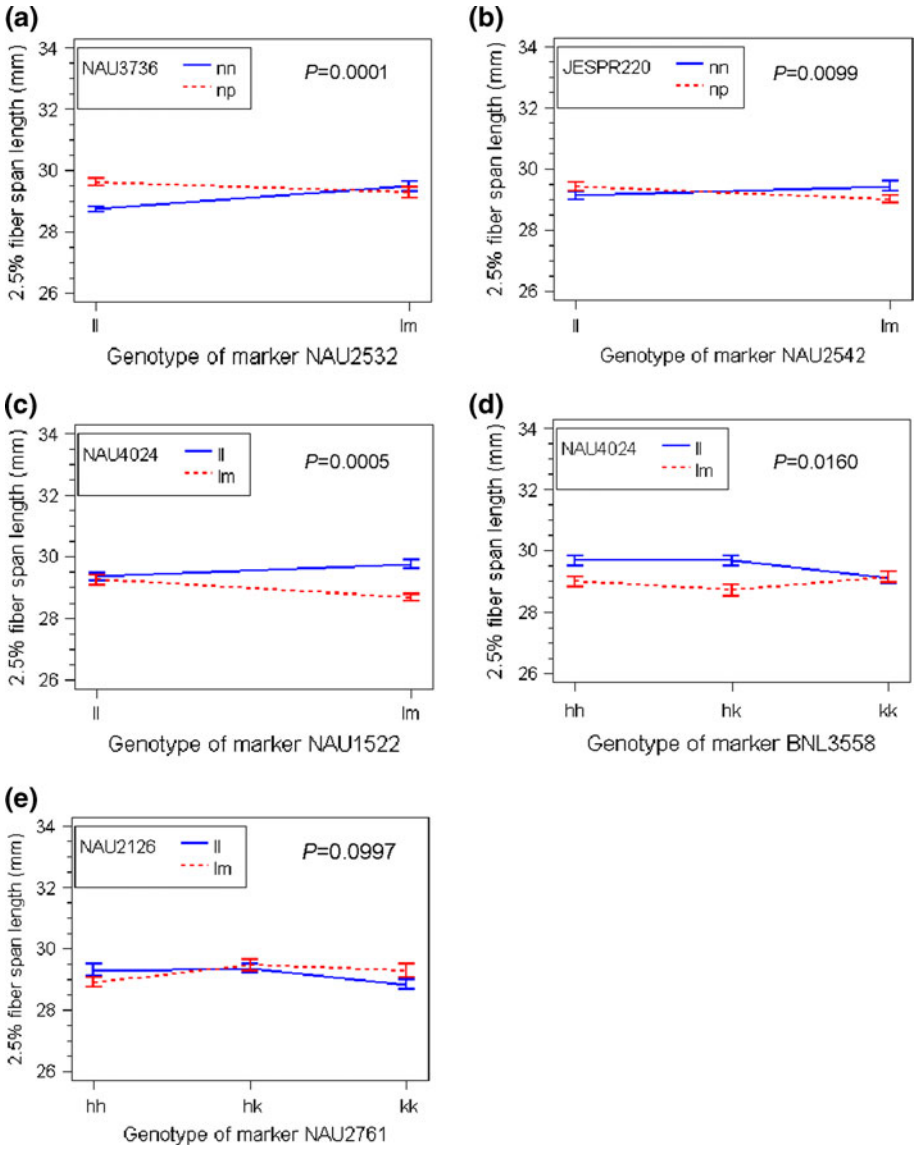

Fig. 2 The graphical representation of the phenotypic effects of identified epistatic QTL responsible for 2.5% fiber span length in cotton. The segregation type and phase were <lm × ll> and {0-} for markers NAU2532, NAU2542 and NAU4024; <lm × ll> and {1-} for markers NAU1522 and NAU2126; <nn × np> and {-0} for markers NAU3736 and JESPR220; <hk × hk> and {00} for marker BNL3558; and <hk × hk> and {01} for marker NAU2761. *P* was *P* value for two-way interaction between paired markers under *F* test using ANOVA method

