

Research Article

Mapping Quantitative Trait Loci Using Distorted Markers

Shizhong Xu and Zhiqiu Hu

Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, USA

Correspondence should be addressed to Shizhong Xu, shizhong.xu@ucr.edu

Received 21 September 2009; Accepted 13 November 2009

Recommended by Manuel Talon

Quantitative trait locus (QTL) mapping is usually performed using markers that follow a Mendelian segregation ratio. We developed a new method of QTL mapping that can use markers with segregation distortion (non-Mendelian markers). An EM (expectation-maximization) algorithm is used to estimate QTL and SDL (segregation distortion loci) parameters. The joint analysis of QTL and SDL is particularly useful for selective genotyping. Application of the joint analysis is demonstrated using a real life data from a wheat QTL mapping experiment.

Copyright © 2009 S. Xu and Z. Hu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Segregation distortion is a phenomenon that the genotypic frequency array of a locus does not follow a typical Mendelian ratio. Depending on the population under investigation, Mendelian ratio of a locus varies from 1 : 1 for a backcross to 1 : 2 : 1 for an F_2 and to 1 : 1 : 1 : 1 for a four-way cross. These ratios hold for codominant markers. For some reasons, a marker may not follow a typical Mendelian ratio. Such a marker is called a distorted marker. For a long period of time, the effects of distorted markers on the result of QTL mapping were not known. For the reason of precaution, people simply discarded all the distorted markers in QTL mapping. Recently, we found that distorted markers can be safely used for QTL mapping with no detrimental effect on the result of QTL mapping [1]. This finding can help QTL mappers save tremendous resources by using all available markers, regardless whether they are Mendelian or not. We also found that if distorted markers are handled properly, they can be beneficial to QTL mapping.

Marker segregation distortion is only a phenomenon. The reason behind the distortion is due to one or more segregation distortion loci (SDL). These loci are subject to gametic selection [2], zygotic selection [3], or both and their (unobservable) distorted segregation causes the observed markers to deviate from the Mendelian ratio. Several investigators [4–11] have attempted to map these segregation distortion loci using molecular markers. It is natural to consider mapping QTL and SDL jointly in the same population.

Agricultural scientists are interested in mapping QTL for economically important traits while evolutionary biologists are interested in mapping SDL that respond to natural selection. Combining the two mapping strategies into one is beneficial to both communities. Performing such a joint mapping strategy is the main objective of this study. Since the theory of segregation distortion has been introduced and discussed in previous studies [7, 8] and our own research [1], this study only presents the EM (expectation-maximization) implementation of the statistical method. The variance-covariance matrix of estimated parameters under the EM algorithm is also derived and presented in Appendix A for interested readers.

2. Methods

We only investigate interval mapping where a model contains a single QTL at a time and the entire genome is scanned through repeated calling of the same program for different locations of the genome. The technical difference between the joint mapping and QTL mapping occurs only in one place. In the traditional interval mapping of QTL, the conditional probabilities of genotypes for a QTL are calculated using flanking marker genotypes with the prior probabilities of QTL genotypes being substituted by the Mendelian ratio. For the joint mapping, the genotypic frequencies (segregation ratios) are treated as unknown parameters that are subject to estimation. We use an F_2 population as an

example to demonstrate the method. Extension to other population is discussed subsequently.

2.1. The Likelihood of Markers. Let M and N be the left and right flanking markers bracketing the QTL (denoted by G for short). The interval of the genome carrying the three loci is labeled by a segment MGN . The marker linkage phases are known for line crosses derived from inbred lines. The three genotypes of the QTL are denoted by G_1G_1 , G_1G_2 , and G_2G_2 , respectively. Similar notation also applies to the genotypes of the flanking markers. The interval defined by markers M and N is divided into two segments. Let r_1 and r_2 be the recombination fractions for segments MG and GN , respectively. The joint distribution of the marker genotypes conditional on the QTL genotype can be derived using the Markov chain property under the assumption of no interference between consecutive loci in segregation. Let us denote the three ordered genotypes, G_1G_1 , G_1G_2 , and G_2G_2 , by genotypes 1, 2, and 3, respectively. If individual j takes the κ th genotype for the QTL, we denote the event by $G_j = \kappa$, for all $\kappa = 1, 2, 3$. The joint probability of the two markers conditional on the genotype of the QTL is

$$\begin{aligned} \Pr(M_j = \xi, N_j = \zeta | G_j = \kappa) \\ = \Pr(M_j = \xi | G_j = \kappa) \Pr(N_j = \zeta | G_j = \kappa) \end{aligned} \quad (1)$$

for all $\kappa, \xi, \zeta = 1, 2, 3$, where $\Pr(M_j = \xi | G_j = \kappa) = \Gamma_1(\kappa, \xi)$ and $\Pr(N_j = \zeta | G_j = \kappa) = \Gamma_2(\kappa, \zeta)$. We use $\Gamma_i(\kappa, \xi)$ to denote the κ th row and the ξ th column of the following transition matrix:

$$\Gamma_i = \begin{bmatrix} (1-r_i)^2 & 2r_i(1-r_i) & r_i^2 \\ r_i(1-r_i) & (1-r_i)^2 + r_i^2 & r_i(1-r_i) \\ r_i^2 & 2r_i(1-r_i) & (1-r_i)^2 \end{bmatrix}, \quad \forall i = 1, 2. \quad (2)$$

For example,

$$\begin{aligned} \Pr(M_j = 1, N_j = 2 | G_j = 3) \\ = \Pr(M_j = 1 | G_j = 3) \Pr(N_j = 2 | G_j = 3) \\ = \Gamma_1(3, 1) \Gamma_2(3, 2) = 2r_1^2 r_2 (1-r_2). \end{aligned} \quad (3)$$

Let $\omega_\kappa = \Pr(G = \kappa)$, for all $\kappa = 1, 2, 3$, be the probability that a randomly sampled individual from the F_2 family has a genotype κ . We use a generic notation p for probability so that $p(G_j = \kappa)$ represents $\Pr(G_j = \kappa)$ and $p(M_j, N_j | G_j = \kappa)$ stands for $\Pr(M_j, N_j | G_j = \kappa)$. The log likelihood function of the flanking marker genotypes in the F_2 population is

$$\begin{aligned} L(\omega | m) &= \sum_{j=1}^n \ln \left[\sum_{\kappa=1}^3 p(G_j = \kappa) p(M_j, N_j | G_j = \kappa) \right] \\ &= \sum_{j=1}^n \ln \left[\sum_{\kappa=1}^3 \omega_\kappa \Gamma_1(\kappa, M_j) \Gamma_2(\kappa, N_j) \right], \end{aligned} \quad (4)$$

where $\omega^T = [\omega_1 \ \omega_2 \ \omega_3]$ is a vector of parameters with constraint $\sum_{\kappa=1}^3 \omega_\kappa = 1$, where m in $L(\omega | m)$ stands for marker information. Note that without any prior information, $p(G_j = \kappa) = \omega_\kappa$, for all $j = 1, \dots, n$. Under the assumption of Mendelian segregation, $\omega = \phi$ where $\phi = [\phi_1 \ \phi_2 \ \phi_3]^T = [1/4 \ 1/2 \ 1/4]^T$. However, we treat ω as unknown parameters. Because we are dealing with the genotypic frequencies, the segregation distortion is called zygotic distortion. Segregation distortion due to gametic selection will be discussed later. We postulate that deviation of ω from ϕ causes a marker linked to locus G to show distorted segregation. This likelihood function is the one used in mapping viability loci [10].

