

Multi-QTL Mapping for Quantitative Traits Using Epistatic Distorted Markers

Shang-Qian Xie, Jia Wen, Yuan-Ming Zhang*

Statistical Genomics Group, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Department of Crop Genetics and Breeding, Nanjing Agricultural University, Nanjing, Jiangsu, China

Abstract

The interaction between segregation distortion loci (SDL) has been often observed in all kinds of mapping populations. However, little has been known about the effect of epistatic SDL on quantitative trait locus (QTL) mapping. Here we proposed a multi-QTL mapping approach using epistatic distorted markers. Using the corrected linkage groups, epistatic SDL was identified. Then, these SDL parameters were used to correct the conditional probabilities of QTL genotypes, and these corrections were further incorporated into the new QTL mapping approach. Finally, a set of simulated datasets and a real data in 304 mouse F_2 individuals were used to validate the new method. As compared with the old method, the new one corrects genetic distance between distorted markers, and considers epistasis between two linked SDL. As a result, the power in the detection of QTL is higher for the new method than for the old one, and significant differences for estimates of QTL parameters between the two methods were observed, except for QTL position. Among two QTL for mouse weight, one significant difference for QTL additive effect between the above two methods was observed, because epistatic SDL between markers C66 and T93 exists ($P = 2.94e-4$).

Citation: Xie S-Q, Wen J, Zhang Y-M (2013) Multi-QTL Mapping for Quantitative Traits Using Epistatic Distorted Markers. *PLoS ONE* 8(7): e68510. doi:10.1371/journal.pone.0068510

Editor: Tongming Yin, Nanjing Forestry University, China

Received: April 21, 2013; **Accepted:** May 31, 2013; **Published:** July 9, 2013

Copyright: © 2013 Xie et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the National Basic Research Program of China grant 2011CB109306; the Fundamental Research Funds for the Central Universities grants KYT201002 and KYZ201202-9; a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions and the Specialized Research Fund for the Doctoral Program of Higher Education grants 20100097110035 and 20120097110023. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: soyzhang@njau.edu.cn

Introduction

Quantitative trait locus (QTL) mapping has become a routine approach in genetic studies of quantitative traits [1–5]. Most QTL mapping approaches usually make use of markers that follow a Mendelian segregation ratio. However, some markers often show distorted segregation ratios from Mendelian expectations in actual QTL mapping populations [6–8]. This segregation distortion often affects linkage group construction and QTL mapping results [9–11]. Therefore, how to use distorted markers in QTL mapping needs to be addressed.

Segregation distortion is a commonly encountered phenomenon [12]. Several mechanisms or approaches have been proposed to explain this phenomenon [10,13]. Previous studies on the influence of distorted markers mainly focus on two aspects. First, this distortion may lead to biased estimate of recombination fraction between distorted markers. To solve this issue, Lorieux et al. [14–15] adopted two-point method to correct genetic distance between distorted markers. Zhu et al [9] extended multipoint analysis method to more general situations, considering distorted, dominant and missing markers at the same time. Recently, Xie [11] considers epistasis between two linked SDL in the construction of linkage groups. Second, this distortion affects QTL mapping results, for example, QTL detection power [16–17] and QTL additive effect [10]. To improve the precision of QTL mapping, Xu and Hu [18] developed an interval mapping approach for joint analysis of both QTL and SDL. Recently,

Wen et al. [10] further extended a multi-QTL mapping approach. However, the above-mentioned QTL mapping approaches ignore epistatic SDL and linkage group correction. The two issues need to be addressed in QTL mapping.

SDL epistasis is a type of gene interaction [19]. Törjék et al. [20] indicated that marker segregation distortion is due to reduced fertility caused by epistatic interaction. Recently, some similar results have been reported in *Drosophila* [21], alfalfa [22] and rice [23–25]. Therefore, SDL epistasis should be considered in QTL mapping methodology. Multi-QTL mapping is now the state-of-the-art method. However, it is difficult to implement under the maximum-likelihood framework. At present the Bayesian method implemented via expectation-maximization (EM) algorithm [26] is specialized to handle complicated models and thus it is the ideal tool for mapping multiple QTL in this study. Accordingly, there is a critical need for an in-depth study of the methodology for multi-QTL mapping using epistatic distorted markers.

The purpose of this study was to develop a statistical method for mapping QTL of quantitative traits using epistatic distorted markers. First, marker information was used to detect epistatic SDL using a maximum likelihood approach. The detected SDL parameters were used to correct linkage groups. Second, the information about the detected epistatic SDL along with the corrected linkage groups was incorporated into a multi-QTL mapping approach. Finally, the proposed method was validated by Monte Carlo simulation studies and real data analysis.

Methods

Based on the corrected genetic groups [11], molecular marker information from all the n F_2 individuals is used to detect epistatic SDL. We hypothesize that marker segregation distortion is subject to the gametic selection.

Two SDL under consideration are linked, with recombination fraction of r . The viabilities of male gametes Ab , aB and ab relatives to AB are v , u and x , respectively. The expected frequencies π_i ($i = 1, 2 \dots 9$) of nine genotypes after gametic selection are listed in **Table 1**.

Mapping Epistatic SDL

Let n_i ($i = 1, 2 \dots 9$) be the observed number of the i th genotype for two SDL A and B , and $n = \sum_{i=1}^9 n_i$ be the total number of individuals. If two linked SDL that exist interaction are resided at marker positions, the marker information is used to detect epistatic SDL. Therefore, the complete data log-likelihood function in F_2 can be expressed as

$$L = \sum_{i=1}^9 n_i \ln \pi_i \tag{1}$$

where all the π_i are listed in **Table 1**. Note that r in π_i is obtained from the corrected map by Kosambi [27] function. All the SDL parameters u , v and x in π_i can be estimated by

$$\begin{aligned} u &= \frac{(n_{41} + n_{53} + n_7 + n_{81})(1-r)}{r(n_1 + n_{22} + n_{52} + n_{42})} \\ v &= \frac{(n_{21} + n_3 + n_{54} + n_{61})(1-r)}{r(n_1 + n_{22} + n_{52} + n_{42})} \\ x &= \frac{(n_{51} + n_{62} + n_{82} + n_9)}{(n_1 + n_{22} + n_{52} + n_{42})} \end{aligned} \tag{2}$$