belonged to additive-by-additive. Furthermore, for each interaction, one QTL in one of the linkage groups, D1(3), D2(1), D3 and D4(2), had a main effect, but the other, in one of the linkage groups A10, A13, D5(2) and D13(1), did not. The heritability for a single QTL varied from 0.88 to 6.81%, and the total heritability of the detected main-effect and epistatic QTL was 47.74%. Obvious differences between the new method and interval mapping were that

the new method (1) could detect epistatic QTL and (2) seemed to be more powerful in the detection of main-effect QTL than interval mapping.

Discussion

Most economically important traits are quantitative. The genetic architecture of quantitative traits of interest is the keystone for genetic improvement of animals and plants. In the inheritance analysis of complex traits, a 4WC population is more valuable than a segregating population derived from crosses between two inbred lines because the combination of QTL studies and breeding practices can ensure an expected selection response in the breeding program. Therefore, it is important to detect epistatic QTL in a 4WC population. The new approach is also applicable to human, tree and animal genetics.

For the current study it is assumed that complete marker information is available. However, the model (2) is only "approximative'' for the incidence matrix of epistatic effects while there are consecutive markers on the genetic map for missing or ambiguous QTL genotypes. This is because that we take the products of the single locus conditional QTL probabilities as joint conditional QTL probabilities and the products of conditional QTL probabilities are probably a reasonable approximation for epistatic regressors for pairs of loci that are far apart, but less for close pairs of loci. In this study the consecutive markers, with incomplete information and marker interval length of 15 cM, had been created while the proportion of dominant marker was set at 25 and 37.5% in Table 5. Results showed that the proposed method here works fine for loci at a distance. In my opinion, the new approach works well if the marker interval length is about 5 cM. However, we need to delete some closely linked markers if the interval length is less than 1 cM.

There are two types of errors in QTL mapping practice: type I and type II errors. If a QTL is identified at a location with no actual QTL, the first type appears. If an existing QTL is not detected, the second type occurs (Jansen 1994). Although we cannot avoid the two errors in real data analysis, there are ways to reduce them. First, the 4WC design can reduce or potentially eliminate the type II error caused by genetic drift because the probability that drift error is prevented in the 4WC design is larger than that in a backcross design (Xu 1996). Second, the correct genetic model is very important in reducing the probability of both type I and type II errors (Jansen 1994). Our multi-QTL genetic method described in Table 7 seemed to be more powerful (less type II error) than the single-QTL genetic model presented in Table 5 (Qin et al. 2008) for mapping main-effect QTL. The theoretical explanation was that multi-QTL genetic method could effectively reduce the residual variance (Zeng 1994; Jansen and Stam 1994; Jansen 1994; Kao et al. 1999). More importantly, a full model including epistatic effects is very important for reducing type II errors (see Table 8). Third, the 4WC effectively reduces the FPR by means of the estimation method, i.e., the PML method (Zhang and Xu 2005; He and Zhang 2008; Lü et al. 2009). Fourth, techniques for decreasing the probability of type I error, i.e., increasing the sample size (Table 4), using codominant markers, avoiding missing markers (Table 5) and enhancing the heritability of individual QTL (Table 3), are available. Finally, it is feasible to implement imputation experiments. In real data analysis, we can generate 20–50 imputed data sets (Sen and Churchill 2001; Zhang and Xu 2005), run mapping programs for each to summarize their results, and report only those QTL with high relative frequency. Thus, the true QTL can be identified because it frequently appears at a fixed location. As a result, the genome-wide FPR or empirical probability of type I error, computed from our simulation experiments (Tables 3, 4, 5), was generally less than 0.3%. In addition, it should be noted that the empirical probability of type I errors, reported in our previous work on mapping epistatic QTL for endosperm traits (He and Zhang 2008), was less than that in the present paper. This is probably because average family data were used in our previous work and individual plant data were used in the present study.

