Methodology Report

EM Algorithm for Mapping Quantitative Trait Loci in Multivalent Tetraploids

Jiahua Li,¹ Kiranmoy Das,¹ Guifang Fu,¹ Chunfa Tong,¹ Yao Li,²
Christian Tobias,³ and Rongling Wu¹

¹ Center for Statistical Genetics, Pennsylvania State University, Hershey, PA 17033, USA

² Department of Statistics, West Virginia University, Morgantown, WV 26506, USA

³ Genomics and Gene Discovery Research Unit, USDA-ARS Western Regional Research Center, Albany, CA 94710, USA

Correspondence should be addressed to Rongling Wu, rwu@hes.hmc.psu.edu

Received 2 September 2010; Accepted 16 November 2010

Academic Editor: Nengjun Yi

Copyright © 2010 Jiahua Li et al. 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.

Multivalent tetraploids that include many plant species, such as potato, sugarcane and rose, are of paramount importance to agricultural production and biological research. Quantitative trait locus (QTL) mapping in multivalent tetraploids is challenged by their unique cytogenetic properties, such as double reduction. We develop a statistical method for mapping multivalent tetraploid QTLs by considering these cytogenetic properties. This method is built in the mixture model-based framework and implemented with the EM algorithm. The method allows the simultaneous estimation of QTL positions, QTL effects, the chromosomal pairing factor and the degree of double reduction as well as the assessment of the estimation precision of these parameters. We used simulated data to examine the statistical properties of the method and validate its utilization. The new method and its software will provide a useful tool for QTL mapping in multivalent tetraploids that undergo double reduction.

1. Introduction

Genetic analysis in polyploids has received considerable interest in recent years because of the biological and economic importance [1–3]. Genetic linkage maps constructed from molecular markers have been published for several major polyploids [4–10]. Statistical models for linkage analysis and map construction that consider unique biological properties of polyploids have been developed [11–14]. For bivalent polyploids, Wu et al. [15, 16] incorporated the so-called chromosomal pairing preference [17] into the linkage analysis framework, to increase the biological relevance of linkage mapping models. There have been several statistical models developed to map quantitative trait loci (QTLs) in bivalent polyploids [18, 19].

There is also a group of polyploids, called multivalent polyploids, in which chromosomes pair among more than two homologous copies at meiosis, rather than only two copies as like in bivalent polyploids. The origin of multivalent polyploids is mostly from the duplication of similar

genomes and, for this reason, they are called autopolyploids [20, 21]. The consequence of multivalent pairing in autopolyploids is the occurrence of double reduction, that is, two sister chromatids of a chromosome sort into the same gamete [22]. Fisher [23] proposed a conceptual model for characterizing the individual probabilities of 11 different modes of gamete formation for a quadrivalent polyploid in terms of the recombination fraction between two different loci and their double reductions. Wu et al. [24] used Fisher's model to derive the EM algorithm for the estimation of the linkage between fully informative markers. Wu and Ma [25] extended this model to analyze any type of markers, regardless of their informativeness and dominant or codominant nature. The significant advantage of the models by Wu and colleagues directly lies in their generality, flexibility, and robustness.

In this paper, we develop a statistical method for QTL mapping in multivalent tetraploids by considering Fisher's [23] 11 classifications of gamete formation. The method allows the estimation and test of not only the QTL-marker

linkage, but also the extent of double reduction of the QTL. Because of the inherent complexity of classification analyses of gamete formation, we will focus on the modeling and analysis of one-marker/one-QTL associations. A two-stage hierarchical model is derived to estimate the probabilities of gamete formation modes and therefore double reduction in the upper hierarchy and estimate the marker-QTL recombination fraction in the lower hierarchy within the maximum likelihood context implemented with the EM algorithm. The method is used to analyze a simulated data set, with results demonstrating statistical properties of the method and its analytical and biological merits.

2. Method

2.1. Genetic Design. Consider a heterozygous multivalent tetraploid line crossed with a homozygous line to generate a so-called pseudotest backcross population. For such a population, the genotypes of progeny are consistent with the genotypes of gametes produced by the heterozygous parent and, therefore, the derivation of mapping models can be based on the segregation of gametes. Suppose there are n individuals in the pseudotest backcross population. A panel of codominant markers is typed for each individual, with which a linkage map is constructed. All the pseudotest backcross individuals are phenotyped for a quantitative trait that is assumed to be controlled by QTLs.

2.2. Tetrasomic Co-Inheritance. To simplify our analysis, we assume that the QTLs underlying the trait are mapped with single markers. Let $M_1, \dots, M_4$ be the four alleles at a marker $\mathbf{M}$, and let $Q_1, \dots, Q_4$ be the four alleles at a QTL linked with the marker. The marker and QTL are linked with a recombination fraction of r . Because of the double reduction, the multivalent tetraploid generates 10 diploid gametes for locus, which are arrayed as $(M_1M_1, M_2M_2, M_3M_3, M_4M_4, M_1M_2, M_1M_3, M_1M_4, M_2M_3, M_2M_4, M_3M_4)$ for the markers, and $(Q_1Q_1, Q_2Q_2, Q_3Q_3, Q_4Q_4, Q_1Q_2, Q_1Q_3, Q_1Q_4, Q_2Q_3, Q_2Q_4, Q_3Q_4)$ for the QTL. In each case, the first four gametes are derived from the double reduction, whereas the second six gametes are derived from the chromosome pairing. Let α and β be the frequencies of double reduction at the marker and QTL, respectively. The frequency of double reduction is a constant for any given locus, with the value depending on its distance from the centromere.

When the marker and QTL are co-segregating in a multivalent tetraploid, a total of 136 diploid gamete formation mechanisms are generated although there are only 100 gamete genotypes that are observable. Based on the presence/absence of double reduction and the number of recombinant events, Fisher [23] classified these 136 formation mechanisms into 11 gamete modes. Of these 11 gamete modes, however, only nine can be observed each with a frequency denoted by g_h ($h = 1, \dots, 9$). These 9 observable gamete modes were rearranged by Wu et al. [24] in matrix form expressed as