2.2. The Likelihood of Phenotypes. Let y_j be the phenotypic value of a quantitative trait measured from individual j . The probability density of y_j conditional on the genotype of individual j is normal with mean

$$\mu_\kappa = X_j \beta + H_\kappa \gamma \quad (5)$$

and variance σ^2 , that is,

$$p(y_j | G_j = \kappa) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left[-\frac{1}{2\sigma^2} (y_j - X_j \beta - H_\kappa \gamma)^2 \right], \quad (6)$$

where H_κ is the κ th row of matrix H and

$$H = \begin{bmatrix} +1 & -1 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}. \quad (7)$$

This H matrix can be defined in a different scale, for example,

$$H = \begin{bmatrix} +1 & 0 \\ 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad (8)$$

which does not affect the significance test. The advantage of choosing the scale in (7) is that the expectation of the dominance indicator is zero. Vector $\gamma = [a \ d]^T$ contains the additive and dominance effects. The design matrix X_j and the regression coefficients β capture non-QTL effects, for example, field location effects, year effects, and so on. The likelihood function of the phenotypic values in the F_2 population is

$$\begin{aligned} L(\beta, \gamma, \sigma^2, \omega | y) \\ = \sum_{j=1}^n \ln \left[\sum_{\kappa=1}^3 p(G_j = \kappa) p(y_j | G_j = \kappa) \right] \\ = -\frac{n}{2} \ln(\sigma^2) + \sum_{j=1}^n \ln \left[\sum_{\kappa=1}^3 \omega_\kappa \exp \left[-\frac{1}{2\sigma^2} (y_j - \mu_\kappa)^2 \right] \right], \end{aligned} \quad (9)$$

where letter y in $L(\beta, \gamma, \sigma^2, \omega | y)$ stands for the phenotype. This likelihood function is the one used in segregation analysis of quantitative traits [12] because no marker information is required.

2.3. *Joint Likelihood of Markers and Phenotypes.* Let $\theta = [\beta^T \gamma^T \sigma^2 \omega^T]^T$ be a vector of all parameters in the joint analysis. The likelihood function is obtained by combining (4) and (9):

$$\begin{aligned} L(\theta | m, y) &= \sum_{j=1}^n \ln \left[\sum_{\kappa=1}^3 p(G_j = \kappa) p(y_j | G_j = \kappa) p(M_j, N_j | G_j = \kappa) \right] \\ &= \sum_{j=1}^n \ln \left\{ \sum_{\kappa=1}^3 \omega_{\kappa} \exp \left[-\frac{1}{2\sigma^2} (y_j - \mu_{\kappa})^2 \right] \Gamma_1(\kappa, M_j) \Gamma_2(\kappa, N_j) \right\} \\ &\quad - \frac{n}{2} \ln(\sigma^2). \end{aligned} \quad (10)$$

For QTL mapping under segregation distortion, this log likelihood function is the one to be maximized. The previous two likelihood functions (for markers and for phenotypes) were presented as background information to introduce this joint log likelihood function.

2.4. *EM Algorithm for the Joint Analysis.* The MLE (maximum likelihood estimate) of the parameters is solved via an EM algorithm [13]. We need to rewrite the likelihood function in a form of complete data. Let us define a delta function as

$$\delta(G_j, \kappa) = \begin{cases} 1 & \text{if } G_j = \kappa, \\ 0 & \text{if } G_j \neq \kappa, \end{cases} \quad (11)$$

If the genotypes of all individuals are known, that is, given $\delta(G_j, \kappa)$ for all $j = 1, \dots, n$ and $\kappa = 1, 2, 3$, the complete-data log likelihood is

$$L(\theta, \delta) = \sum_{j=1}^n \ln [p(y_j | G_j) p(M_j, N_j | G_j) p(G_j)], \quad (12)$$

where

$$\begin{aligned} p(y_j | G_j) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{\kappa=1}^3 \delta(G_j, \kappa) (y_j - \mu_{\kappa})^2 \right], \\ p(M_j, N_j | G_j) &= \prod_{\kappa=1}^3 p(M_j, N_j | G_j = \kappa)^{\delta(G_j, \kappa)} \\ &= \prod_{\kappa=1}^3 [\Gamma_1(\kappa, M_j) \Gamma_2(\kappa, N_j)]^{\delta(G_j, \kappa)}, \\ p(G_j) &= \prod_{\kappa=1}^3 \omega_{\kappa}^{\delta(G_j, \kappa)}. \end{aligned} \quad (13)$$

The delta variables are missing values. Therefore, we need to take expectation of the log likelihood with respect to δ . The

expected complete-data log likelihood function is

$$L(\theta | \theta^{(t)}) = E_{\delta} [L(\theta, \delta) | \theta^{(t)}] = \psi_0 + \psi_1(\theta) + \psi_2(\theta). \quad (14)$$

Note that $E_{\delta} [L(\theta, \delta) | \theta^{(t)}]$ stands for the expectation of $L(\theta, \delta)$ with respect to δ conditional on parameters at the current state $\theta^{(t)}$ and the data (the symbol for data is suppressed for simplicity). The three components of (14) are

$$\begin{aligned} \psi_0 &= \sum_{j=1}^n \sum_{\kappa=1}^3 E_{\delta} [\delta(G_j, \kappa) | \theta^{(t)}] \\ &\quad \times [\ln \Gamma_1(\kappa, M_j) + \ln \Gamma_2(\kappa, N_j)], \\ \psi_1(\theta) &= -\frac{n}{2} \ln(\sigma^2) \\ &\quad - \frac{1}{2\sigma^2} \sum_{j=1}^n \sum_{\kappa=1}^3 E_{\delta} [\delta(G_j, \kappa) | \theta^{(t)}] (y_j - \mu_{\kappa})^2, \\ \psi_2(\theta) &= \sum_{j=1}^n \sum_{\kappa=1}^3 E_{\delta} [\delta(G_j, \kappa) | \theta^{(t)}] \ln \omega_{\kappa}. \end{aligned} \quad (15)$$

The first component ψ_0 is a function of $\theta^{(t)}$ but not a function of θ . Therefore, it is considered as a constant.

2.4.1. *Expectation (E-Step).* The expectation step of the EM algorithm requires computing the expectation of δ conditional on the data and $\theta^{(t)}$. Because δ is a Bernoulli variable, the expectation is simply the probability of $\delta = 1$, that is,

$$\begin{aligned} E_{\delta} [\delta(G_j, \kappa) | \theta^{(t)}] &= \Pr [\delta(G_j, \kappa) = 1 | \theta^{(t)}, m, y] \\ &= \frac{p(G_j = \kappa) p(y_j | G_j = \kappa) p(M_j, N_j | G_j = \kappa)}{\sum_{\xi=1}^3 p(G_j = \xi) p(y_j | G_j = \xi) p(M_j, N_j | G_j = \xi)} \\ &= \frac{\omega_{\kappa} \exp \left[-(1/2\sigma^2) (y_j - \mu_{\kappa})^2 \right] \Gamma_1(\kappa, M_j) \Gamma_2(\kappa, N_j)}{\sum_{\xi=1}^3 \omega_{\xi} \exp \left[-(1/2\sigma^2) (y_j - \mu_{\xi})^2 \right] \Gamma_1(\xi, M_j) \Gamma_2(\xi, N_j)}. \end{aligned} \quad (16)$$