[11], where $n_{21} = \frac{vn_2}{v+1}$, $n_{22} = \frac{n_2}{v+1}$, $n_{41} = \frac{un_4}{u+1}$, $n_{42} = \frac{n_4}{u+1}$, $n_{51} = \frac{(1-r)^2 xn_5}{(1-r)^2(x+1) + r^2(u+v)}$, $n_{52} = \frac{(1-r)^2 n_5}{(1-r)^2(x+1) + r^2(u+v)}$, $n_{53} = \frac{r^2 un_5}{(1-r)^2(x+1) + r^2(u+v)}$, $n_{54} = \frac{r^2 vn_5}{(1-r)^2(x+1) + r^2(u+v)}$, $n_{61} = \frac{vn_6}{v+x}$, $n_{62} = \frac{xn_6}{v+x}$, $n_{81} = \frac{un_8}{u+x}$ and $n_{82} = \frac{xn_8}{u+x}$. The EM algorithm can be used to estimate all the above parameters [11].

Table 1. Expected frequencies of nine genotypes under gametic selection.

Genotype	Expected frequency	Genotype	Expected frequency
<i>AABB</i>	$(1-r)^2/2d$	<i>Aabb</i>	$r(1-r)(v+x)/2d$
<i>AABb</i>	$r(1-r)(v+1)/2d$	<i>aabb</i>	$r^2u/2d$
<i>AAbb</i>	$r^2v/2d$	<i>aaBb</i>	$r(1-r)(u+x)/2d$
<i>AaBB</i>	$r(1-r)(u+1)/2d$	<i>aabb</i>	$(1-r)^2x/2d$
<i>AaBb</i>	$(1-r)^2(x+1) + r^2(u+v)/2d$		

$d = (1-r)(x+1) + r(u+v)$
doi:10.1371/journal.pone.0068510.t001

If two linked SDL that exist interaction are not resided at marker positions, the count data n_i are not observable and we need to substitute them by their expectations. At this case, the EM algorithm is also used to estimate all the above SDL parameters.

In the E-step, the expected numbers of the SDL genotypes can be obtained by the below equations,

$$\begin{aligned} E(n_1) &= \sum_{j=1}^n P_j^*(AABB) = \sum_{j=1}^n \frac{P_j(AABB)\pi_1}{\bar{\pi}_j} \\ E(n_2) &= \sum_{j=1}^n P_j^*(AABb) = \sum_{j=1}^n \frac{P_j(AABb)\pi_2}{\bar{\pi}_j} \\ E(n_3) &= \sum_{j=1}^n P_j^*(AAbb) = \sum_{j=1}^n \frac{P_j(AAbb)\pi_3}{\bar{\pi}_j} \\ E(n_4) &= \sum_{j=1}^n P_j^*(AaBB) = \sum_{j=1}^n \frac{P_j(AaBB)\pi_4}{\bar{\pi}_j} \\ E(n_5) &= \sum_{j=1}^n P_j^*(AaBb) = \sum_{j=1}^n \frac{P_j(AaBb)\pi_5}{\bar{\pi}_j} \\ E(n_6) &= \sum_{j=1}^n P_j^*(Aabb) = \sum_{j=1}^n \frac{P_j(Aabb)\pi_6}{\bar{\pi}_j} \\ E(n_7) &= \sum_{j=1}^n P_j^*(aaBB) = \sum_{j=1}^n \frac{P_j(aaBB)\pi_7}{\bar{\pi}_j} \\ E(n_8) &= \sum_{j=1}^n P_j^*(aaBb) = \sum_{j=1}^n \frac{P_j(aaBb)\pi_8}{\bar{\pi}_j} \\ E(n_9) &= \sum_{j=1}^n P_j^*(aabb) = \sum_{j=1}^n \frac{P_j(aabb)\pi_9}{\bar{\pi}_j} \end{aligned} \tag{3}$$

where $\bar{\pi}_j = P_j(AABB)\pi_1 + \frac{1}{2}P_j(AABb)\pi_2 + P_j(AAbb)\pi_3 + \frac{1}{2}P_j(AaBB)\pi_4 + \frac{1}{4}P_j(AaBb)\pi_5 + \frac{1}{4}P_j(Aabb)\pi_6 + P_j(aaBB)\pi_7 + \frac{1}{2}P_j(aaBb)\pi_8 + P_j(aabb)\pi_9$; $P_j^*(\bullet)$ is the posterior probability that incorporates SDL parameters; and $P_j(\bullet)$ is the probability of the SDL genotypes for individual j conditional on marker information. Note that the explanation for the coefficients 1/2 and 1/4 in $\bar{\pi}_j$ is similar to that in Luo et al. [28]; and $P_j(\bullet)$ under two SDL is difficult to calculate, fortunately, the method of Kao et al. [29] is available, that is

$$P_j(AABB) = P_j(AA)P_j(BB) \tag{4}$$

where $P_j(AA)$ or $P_j(BB)$ can be calculated from multi-point approach [30].

In the M-step, all the SDL parameters u , v and x can be updated by equations (2).

Repeating E-step and M-step until a certain criterion of convergence is satisfied.

Interval QTL Mapping using Epistatic Distorted Markers

Let y_j be the observation of quantitative trait for individual j in F_2 . For single QTL, the genetic model for quantitative trait may be described as follows

$$y_j = \mu + s_{1j}a_1 + s_{2j}d_1 + \varepsilon_j \tag{5}$$

where μ is population mean; s_{1j} and s_{2j} are dummy variables defined as $s_{1j} = 1$ and $s_{2j} = 0$ for QQ , $s_{1j} = 0$ and $s_{2j} = 1$ for Qq , and $s_{1j} = -1$ and $s_{2j} = 0$ for qq , respectively; a_1 is additive effect; d_1 is dominant effect; and $\varepsilon_i \sim N(0, \sigma_e^2)$.