The PML analysis was multi-marker-based, and the fine positions of the inferred QTL were not estimated. The QTL effects were slightly underestimated when the QTL was not close to the marker's position (Table 6, Appendix 2). However, this shortcoming was not detrimental. With new techniques for screening molecular markers, saturated linkage maps (e.g., average interval length 5 or 10 cM) can be easily obtained, and the accuracy was adequate for breeding purposes. Furthermore, we may use the two-stage method to map epistatic QTL. In the first stage, the PML was used to capture the correct genetic model (QTL number and their marker intervals). In the second stage, the effects and positions of QTL were refined. The one-by-one method (only one QTL per interval was moving at any given time; all other QTL were fixed at their current positions), e.g., the MIM (Kao et al. 1999), or a synchronization method (the positions of all QTL were updated at the same time), e.g., Bayesian shrinkage estimation (Wang et al. 2005), may be used.

Epistatic effects are difficult to detect because an epistatic genetic model potentially contains a large number of model effects (Cheverud and Routman 1995; Zhang and Xu 2005). The variable selection technique, which excluded interactions with negligible effects, may overlook some important interaction effects due to incomplete exploration of the extremely large parameter space of some models (Zhang and Xu 2005). The PML, which was

adopted in this paper, can simultaneously fit all potential main and epistatic QTL on the whole genome, such that spurious QTL effects would be reduced toward zero and QTL with large effects could be estimated without significant reductions. Therefore, all important interaction effects can be identified. Another advantage of the PML is its time-saving characteristics, which are very important for real data analysis. The time cost of our real data (262 individuals and 184 markers) was ~28 h/imputed sample on our personal computer (CPU: Core2 DUO 3.0; Memory: 2 M), which is acceptable for most practical research purposes compared with ~135 days to run the Bayesian shrinkage estimation (Wang et al. 2005).

The proposed method differs from the random walk-based method proposed by Hanlon and Lorenz (2005) in several ways. First, the epistatic effect was defined as the difference between the theoretical and observed effects of genotypes at two or more markers in Hanlon and Lorenz (2005) while in the new method the interaction effect was partitioned into additive-by-additive, additive-by-dominant, dominant-by-additive and dominant-by-dominant effects according to Fisher's (1918) definition. Second, the ANOVA test in the second step in Hanlon et al. (2006) was not built on a full model that includes all main and epistatic effects while in the proposed method the full genetic model was considered and the corresponding estimates were unbiased (Table 3, 4, 5). Finally, the space of sums of epistatic effects was not fully explored in the 'random walk' step in Hanlon and Lorenz's (2005) approach and some important epistatic effects might be missing under the situations that the noise level was high and epistatic effects were overlapping (shared markers). Our new approach fully explored the extremely large parameter space, and all important epistatic effects could be captured, even with low heritability (Tables 3, 4, 5, 6). Although the three- or four-way interactions might be incorporated into the full model in the new method, the model including all the potential interaction effects is saturated quickly as the number of loci increases. To solve this issue, one way is to develop new algorithm and one way is to increase sample size and to decrease the number of parameters.

In the real data analysis in cotton, the five epistatic QTL responsible for 2.5% fiber span length shared several common phenomena. First, in the epistatic QTL pair, one QTL had main effects and another QTL did not. This phenomenon was often found in other experiments. In the three-locus epistatic combination (*D5Mit95*p, *D17Mit185*m, and *D18Mit55*p) responsible for mouse insulin-like growth factor-I, only the one linked to *D17Mit185*m had main effect (Hanlon et al. 2006). Of five epistatic QTL pairs for the trait HED, there was only one significant main effect for one epistatic QTL pair and there were no main effects for the remaining four epistatic QTL pairs (Xu and Jia 2007). This increases the difficulty for the detection of epistasis. However, the new method does not depend on whether or not the two loci both have main effects. Second, the two epistatic QTL pairs, NAU1522-NAU4024 and NAU4024-BNL3558, shared common locus NAU4024 (Table 8). We deduce that the above three loci may construct a QTL network. And NAU4024 might be linked to a downstream gene, regulated by the genes close to NAU1522 and BNL3558, according to the results in Gjuvsland et al. (2007). However, the true metabolic pathway among these genes will be further studied. Finally, additive-by-additive interaction was the most frequently significant type of interaction (Table 8). This phenomenon is similar to that in Gjuvsland et al. (2007).