2.4.2. *Maximization (M-Step).* The maximization step of the EM algorithm requires taking the partial derivatives of $L(\theta | \theta^{(t)})$ with respect to θ , setting the partial derivatives equal to zero, and solving for the parameters:

$$\frac{\partial}{\partial \theta} L(\theta | \theta^{(t)}) = \frac{\partial}{\partial \theta} \psi_1(\theta) + \frac{\partial}{\partial \theta} \psi_2(\theta) = 0. \quad (17)$$

The solutions of the parameters are

$$\begin{aligned}\beta &= \left[\sum_{j=1}^n X_j^T X_j \right]^{-1} \left[\sum_{j=1}^n \sum_{\kappa=1}^3 E[\delta(G_j, \kappa)] X_j^T (y_j - H_\kappa \gamma) \right], \\ \gamma &= \left[\sum_{j=1}^n \sum_{\kappa=1}^3 E[\delta(G_j, \kappa)] (H_\kappa^T H_\kappa) \right]^{-1} \\ &\quad \times \left[\sum_{j=1}^n \sum_{\kappa=1}^3 E[\delta(G_j, \kappa)] H_\kappa^T (y_j - X_j \beta) \right], \\ \sigma^2 &= \frac{1}{n} \sum_{j=1}^n \sum_{\kappa=1}^3 E[\delta(G_j, \kappa)] (y_j - X_j \beta - H_\kappa \gamma)^2, \\ \omega_\kappa &= \frac{1}{n} \sum_{j=1}^n E[\delta(G_j, \kappa)], \quad \forall \kappa = 1, 2, 3.\end{aligned}\tag{18}$$

2.5. Hypothesis Tests. Hypothesis tests are complicated when QTL segregate in a non-Mendelian fashion. There are many different hypotheses we can test here. Although the Wald test can be performed for testing the presence of QTL, such a test is not justified for testing the null hypothesis of Mendelian segregation. Therefore, the likelihood ratio tests are more justifiable. Regardless what hypothesis is tested, the full model joint log likelihood function given in (10) is required. Let us reintroduce this joint log likelihood function using a different notation so that different likelihood ratio tests are easily interpreted. The joint likelihood is rewritten as

$$\begin{aligned}L_{QS}(\gamma, \omega) &= -\frac{n}{2} \log(\sigma^2) \\ &\quad + \sum_{j=1}^n \ln \left\{ \sum_{\kappa=1}^3 \omega_\kappa \exp \left[-\frac{1}{2\sigma^2} (y_j - \mu_\kappa)^2 \right] \Gamma_1(\kappa, M_j) \Gamma_2(\kappa, N_j) \right\},\end{aligned}\tag{19}$$

where γ represents QTL effects and ω stands for non-Mendelian segregation. The null hypothesis for QTL detection is $H_{QTL} : \gamma = 0$ while the null hypothesis for detecting segregation distortion is $H_{SDL} : \omega = \phi$.

2.5.1. Testing the Presence of QTL. The null hypothesis is $H_{QTL} : \gamma = 0$. The likelihood ratio test statistic is

$$\lambda_{QTL} = -2[L_S(0, \hat{\omega}) - L_{QS}(\hat{\gamma}, \hat{\omega})],\tag{20}$$

where $L_S(0, \hat{\omega})$ is the log likelihood value under the null model $\gamma = 0$, which is

$$L_S(0, \hat{\omega}) = L(\hat{\beta}, \hat{\sigma}^2 | y) + L(\hat{\omega} | m),\tag{21}$$

where

$$\begin{aligned}L(\hat{\beta}, \hat{\sigma}^2 | y) &= -\frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} \sum_{j=1}^n (y_j - X_j \hat{\beta})^2, \\ L(\hat{\omega} | m) &= \sum_{j=1}^n \ln \left[\sum_{\kappa=1}^3 \hat{\omega}_\kappa \Gamma_1(\kappa, M_j) \Gamma_2(\kappa, N_j) \right].\end{aligned}\tag{22}$$

The estimated parameters in (22) are obtained by maximizing the corresponding likelihood functions (see Appendix B for the estimation).

2.5.2. Testing Non-Mendelian Segregation. The null hypothesis is $H_{SDL} : \omega = \phi$. The likelihood ratio test statistic is

$$\lambda_{SDL} = -2[L_Q(\hat{\gamma}, \phi) - L_{QS}(\hat{\gamma}, \hat{\omega})],\tag{23}$$

where

$$\begin{aligned}L_Q(\hat{\gamma}, \phi) &= -\frac{n}{2} \ln(\hat{\sigma}^2) \\ &\quad + \sum_{j=1}^n \ln \left\{ \sum_{\kappa=1}^3 \phi_\kappa \exp \left[-\frac{1}{2\hat{\sigma}^2} (y_j - \hat{\mu}_\kappa)^2 \right] \Gamma_1(\kappa, M_j) \Gamma_2(\kappa, N_j) \right\}.\end{aligned}\tag{24}$$

Again, the MLEs of the parameters in (24) are obtained by maximizing this likelihood function (see Appendix B for the estimation of parameters under the null model).

2.5.3. Testing Both QTL and SDL. The null hypothesis is $H_0 : \gamma = 0$ and $\omega = \phi$. The likelihood ratio test statistic is

$$\lambda_{QS} = -2[L(0, \phi) - L_{QS}(\hat{\gamma}, \hat{\omega})],\tag{25}$$

where

$$L(0, \phi) = L(\hat{\beta}, \hat{\sigma}^2 | y) + L(\phi | m).\tag{26}$$

The two components of (26) are

$$\begin{aligned}L(\hat{\beta}, \hat{\sigma}^2 | y) &= -\frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} \sum_{j=1}^n (y_j - X_j \hat{\beta})^2, \\ L(\phi | m) &= \sum_{j=1}^n \ln \left[\sum_{\kappa=1}^3 \phi_\kappa \Gamma_1(\kappa, M_j) \Gamma_2(\kappa, N_j) \right].\end{aligned}\tag{27}$$

The parameters involved in these log likelihood functions are estimated using formulas given in Appendix B. This hypothesis is rejected if either $\gamma \neq 0$ or $\omega \neq \phi$ or both inequalities hold. The QTL effects and the segregation distortion are confounded. This hypothesis test is useful in the following situation. Suppose that, for some reason, we know for sure that the population from which the sample is drawn is a Mendelian population. The sample drawn from this population is selected based on extreme

phenotypes (selective genotyping). The sample is then non-Mendelian regarding the QTL that control the trait subject to phenotypic selection. Rejecting this hypothesis is equivalent to rejecting the null hypothesis of QTL. The reason is that segregation distortion in the sample is solely caused by selective genotyping. Therefore, this joint test can be used to detect QTL under selective genotyping.