$$\mathbf{g} = \begin{matrix} & Q_1Q_1 & Q_2Q_2 & Q_3Q_3 & Q_4Q_4 & Q_1Q_2 & Q_1Q_3 & Q_1Q_4 & Q_2Q_3 & Q_2Q_4 & Q_3Q_4 \\ \begin{matrix} M_1M_1 \\ M_2M_2 \\ M_3M_3 \\ M_4M_4 \\ \frac{M_1M_2}{M_2M_1} \\ \frac{M_1M_3}{M_3M_1} \\ \frac{M_1M_4}{M_4M_1} \\ \frac{M_2M_3}{M_3M_2} \\ \frac{M_2M_4}{M_4M_2} \\ \frac{M_3M_4}{M_4M_3} \end{matrix} & \begin{bmatrix} \frac{1}{4}g^1 & \frac{1}{12}g^2 & \frac{1}{12}g^2 & \frac{1}{12}g^2 & \frac{1}{12}g^5 & \frac{1}{12}g^5 & \frac{1}{12}g^5 & \frac{1}{12}g^6 & \frac{1}{12}g^6 & \frac{1}{12}g^6 \\ \frac{1}{12}g^2 & \frac{1}{4}g^1 & \frac{1}{12}g^2 & \frac{1}{12}g^2 & \frac{1}{12}g^5 & \frac{1}{12}g^6 & \frac{1}{12}g^6 & \frac{1}{12}g^5 & \frac{1}{12}g^5 & \frac{1}{12}g^6 \\ \frac{1}{12}g^2 & \frac{1}{12}g^2 & \frac{1}{4}g^1 & \frac{1}{12}g^2 & \frac{1}{12}g^6 & \frac{1}{12}g^5 & \frac{1}{12}g^6 & \frac{1}{12}g^5 & \frac{1}{12}g^6 & \frac{1}{12}g^5 \\ \frac{1}{12}g^2 & \frac{1}{12}g^2 & \frac{1}{12}g^2 & \frac{1}{4}g^1 & \frac{1}{12}g^6 & \frac{1}{12}g^6 & \frac{1}{12}g^5 & \frac{1}{12}g^6 & \frac{1}{12}g^5 & \frac{1}{12}g^5 \\ \frac{1}{12}g^3 & \frac{1}{12}g^3 & \frac{1}{12}g^4 & \frac{1}{12}g^4 & \frac{1}{6}g^7 & \frac{1}{24}g^8 & \frac{1}{24}g^8 & \frac{1}{24}g^8 & \frac{1}{24}g^8 & \frac{1}{6}g^9 \\ \frac{1}{12}g^3 & \frac{1}{12}g^4 & \frac{1}{12}g^3 & \frac{1}{12}g^4 & \frac{1}{24}g^8 & \frac{1}{6}g^7 & \frac{1}{24}g^8 & \frac{1}{24}g^8 & \frac{1}{6}g^9 & \frac{1}{24}g^8 \\ \frac{1}{12}g^3 & \frac{1}{12}g^4 & \frac{1}{12}g^4 & \frac{1}{12}g^3 & \frac{1}{24}g^8 & \frac{1}{24}g^8 & \frac{1}{6}g^7 & \frac{1}{6}g^9 & \frac{1}{24}g^8 & \frac{1}{24}g^8 \\ \frac{1}{12}g^4 & \frac{1}{12}g^3 & \frac{1}{12}g^3 & \frac{1}{12}g^4 & \frac{1}{24}g^8 & \frac{1}{24}g^8 & \frac{1}{6}g^9 & \frac{1}{6}g^7 & \frac{1}{24}g^8 & \frac{1}{24}g^8 \\ \frac{1}{12}g^4 & \frac{1}{12}g^3 & \frac{1}{12}g^4 & \frac{1}{12}g^3 & \frac{1}{24}g^8 & \frac{1}{6}g^9 & \frac{1}{24}g^8 & \frac{1}{24}g^8 & \frac{1}{6}g^7 & \frac{1}{24}g^8 \\ \frac{1}{12}g^4 & \frac{1}{12}g^4 & \frac{1}{12}g^3 & \frac{1}{12}g^3 & \frac{1}{6}g^9 & \frac{1}{24}g^8 & \frac{1}{24}g^8 & \frac{1}{24}g^8 & \frac{1}{24}g^8 & \frac{1}{6}g^7 \end{bmatrix} \end{matrix}, \quad (1)$$

where g_1 and g_2 are associated with double reductions at both the marker and QTL, g_3 and g_4 with double reductions only at QTL Q , g_5 and g_6 with double reductions only at marker M , and $g_7 - g_9$ with nondouble reductions. From matrix (1), we see that there are no, one and two recombinant events in the cells (g_1), (g_3, g_5), and (g_2, g_4, g_6, g_9), respectively. The cells (g_7) and (g_8) are each a mixture of two different gamete formation mechanisms or configurations (A and B), that is, $g_7 = g_{7A} + g_{7B}$ and $g_8 = g_{8A} + g_{8B}$, with

relative proportions determined by r . Because different configurations contain different numbers of recombination events, the expected number of recombination events in each cell, that is, an observable gamete genotype, should be the weighted average of the number of recombination events for each configuration. Wu et al. [24] used a matrix form (e) to count the expected number of recombination events for each observable gamete genotype expressed as

$$\mathbf{c} = \begin{matrix} & Q_1Q_1 & Q_2Q_2 & Q_3Q_3 & Q_4Q_4 & Q_1Q_2 & Q_1Q_3 & Q_1Q_4 & Q_2Q_3 & Q_2Q_4 & Q_3Q_4 \\ \begin{matrix} M_1M_1 \\ M_2M_2 \\ M_3M_3 \\ M_4M_4 \\ \hline \frac{M_1M_2}{M_2M_1} \\ \hline \frac{M_1M_3}{M_3M_1} \\ \hline \frac{M_1M_4}{M_4M_1} \\ \hline \frac{M_2M_3}{M_3M_2} \\ \hline \frac{M_2M_4}{M_4M_2} \\ \hline \frac{M_3M_4}{M_4M_3} \end{matrix} & \begin{bmatrix} 0 & 2 & 2 & 2 & 1 & 1 & 1 & 2 & 2 & 2 \\ 2 & 0 & 2 & 2 & 1 & 2 & 2 & 1 & 1 & 2 \\ 2 & 2 & 0 & 2 & 2 & 1 & 2 & 1 & 2 & 1 \\ 2 & 2 & 2 & 0 & 2 & 2 & 1 & 2 & 1 & 1 \\ \hline 1 & 1 & 2 & 2 & 2\phi & 1+\psi & 1+\psi & 1+\psi & 1+\psi & 2 \\ \hline 1 & 2 & 1 & 2 & 1+\psi & 2\phi & 1+\psi & 1+\psi & 2 & 1+\psi \\ \hline 1 & 2 & 2 & 1 & 1+\psi & 1+\psi & 2\phi & 2 & 1+\psi & 1+\psi \\ \hline 2 & 1 & 1 & 2 & 1+\psi & 1+\psi & 2 & 2\phi & 1+\psi & 1+\psi \\ \hline 2 & 1 & 2 & 1 & 1+\psi & 2 & 1+\psi & 1+\psi & 2\phi & 1+\psi \\ \hline 2 & 2 & 1 & 1 & 2 & 1+\psi & 1+\psi & 1+\psi & 1+\psi & 2\phi \end{bmatrix} \end{matrix}, \quad (2)$$

where

$$\phi = \frac{r^2}{10r^2 - 18r + 9}, \quad (3)$$

$$\psi = \frac{r}{3 - 2r}.$$

Based on matrices (1) and (2), the expressions for the frequencies of double reduction (α and β) and the recombination fraction r can be expressed in terms of g_i as

$$\alpha = g_1 + g_2 + g_3 + g_4,$$

$$\beta = g_1 + g_2 + g_5 + g_6,$$

$$r = \frac{1}{2} [g_3 + g_5 + 2(g_2 + g_4 + g_6 + g_9) + 2\phi g_7 + (1 + \psi)g_8]. \quad (4)$$