The posterior probabilities of QTL genotypes conditional on normal markers for the j th F_2 individual are calculated as

$$\begin{aligned}
 P_{j(QQ)}^* &= \frac{P_{j(QQ)}\phi(y_j; \mu + a_1, \sigma_e^2)}{P_{j(QQ)}\phi(y_j; \mu + a_1, \sigma_e^2) + P_{j(Qq)}\phi(y_j; \mu + d_1, \sigma_e^2) + P_{j(qq)}\phi(y_j; \mu - a_1, \sigma_e^2)} \\
 P_{j(Qq)}^* &= \frac{P_{j(Qq)}\phi(y_j; \mu + d_1, \sigma_e^2)}{P_{j(QQ)}\phi(y_j; \mu + a_1, \sigma_e^2) + P_{j(Qq)}\phi(y_j; \mu + d_1, \sigma_e^2) + P_{j(qq)}\phi(y_j; \mu - a_1, \sigma_e^2)} \quad (6) \\
 P_{j(qq)}^* &= \frac{P_{j(qq)}\phi(y_j; \mu - a_1, \sigma_e^2)}{P_{j(QQ)}\phi(y_j; \mu + a_1, \sigma_e^2) + P_{j(Qq)}\phi(y_j; \mu + d_1, \sigma_e^2) + P_{j(qq)}\phi(y_j; \mu - a_1, \sigma_e^2)}
 \end{aligned}$$

where $\phi(y; m, e)$ represents normal density function of variable y with mean m and variance e ; and $p_j(\bullet)$ and $p_j^*(\bullet)$ for QTL are similar to $P_j(\bullet)$ and $P_j^*(\bullet)$ for SDL, respectively.

If epistatic SDL exist, the conditional probabilities $p_j(\bullet)$ in equation (6) are biased. In this case, the estimates of SDL parameters are used to correct the probabilities. For one locus of epistatic SDL, the expected frequencies of four genotypes are deduced from **Table 1**, and the relative fitness of each genotype is defined as $w_{AA} = 1$, $w_{Aa} = t_1$, $w_{aA} = t_2$ and $w_{aa} = t_3$. The above two results are listed in **Table 2**. Therefore, their relationship can be found in the below equations

$$\begin{aligned}
 \frac{1}{1 + t_1 + t_2 + t_3} &= \frac{1 - r + rv}{2d} \\
 \frac{t_1}{1 + t_1 + t_2 + t_3} &= \frac{1 - r + rv}{2d} \\
 \frac{t_2}{1 + t_1 + t_2 + t_3} &= \frac{ru + x(1 - r)}{2d} \\
 \frac{t_3}{1 + t_1 + t_2 + t_3} &= \frac{ru + x(1 - r)}{2d} \quad (7)
 \end{aligned}$$

As a result, $t_1 = 1$ and $t_2 = t_3 = [ru + x(1 - r)] / (1 - r + rv)$, indicating $w_{AA} = w_{Aa} = 1$ and $w_{aA} = w_{aa} = \frac{ru + x(1 - r)}{1 - r + rv}$. If the SDL is overlapped with the QTL under study, the above fitnesses are same as those for QTL. Therefore, the conditional probabilities $p_j(\bullet)$ in equation (6) can be corrected using the below equations

$$\begin{aligned}
 P'_{j(QQ)} &= \frac{w_{AA}P_{j(QQ)}}{w_{AA}P_{j(QQ)} + w_{Aa}P_{j\{Qq\}} + w_{aA}P_{j\{qQ\}} + w_{aa}P_{j(qq)}} \\
 &= \frac{(1 - r + rv)(ru + rv - rx - r + x + 1)p_{j(QQ)}}{(1 - r + rv)(ru + rv - rx - r + x + 1)p_{j(QQ)} + [(1 - r + rv)^2 + (x - rx + ru)^2]p_{j(Qq)} + (ru + rv - rx - r + x + 1)(ru + x - rx)p_{j(qq)}} \\
 P'_{j(Qq)} &= P'_{j\{Qq\}} + P'_{j\{qQ\}} = \frac{w_{Aa}P_{j\{Qq\}}}{w_{AA}P_{j(QQ)} + w_{Aa}P_{j\{Qq\}} + w_{aA}P_{j\{qQ\}} + w_{aa}P_{j(qq)}} + \frac{w_{aA}P_{j\{qQ\}}}{w_{AA}P_{j(QQ)} + w_{Aa}P_{j\{Qq\}} + w_{aA}P_{j\{qQ\}} + w_{aa}P_{j(qq)}} \\
 &= \frac{[(1 - r + rv)^2 + (x - rx + ru)^2]p_{j(Qq)}}{(1 - r + rv)(ru + rv - rx - r + x + 1)p_{j(QQ)} + [(1 - r + rv)^2 + (x - rx + ru)^2]p_{j(Qq)} + (ru + rv - rx - r + x + 1)(ru + x - rx)p_{j(qq)}} \\
 P'_{j(qq)} &= \frac{w_{aa}P_{j(qq)}}{w_{AA}P_{j(QQ)} + w_{Aa}P_{j\{Qq\}} + w_{aA}P_{j\{qQ\}} + w_{aa}P_{j(qq)}} \\
 &= \frac{(ru + rv - rx - r + x + 1)(ru + x - rx)p_{j(qq)}}{(1 - r + rv)(ru + rv - rx - r + x + 1)p_{j(QQ)} + [(1 - r + rv)^2 + (x - rx + ru)^2]p_{j(Qq)} + (ru + rv - rx - r + x + 1)(ru + x - rx)p_{j(qq)}} \quad (8)
 \end{aligned}$$

Table 2. Expected frequencies and relative fitness for one locus of epistatic SDL.

Genotype	Expected frequency		Relative fitness
	Before selection	After selection	
AA	1/4	$(1 - r + rv)/2d$	1
Aa	1/4	$(1 - r + rv)/2d$	t_1
aA	1/4	$[ru + x(1 - r)]/2d$	t_2
aa	1/4	$[ru + x(1 - r)]/2d$	t_3

doi:10.1371/journal.pone.0068510.t002

where $P_{j\{Qq\}} = \frac{(1 - r) + rv}{(1 - r)(x + 1) + r(u + v)}P_{j(Qq)}$ and $P_{j\{qQ\}} = \frac{(1 - r)x + ru}{(1 - r)(x + 1) + r(u + v)}P_{j(Qq)}$.