Acknowledgments We are grateful to the Associate Editor Dr. Fred van Eeuwijk, and three anonymous reviewers for their constructive comments and suggestions that significantly improve the presentation of the manuscript. The work was supported in part by the National Basic Research Program of China (2006CB101708), the National Natural Science Foundation of China (30900842, 30971848, 30471114), Jiangsu Natural Science Foundation (BK2008335), NCET (NCET-05-0489), the 111 Project (B08025) and NAU Youth Sci-Tech Innovation Fund (KJ08001).

Appendix 1. The genetic variance components for a four-way cross population

For simplicity, assume that a quantitative trait was controlled by two QTL, Q_1 and Q_2, and the recombination

Table 9 Genotypic probability for two linked loci in a four-way cross population

Genotype for Q_1 locus	Genotype for Q_2 locus				Marginal probability
	$Q_{21}Q_{23}$	$Q_{21}Q_{24}$	$Q_{22}Q_{23}$	$Q_{22}Q_{24}$	
$Q_{11}Q_{13}$	$\frac{1}{4}(1-r)^2$	$\frac{1}{4}r(1-r)$	$\frac{1}{4}r(1-r)$	$\frac{1}{4}r^2$	$\frac{1}{4}$
$Q_{11}Q_{14}$	$\frac{1}{4}r(1-r)$	$\frac{1}{4}(1-r)^2$	$\frac{1}{4}r^2$	$\frac{1}{4}r(1-r)$	$\frac{1}{4}$
$Q_{12}Q_{13}$	$\frac{1}{4}r(1-r)$	$\frac{1}{4}r^2$	$\frac{1}{4}(1-r)^2$	$\frac{1}{4}r(1-r)$	$\frac{1}{4}$
$Q_{12}Q_{14}$	$\frac{1}{4}r^2$	$\frac{1}{4}r(1-r)$	$\frac{1}{4}r(1-r)$	$\frac{1}{4}(1-r)^2$	$\frac{1}{4}$
Marginal probability	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1

Table 10 The conditional probability of four QTL genotypes on the observed marker genotype and expected genotypic value for each marker genotype group in a four-way cross population

Genotype for marker locus M	Conditional probability of each QTL genotype for the observed marker genotype				Expected genotypic value for each marker genotype group
	Q_1Q_3	Q_1Q_4	Q_2Q_3	Q_2Q_4	
M_1M_3	$(1-r)^2$	$r(1-r)$	$r(1-r)$	r^2	$\overline{M_1M_3} = \mu + (1-2r)(a_1+a_2) + (1-2r)^2 d$
M_1M_4	$r(1-r)$	$(1-r)^2$	r^2	$r(1-r)$	$\overline{M_1M_4} = \mu + (1-2r)(a_1-a_2) - (1-2r)^2 d$
M_2M_3	$r(1-r)$	r^2	$(1-r)^2$	$r(1-r)$	$\overline{M_2M_3} = \mu + (1-2r)(-a_1+a_2) - (1-2r)^2 d$
M_2M_4	r^2	$r(1-r)$	$r(1-r)$	$(1-r)^2$	$\overline{M_2M_4} = \mu + (1-2r)(-a_1-a_2) + (1-2r)^2 d$

fraction between them was r. For locus Q_1, there were four genotypes: $Q_{11}Q_{13}$, $Q_{11}Q_{14}$, $Q_{12}Q_{13}$, and $Q_{12}Q_{14}$. For locus Q_2, there were four genotypes: $Q_{21}Q_{23}$, $Q_{21}Q_{24}$, $Q_{22}Q_{23}$, and $Q_{22}Q_{24}$. Thus, there were $4 \times 4 = 16$ genotypes in the population. The probability for each genotype is listed in Table 9.