2.3. Quantitative Genetic Model. For a given QTL, there are 10 different QTL gamete genotypes in the multivalent

tetraploid, whose values can be partitioned into additive and dominance genetic effects of different types, expressed as

$$\begin{aligned} \mu_{11} &= \mu + a_1, & \text{for } Q_1Q_1, \\ \mu_{22} &= \mu + a_2, & \text{for } Q_2Q_2, \\ \mu_{33} &= \mu + a_3, & \text{for } Q_3Q_3, \\ \mu_{44} &= \mu - a_1 - a_2 - a_3, & \text{for } Q_4Q_4, \\ \mu_{12} &= \mu + a_1 + a_2 + d_{12}, & \text{for } Q_1Q_2, \\ \mu_{13} &= \mu + a_1 + a_3 + d_{13}, & \text{for } Q_1Q_3, \\ \mu_{14} &= \mu - a_2 - a_3 + d_{14}, & \text{for } Q_1Q_4, \\ \mu_{23} &= \mu + a_2 + a_3 + d_{23}, & \text{for } Q_2Q_3, \\ \mu_{24} &= \mu - a_1 - a_3 + d_{24}, & \text{for } Q_2Q_4, \\ \mu_{34} &= \mu - a_1 - a_2 + d_{34}, & \text{for } Q_3Q_4, \end{aligned} \quad (5)$$

where μ is the overall mean, a_1 , a_2 , and a_3 are the additive genetic effects of alleles Q_1 , Q_2 , and Q_3 relative to allele Q_4 , and d_{12} , d_{13} , d_{14} , d_{23} , d_{24} , and d_{34} are the dominant genetic effects due to interactions between different alleles Q_1 and Q_2 , Q_1 and Q_3 , Q_1 and Q_4 , Q_2 and Q_3 , Q_2 and Q_4 , and Q_3 and Q_4 , respectively.

From expression (5), we can solve the overall mean and additive and dominant effects as

$$\mu = \frac{1}{4}(\mu_{11} + \mu_{22} + \mu_{33} + \mu_{44}), \quad (6)$$

$$a_1 = \frac{1}{4}(3\mu_{11} - \mu_{22} - \mu_{33} - \mu_{44}), \quad (7)$$

$$a_2 = \frac{1}{4}(3\mu_{22} - \mu_{11} - \mu_{33} - \mu_{44}), \quad (8)$$

$$a_3 = \frac{1}{4}(3\mu_{33} - \mu_{11} - \mu_{22} - \mu_{44}), \quad (9)$$

$$d_{12} = \mu_{12} - \frac{1}{4}(3\mu_{33} + 3\mu_{44} - \mu_{11} - \mu_{22}), \quad (10)$$

$$d_{13} = \mu_{13} - \frac{1}{4}(3\mu_{22} + 3\mu_{44} - \mu_{11} - \mu_{33}), \quad (11)$$

$$d_{14} = \mu_{14} - \frac{1}{4}(3\mu_{22} + 3\mu_{33} - \mu_{11} - \mu_{44}), \quad (12)$$

$$d_{23} = \mu_{23} - \frac{1}{4}(3\mu_{11} + 3\mu_{44} - \mu_{22} - \mu_{33}), \quad (13)$$

$$d_{24} = \mu_{24} - \frac{1}{4}(3\mu_{11} + 3\mu_{33} - \mu_{22} - \mu_{44}), \quad (14)$$

$$d_{34} = \mu_{34} - \frac{1}{4}(3\mu_{11} + 3\mu_{22} - \mu_{33} - \mu_{44}). \quad (15)$$

2.4. EM Algorithm. Ignoring the effects of other covariates, the phenotypic value, y_i , for individual i in the pseudotest backcross can be expressed in terms of the QTL effect and residual error as

$$y_i = \sum_{j_1 \leq j_2=1}^4 \xi_{j_1 j_2 | i} \mu_{j_1 j_2} + e_i, \quad (16)$$

where $\xi_{j_1 j_2 | i}$ is the indicator variable that is defined as 1 if individual i has a QTL genotype $j_1 j_2$ ($j_1 \leq j_2 = Q_1, Q_2, Q_3, Q_4$), and 0 otherwise, $\mu_{j_1 j_2}$ is the genotypic value of QTL genotype $j_1 j_2$ as defined in (5), and e_i is the residual error assumed to be normally distributed with mean zero and variance σ^2 . We use Ω to denote the unknown vector $(\mu_{11}, \mu_{22}, \mu_{33}, \mu_{44}, \mu_{12}, \mu_{13}, \mu_{14}, \mu_{23}, \mu_{24}, \mu_{34}, \sigma^2)$.

For a QTL mapping experiment, marker genotypes are observable. Let $n_{l_1 l_2}$ be the observation of marker genotype $l_1 l_2$ ($l_1 \leq l_2 = M_1, M_2, M_3, M_4$). The likelihood of the phenotypic ($\mathbf{y}$) and marker data ($\mathbf{M}$) is constructed, within

the mixture model framework, as

$$L(\Omega | \mathbf{y}, \mathbf{M}) = \prod_{i=1}^n \prod_{l_1=M_1}^{M_4} \prod_{l_2=M_1}^{M_4} \sum_{j_1=Q_1}^{Q_4} \sum_{j_2=Q_1}^{Q_4} \pi_{j_1 j_2 | l_1 l_2} f_{j_1 j_2}(y_i), \quad (17)$$

$l_1 \leq l_2, j_1 \leq j_2,$

where $\pi_{j_1 j_2 | l_1 l_2}$ is the conditional probability of QTL genotype $j_1 j_2$ given marker genotype $l_1 l_2$, and $f_{j_1 j_2}(y_i)$ is assumed to follow a normal distribution with mean $\mu_{j_1 j_2}$ and variance σ^2 . Prior conditional probability $\pi_{j_1 j_2 | l_1 l_2}$ is calculated as the frequency of joint marker-QTL genotype $l_1 l_2 j_1 j_2$, expressed in terms of nine g probabilities in matrix (1), divided by the frequency of marker genotype $l_1 l_2$. Marker genotype frequencies are $\alpha/4$ for each of double reduction gametes $M_1 M_1$, $M_2 M_2$, $M_3 M_3$, and $M_4 M_4$, and $(1 - \alpha)/6$ for each of nondouble reduction gametes $M_1 M_2$, $M_1 M_3$, $M_1 M_4$, $M_2 M_3$, $M_2 M_4$, and $M_3 M_4$.

The estimates of unknown parameters that maximize the likelihood (17) can be obtained by implementing the EM algorithm. In step E, we calculate the posterior probability of a QTL genotype given a specific marker genotype of individual i by

$$\Theta_{j_1 j_2 | l_1 l_2 i} = \frac{\pi_{j_1 j_2 | l_1 l_2} f_{j_1 j_2}(y_i)}{\sum_{j'_1=Q_1}^{Q_4} \sum_{j'_2=Q_1}^{Q_4} \pi_{j'_1 j'_2 | l_1 l_2} f_{j'_1 j'_2}(y_i)}. \quad (18)$$