Multi-QTL Mapping using Epistatic Distorted Markers

When multiple QTL are taken into account, the model (5) is extended as

$$y_j = \mu + \sum_{l=1}^p (s_{2l-1,j}a_l + s_{2l,j}d_l) + \varepsilon_j \quad (9)$$

where a_l and d_l are additive and dominant effects of the l th QTL ($l = 1, \dots, p$), respectively; p is the number of putative QTL; and $s_{2l-1,j}$ and $s_{2l,j}$ are dummy variables, which are similar to s_{1j} and s_{2j} in model (5). This model can be expressed in matrix form

$$y = X\beta + Z\gamma + \varepsilon \quad (10)$$

where $y = (y_1, y_2, \dots, y_n)^T$; $X = (1, 1, \dots, 1)^T$; $\beta = \mu$; $Z = (Z_1, Z_2, \dots, Z_{2p})$; $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{2p})^T$; and $\varepsilon \sim N(0, I\sigma^2)$.

At present, there are several methods available for estimating the parameters in model (10). There we adopt an empirical Bayes approach [26], and employ normal prior $N(0, \sigma_l^2)$ for QTL effect γ_l and the scaled inverse χ^2 prior $p(\sigma_l^2 | \tau, \omega)$ for σ_l^2 , $p(\sigma_l^2 | \tau, \omega) = \text{Inv-}\chi^2(\sigma_l^2 | \tau, \omega) \propto (\sigma_l^2)^{-\frac{1}{2}(\tau + 2)} \exp(-\frac{\omega}{2\sigma_l^2})$, where $(\tau, \omega) = (0, 0)$ [26]. This procedure for parameter estimation is as follows.

Table 3. Simulated parameters in all the Monte Carlo simulation experiments.

Case	Position (cM)			Distance of QTL and SDL (cM)	Heritability (%)		Sample size
	SDL ₁	QTL	SDL ₂		SDL ₁ =SDL ₂	QTL	
1	20	25	30	5	5, 10, 15	10	300
2	20	25	30	5	10	5, 10, 15	300
3	20	25	30	5	10	10	100, 200, 300
4	20	22, 25, 30	24, 30, 40	2, 5, 10	15	10	300

doi:10.1371/journal.pone.0068510.t003

- (1) Setting initial values: $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_p^2 = 1$, $\beta = (X^T X)^{-1} X^T y$, $\sigma^2 = \frac{1}{2n} (y - X\beta)^T (y - X\beta)$.
- (2) E-step: QTL effect can be predicted by $E(\gamma_l) = \sigma_l^2 Z_l^T V^{-1} (y - X\beta)$, where $V = \sum_{l=1}^p Z_l Z_l^T \sigma_l^2 + I\sigma^2$, $E(\gamma_l^T \gamma_l) = E(\gamma_l^T) E(\gamma_l) + \text{tr}[\text{var}(\gamma_l)]$, and $\text{var}(\gamma_l) = I\sigma_l^2 - \sigma_l^2 Z_l^T V^{-1} Z_l \sigma_l^2$.
- (3) M-step: update parameters σ_l^2 , β and residual error variance

$$\begin{aligned} \sigma_l^2 &= \frac{E(\gamma_l^T \gamma_l) + \omega}{\tau + 2 + 1} \\ \beta &= (X^T V^{-1} X)^{-1} (X^T V^{-1} y) \\ \sigma^2 &= \frac{1}{n} (y - X\beta)^T \left[y - X\beta - \sum_{l=1}^p Z_l E(\gamma_l) \right] \end{aligned} \tag{11}$$

Repeat E-step and M-step until a certain criterion of convergence is satisfied.

In summary, the proposed approach includes three steps. First, marker information is used to detect epistatic SDL. Then, the estimates of SDL parameters are used to correct conditional probabilities of QTL genotypes. Finally, the corrected probabilities are incorporated into a multi-QTL mapping approach.

Results

Monte Carlo Simulation Studies

The purpose of the simulation experiment was to evaluate the statistical performance of the proposed approach by changing SDL heritability, QTL heritability, sample size, and genetic distance between SDL and QTL, respectively.

Effect of SDL heritability on new method. In the first simulated experiment, the simulated genome consisted of one chromosome, and twenty-one evenly spaced co-dominant markers covered the chromosome with an average marker interval of 5.0 cM. We simulated a single QTL and two SDL, all of which overlapped with markers. The single QTL with 0.10 heritability was located at marker position 25 cM, and two SDL each with heritabilities of 0.05, 0.10 and 0.15 were placed at marker positions 20 cM and 30 cM, respectively. The genetic parameters in F₂ population with a sample size of 300 were as follows: $a = d = 0.3849$ (QTL effects), $\sigma^2 = 1$, and $\mu = 100$. The phenotypic values for quantitative trait are simulated by model (9). The parameters for viability selection were set at $u = v = x = 0.5141$ (5%), 0.3615 (10%) and 0.1617 (15%), respectively. These parameters could be transferred into SDL effects [11]. These SDL genotypic effects along with random error were used to simulate phenotypic values of viability selection. If the value is larger than zero, this individual remains, or it is deleted from the simulated population. Each treatment was replicated 200 times. In the analyses of each simulated dataset, two approaches were