3. Applications

This example demonstrates the application of the method to joint mapping of QTL and SDL in a wheat QTL mapping experiment. The experiment was conducted by Dou et al. [14] who made the data available to us for this analysis. A female sterile line XND126 and an elite cultivar Gaocheng 8901 with normal fertility were crossed for genetic analysis of female sterility measured as a ratio of the number of seeded spikelets to the total number of spikelets per plant. The parents and their F_1 and F_2 progeny were planted in the Huaian experimental station in China for the 2006-2007 growing season under the normal autumn sowing condition. The mapping population was the F_2 family consisting of 234 individual plants. The sterility trait was transformed using the angular sine transformation, $y = \arcsin(x)$, where x is the phenotypic value expressed as ratio. A total of 28 SSR markers were used in this experiment. These markers covered five chromosomes of the wheat genome with an average genome marker density of 15.5 cM per marker interval. The five chromosomes are only part of the wheat genome. The model for the female sterility is

$$y_j = X_j\beta + Z_{j1}\gamma_1 + Z_{j2}\gamma_2 + \varepsilon_j, \quad (28)$$

where $X_j = 1$ for all j , β is the intercept, $Z_{j1} = \{+1, 0, -1\}$ is the genotype indicator variable for the additive effect $\gamma_1 = a$, and $Z_{j2} = \{-1, 1, -1\}$ is the genotype indicator variable for the dominance effect $\gamma_2 = d$. We used an interval mapping approach to scanning the entire genome. Therefore, the model contains one QTL at a time. With the interval mapping, $Z_j = [Z_{j1} \ Z_{j2}]$ is missing and can take one of three values denoted by H_κ for $\kappa = 1, 2, 3$ (see definition of H_κ) in Section 2.

The likelihood ratio test statistics were divided by 4.61 to obtain their corresponding LOD scores. The LOD score profiles across the genome are shown in Figure 1. The top panel (a) shows the LOD profile for QTL detection regardless whether there is segregation distortion or not. One major QTL was detected in the second chromosome. This chromosome segregates normally without distortion. Figure 3(b) shows the LOD profile for testing segregation distortion, regardless whether the QTL is present or absent. One major SDL was found on chromosome five. Figure 3(c) is the joint LOD score for both QTL and SDL. We can see that both the major QTL and the SDL were detected. These two major loci have very high LOD scores. We used the quick method of Piepho [15] to calculate the genome wide critical value of the LOD for significance test. We found that the genome wide critical value was slightly less than LOD = 3 criterion (data not shown). Therefore, we set LOD = 3 as

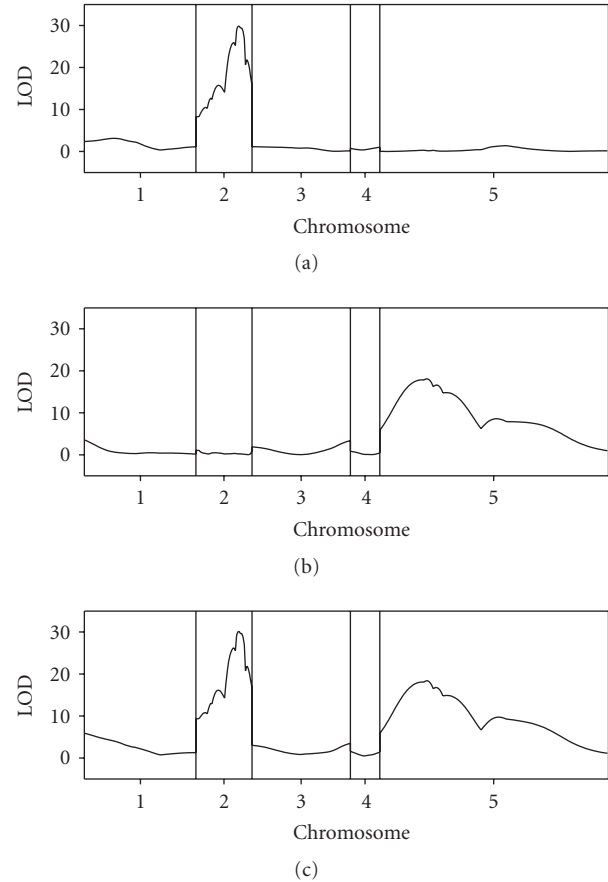


FIGURE 1: LOD score profiles for the wheat genome. The 5 chromosomes of the genome are separated by the gray reference lines. (a) The top panel represents the LOD profile for testing significance of QTL for the female sterility of wheat (regardless whether segregation is distorted or not). (b) The panel in the middle represents the LOD profile for testing significance of SDL (regardless whether a QTL is present or not). (c) The panel at the bottom represents the LOD profile for testing both QTL and SDL (joint test and the null model being no QTL and no SDL).

the criterion. In addition to the two major loci, several other regions of the genome also showed significant peaks of the LOD score profile. The estimated parameters are listed in Table 1. Overall we detected eight loci, four QTLs and four SDLs. Among the detected QTL, each explains from 10% to 47% of the phenotypic variance (listed as heritability denoted by h^2 in Table 1). Among the four SDL detected, all showed bias in favor of the XND126 parent, that is, the homozygote of XND126 allele was over represented at the cost of low representation of the other parent. The largest SDL locus is located in chromosome five at position 32.11 cM. The frequency of the heterozygote was close to the Mendelian frequency of 0.5, but the homozygote of Gaocheng 8901 allele was almost wiped out. The estimated genotypic frequencies are plotted against the genome location as shown in Figure 2. The deviation from Mendelian segregation is quite obvious for chromosome five.

One of the major theoretical contributions of this study is the development of the variance-covariance matrix of the

TABLE 1: Estimated parameters for eight loci (QTL/SDL) of the wheat QTL and SDL analysis using an F₂ family derived from two inbred lines of the wheat.

Locus	1	2	3	4	5	6	7	8
Type	SDL	QTL	QTL	QTL	QTL	SDL	SDL	SDL
Chromosome	1	1	2	2	2	3	5	5
Position (cM)	0.00	19.8	15.78	29.57	34.79	67.32	32.11	79.79
Interval ^a	0.00–5.45	0.00–35.15	12.40–18.18	27.41–31.74	34.35–36.10	59.21–67.32	22.00–35.25	72.56–102.84
LOD score	3.5	3.13	15.77	29.87	21.86	3.35	18.12	8.6
ω_1	0.2099	0.1986	0.2334	0.2684	0.2637	0.1852	0.0763	0.1071
ω_2	0.4239	0.5083	0.5456	0.5054	0.494	0.4568	0.4291	0.4801
ω_3	0.3663	0.2931	0.221	0.2262	0.2423	0.358	0.4946	0.4128
a	0.1435	0.2076	0.3856	0.4438	0.3963	0.0307	−0.0539	0.0882
d	0.0494	0.1594	0.3071	0.4159	0.3542	0.0269	−0.0036	−0.2086
σ^2	0.2796	0.2622	0.2032	0.16	0.1847	0.2917	0.2913	0.2837
h^2	0.0375	0.0962	0.3252	0.4697	0.373	0.0022	0.005	0.0495
β	1.1247	1.0619	0.9509	0.8943	0.9398	1.1163	1.1023	1.2503

^aInterval means one LOD drop supporting interval.