In step M, we calculate the frequencies of nine observable gamete modes based on the calculated posterior probabilities using the following:

$$\begin{aligned} \hat{g}_1 &= \frac{1}{2(n_{11} + n_{22} + n_{33} + n_{44})} \\ &\times \left(\sum_{i=1}^{n_{11}} \Theta_{11|11i} + \sum_{i=1}^{n_{22}} \Theta_{22|22i} + \sum_{i=1}^{n_{33}} \Theta_{33|33i} + \sum_{i=1}^{n_{44}} \Theta_{44|44i} \right), \\ \hat{g}_2 &= \frac{1}{2(n_{11} + n_{22} + n_{33} + n_{44})} \\ &\times \left[\sum_{i=1}^{n_{11}} (\Theta_{22|11i} + \Theta_{33|11i} + \Theta_{44|11i}) \right. \\ &\quad + \sum_{i=1}^{n_{22}} (\Theta_{11|22i} + \Theta_{33|22i} + \Theta_{44|22i}) \\ &\quad + \sum_{i=1}^{n_{33}} (\Theta_{11|33i} + \Theta_{22|33i} + \Theta_{44|33i}) \\ &\quad \left. + \sum_{i=1}^{n_{44}} (\Theta_{11|44i} + \Theta_{22|44i} + \Theta_{33|44i}) \right], \end{aligned}$$

$$\hat{g}_3 = \frac{1}{2(n_{12} + n_{13} + n_{14} + n_{23} + n_{24} + n_{34})} \times \left[\sum_{i=1}^{n_{12}} (\Theta_{11|12i} + \Theta_{22|12i}) + \sum_{i=1}^{n_{13}} (\Theta_{11|13i} + \Theta_{33|13i}) + \sum_{i=1}^{n_{14}} (\Theta_{11|14i} + \Theta_{44|14i}) + \sum_{i=1}^{n_{23}} (\Theta_{22|23i} + \Theta_{33|23i}) + \sum_{i=1}^{n_{24}} (\Theta_{22|24i} + \Theta_{44|24i}) + \sum_{i=1}^{n_{34}} (\Theta_{33|34i} + \Theta_{44|34i}) \right],$$

$$\hat{g}_4 = \frac{1}{2(n_{12} + n_{13} + n_{14} + n_{23} + n_{24} + n_{34})} \times \left[\sum_{i=1}^{n_{12}} (\Theta_{33|12i} + \Theta_{44|12i}) + \sum_{i=1}^{n_{13}} (\Theta_{22|13i} + \Theta_{44|13i}) + \sum_{i=1}^{n_{14}} (\Theta_{22|14i} + \Theta_{33|14i}) + \sum_{i=1}^{n_{23}} (\Theta_{11|23i} + \Theta_{44|23i}) + \sum_{i=1}^{n_{24}} (\Theta_{11|24i} + \Theta_{33|24i}) + \sum_{i=1}^{n_{34}} (\Theta_{11|34i} + \Theta_{22|34i}) \right],$$

$$\hat{g}_5 = \frac{1}{2(n_{11} + n_{22} + n_{33} + n_{44})} \times \left[\sum_{i=1}^{n_{11}} (\Theta_{12|11i} + \Theta_{13|11i} + \Theta_{14|11i}) + \sum_{i=1}^{n_{22}} (\Theta_{12|22i} + \Theta_{23|22i} + \Theta_{24|22i}) + \sum_{i=1}^{n_{33}} (\Theta_{13|33i} + \Theta_{23|33i} + \Theta_{34|33i}) + \sum_{i=1}^{n_{44}} (\Theta_{14|44i} + \Theta_{24|44i} + \Theta_{34|44i}) \right],$$

$$\hat{g}_6 = \frac{1}{2(n_{11} + n_{22} + n_{33} + n_{44})} \times \left[\sum_{i=1}^{n_{11}} (\Theta_{23|11i} + \Theta_{24|11i} + \Theta_{34|11i}) + \sum_{i=1}^{n_{22}} (\Theta_{13|22i} + \Theta_{14|22i} + \Theta_{34|22i}) + \sum_{i=1}^{n_{33}} (\Theta_{12|33i} + \Theta_{14|33i} + \Theta_{24|33i}) + \sum_{i=1}^{n_{44}} (\Theta_{12|44i} + \Theta_{13|44i} + \Theta_{23|44i}) \right],$$

$$\hat{g}_7 = \frac{1}{2(n_{12} + n_{13} + n_{14} + n_{23} + n_{24} + n_{34})} \times \left(\sum_{i=1}^{n_{12}} \Theta_{12|12i} + \sum_{i=1}^{n_{13}} \Theta_{13|13i} + \sum_{i=1}^{n_{14}} \Theta_{14|14i} + \sum_{i=1}^{n_{23}} \Theta_{23|23i} + \sum_{i=1}^{n_{24}} \Theta_{24|24i} + \sum_{i=1}^{n_{34}} \Theta_{34|34i} \right),$$

$$\hat{g}_8 = \frac{1}{2(n_{12} + n_{13} + n_{14} + n_{23} + n_{24} + n_{34})} \times \left[\sum_{i=1}^{n_{12}} (\Theta_{13|12i} + \Theta_{14|12i} + \Theta_{23|12i} + \Theta_{24|12i}) + \sum_{i=1}^{n_{13}} (\Theta_{12|13i} + \Theta_{14|13i} + \Theta_{23|13i} + \Theta_{34|13i}) + \sum_{i=1}^{n_{14}} (\Theta_{12|14i} + \Theta_{13|14i} + \Theta_{24|14i} + \Theta_{34|14i}) + \sum_{i=1}^{n_{23}} (\Theta_{12|23i} + \Theta_{13|23i} + \Theta_{24|23i} + \Theta_{34|23i}) + \sum_{i=1}^{n_{24}} (\Theta_{12|24i} + \Theta_{14|24i} + \Theta_{23|24i} + \Theta_{34|24i}) + \sum_{i=1}^{n_{34}} (\Theta_{13|34i} + \Theta_{14|34i} + \Theta_{23|34i} + \Theta_{24|34i}) \right],$$

$$\hat{g}_9 = \frac{1}{2(n_{12} + n_{13} + n_{14} + n_{23} + n_{24} + n_{34})} \times \left(\sum_{i=1}^{n_{12}} \Theta_{34|12i} + \sum_{i=1}^{n_{13}} \Theta_{24|13i} + \sum_{i=1}^{n_{14}} \Theta_{23|14i} + \sum_{i=1}^{n_{23}} \Theta_{14|23i} + \sum_{i=1}^{n_{24}} \Theta_{13|24i} + \sum_{i=1}^{n_{34}} \Theta_{12|34i} \right) \quad (19)$$

which lead to the estimates of the frequencies of double reduction as

$$\begin{aligned} \hat{\alpha} &= \hat{g}_1 + \hat{g}_2 + \hat{g}_3 + \hat{g}_4 \\ &= \frac{1}{n} (n_{11} + n_{22} + n_{33} + n_{44}), \\ \hat{\beta} &= \hat{g}_1 + \hat{g}_2 + \hat{g}_5 + \hat{g}_6, \\ \hat{r} &= \frac{1}{2} [\hat{g}_3 + \hat{g}_5 + 2(\hat{g}_2 + \hat{g}_4 + \hat{g}_6 + \hat{g}_9) \\ &\quad + 2\phi\hat{g}_7 + (1 + \psi)\hat{g}_8]. \end{aligned} \quad (20)$$

TABLE 1: Maximum likelihood estimates of the double reduction, recombination fraction, overall mean, additive effects, and dominant effects under different simulation scenarios in a mapping population of size 100. Numbers in parentheses are the standard errors of the estimates.