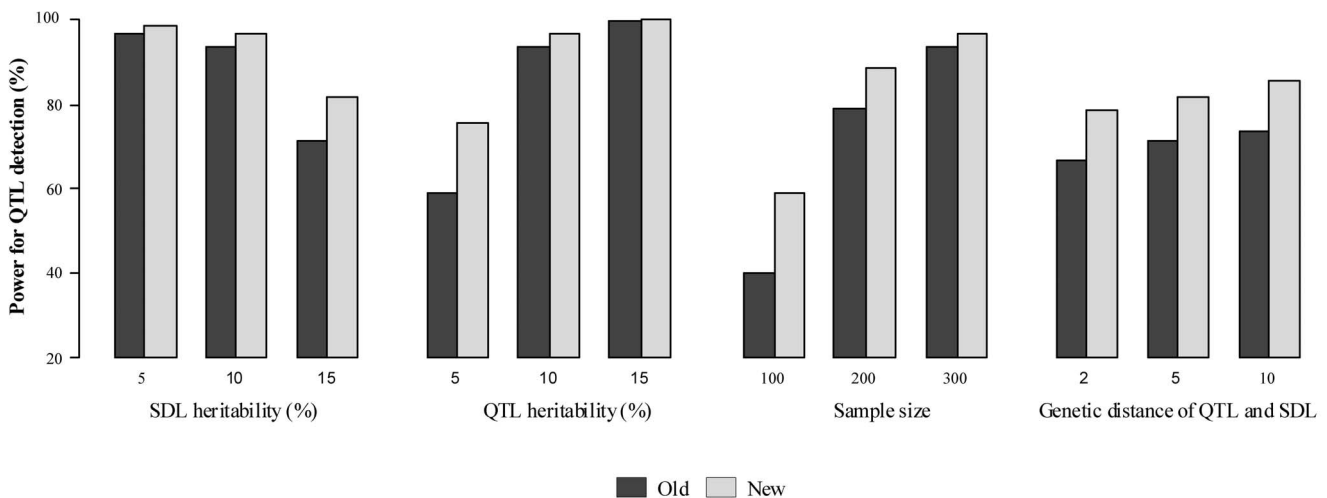


Figure 1. Power of QTL detection under various situations. Old: the modified method of Wen et al. (2013) with the corrected linkage groups; New: the proposed method in this study. doi:10.1371/journal.pone.0068510.g001

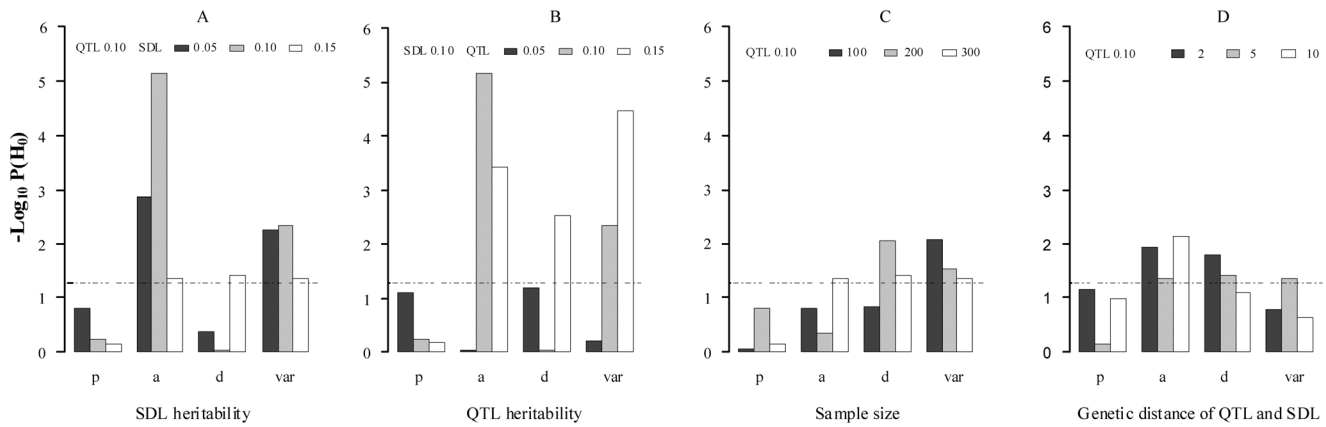


Figure 2. $-\log_{10}P(H_0)$ for QTL parameters in the paired *t* test between the old and new methods. *p*: QTL position; *a*: additive effect of QTL; *d*: dominant effect of QTL; *var*: residual variance. The dashed line represents the critical value at the 0.05 level of significance. doi:10.1371/journal.pone.0068510.g002

adopted: 1) New method, the proposed method in this study; and 2) Old method, the modified method of Wen et al. [10] with linkage groups corrected by epistatic distorted markers. The mean, absolute bias and standard deviation among the estimates obtained from 200 replicates were used to indicate the precision, and paired *t* test was used to compare the above two methods. All simulation parameters were given in **Table 3**.

All the results are listed in **Table S1** and **Figs. 1, 2A, 3A and 3E**. Results showed that the ability to detect QTL decreased as SDL heritability increased in both two methods (**Table S1**). However, some differences between the two methods were observed if SDL heritability was fixed. For example, the power for QTL detection is higher for new method than for old one (**Fig. 1**); and significant differences for QTL parameter estimates, QTL additive effect and residual variance, exist (**Fig. 2A**). The

standard deviations and absolute biases for the estimates of QTL additive and dominance effects are all small, although the accuracy of the new method decreases as SDL heritability increases (**Fig. 3A** and 3E). The unbiasedness is better for the new method than for old one, although the difference of SD between the two methods is not obvious (**Table S1**).

Effect of QTL heritability on new method. In the second simulation experiment, the effect of QTL size on the new method was assessed by setting QTL heritability as 0.05, 0.10 and 0.15, indicating $a = d = 0.2649, 0.3849$ and 0.4851 , respectively. SDL heritability was set as 0.10. Other parameters were the same as those in the first simulation experiment (**Table 3**). All the results were listed in **Table S2** and **Figs. 1, 2B, 3B and 3F**. Results showed the similar trends in the first simulated experiment. In addition, the difference of QTL detection power between the