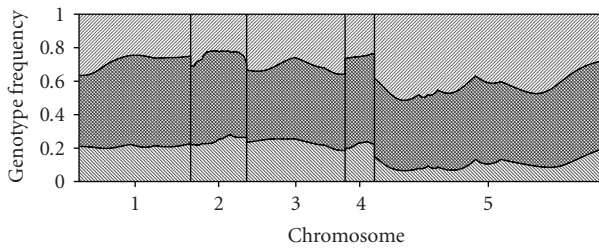


FIGURE 2: Estimated genotypic frequencies for the wheat genome. Frequencies of the three genotypes are represented by areas with different patterns. Chromosomes are separated by the gray reference lines.

estimated QTL-SDL parameters. The covariances between pairs of estimated parameters are not of interest, but the variances of the estimated parameters are important. We reported the standard errors for two selected loci, locus 4 (QTL) and locus 7 (SDL). These standard errors are listed in Table 2. The standard errors are the square roots of the variances obtained from the EM algorithm. The variance-covariance matrix of the estimated parameters takes the inverse of the information matrix. As a result, they are approximate and biased downwards (Louis 1982). The approximation is close to the true variance only in large samples. We also performed a bootstrap analysis (1000 bootstrap samples) to provide more accurate estimation of the variance. The results of the bootstrap estimates of the standard errors for the two loci are also listed in Table 2 for comparison. The approximate standard errors from the EM algorithm are indeed biased downward, especially for locus 4 (QTL). The approximation is much better for locus 7 (SDL). In practice, the bootstraps method is recommended for obtaining more accurate estimates of the standard errors if the sample size is small.

4. Discussion

Statistical methods for mapping quantitative trait loci are well developed for Mendelian populations. Methods also are available for mapping viability loci or segregation distortion loci when markers do not segregate in a typical Mendelian ratio [4–6, 9–11]. However, QTL mapping and SDL mapping have never been combined in a single analysis. This study is the first attempt to combine the two seemingly different analyses into a joint one. When QTL and SDL are loosely linked or not linked, the joint analysis does not offer much advantage over the separate analyses. When they do overlap, a phenomenon called pleiotropy, joint mapping does offer some advantage. Unfortunately, the wheat experiment introduced here is not a good example to demonstrate the advantage of joint analysis because the QTL and SDL detected do not overlap.

An obvious situation where the joint analysis can be more powerful is QTL mapping with selective genotyping. In most designed selective genotyping experiments, two groups of extreme phenotypes are selected for genotyping. The power increase under selective genotyping has been demonstrated [16]. Directional (one-tailed) selection is rarely used for selective genotyping because it artificially reduces the variation of the trait and thus reduces the statistical power of QTL detection. However, with the joint analysis, the power can increase under directional selection. Such a directional selection is common in breeding populations. We now use a simulated example to demonstrate this power increase. We simulated 500 F₂ individuals for a single chromosome of 300 cM long. We placed 16 markers evenly over the chromosome with 20 cM per marker interval. A QTL was placed at position 150 cM of the chromosome. The QTL explains 5% of the phenotypic variance with additive effect only. This QTL is considered small or modest. We selected 300 smallest individuals out of the 500 (one-tailed selection)

TABLE 2: Standard errors of the estimated parameters for loci 4 (QTL) and 7 (SDL) of the wheat F_2 mapping population (see Table 1 for detailed information about loci 4 and 7). The StdErr (EM) and StdErr (Boots) represent the standard errors obtained from the EM algorithm and the bootstrap method, respectively.

Parameter	Locus 4(QTL)			Locus 7(SDL)		
	Estimate	StdErr (EM)	StdErr (Boots)	Estimate	StdErr (EM)	StdErr (Boots)
β	0.8943	0.03705	0.06233	1.1023	0.06934	0.07434
a	0.4438	0.03721	0.07086	-0.0539	0.06934	0.07653
d	0.4159	0.05258	0.08066	-0.0036	0.09138	0.09763
σ^2	0.1600	0.01515	0.04082	0.2913	0.02644	0.02771
ω_1	0.2684	0.02905	0.02857	0.0763	0.01767	0.01749
ω_2	0.5054	0.03274	0.03437	0.4291	0.03336	0.03511

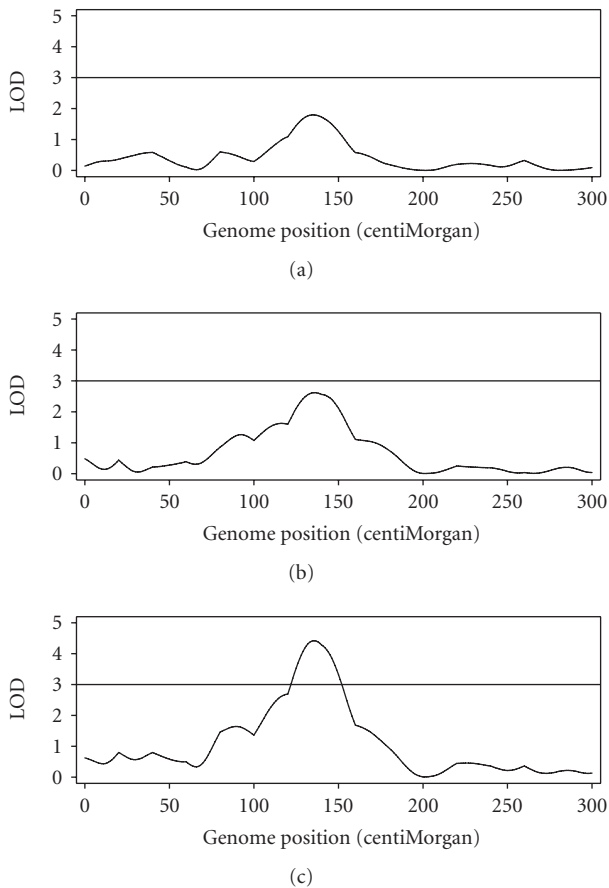


FIGURE 3: LOD score profiles for the simulated genome (single chromosome). The horizontal line at $LOD = 3$ represents the threshold. (a) The top panel represents the LOD profile for testing significance of QTL for the simulated trait (regardless whether segregation is distorted or not). (b) The panel in the middle represents the LOD profile for testing significance of SDL (regardless whether a QTL is present or not). (c) The panel at the bottom represents the LOD profile for testing both QTL and SDL (joint test and the null model being no QTL and no SDL).

for mapping. The LOD test statistic profiles are depicted in Figure 3. Figure 3(a) shows the test statistic for QTL regardless whether segregation is distorted or not. The panel in the middle (b) shows the LOD test statistic for segregation distortion, regardless whether a QTL is present or not. The

panel at the bottom (c) shows the LOD score profile of the joint analysis where the null model is no QTL and no SDL. If $LOD = 3$ is the threshold value, neither of the separate analyses is significant. However, the joint analysis has a LOD score as high as 4.5, indicating a significant QTL in the middle of the chromosome. This example clearly demonstrated the advantage of joint analysis over separate analyses.

Segregation distortion may be caused by gametic selection, zygotic selection, or both. Our model was developed under zygotic selection because we are dealing with the genotypic frequencies. However, if the true cause of segregation distortion is gametic selection, we can still detect segregation distortion as long as the gametic selection leads to the genotypic frequencies deviating from the expected Mendelian ratio. A model particularly handling gametic selection has not been developed yet, but it is not difficult. Similar to the zygotic selection model, gametic selection requires known marker linkage phases. Let us take the F_2 population as an example to show the gametic selection model. Denote the frequencies of G_1 and G_2 alleles from the female parent by ν_1^f and ν_2^f , respectively, for $\nu_1^f + \nu_2^f = 1$, and the corresponding allele frequencies from the male parent by ν_1^m and ν_2^m for $\nu_1^m + \nu_2^m = 1$. Under Mendelian segregation, $\nu_1^f = \nu_2^f = \nu_1^m = \nu_2^m = 1/2$. When gametic selection occurs, we treat ν_1^f and ν_1^m as unknown parameters for estimation. The genotypic frequencies are simply functions of the two unknown parameters, as given by $\omega_1 = \nu_1^f \nu_1^m$, $\omega_2 = \nu_1^f \nu_2^m + \nu_2^f \nu_1^m$ and $\omega_3 = \nu_2^f \nu_2^m$. QTL parameters are estimated using the same algorithm as described in the zygotic selection model because they depend only on the genotypic frequencies. Estimating parameters ν_1^f and ν_1^m requires a modified algorithm. Under the gametic selection model, we can test whether the segregation distortion is caused by the distortion of female gametes, male gametes, or both. This is an interesting topic that deserves further investigation.