True value			α	r	H^2	$\hat{\alpha}$	$\hat{r}$	μ	a_1	a_2	a_3	d_{12}	d_{13}	d_{14}	d_{23}	d_{24}	d_{34}
								1	0.6	0.6	0.6	0.5	0.5	0.5	0.5	0.5	0.5
0.3	0.25	0.1	0.295 (0.047)	0.314 (0.120)	1.153 (1.139)	0.047 (2.339)	0.116 (2.651)	0.037 (2.496)	2.228 (4.013)	3.007 (4.237)	-1.953 (4.499)	2.375 (4.349)	-2.661 (4.453)	-2.107 (4.407)			
0.3	0.25	0.4	0.301 (0.048)	0.251 (0.098)	1.095 (0.540)	0.556 (0.986)	0.496 (1.003)	0.546 (1.133)	0.677 (1.744)	0.625 (1.861)	0.313 (2.143)	0.482 (1.947)	0.181 (2.103)	0.138 (1.931)			
0.3	0.05	0.1	0.297 (0.041)	0.142 (0.114)	0.961 (0.767)	0.659 (1.330)	0.580 (1.293)	0.512 (1.303)	1.216 (1.967)	1.130 (2.208)	-0.036 (2.301)	1.286 (2.129)	0.119 (2.193)	-0.084 (2.102)			
0.3	0.05	0.4	0.296 (0.045)	0.057 (0.051)	0.998 (0.296)	0.642 (0.461)	0.643 (0.581)	0.533 (0.497)	0.445 (0.868)	0.562 (0.787)	0.428 (0.851)	0.601 (0.973)	0.488 (0.762)	0.594 (0.687)			
0.15	0.25	0.1	0.150 (0.034)	0.305 (0.155)	1.058 (1.067)	0.244 (1.800)	0.821 (1.794)	0.945 (1.965)	1.232 (2.707)	1.334 (2.645)	0.167 (2.819)	0.660 (2.820)	-0.150 (2.658)	-0.373 (2.614)			
0.15	0.25	0.4	0.152 (0.031)	0.239 (0.097)	1.061 (0.369)	0.522 (0.777)	0.548 (0.812)	0.598 (0.794)	0.645 (1.082)	0.484 (1.166)	0.449 (1.100)	0.492 (1.203)	0.382 (1.227)	0.230 (1.238)			
0.15	0.05	0.1	0.144 (0.035)	0.168 (0.127)	0.837 (1.108)	0.411 (2.103)	0.540 (1.962)	0.930 (1.635)	1.616 (2.854)	1.242 (2.698)	0.271 (2.488)	1.033 (2.452)	-0.037 (3.147)	-0.313 (2.700)			
0.15	0.05	0.4	0.143 (0.035)	0.057 (0.058)	0.921 (0.488)	0.372 (0.877)	0.736 (0.759)	0.545 (0.775)	0.685 (1.192)	0.965 (1.148)	0.621 (1.085)	0.677 (1.085)	0.282 (1.167)	0.453 (1.147)			
0.05	0.25	0.1	0.052 (0.022)	0.371 (0.180)	0.769 (1.293)	0.356 (2.122)	0.300 (2.576)	0.455 (2.392)	2.239 (3.425)	2.178 (3.478)	-0.691 (3.104)	2.031 (2.787)	-0.426 (2.998)	-0.588 (3.563)			
0.05	0.25	0.4	0.051 (0.022)	0.253 (0.111)	0.643 (0.699)	0.440 (1.067)	0.406 (1.003)	0.458 (1.105)	1.243 (1.522)	1.202 (1.605)	0.500 (1.341)	1.224 (1.541)	0.553 (1.483)	0.362 (1.482)			
0.05	0.05	0.1	0.052 (0.025)	0.193 (0.136)	1.008 (1.346)	0.056 (2.119)	1.012 (2.537)	0.362 (2.121)	1.552 (3.110)	1.977 (2.718)	-0.065 (3.191)	1.068 (3.427)	-1.077 (3.354)	-0.438 (3.391)			
0.05	0.05	0.4	0.052 (0.024)	0.064 (0.059)	0.762 (0.736)	0.525 (1.005)	0.361 (0.959)	0.337 (1.076)	1.119 (1.480)	1.175 (1.482)	0.193 (1.542)	1.255 (1.636)	0.303 (1.473)	0.351 (1.285)			

The genotypic value of QTL genotype $j_1 j_2$ and residual variance are estimated by

$$\hat{\mu}_{j_1 j_2} = \frac{\sum_{i=1}^n \sum_{l_1=M_1}^{M_4} \sum_{l_2=M_1}^{M_4} \Theta_{j_1 j_1 | l_1 l_2 i} y_i}{\sum_{l_1=M_1}^{M_4} \sum_{l_2=M_1}^{M_4} \Theta_{j_1 j_1 | l_1 l_2 i}}, \quad (21)$$

$$l_1 \leq l_2; j_1 \leq j_2,$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \sum_{l_1=M_1}^{M_4} \sum_{l_2=M_1}^{M_4} \sum_{j_1=Q_1}^{Q_4} \sum_{j_2=Q_1}^{Q_4} \Theta_{j_1 j_1 | l_1 l_2 i} (y_i - \hat{\mu}_{j_1 j_2})^2, \quad (22)$$

$$l_1 \leq l_2; j_1 \leq j_2.$$

The iteration is repeated between the E step, (3) and (18), and M step, (19)–(22), until stable estimates are obtained. The stable estimates are the maximum likelihood estimates (MLEs) of parameters.

2.5. Hypothesis Testing. Following parameter estimation, several hypotheses should be tested. The hypothesis about

the presence of a QTL segregating in the pseudotest backcrosses is formulated as

$$H_0: a_1 = a_2 = a_3 = d_{12} = d_{13} = d_{14} = d_{23} = d_{24} = d_{34} = 0,$$

$$H_1: \text{at least one of them is not equal to zero.} \quad (23)$$

The difference between the log-likelihood functions under the null and alternative hypotheses are calculated. But the distribution of this log-likelihood ratio (LR) is not known because of the violation of regularity conditions for the mixture model (1). For this reason, a commonly used empirical approach based on permutation tests by reshuffling the relationships between the marker genotypes and phenotypes [26] is used to determine the critical threshold, in order to judge whether there is a QTL for the trait.

After a significant QTL is detected, the next hypothesis is about the additive genetic effect of the QTL. This can be tested by formulating the null hypothesis,

$$H_0: a_1 = a_2 = a_3 = 0, \quad (24)$$

under which the estimates of genotypic values of QTL genotypes can be obtained with the EM algorithm as described above, but posing three constraints derived from

TABLE 2: Maximum likelihood estimates of the double reduction, recombination fraction, overall mean, additive effects, and dominant effects under different simulation scenarios in a mapping population of size 200. Numbers in parentheses are the standard errors of the estimates.