The population mean is $\mu = \sum f_i g_i = (1-2r)(i_{a_{11}a_{21}} + i_{a_{12}a_{22}}) + (1-2r)^2 i_{d_1d_2}$, where f_i and g_i $(i = 1, 2, \ldots, 16)$ are genotypic probability (Table 9) and genotypic value [Model (1) and (2)] for the 16 two loci genotypes. If the two QTL are unlinked, $\mu = 0$. The genetic variance components for a 4WC population are

$$\begin{aligned}
\sigma_g^2 &= \sum f_i g_i^2 - \left(\sum f_i g_i\right)^2 \\
&= a_{11}^2 + a_{12}^2 + d_1^2 + a_{21}^2 + a_{22}^2 + d_2^2 \\
&\quad + 4r(1-r) i_{a_{11}a_{21}}^2 + i_{a_{11}a_{22}}^2 + i_{a_{11}d_2}^2 + i_{a_{12}a_{21}}^2 + 4r(1-r) i_{a_{12}a_{22}}^2 \\
&\quad + i_{a_{12}d_2}^2 + i_{d_1a_{21}}^2 + i_{d_1a_{22}}^2 + 8r(1-r)[r^2 + (1-r)^2] i_{d_1d_2}^2 \\
&\quad + 2(1-2r)[a_{11}a_{21} + a_{12}a_{22} + a_{11} i_{d_1a_{22}} + a_{12} i_{d_1a_{21}} \\
&\quad + a_{21} i_{a_{12}d_2} + a_{22} i_{a_{11}d_2} + (d_1 + d_2)(i_{a_{11}a_{22}} + i_{a_{12}a_{21}}) \\
&\quad + i_{a_{11}d_2} i_{d_1a_{21}} + i_{a_{12}d_2} i_{d_1a_{22}}] + 2(1-2r)^2 (d_1d_2 + a_{11} i_{a_{12}d_2} \\
&\quad + a_{12} i_{a_{11}d_2} + a_{21} i_{d_1a_{22}} + a_{22} i_{d_1a_{21}} + i_{a_{11}a_{22}} i_{a_{12}a_{21}}) \\
&\quad + 8r(1-r)(1-2r)(i_{a_{11}a_{21}} i_{d_1d_2} + i_{a_{12}a_{22}} i_{d_1d_2})
\end{aligned}$$

If the two QTL are unlinked, the genetic variance is

$$\begin{aligned}
\sigma_g^2 &= a_{11}^2 + a_{12}^2 + d_1^2 + a_{21}^2 + a_{22}^2 + d_2^2 + i_{a_{11}a_{21}}^2 \\
&\quad + i_{a_{11}a_{22}}^2 + i_{a_{11}d_2}^2 + i_{a_{12}a_{21}}^2 + i_{a_{12}a_{22}}^2 + i_{a_{12}d_2}^2 \\
&\quad + i_{d_1a_{21}}^2 + i_{d_1a_{22}}^2 + i_{d_1d_2}^2
\end{aligned}$$

Appendix 2. The expectation of main effects of QTL in the marker analysis in a four-way cross

Assume that the recombination fraction between an arbitrary quantitative trait locus (QTL) Q and its nearest marker locus M is r. The conditional probability of the four QTL genotypes for the observed marker genotype is easily obtained from Table 9 and is listed in Table 10. As described in model (1), the genotypic value of each QTL genotype is defined as

$$\begin{pmatrix} \overline{Q_1Q_3} \\ \overline{Q_1Q_4} \\ \overline{Q_2Q_3} \\ \overline{Q_2Q_4} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ a_1 \\ a_2 \\ d \end{pmatrix}$$

The expected genotypic value for each marker genotype group was calculated and is also listed in Table 10.