The joint analysis developed in this study only applies to line crossing data where the marker linkage phases are known. It cannot be applied to pedigree data analysis. Application of the method to pedigrees warrants further investigation and it is not obvious to us at this moment. However, extension to other line crossing families is possible. We have already extended the method to BC (backcrosses), RIL (recombinant inbred lines), DH (double haploids), and

FW (four way crosses) and incorporated them into our QTL mapping program that is described in the next paragraph. The extension also includes dominance markers and missing marker genotypes.

The proposed joint mapping applies to interval mapping only. Extension to multiple QTL/SDL mapping is difficult. However, interval mapping is still the quickest method of QTL mapping, even though multiple QTL mapping programs are available. Compared with traditional QTL interval mapping, this joint analysis involves one additional step of updating the genotypic frequencies. This additional step presents a complication where the conditional genotypic frequencies given flanking marker genotypes cannot be calculated prior to QTL mapping. They must be calculated with the phenotypic values along with the flanking marker genotypes. This complication makes modification of existing QTL mapping programs difficult. Fortunately, we have incorporated the joint QTL/SDL mapping into our QTL mapping program. This program is a SAS procedure called PROC QTL [17]. The METHOD = "ML" option in the PROC QTL statement must be turned on with an additional suboption /DISTORTION to invoke the joint mapping procedure. Without the /DISTORTION option, the ML analysis simply assumes Mendelian segregation. PROC QTL is available on our website (<http://www.statgen.ucr.edu/software.html>) and users can download the program with no charge. The program is also accompanied with a detailed user manual.

Appendices

A. Standard Errors of the Estimated Parameters

Let us define the individual-wise complete-data log likelihood for plant j as

$$\begin{aligned}
 L_j(\theta, \delta) &= \left[-\frac{1}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{\kappa=1}^3 \delta(G_j, \kappa) (y_j - X_j\beta - H_\kappa\gamma)^2 \right] \\
 &+ \sum_{\kappa=1}^3 \delta(G_j, \kappa) \left[\ln(T_1(\kappa, M_j)) + \ln(T_2(\kappa, N_j)) \right] \\
 &+ \sum_{\kappa=1}^3 \delta(G_j, \kappa) \ln \omega_\kappa,
 \end{aligned} \tag{A.1}$$

where $\omega_3 = 1 - \omega_1 - \omega_2$ so that ω_3 is excluded from the parameter vector. To derive the variance covariance matrix of the estimated parameters, we need to define the score vector $S_j(\theta, \delta)$ and the Hessian matrix $H_j(\theta, \delta)$ for the individual-wise complete-data log likelihood function. The Louis [18] information matrix of the parameters under the EM algorithm is

$$I(\hat{\theta}) = -\sum_{j=1}^n E[H_j(\hat{\theta}, \delta)] - \sum_{j=1}^n \text{var}[S_j(\hat{\theta}, \delta)], \tag{A.2}$$

where the expectation and variance are taken with respect to the missing values δ . Once the information matrix is defined, the variance matrix of the estimated parameters simply takes

$$\text{var}(\hat{\theta}) \approx I^{-1}(\hat{\theta}). \tag{A.3}$$

The standard error of each parameter takes the square root of each diagonal element of the above matrix. The approximation performs better for large sample sizes.

We now present the score vector and the Hessian matrix. The score vector is denoted by $S_j(\theta, \delta) = \partial L_j(\theta, \delta) / \partial \theta$, which consists of five blocks, as follows:

$$\begin{aligned}
 S_j(\beta, \delta) &= \frac{\partial L_j(\theta, \delta)}{\partial \beta} = \frac{1}{\sigma^2} \sum_{\kappa=1}^3 \delta(G_j, \kappa) X_j^T (y_j - X_j\beta - H_\kappa\gamma), \\
 S_j(\gamma, \delta) &= \frac{\partial L_j(\theta, \delta)}{\partial \gamma} = \frac{1}{\sigma^2} \sum_{\kappa=1}^3 \delta(G_j, \kappa) H_\kappa^T (y_j - X_j\beta - H_\kappa\gamma), \\
 S_j(\sigma^2, \delta) &= \frac{\partial L_j(\theta, \delta)}{\partial \sigma^2} = -\frac{1}{2\sigma^2} \\
 &+ \frac{1}{2\sigma^4} \sum_{\kappa=1}^3 \delta(G_j, \kappa) (y_j - X_j\beta - H_\kappa\gamma)^2, \\
 S_j(\omega_1, \delta) &= \frac{\partial L_j(\theta, \delta)}{\partial \omega_1} = \delta(G_j, 1) \frac{1}{\omega_1} - \delta(G_j, 3) \frac{1}{1 - \omega_1 - \omega_2}, \\
 S_j(\omega_2, \delta) &= \frac{\partial L_j(\theta, \delta)}{\partial \omega_2} = \delta(G_j, 2) \frac{1}{\omega_2} - \delta(G_j, 3) \frac{1}{1 - \omega_1 - \omega_2}.
 \end{aligned} \tag{A.4}$$

Concatenating the above five scores vertically, we can get the score vector:

$$S_j(\theta, \delta) = \begin{bmatrix} S_j(\beta, \delta) \\ S_j(\gamma, \delta) \\ S_j(\sigma^2, \delta) \\ S_j(\omega_1, \delta) \\ S_j(\omega_2, \delta) \end{bmatrix}. \tag{A.5}$$

The Hessian matrix is denoted by $H_j(\theta, \delta) = \partial^2 L_j(\theta, \delta) / \partial \theta \partial \theta^T$, which is block diagonal with non-zero blocks given as follows:

$$\begin{aligned}
 H_j(\beta\beta) &= \frac{\partial^2 L_j(\theta, \delta)}{\partial \beta \partial \beta^T} = -\frac{1}{\sigma^2} X_j^T X_j, \\
 H_j(\beta\gamma) &= \frac{\partial^2 L_j(\theta, \delta)}{\partial \beta \partial \gamma^T} = -\frac{1}{\sigma^2} \sum_{\kappa=1}^3 \delta(G_j, \kappa) X_j^T H_\kappa,
 \end{aligned}$$