True value														
α	r	H^2	$\hat{\alpha}$	$\hat{r}$	μ	a_1	a_2	a_3	d_{12}	d_{13}	d_{14}	d_{23}	d_{24}	d_{34}
					1	0.6	0.6	0.6	0.5	0.5	0.5	0.5	0.5	0.5
0.3	0.25	0.1	0.302 (0.031)	0.313 (0.093)	1.181 (1.207)	0.397 (1.708)	0.218 (1.697)	0.126 (1.443)	2.303 (2.844)	2.553 (3.144)	-1.649 (3.576)	2.272 (3.048)	-1.471 (3.454)	-1.268 (3.526)
0.3	0.25	0.4	0.301 (0.034)	0.234 (0.085)	1.091 (0.508)	0.582 (0.781)	0.553 (0.730)	0.677 (0.727)	0.493 (1.432)	0.317 (1.402)	0.340 (1.438)	0.646 (1.416)	0.459 (1.435)	0.262 (1.362)
0.3	0.05	0.1	0.302 (0.032)	0.093 (0.093)	1.013 (0.483)	0.545 (0.789)	0.551 (0.829)	0.660 (0.753)	0.739 (1.432)	0.703 (1.561)	0.206 (1.309)	0.924 (1.355)	0.114 (1.453)	0.089 (1.486)
0.3	0.05	0.4	0.297 (0.032)	0.047 (0.037)	0.999 (0.190)	0.596 (0.362)	0.585 (0.346)	0.574 (0.342)	0.522 (0.570)	0.531 (0.645)	0.464 (0.666)	0.509 (0.611)	0.420 (0.577)	0.530 (0.569)
0.15	0.25	0.1	0.152 (0.028)	0.277 (0.117)	1.025 (0.718)	0.692 (1.168)	0.563 (1.225)	0.587 (1.215)	0.953 (1.866)	0.968 (1.930)	-0.232 (1.862)	1.224 (1.900)	-0.034 (1.780)	-0.033 (1.692)
0.15	0.25	0.4	0.147 (0.024)	0.234 (0.078)	0.982 (0.287)	0.581 (0.520)	0.675 (0.471)	0.580 (0.530)	0.420 (0.838)	0.551 (0.724)	0.561 (0.775)	0.386 (0.822)	0.484 (0.745)	0.617 (0.736)
0.15	0.05	0.1	0.149 (0.024)	0.140 (0.106)	0.971 (0.633)	0.475 (1.251)	0.548 (1.257)	0.723 (1.479)	1.232 (1.814)	0.985 (1.719)	0.246 (1.990)	0.863 (1.968)	0.048 (1.868)	-0.325 (1.824)
0.15	0.05	0.4	0.149 (0.026)	0.055 (0.047)	1.026 (0.299)	0.691 (0.558)	0.573 (0.511)	0.558 (0.585)	0.394 (0.684)	0.380 (0.799)	0.478 (0.800)	0.536 (0.862)	0.485 (0.754)	0.523 (0.745)
0.05	0.25	0.1	0.049 (0.015)	0.326 (0.143)	1.074 (1.209)	0.750 (2.486)	0.325 (2.232)	0.557 (1.868)	1.235 (2.857)	1.190 (2.619)	-0.420 (3.332)	1.392 (3.113)	-0.127 (3.023)	-0.331 (2.775)
0.05	0.25	0.4	0.048 (0.013)	0.241 (0.075)	0.932 (0.499)	0.619 (1.038)	0.345 (1.037)	0.528 (0.932)	0.885 (1.365)	0.658 (1.305)	0.234 (1.282)	0.879 (1.317)	0.498 (1.190)	0.242 (1.276)
0.05	0.05	0.1	0.050 (0.015)	0.196 (0.127)	0.914 (1.209)	0.490 (1.941)	0.294 (2.134)	0.612 (2.157)	1.755 (2.902)	1.435 (2.792)	-0.487 (2.747)	1.713 (2.872)	-0.146 (2.758)	-0.379 (2.855)
0.05	0.05	0.4	0.049 (0.017)	0.059 (0.049)	0.818 (0.541)	0.609 (0.978)	0.351 (0.988)	0.697 (0.967)	0.952 (1.130)	0.646 (1.527)	0.548 (1.419)	0.908 (1.285)	0.737 (1.279)	0.412 (1.001)

(7), (8), and (9). Similarly, the dominant genetic effects can be tested with the null hypothesis,

$$H_0: d_{12} = d_{13} = d_{14} = d_{23} = d_{24} = d_{34} = 0, \quad (25)$$

with estimates of genotypic values under the constraints derived from (10)–(15). All these genetic effects can be tested individually.

2.6. Application to Simulated Data. A pseudotest backcross for a multivalent tetraploid was hypothesized, in which a marker is assumed to be linked with a QTL that affects a quantitative trait. Marker and QTL genotypes were simulated for the pseudotest backcross of different sample sizes ($n = 100, 200, 400$) based on a range of double reduction (0.05, 0.15, 0.30) and recombination fraction (0.05, 0.25). We assume the same frequency of double reduction between the marker and QTL. The phenotypic value of an individual is expressed as the summation of genotypic values of a QTL genotype carried by this individual and a normally distributed error. The genotypic values of a QTL genotype are calculated by (5), where the overall mean μ is assigned as 1, and the additive and dominant effects assigned as $a_1 = a_2 = a_3 = 0.6$ and $d_{12} = d_{13} = d_{14} = d_{23} = d_{24} = d_{34} = 0.5$.

The error variance is determined according to the heritability of $H^2 = 0.1$ and 0.4, respectively.

In this simulation study, fully informative markers and QTL are assumed and, thus, the double reduction at the marker can be estimated analytically. The estimates of the parameters converge to stable values at a rapid rate given that there are closed forms for parameter estimators in the EM framework. We evaluate the estimation of the other parameters related to QTL segregation, effects, and position. The means of the MLEs of the QTL-related parameters and their standard errors based on 1000 simulation replicates are illustrated in Tables 1, 2, and 3. With a small sample size (100), the double reduction of the QTL was accurately estimated, with the precision of estimation relatively independent of the magnitude of heritability and the degree of QTL-marker linkage (Table 1). The most significant factor that affected the estimate of QTL position (in terms of its recombination with the marker) was the heritability, followed by sample size and the degree of QTL-marker linkage. In general, at least a sample size of 200 was required to reliably estimate the QTL position for a major gene that explains about 20%–30% of the phenotypic variance.

The estimation precision of the QTL effects depended on the heritability, sample size, and degree of QTL-marker

TABLE 3: Maximum likelihood estimates of the double reduction, recombination fraction, overall mean, additive effects, and dominant effects under different simulation scenarios in a mapping population of size 400. Numbers in parentheses are the standard errors of the estimates.