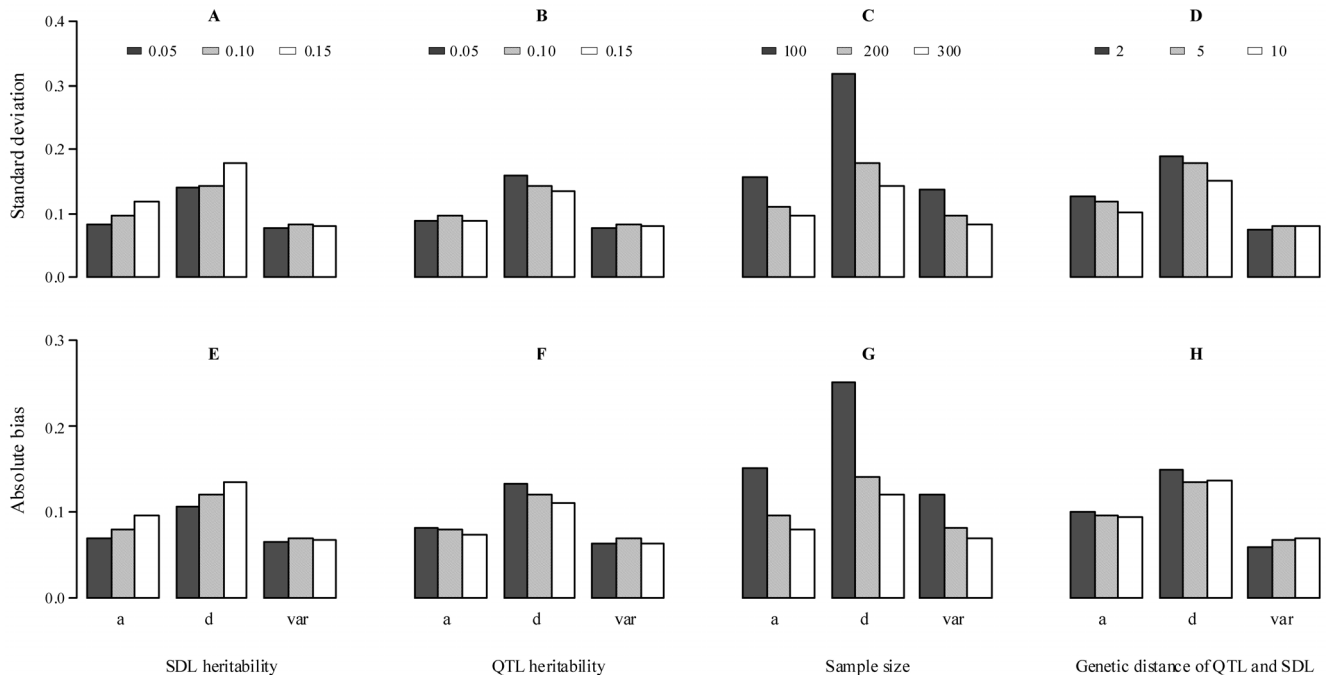


Figure 3. Standard deviation and absolute bias for QTL estimates in the new method. *a*: additive effect; *d*: dominant effect; and *var*: residual variance. doi:10.1371/journal.pone.0068510.g003

above two methods decrease as the increase of QTL heritability (**Fig. 1**), and significant difference for QTL dominant effect exists under QTL heritability of 0.15 (**Fig. 2B**). As for the accuracy of the new method, a general trend is observed (**Figs. 3B and 3F**). In addition, the unbiasedness is also better for the new method than for old one (Table S2).

Effect of sample size on new method. The third simulation experiment was designed to investigate the effects of sample size on the new method by setting sample size as 100, 200 and 300. Heritability of each SDL and QTL was set as 0.10. Other parameters were the same as those in the first simulation experiment (Table 3). All the results were listed in Table S3 and **Figs. 1, 2C, 3C and 3G**. Results showed a general trend in QTL mapping. In addition, significant differences exist in QTL parameter estimates, i.e., QTL dominant effect and residual

variance (**Fig. 2C**). As for the accuracy for parameter estimation, the results are similar to those in the second simulation experiment (**Fig. 3C and 3G**).

Effect of genetic distance between QTL and SDL on new method. The last simulation was performed to evaluate the effect of genetic distance between SDL and QTL on the new method by setting the distance as 2, 5 and 10 cM. All the parameters were shown in Table 3, and all the results were showed in Table S4 and **Figs. 1, 2D, 3D and 3H**. Results showed a general trend in QTL mapping. In addition, significant differences exist in QTL parameter estimates, i.e., QTL additive and dominant effects (**Fig. 2D**). The standard deviations of QTL additive and dominant effects decrease as the genetic distance between QTL and SDL increases (**Fig. 3D**), the absolute bias for