Let μ', a_1', a_2', and d' be estimates of the main effects of QTL in the marker analysis. Then,

$$\begin{pmatrix} \overline{M_1M_3} \\ \overline{M_1M_4} \\ \overline{M_2M_3} \\ \overline{M_2M_4} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} \mu' \\ a_1' \\ a_2' \\ d' \end{pmatrix}.$$

The expectations of QTL main effects during marker analysis are

$$\begin{aligned}
\begin{pmatrix} \mu' \\ a_1' \\ a_2' \\ d' \end{pmatrix} &= \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \overline{M_1M_3} \\ \overline{M_1M_4} \\ \overline{M_2M_3} \\ \overline{M_2M_4} \end{pmatrix} \\
&= \begin{pmatrix} \mu \\ (1-2r)a_1 \\ (1-2r)a_2 \\ (1-2r)^2 d \end{pmatrix}
\end{aligned}$$

This results show that all QTL main effects were confused with recombination, but not with each other, and the recombination fraction had a greater influence on the dominant effect than on the additive effects.

References

Blanc G, Charcosset A, Mangin B, Gallais A, Moreau L (2006) Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. Theor Appl Genet 113:206–224

Carlborg Ö, Haley CS (2004) Epistasis: too often neglected in complex trait studies. Nat Rev Genet 5:618–625

Cheverud JM, Routman EJ (1995) Epistasis and its contribution to genetic variance components. Genetics 139:1455–1461

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. R Soc (Edinburgh) Trans 52:399–433

Gjuvsland AB, Hayes BJ, Omholt SW, Carlborg Ö (2007) Statistical epistasis is a generic feature of gene regulatory networks. Genetics 175:411–420

Groover A, Devey M, Fiddler T, Lee J, Megraw R, Mitchel-Olds T, Sherman B, Vujcic S, Williams C, Neale D (1994) Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. Genetics 138:1293–1300

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Haley CS, Knott SA, Elsen JM (1994) Mapping quantitative trait loci in crosses between outbred lines using least squares. Genetics 136:1195–1207

Hanlon P, Lorenz WA (2005) A computational method to detect epistatic effects contributing to a quantitative trait. J Theor Biol 235:350–364

Hanlon P, Lorenz WA, Shao Z, Harper JM, Galecki AT, Miller RA, Burke DT (2006) Three-locus and four-locus QTL interactions influence mouse insulin-like growth factor-I. Physiol Genomics 26:46–54

Harmegnies N, Davin F, De Smet S, Buys N, Georges M, Coppieters W (2006) Results of a whole-genome quantitative trait locus scan for growth, carcass composition and meat quality in a porcine four-way cross. Anim Genet 37:543–553

He XH, Zhang YM (2008) Mapping epistatic QTL underlying endosperm traits using all markers on the entire genome in random hybridization design. Heredity 101:39–47

Hoeschele I, VanRanden PM (1993a) Bayesian analysis of linkage between genetic markers and quantitative trait loci I. Prior knowledge. Theor Appl Genet 85:953–960

Hoeschele I, Vanranden PM (1993b) Bayesian analysis of linkage between genetic markers and quantitative trait loci. II. Combining prior knowledge with experimental evidence. Theor Appl Genet 85:946–952

Jansen RC (1994) Controlling the type I and type II errors in mapping quantitative trait loci. Genetics 138:871–881

Jansen RC, Stam P (1994) High-resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447–1455

Jiang C, Zeng ZB (1997) Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. Genetica 101:47–58

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152:1203–1216

Knott SA, Neale DB, Sewell MM, Haley CS (1997) Multiple marker mapping of quantitative trait loci in an outbred pedigree of loblolly pine. Theor Appl Genet 94:810–820