True value														
α	r	H^2	$\hat{\alpha}$	$\hat{r}$	μ 1	a_1 0.6	a_2 0.6	a_3 0.6	d_{12} 0.5	d_{13} 0.5	d_{14} 0.5	d_{23} 0.5	d_{24} 0.5	d_{34} 0.5
0.3	0.25	0.1	0.300 (0.023)	0.300 (0.082)	1.208 (0.933)	0.182 (1.405)	0.435 (1.400)	0.208 (1.074)	1.800 (2.357)	2.002 (2.429)	-1.236 (2.829)	1.546 (2.130)	-1.574 (2.744)	-1.318 (2.906)
0.3	0.25	0.4	0.301 (0.023)	0.231 (0.078)	1.066 (0.358)	0.585 (0.552)	0.557 (0.497)	0.603 (0.532)	0.508 (1.050)	0.451 (0.984)	0.361 (0.977)	0.494 (1.115)	0.463 (1.060)	0.309 (0.964)
0.3	0.05	0.1	0.298 (0.023)	0.097 (0.077)	1.026 (0.389)	0.577 (0.603)	0.646 (0.637)	0.629 (0.580)	0.704 (1.067)	0.818 (1.138)	0.245 (1.206)	0.689 (1.028)	0.202 (1.120)	0.208 (1.071)
0.3	0.05	0.4	0.299 (0.025)	0.050 (0.027)	1.014 (0.127)	0.626 (0.245)	0.553 (0.253)	0.639 (0.250)	0.525 (0.388)	0.442 (0.435)	0.458 (0.405)	0.507 (0.403)	0.575 (0.388)	0.414 (0.350)
0.15	0.25	0.1	0.150 (0.016)	0.261 (0.135)	0.928 (0.510)	0.641 (0.684)	0.463 (0.904)	0.611 (0.737)	0.975 (1.372)	0.860 (1.476)	0.192 (1.423)	0.995 (1.286)	0.265 (1.458)	0.202 (1.353)
0.15	0.25	0.4	0.150 (0.017)	0.250 (0.060)	1.012 (0.180)	0.584 (0.329)	0.624 (0.358)	0.565 (0.326)	0.495 (0.511)	0.552 (0.526)	0.479 (0.510)	0.516 (0.558)	0.452 (0.478)	0.424 (0.419)
0.15	0.05	0.1	0.150 (0.017)	0.113 (0.095)	1.075 (0.485)	0.583 (0.866)	0.583 (0.846)	0.706 (0.838)	0.740 (1.341)	0.705 (1.254)	0.218 (1.322)	0.667 (1.527)	0.170 (1.272)	0.058 (1.218)
0.15	0.05	0.4	0.150 (0.017)	0.049 (0.034)	1.014 (0.197)	0.526 (0.307)	0.654 (0.382)	0.607 (0.366)	0.531 (0.557)	0.572 (0.511)	0.547 (0.551)	0.406 (0.516)	0.425 (0.506)	0.482 (0.483)
0.05	0.25	0.1	0.050 (0.010)	0.322 (0.113)	0.949 (0.858)	0.516 (1.819)	0.726 (1.626)	0.629 (1.773)	1.256 (2.170)	1.327 (2.180)	0.187 (2.155)	1.026 (2.213)	-0.191 (2.437)	-0.037 (2.459)
0.05	0.25	0.4	0.050 (0.012)	0.250 (0.057)	1.007 (0.423)	0.610 (0.664)	0.652 (0.680)	0.551 (0.624)	0.404 (0.873)	0.562 (0.789)	0.481 (0.867)	0.480 (0.916)	0.457 (0.909)	0.545 (0.991)
0.05	0.05	0.1	0.050 (0.011)	0.167 (0.109)	1.002 (0.830)	0.656 (1.548)	0.343 (1.553)	0.723 (1.581)	1.156 (2.032)	0.796 (2.084)	-0.164 (2.157)	1.147 (2.213)	0.111 (1.997)	-0.211 (2.140)
0.05	0.05	0.4	0.051 (0.011)	0.054 (0.037)	1.025 (0.347)	0.591 (0.637)	0.608 (0.624)	0.635 (0.698)	0.512 (0.771)	0.459 (0.895)	0.518 (0.894)	0.469 (0.931)	0.477 (0.800)	0.450 (0.770)

linkage. As heritability, sample size, and linkage degree increased, the estimates of various QTL effects were more precise. As compared with the dominant genetic effects, the estimates of the additive genetic effects required a larger sample size, more precise phenotypic measurements (leading to a higher heritability), and a denser linkage map (with a stronger degree of QTL-marker linkage). We found that the estimates of QTL effects were influenced by the frequency of QTL double reduction. At low frequencies of double reduction, the effects of QTL were more accurately estimated than at higher frequencies. For a QTL undergoing a strong double reduction (say $\alpha = 0.3$), a sample size of at least 400 is required even if the QTL explains a large proportion of the phenotypic variance (0.4). For a modest-sized QTL, a much larger sample size was required.

We performed a simulation study to test how the misspecification of double reduction affects the estimate of QTL-related parameters. This was done by using traditional mapping models (without considering double reduction) to analyze the simulated data of QTL genotypes with different degrees of double reduction. When a QTL undergoes double reduction, traditional models that do not consider double reduction provided misleading results about the estimates

of QTL effects and position (data not shown). Furthermore, increasing heritabilities and sample sizes did not improve the estimates. In this case, the power of QTL detection was reduced.

3. Discussion

A statistical method for genetic mapping of quantitative trait loci (QTLs) in a multivalent tetraploid undergoing a double reduction process is described. As an important cytological characteristic of polyploids, double reduction may play a significant role in plant evolution and maintenance of genetic polymorphism in natural populations. Also, because double reduction affects the result of linkage analysis through the crossing-over events between different chromosomes [24, 25], it is important to incorporate double reduction into a QTL mapping framework. This method provides a powerful tool for QTL mapping and understanding the genetic control of a quantitative trait in a multivalent tetraploid.

The method capitalizes on 11 different classifications of two-locus gamete formations, derived by Fisher [23], during multivalent tetraploid meiosis and has proven to be powerful for simultaneous estimation of the frequencies of

double reduction and the recombination fraction between different loci. Although a couple of statistical approaches have been proposed to map multivalent tetraploid QTLs [26, 27], this method has for the first time incorporated Fisher's tetrasomic inheritance into the mapping framework, thus enhancing the cytological relevance of QTL detection. Results from simulation studies showed that the method can be used to map QTLs in a controlled cross of multivalent tetraploids when the mapping population is adequately large (say 400). When a QTL undergoes double reduction, traditional mapping approaches will incorrectly estimate the position and effects of the QTL, proportional to the degree of double reduction. The new method can estimate the double reduction of a QTL, an important parameter related to the genetic diversity and evolution of polyploids [28, 29].

Because of the high complexity of the mixture model implemented with tetrasomic inheritance, we only considered a one-marker model for QTL mapping. Interval mapping, which localizes a QTL with two flanking markers, has proven to be more advantageous in parameter estimation over the one-marker model [30]. It will be worthwhile to integrate components of our model into the interval mapping framework to fully explore the statistical merits of interval mapping for QTL mapping in multivalent tetraploids. Furthermore, the model proposed in this article assumes the segregation of fully informative codominant loci, each with 10 distinct genotypes, in a controlled cross of multivalent tetraploids. For partially informative codominant markers, a two-stage hierarchical mixture model will be needed to model the different allelic configurations for a phenotypically identical genotype. Although molecular marker technologies have improved in recent years, dominant markers may still be used in genetic mapping projects of some underrepresentative species including polyploids. Thus, it is also important to extend our model to map QTLs with dominant markers. For partially informative loci, the number of QTL genotypes may be unknown and, thus, a model selection procedure should be incorporated to determine the optimal number of genotypes at a QTL.

The genetic mapping of polyploids is complex because of their complex inheritance modes. Sophisticated statistical models are required to tackle genetic problems hidden in the polysomic inheritance of polyploids. Currently, there are some debates on the optimal modeling of tetrasomic inheritance in linkage analysis [13, 25] and QTL mapping [18, 31] partly because of our limited knowledge about these fascinating species. Before a detailed understanding of the cytological mechanisms for meiosis in multivalent polyploids is obtained, this type of debate will continue. In any case, the development of powerful statistical models for polyploid mapping continues to be a pressing need. The application of these models to real-world data will not only test their usefulness, but also provide an unprecedented opportunity to understand the genetic differentiation among polyploid genomes and characterize the genetic architecture of quantitatively inherited traits for this unique group of species. Software for the method described is available at <http://statgen.psu.edu/>.

Acknowledgments

This work is supported by Joint grant DMS/NIGMS-0540745. Additional support is through the Office of Sciences (BER), U.S. Department of Energy, Interagency Agreement no. DE-A102-07ER64453.