$$\begin{aligned}
H_j(\beta\sigma^2) &= \frac{\partial^2 L_j(\theta, \delta)}{\partial \beta \partial \sigma^2} \\
&= -\frac{1}{\sigma^4} \sum_{\kappa=1}^3 \delta(G_j, \kappa) X_j^T (y_j - X_j \beta - H_\kappa \gamma), \\
H_j(\gamma\gamma) &= \frac{\partial^2 L_j(\theta, \delta)}{\partial \gamma \partial \gamma^T} = -\frac{1}{\sigma^2} \sum_{\kappa=1}^3 \delta(G_j, \kappa) H_\kappa^T H_\kappa, \\
H_j(\gamma\sigma^2) &= \frac{\partial^2 L_j(\theta, \delta)}{\partial \gamma \partial \sigma^2} \\
&= -\frac{1}{\sigma^4} \sum_{\kappa=1}^3 \delta(G_j, \kappa) H_\kappa^T (y_j - X_j \beta - H_\kappa \gamma), \\
H_j(\sigma^2\sigma^2) &= \frac{\partial^2 L_j(\theta, \delta)}{\partial \sigma^2 \partial \sigma^2} \\
&= \frac{1}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{\kappa=1}^3 \delta(G_j, \kappa) (y_j - X_j \beta - H_\kappa \gamma)^2, \\
H_j(\omega_1\omega_1) &= \frac{\partial^2 L_j(\theta, \delta)}{\partial \omega_1 \partial \omega_1} = -\delta(G_j, 1) \frac{1}{\omega_1^2} - \delta(G_j, 3) \frac{1}{\omega_3^2}, \\
H_j(\omega_1\omega_2) &= \frac{\partial^2 L_j(\theta, \delta)}{\partial \omega_1 \partial \omega_2} = -\delta(G_j, 3) \frac{1}{\omega_3^2}, \\
H_j(\omega_2\omega_2) &= \frac{\partial^2 L_j(\theta, \delta)}{\partial \omega_2 \partial \omega_2} = -\delta(G_j, 2) \frac{1}{\omega_2^2} - \delta(G_j, 3) \frac{1}{\omega_3^2}.
\end{aligned} \tag{A.6}$$

The Hessian matrix is obtained through

$$\begin{aligned}
H_j(\theta, \delta) &= \begin{bmatrix} H_j(\beta\beta) & H_j(\beta\gamma) & H_j(\beta\sigma^2) & 0 & 0 \\ H_j^T(\beta\gamma) & H_j(\gamma\gamma) & H_j(\gamma\sigma^2) & 0 & 0 \\ H_j^T(\beta\sigma^2) & H_j^T(\gamma\sigma^2) & H_j(\sigma^2\sigma^2) & 0 & 0 \\ 0 & 0 & 0 & H_j(\omega_1\omega_1) & H_j(\omega_1\omega_2) \\ 0 & 0 & 0 & H_j^T(\omega_1\omega_2) & H_j(\omega_2\omega_2) \end{bmatrix}.
\end{aligned} \tag{A.7}$$

The expectation of the Hessian matrix $E[H_j(\theta, \delta)]$ and the variance matrix of the score vector $\text{var}[S_j(\theta, \delta)]$ can be expressed explicitly because both the Hessian matrix and the score vector are linear functions of the missing value

$$\delta_j = [\delta(G_j, 1) \quad \delta(G_j, 2) \quad \delta(G_j, 3)]^T. \tag{A.8}$$

Therefore, $E[H_j(\theta, \delta)]$ and $\text{var}[S_j(\theta, \delta)]$ can eventually be expressed as functions of the expectation and variance of δ_j , which have simple expressions because δ_j is a multinomial variable. The Hessian matrix $H_j(\theta, \delta)$ is already expressed in linear function of δ_j and thus the expectation can be obtained straightforwardly by replacing δ_j by $E(\delta_j)$. An explicit linearity for the score function is not obvious. The

following gives the linear relationship using matrix notation. Let us define the following matrices:

$$\begin{aligned}
C_j(\beta) &= \begin{bmatrix} \frac{1}{\sigma^2} X_j^T (y_j - X_j \beta - H_1 \gamma) \\ \frac{1}{\sigma^2} X_j^T (y_j - X_j \beta - H_2 \gamma) \\ \frac{1}{\sigma^2} X_j^T (y_j - X_j \beta - H_3 \gamma) \end{bmatrix}, \\
C_j(\gamma) &= \begin{bmatrix} \frac{1}{\sigma^2} H_1^T (y_j - X_j \beta - H_1 \gamma) \\ \frac{1}{\sigma^2} H_2^T (y_j - X_j \beta - H_2 \gamma) \\ \frac{1}{\sigma^2} H_3^T (y_j - X_j \beta - H_3 \gamma) \end{bmatrix}, \\
C_j(\sigma^2) &= \begin{bmatrix} \frac{1}{2\sigma^4} (y_j - X_j \beta - H_1 \gamma)^2 \\ \frac{1}{2\sigma^4} (y_j - X_j \beta - H_2 \gamma)^2 \\ \frac{1}{2\sigma^4} (y_j - X_j \beta - H_3 \gamma)^2 \end{bmatrix}, \\
C_j(\omega_1) &= \begin{bmatrix} \frac{1}{\omega_1} \\ 0 \\ -\frac{1}{\omega_3} \end{bmatrix}, \quad C_j(\omega_2) = \begin{bmatrix} 0 \\ \frac{1}{\omega_2} \\ -\frac{1}{\omega_3} \end{bmatrix}.
\end{aligned} \tag{A.9}$$

The score vector in matrix notation is

$$S_j(\theta, \delta) = \frac{\partial L_j(\theta, \delta)}{\partial \theta} = \begin{bmatrix} C_j^T(\beta) \\ C_j^T(\gamma) \\ C_j^T(\sigma^2) \\ C_j^T(\omega_1) \\ C_j^T(\omega_2) \end{bmatrix} \delta_j + \begin{bmatrix} 0 \\ 0 \\ -\frac{1}{2\sigma^2} \\ 0 \\ 0 \end{bmatrix} = C_j \delta_j + D. \tag{A.10}$$

As a result,

$$\text{var}[S_j(\theta, \delta)] = C_j \text{var}(\delta_j) C_j^T, \tag{A.11}$$

where

$$\text{var}(\delta_j) = \text{diag}[E(\delta_j)] - E(\delta_j)E(\delta_j^T) \tag{A.12}$$

is the variance-covariance matrix of the multinomial variable δ_j .

B. Estimation of Parameters under the Null Models

This appendix provides methods for estimating parameters under various null models that are required for constructing likelihood ratio test statistics.

B.1. Null Model for Testing QTL. The log likelihood for the null model is

$$L_S(0, \hat{\omega}) = L(\hat{\beta}, \hat{\sigma}^2 | y) + L(\hat{\omega} | m), \quad (\text{B.1})$$

where

$$L(\hat{\beta}, \hat{\sigma}^2 | y) = -\frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} \sum_{j=1}^n (y_j - X_j \hat{\beta})^2, \quad (\text{B.2})$$

$$L(\hat{\omega} | m) = \sum_{j=1}^n \ln \left[\sum_{\kappa=1}^3 \hat{\omega}_\kappa \Gamma_1(\kappa, M_j) \Gamma_2(\kappa, N_j) \right].$$

The MLEs of the population parameters are obtained explicitly without iteration:

$$\hat{\beta} = \left[\sum_{j=1}^n X_j^T X_j \right]^{-1} \left[\sum_{j=1}^n X_j^T y_j \right], \quad (\text{B.3})$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n (y_j - X_j \hat{\beta})^2.$$

The estimated genotypic frequencies, however, require the following iterations:

$$\omega_\kappa^{(t+1)} = \frac{1}{n} \sum_{j=1}^n E[\delta(G_j, \kappa)], \quad \forall \kappa = 1, 2, 3, \quad (\text{B.4})$$

where

$$E[\delta(G_j, \kappa)] = \frac{\omega_\kappa^{(t)} \Gamma_1(\kappa, M_j) \Gamma_2(\kappa, N_j)}{\sum_{\xi=1}^3 \omega_\xi^{(t)} \Gamma_1(\xi, M_j) \Gamma_2(\xi, N_j)}. \quad (\text{B.5})$$