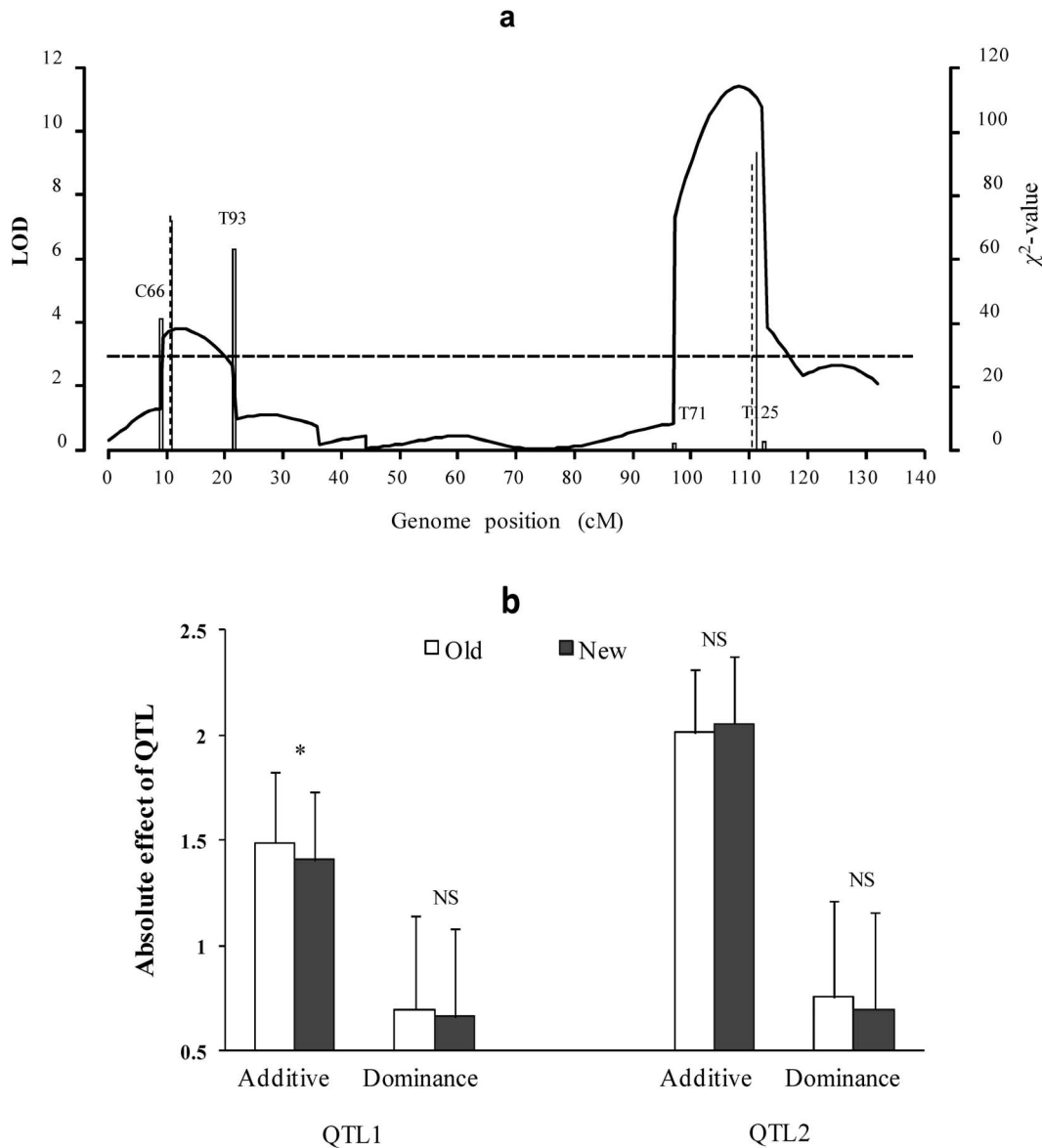


Figure 4. Mapping QTL for weight in 333 mouse F₂ individuals. (a) LOD scores using composite interval mapping (CIM, curve), old (solid vertical line) and new (dashed vertical line) methods. Dashed horizontal line represents critical value for significant QTL. Hollow vertical line indicates χ^2 value of segregation distortion test for marker interval of QTL; (b) QTL absolute effects, detected by old and new methods. NS and star indicate no difference and significant difference at the 0.01 level between old and new methods, respectively. doi:10.1371/journal.pone.0068510.g004

QTL additive effect has a same trend as standard deviation (**Fig. 3H**).

Real Data Analysis

The mouse dataset of 304 F_2 individuals, derived from MapMaker/Exp v3.0b [31], was used for the demonstration. Twelve RFLP markers were divided into two linkage groups. Each linkage group was corrected by the method of Xie [11], and the corrected length was 132.36 cM. Using the corrected linkage groups, QTL for weight in mouse was detected by three approaches, the above two methods and composite interval mapping [32]. When the marker density is too sparse (>1 cM), a virtual marker (treated as missing data) may be inserted. In the case of incomplete marker information, the corrected conditional probabilities were calculated by equation (8) and these probabilities were used to calculate dummy variables for all putative QTL, i.e., $x_{2l-1j} = p'_{j(QQ_l)} - p'_{j(qq_l)}$ and $x_{2l,j} = p'_{j(Qq_l)}$. The LOD threshold was set at 3.0 for QTL detection.

All the results are listed in **Table S5** and shown in **Fig. 4**. Among two QTL detected by all the three approaches, no significant difference for estimates of the second QTL is observed between the old and new methods (**Fig. 4b**). However, significant difference for additive effect of the first QTL is found (**Fig. 4b**). This is because that the first QTL is located between epistatic distorted markers C66 and T93 ($u=0.71$, $v=0.04$, $x=0.21$, $\lambda=2[L(u,v,x) - L(u,v,x=1)] = 13.11 \chi^2_{df=1}$, $p=2.94E-4$), and the second QTL is located at marker region with normal segregation. The result further confirms the conclusions of the simulation studies.

Discussion

Marker segregation distortion is a common phenomenon observed in QTL analysis [6], thus Wen et al. [10] proposed multi-QTL mapping using distorted markers. However, this work needs to be addressed in two aspects. First, linkage groups in Wen et al. [10] are not corrected by making use of distorted marker information. Second, epistasis between two SDL is not considered although the epistasis is very important. To overcome the above shortcomings, a new approach was proposed in this study. In the simulated data analyses, some new results were found, for example, QTL detection power is slightly higher for the new method than for the above old method and significant difference for dominant effect of QTL is observed between the above two methods (**Fig. 2**). In real data analyses, epistatic distorted markers were detected. Therefore, the new approach is valuable. To further validate the new method, we assume no epistasis between two linked SDL. Results from a Monte Carlo simulation experiment show that similar results for QTL detection power and parameter estimates between the old and new methods were observed (**Table S7**), indicating that the new method works well.

In this study new method is based on gametic selection. This is because gametic selection caused by epistatic SDL is often reported [21,23–24,33]. In this case, multi-QTL mapping approach can be set up by incorporating viability coefficients of male gametes (u , v and x) into QTL mapping approach. Note that the fitness model can be linked with the quantitative genetics model for viability selection [11,28]. Therefore, epistatic SDL effect and size can be easily calculated. As for zygotic selection, once epistatic SDL is identified by the method of Xie [11], the

epistatic SDL information can be used to correct conditional probabilities of QTL genotypes in multi-QTL mapping approach. Therefore, multi-QTL mapping approach under zygotic selection can be easily set up.

In real data analysis, corrected linkage groups are also useful. To explain this issue, we re-analyze real dataset in Wen et al. [10]. As compared with the results from Wen et al. [10], one additional QTL was further detected by the old and new methods (**Table S6**), indicating that the corrected linkage groups using distorted markers can increase the power of QTL mapping. This result is consistent with that in Monte Carlo simulation studies in this study. As for the same QTL detected by the old and new methods, no significant differences are identified. This is because no epistasis SDL was mapped. As for the third QTL, the position from Wen et al. [10] is different from those from the old and new methods; and as for the fourth QTL, the additive effect estimate from Wen et al. [10] is different from those from the old and new methods. The possible reason is derived from the corrected linkage groups.

This method may be extended to the QTL-by-QTL interaction detection and additional bi-parental populations, e.g. backcross, recombination inbred lines and doubled haploid lines. Although co-dominant markers were adopted in this study, missing and dominant markers are also available, as shown in the real data analysis in this study. The related results can be found in Xie [11]. If more than two distorted markers exist in a same linkage group, a two-dimensional scanning approach in this study was also used to detect the epistasis. The source codes for R program and the much more user-friendly interface software will be available soon.