References

- [1] P. S. Soltis and D. E. Soltis, "The role of genetic and genomic attributes in the success of polyploids," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 13, pp. 7051–7057, 2000.
- [2] J. G. Robins, D. Luth, T. A. Campbell et al., "Genetic mapping of biomass production in tetraploid alfalfa," *Crop Science*, vol. 47, no. 1, pp. 1–10, 2007.
- [3] M. Stift, C. Berenos, P. Kuperus, and P. H. Van Tienderen, "Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to rorippa (yellow cress) microsatellite data," *Genetics*, vol. 179, no. 4, pp. 2113–2123, 2008.
- [4] J. A. G. da Silva, M. E. Sorrells, W. L. Burnquist, and S. D. Tanksley, "RFLP linkage map and genome analysis of *Saccharum spontaneum*," *Genome*, vol. 36, no. 4, pp. 782–791, 1993.
- [5] R. Ming, S. C. Liu, Y. R. Lin et al., "Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes," *Genetics*, vol. 150, no. 4, pp. 1663–1682, 1998.
- [6] R. Ming, S. C. Liu, J. E. Irvine et al., "Comparative QTL analysis in a complex autopolyploid: candidate genes for determinants of sugar content in sugarcane," *Genome Research*, vol. 11, pp. 2075–2084, 2001.
- [7] R. C. Meyer, D. Milbourne, C. A. Hackett, J. E. Bradshaw, J. W. McNichol, and R. Waugh, "Linkage analysis in tetraploid potato and association of markers with quantitative resistance to late blight (*Phytophthora infestans*)," *Molecular and General Genetics*, vol. 259, no. 2, pp. 150–160, 1998.
- [8] D. J. Brouwer and T. C. Osborn, "A molecular marker linkage map of tetraploid alfalfa (*Medicago sativa* L.)," *Theoretical and Applied Genetics*, vol. 99, no. 7–8, pp. 1194–1200, 1999.
- [9] Z. W. Luo, C. A. Hackett, J. E. Bradshaw, J. W. McNicol, and D. Milbourne, "Construction of a genetic linkage map in tetraploid species using molecular markers," *Genetics*, vol. 157, no. 3, pp. 1369–1385, 2001.
- [10] B. Julier, S. Flajoulot, P. Barre et al., "Construction of two genetic linkage maps in cultivated tetraploid alfalfa (*Medicago sativa*) using microsatellite and AFLP markers," *BMC Plant Biology*, vol. 3, article 9, 2003.
- [11] C. A. Hackett, J. E. Bradshaw, R. C. Meyer, J. W. McNicol, D. Milbourne, and R. Waugh, "Linkage analysis in tetraploid species: a simulation study," *Genetical Research*, vol. 71, no. 2, pp. 143–154, 1998.
- [12] M. I. Ripol, G. A. Churchill, J. A. G. da Silva, and M. Sorrells, "Statistical aspects of genetic mapping in autopolyploids," *Gene*, vol. 235, no. 1–2, pp. 31–41, 1999.
- [13] Z. W. Luo, R. M. Zhang, and M. J. Kearsey, "Theoretical basis for genetic linkage analysis in autotetraploid species," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 18, pp. 7040–7045, 2004.
- [14] Z. W. Luo, Z. E. Zhang, L. Leach, R. M. Zhang, J. E. Bradshaw, and M. J. Kearsey, "Constructing genetic linkage maps under

- a tetrasomic model,” *Genetics*, vol. 172, no. 4, pp. 2635–2645, 2006.
- [15] R. Wu, C. X. Ma, and G. Casella, “A bivalent polyploid model for linkage analysis in outcrossing tetraploids,” *Theoretical Population Biology*, vol. 62, no. 2, pp. 129–151, 2002.
 - [16] R. Wu, C. X. Ma, and G. Casella, “A mixed polyploid model for linkage analysis in outcrossing tetraploids using a pseudo-test backcross design,” *Journal of Computational Biology*, vol. 11, no. 4, pp. 562–580, 2004.
 - [17] J. Sybenga, “Preferential pairing estimates from multivalent frequencies in tetraploids,” *Genome*, vol. 37, no. 6, pp. 1045–1055, 1994.
 - [18] C. X. Ma, G. Casella, Z. J. Shen, T. C. Osborn, and R. Wu, “A unified framework for mapping quantitative trait loci in bivalent tetraploids using single-dose restriction fragments: a case study from alfalfa,” *Genome Research*, vol. 12, no. 12, pp. 1974–1981, 2002.
 - [19] R. Wu, C. X. Ma, and G. Casella, “A bivalent polyploid model for mapping quantitative trait loci in outcrossing tetraploids,” *Genetics*, vol. 166, no. 1, pp. 581–595, 2004.
 - [20] J. D. Bever and F. Felber, “The theoretical population genetics of autopolyploidy,” *Oxford Surveys in Evolutionary Biology*, vol. 8, pp. 185–217, 1992.
 - [21] D. V. Butruille and L. S. Boiteux, “Selection-mutation balance in polysomic tetraploids: impact of double reduction and gametophytic selection on the frequency and subchromosomal localization of deleterious mutations,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 12, pp. 6608–6613, 2000.
 - [22] C. D. Darlington, “Chromosome behaviour and structural hybridity in the *Tradescantiae*,” *Journal of Genetics*, vol. 21, no. 2, pp. 207–286, 1929.
 - [23] R. A. Fisher, “The theory of linkage in polysomic inheritance,” *Philosophical Transactions of the Royal Society B*, vol. 233, pp. 55–87, 1947.
 - [24] S. S. Wu, R. Wu, C. X. Ma, Z. B. Zeng, M. C. K. Yang, and G. Casella, “A multivalent pairing model of linkage analysis in autotetraploids,” *Genetics*, vol. 159, no. 3, pp. 1339–1350, 2001.
 - [25] R. Wu and C. X. Ma, “A general framework for statistical linkage analysis in multivalent tetraploids,” *Genetics*, vol. 170, no. 2, pp. 899–907, 2005.
 - [26] C. A. Hackett, J. E. Bradshaw, and J. W. McNicol, “Interval mapping of quantitative trait loci in autotetraploid species,” *Genetics*, vol. 159, no. 4, pp. 1819–1832, 2001.
 - [27] D. Cao, B. A. Craig, and R. W. Doerge, “A model selection-based interval-mapping method for autopolyploids,” *Genetics*, vol. 169, no. 4, pp. 2371–2382, 2005.
 - [28] K. G. Haynes and D. S. Douches, “Estimation of the coefficient of double reduction in the cultivated tetraploid potato,” *Theoretical and Applied Genetics*, vol. 85, no. 6-7, pp. 857–862, 1993.
 - [29] R. Wu, M. Gallo-Meagher, R. C. Littell, and Z. B. Zeng, “A general polyploid model for analyzing gene segregation in outcrossing tetraploid species,” *Genetics*, vol. 159, no. 2, pp. 869–882, 2001.
 - [30] E. S. Lander and S. Botstein, “Mapping mendelian factors underlying quantitative traits using RFLP linkage maps,” *Genetics*, vol. 121, no. 1, pp. 185–199, 1989.
 - [31] D. Cao, T. C. Osborn, and R. W. Doerge, “Correct estimation of preferential chromosome pairing in autotetraploids,” *Genome Research*, vol. 14, no. 3, pp. 459–462, 2004.