Knott SA, Marklund L, Haley CS, Andersson K, Davies W, Ellegren H, Fredholm M, Hansson I, Hoyheim B, Lundstrom K, Moller M, Andersson L (1998) Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and large white pigs. Genetics 149:1069–1080

Kruglyak L, Lander ES (1995) A nonparametric approach for mapping quantitative trait loci. Genetics 139:1421–1428

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci 84:2363–2367

Lippman ZB, Zamir D (2006) Heterosis: revisiting the magic. Trends Genet 23:60–66

Liu YF, Zeng ZB (2000) A general mixture model approach for mapping quantitative trait loci from diverse cross designs involving multiple inbred lines. Genet Res 75:345–355

Lü HY, Li M, Li GJ, Yao LL, Lin F, Zhang YM (2009) Multiple loci in silico mapping in inbred lines. Heredity 103:346–354

Luo L, Xu S (2003) Mapping viability loci using molecular markers. Heredity 90:459–467

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland

Martínez O, Curnow RN (1992) Estimating the locations and sizes of effects of quantitative trait loci using flanking markers. Theor Appl Genet 85:480–488

Melchinger AE, Utz HF, Piepho HP, Zeng ZB, Schön CC (2007) The role of epistasis in the manifestation of heterosis: a systems-oriented approach. Genetics 177:1815–1825

Moore JH (2005) A global view of epistasis. Nat Genet 37:77–83

Park T, Casella G (2008) The Bayesian Lasso. J Am Stat Assoc 103:681–686

Qin H, Guo W, Zhang YM, Zhang T (2008) QTL mapping of yield and fiber traits based on a four-way cross population in *Gossypium hirsutum* L. Theor Appl Genet 117:883–894

Rao S, Xu S (1998) Mapping quantitative trait loci for ordered categorical traits in four-way crosses. Heredity 81:214–224

Satagopan JM, Yandell BS, Newton MA, Osborn TC (1996) A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. Genetics 144:805–816

Sen S, Churchill GA (2001) A statistical framework for quantitative trait mapping. Genetics 159:371–387

Sillanpää MJ, Arjas E (1998) Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. Genetics 148:1373–1388

Van Ooijen JW, Voorrips RE (2001) JoinMap® 3.0. Software for the calculation of genetic linkage maps. Plant Research International, Wageningen

Verhoeven KJF, Jannink JL, Mcintyre LM (2006) Using mating designs to uncover QTL and the genetic architecture of complex traits. Heredity 96:139–149

Wang H, Zhang YM, Li X, Masinde GL, Mohan S, Baylink DJ, Xu S (2005) Bayesian shrinkage estimation of QTL parameters. Genetics 170:465–480

Xie C, Xu S (1999) Mapping quantitative trait loci with dominant markers in four-way crosses. Theor Appl Genet 98:1014–1021

Xu S (1995) A comment on the simple regression method for interval mapping. Genetics 141:1657–1659

Xu S (1996) Mapping quantitative trait loci using four-way crosses. Genet Res 68:175–181

Xu S (1998) Iteratively reweighted least squares mapping of quantitative trait loci. Behav Genet 28:341–355

Xu S (2007) An empirical Bayes method for estimating epistatic effects of quantitative trait loci. Biometrics 63:513–521

Xu S, Jia Z (2007) Genomewide analysis of epistatic effects for quantitative traits in barley. Genetics 175:1955–1963

Yi N, Banerjee S (2009) Hierarchical generalized linear models for multiple quantitative trait locus mapping. Genetics 181:1101–1113

Yi N, Xu S (2008) Bayesian Lasso for quantitative trait loci mapping. Genetics 179:1045–1055

Yu SB, Li JX, Xu CG, Tan YF, Gao YJ, Li XH, Zhang Q, Saghai Maroof MA (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 94:9226–9231

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

Zhang YM (2006) Advances on methods for mapping QTL in plants. Chin Sci Bull 51:2809–2818

Zhang YM, Xu S (2005) A penalized maximum likelihood method for estimating epistatic effects of QTL. Heredity 95:96–104