Supporting Information

Table S1 Effect of SDL heritability on new method.
(DOC)

Table S2 Effect of QTL heritability on new method.
(DOC)

Table S3 Effect of sample size on new method.
(DOC)

Table S4 Effect of genetic distance between QTL and SDL on new method.
(DOC)

Table S5 Mapping QTL for weight in 333 mouse F_2 individuals using composite interval mapping (CIM), old and new methods.
(DOC)

Table S6 Mapping QTL for dried soymilk in 222 soybean $F_{2,4}$ families using Wen et al. (2013), old and new methods.
(DOC)

Table S7 Estimates of SDL and QTL parameters by old and new methods at the case that epistasis between two linked SDL is absent.
(DOC)

Author Contributions

Conceived and designed the experiments: YMZ. Performed the experiments: SQX JW YMZ. Analyzed the data: SQX JW YMZ. Contributed reagents/materials/analysis tools: SQX. Wrote the paper: YMZ SQX.

References

1. Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, et al. (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. *Genetics* 127: 181–197.
2. Lynch M, Walsh B (1998) *Genetics and analysis of quantitative traits*. Sinauer Associates, Inc, Sunderland, MA.
3. Mackay TFC (2001) Quantitative trait loci in *Drosophila*. *Nat Rev Genet* 2: 11–20.
4. Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. *Nat Rev Genet* 3: 43–52.
5. Xu Y, Li HN, Li G, Wang X, Cheng L, et al. (2011) Mapping quantitative trait loci for seed size traits in soybean (*Glycine max* L. Merr.). *Theor Appl Genet* 122: 581–594.
6. Lyttle TW (1991) Segregation distortion. *Ann Rev Genet* 25: 511–557.
7. Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics* Fourth edition. London: Longman.
8. Carr DE, Dudash MR (2003) Recent approaches into the genetic basis of inbreeding depression in plants. *Philos Trans R Soc London B* 358: 1071–1084.
9. Zhu CS, Wang CM, Zhang YM (2007) Modeling segregation distortion for viability selection I. Reconstruction of genetic linkage maps with distorted markers. *Theor Appl Genet* 114: 295–305.
10. Wen J, Can VT, Zhang YM (2013) Multi-QTL mapping for quantitative traits using distorted markers. *Mol Breed* 31(2): 395–404.
11. Xie SQ (2013) Genetic map correction, QTL mapping and software development under epistatic segregation distortion locus (Ph D Dissertation). Nanjing Agricultural University.
12. Taylor DR, Ingvarsson PK (2003) Common features of segregation distortion in plants and animals. *Genetica* 117: 27–35.
13. Vogl C, Xu S (2000) Multipoint mapping of viability and segregation distorting loci using molecular markers. *Genetics* 155: 1439–1447.
14. Lorieux M, Goffinet B, Perrier X, González-de-León D, Lanaud C (1995a) Maximum-likelihood models for mapping genetic markers showing segregation distortion. 1. Backcross population. *Theor Appl Genet* 90: 73–80.
15. Lorieux M, Perrier X, Goffinet B, Lanaud C, González-de-León D (1995b) Maximum-likelihood models for mapping genetic markers showing segregation distortion. 2. F₂ populations. *Theor Appl Genet* 90: 81–89.
16. Xu S (2008) Quantitative trait locus mapping can benefit from segregation distortion. *Genetics* 180: 2201–2208.
17. Zhang LY, Wang SQ, Li HH, Deng QM, Zheng AP, et al. (2010) Effects of missing markers and segregation distortion on QTL mapping in F₂ populations. *Theor Appl Genet* 121: 1071–1082.
18. Xu S, Hu Z (2009) Mapping quantitative trait loci using distorted markers. *Int J Plant Genomics* Article ID: 410825.
19. Phillips PC (2008) Epistasis: the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 9: 855–867.
20. Törjék O, Witucka-Wall H, Meyer HC, Maria von Korff, Kusterer B, et al. (2006) Segregation distortion in *Arabidopsis* C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. *Theor Appl Genet* 113: 1551–1561.
21. Chang AS, Noor MAF (2010) Epistasis modifies the dominance of loci causing hybrid male sterility in the *Drosophila pseudoobscura* species group. *Evolution* 64: 253–260.
22. Li XH, Wang XJ, Wei YL, Brummer EC (2011) Prevalence of segregation distortion in diploid alfalfa and its implications for genetics and breeding applications. *Theor Appl Genet* 123: 667–679.
23. Kubo T, Yamagata Y, Eguchi M, Yoshimura A (2008) A novel epistatic interaction at two loci causing hybrid male sterility in an inter-subspecific cross of rice (*Oryza sativa* L.). *Genes Genet Syst* 83: 443–453.
24. Kubo T, Yoshimura A, Kurata N (2011) Hybrid male sterility in rice is due to epistatic interactions with a pollen killer locus. *Genetics* 189: 1083–1092.
25. Yang J, Zhao X, Cheng K, Du H, Ouyang Y, et al. (2012) A killer-protector system regulates both hybrid sterility and segregation distortion in rice. *Science* 337(6100): 1336–1340.
26. Xu S (2010) An expectation-maximization algorithm for the Lasso estimation of quantitative trait locus effects. *Heredity* 105: 483–494.
27. Kosambi DD (1944) The estimation of map distances from recombination values. *Ann Eugen* 12: 172–175.
28. Luo L, Zhang YM, Xu S (2005) A quantitative genetics model for viability selection. *Heredity* 94: 347–355.
29. Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. *Genetics* 152: 1203–1216.
30. Jiang CJ, Zeng ZB (1997) Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. *Genetica* 101: 47–58.
31. Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, et al. (1987) MapMaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174–181.
32. Wang S, Basten CJ, Zeng ZB (2012) *Windows QTL Cartographer v2.5*. Department of Statistics, North Carolina State University, Raleigh, NC.
33. Baumbach J, Rogers JP, Slattery RA, Narayanan NN, Xu M, et al. (2012) Segregation distortion in a region containing a male-sterility, female-sterility locus in soybean. *Plant Sci* 195: 151–156.