B.2. Null Model for Testing SDL. The log likelihood function for the null model is

$$L_Q(\hat{y}, \phi) = -\frac{n}{2} \ln(\hat{\sigma}^2) + \sum_{j=1}^n \ln \left\{ \sum_{\kappa=1}^3 \phi_\kappa \exp \left[-\frac{1}{2\hat{\sigma}^2} (y_j - \hat{\mu}_\kappa)^2 \right] \Gamma_1(\kappa, M_j) \Gamma_2(\kappa, N_j) \right\}. \quad (\text{B.6})$$

Finding the solution of the parameters requires iterations as given as follows:

$$\beta^{(t+1)} = \left[\sum_{j=1}^n X_j^T X_j \right]^{-1} \times \left[\sum_{j=1}^n \sum_{\kappa=1}^3 E[\delta(G_j, \kappa)] X_j^T (y_j - H_\kappa \gamma^{(t)}) \right],$$

$$\gamma^{(t+1)} = \left[\sum_{j=1}^n \sum_{\kappa=1}^3 E[\delta(G_j, \kappa)] (H_\kappa^T H_\kappa) \right]^{-1} \times \left[\sum_{j=1}^n \sum_{\kappa=1}^3 E[\delta(G_j, \kappa)] H_\kappa^T (y_j - X_j \beta^{(t)}) \right],$$

$$\sigma^{2(t+1)} = \frac{1}{n} \sum_{j=1}^n \sum_{\kappa=1}^3 E[\delta(G_j, \kappa)] (y_j - X_j \beta^{(t)} - H_\kappa \gamma^{(t)})^2, \quad (\text{B.7})$$

where

$$E[\delta(G_j, \kappa)] = \frac{\phi_\kappa \exp \left[-(1/2\sigma^2) (y_j - \mu_\kappa)^2 \right] \Gamma_1(\kappa, M_j) \Gamma_2(\kappa, N_j)}{\sum_{\xi=1}^3 \phi_\xi \exp \left[-(1/2\sigma^2) (y_j - \mu_\xi)^2 \right] \Gamma_1(\xi, M_j) \Gamma_2(\xi, N_j)}. \quad (\text{B.8})$$

The Mendelian frequencies ϕ_κ 's are constants, not parameters, as $\phi_1 = \phi_3 = (1/2)\phi_2 = 1/4$.

B.3. Null Model for Testing Both QTL and SDL. The log likelihood function for the null model is

$$L(0, \phi) = L(\hat{\beta}, \hat{\sigma}^2 | y) + L(\phi | m), \quad (\text{B.9})$$

where

$$L(\hat{\beta}, \hat{\sigma}^2 | y) = -\frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} \sum_{j=1}^n (y_j - X_j \hat{\beta})^2, \quad (\text{B.10})$$

$$L(\phi | m) = \sum_{j=1}^n \ln \left[\sum_{\kappa=1}^3 \phi_\kappa \Gamma_1(\kappa, M_j) \Gamma_2(\kappa, N_j) \right].$$

The MLEs of parameters are

$$\hat{\beta} = \left[\sum_{j=1}^n X_j^T X_j \right]^{-1} \left[\sum_{j=1}^n X_j^T y_j \right], \quad (\text{B.11})$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n (y_j - X_j \hat{\beta})^2.$$

They are the same as those provided in (B.3). The Mendelian frequencies ϕ_κ 's are constants, not parameters.

Acknowledgments

The authors are grateful to Dr. Dou and his research group for making their wheat QTL mapping data available to them. They also appreciate three anonymous reviewers for their comments and suggestions on an early version of the manuscript. This project was supported by the National Research Initiative (NRI) Plant Genome of the USDA Cooperative State Research, Education and Extension Service (CSREES) 2007-02784 to SX.

References

- [1] S. Xu, "Quantitative trait locus mapping can benefit from segregation distortion," *Genetics*, vol. 180, no. 4, pp. 2201–2208, 2008.
- [2] J. D. Faris, B. Laddomada, and B. S. Gill, "Molecular mapping of segregation distortion loci in *Aegilops tauschii*," *Genetics*, vol. 149, no. 1, pp. 319–327, 1998.
- [3] K. Kärkkäinen, V. Koski, and O. Savolainen, "Geographical variation in the inbreeding depression of scots pine," *Evolution*, vol. 50, no. 1, pp. 111–119, 1996.
- [4] L. Luo and S. Xu, "Mapping viability loci using molecular markers," *Heredity*, vol. 90, no. 6, pp. 459–467, 2003.
- [5] C. Vogl and S. Xu, "Multipoint mapping of viability and segregation distorting loci using molecular markers," *Genetics*, vol. 155, no. 3, pp. 1439–1447, 2000.
- [6] Y. B. Fu and K. Ritland, "On estimating the linkage of marker genes to viability genes controlling inbreeding depression," *Theoretical and Applied Genetics*, vol. 88, no. 8, pp. 925–932, 1994.
- [7] M. Lorieux, X. Perrier, B. Goffinet, C. Lanaud, and D. G. de León, "Maximum-likelihood models for mapping genetic markers showing segregation distortion. 2. F₂ populations," *Theoretical and Applied Genetics*, vol. 90, no. 1, pp. 81–89, 1995.
- [8] M. Lorieux, B. Goffinet, X. Perrier, D. G. de León, and C. Lanaud, "Maximum-likelihood models for mapping genetic markers showing segregation distortion. 1. Backcross populations," *Theoretical and Applied Genetics*, vol. 90, no. 1, pp. 73–80, 1995.
- [9] C. Wang, C. Zhu, H. Zhai, and J. Wan, "Mapping segregation distortion loci and quantitative trait loci for spikelet sterility in rice (*Oryza sativa* L.)," *Genetical Research*, vol. 86, no. 2, pp. 97–106, 2005.
- [10] L. Luo, Y.-M. Zhang, and S. Xu, "A quantitative genetics model for viability selection," *Heredity*, vol. 94, no. 3, pp. 347–355, 2005.
- [11] C. Zhu and Y.-M. Zhang, "An EM algorithm for mapping segregation distortion loci," *BMC Genetics*, vol. 8, article 82, 2007.
- [12] R. C. Elston and J. Stewart, "The analysis of quantitative traits for simple genetic models from parental, F₁ and backcross data," *Genetics*, vol. 73, no. 4, pp. 695–711, 1973.
- [13] A. P. Dempster, M. N. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.
- [14] B. Dou, B. Hou, H. Xu, et al., "Efficient mapping of a female sterile gene in wheat (*Triticum aestivum* L.)," *Genetics Research*, vol. 91, no. 5, pp. 337–343, 2009.
- [15] H.-P. Piepho, "A quick method for computing approximate thresholds for quantitative trait loci detection," *Genetics*, vol. 157, no. 1, pp. 425–432, 2001.
- [16] Y. I. Ronin, A. B. Korol, and J. I. Weller, "Selective genotyping to detect quantitative trait loci affecting multiple traits: interval mapping analysis," *Theoretical and Applied Genetics*, vol. 97, no. 7, pp. 1169–1178, 1998.
- [17] Z. Hu and S. Xu, "PROC QTL—a SAS procedure for mapping quantitative trait loci," in *Plant and Animal Genomes XVII Conference*, San Diego, Calif, USA, January 2009.
- [18] T. A. Louis, "Finding the observed information matrix when using the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 44, no. 2, pp. 226–233, 1982.