A multivalent three-point linkage analysis model of autotetraploids

Yafei Lu*, Xiaoxia Yang*, Chunfa Tong, Xin Li, Sisi Feng, Zhong Wang, Xiaoming Pang, Yaqun Wang, Ningtao Wang, Christian M. Tobias and Rongling Wu

Submitted: 29th May 2012; Received (in revised form): 13th July 2012

Abstract
Because of its widespread occurrence and role in shaping evolutionary processes in the biological kingdom, especially in plants, polyploidy has been increasingly studied from cytological to molecular levels. By inferring gene order, gene distances and gene homology, linkage mapping with molecular markers has proven powerful for investigating genome structure and organization. Here we review and assess a general statistical model for three-point linkage analysis in autotetraploids by integrating double reduction, a phenomenon that commonly occurs in autopolyploids whose chromosomes are derived from a single ancestral species. This model does not require any assumption on the distribution of the occurrence of double reduction and can handle the complexity of multilocus linkage in terms of crossover interference. Implemented with the expectation-maximization (EM) algorithms, the model can estimate and test the recombination fractions between less informative dominant markers, thus facilitating its practical implications for any autopolyploids in most of which inexpensive dominant markers are still used for their genetic and evolutionary studies. The model was applied to reanalyze a published data in tetraploid switchgrass, validating its practical usefulness and utilization.

Keywords: autopolyploids; double reduction; statistical model; three-point analysis

Corresponding author. Rongling Wu, Center for Statistical Genetics, The Pennsylvania State University, Hershey, PA 17033, USA. Tel: +1 717 531 2037; Fax: +1 717 531 0480. Email: rwu@phs.psu.edu

*These authors contributed equally to this work.

Yafei Lu is a Masters student in Forest Genetics and Tree Breeding in the Center for Computational Biology at Beijing Forestry University. His research focuses on the statistical genetics of quantitative traits in forest trees.
Xiaoxia Yang is a PhD candidate in Quantitative Genetics and Bioinformatics in the Center for Computational Biology and College of Science at Beijing Forestry University. She develops statistical and computational models and algorithms for complex trait genetics.
Chunfa Tong was a visiting scholar in the Center for Statistical Genetics, The Pennsylvania State University from 2009 to 2011. He is interested in the genetic mapping for complex traits.
Xin Li is a Masters student in Quantitative Genetics and Bioinformatics in the Center for Computational Biology at Beijing Forestry University. Her thesis focuses on the development of statistical and computational algorithms for bioinformatics.
Sisi Feng is a Masters student in Forest Genetics and Tree Breeding in the Center for Computational Biology at Beijing Forestry University. Her thesis focuses on the use of biotechnologies to study the population genetics of forest trees.
Zhong Wang works as a computer scientist working at The Pennsylvania State University and Beijing Forestry University. He develops computer algorithms for statistical genetic models.
Xiaoming Pang is an Associate Professor of Tree Breeding at Beijing Forestry University. His research interest centers on the utilization of molecular genetics and biotechnologies to study population genetic diversity and map quantitative trait loci in fruit trees.
Yaqun Wang is a PhD student in statistics at The Pennsylvania State University. The focus of his thesis is the dynamic modeling of Bayesian network.
Ningtao Wang is a PhD student in statistics at The Pennsylvania State University. The focus of his thesis is the modeling and dissection of complex phenotypes using differential geometry.
Christian M. Tobias is a Research Molecular Biologist in the Genomics and Gene Discovery Research Unit, Western Regional Research Center, USDA. Part of Dr Tobias’s ongoing research is to develop a comprehensive molecular marker system for switchgrass, a tetraploid forage grass, and develop a saturated linkage map of this species.
Rongling Wu is a Professor of Biostatistics and Bioinformatics and the Director of the Center for Statistical Genetics at The Pennsylvania State University. He found the Center for Computational Biology at Beijing Forestry University. He is interested in quantitative and statistical genetics.
INTRODUCTION

Polyploidy, or whole-genome duplication, has long been recognized as an important driving force for evolution, speciation and adaptation to environmental perturbations in animals, fungi and especially plants [1-10]. One of the important approaches for genetic and evolutionary studies is to use linkage mapping based on molecular markers genotyped throughout the genome. By relocating these markers to their original positions in order and studying their cosegregation and recombination, linkage mapping can provide unique information about genome structure, organization and function [11]. However, linkage analysis, when used for a segregating population of polyploids, has encountered a major challenge, since the pattern of gamete segregation during meiosis is affected by unique cytological properties of these species. Several approaches that integrate polyploid meiotic properties have been developed in a hope to better infer the linkage in the literature [12-17]. For example, for allopolyploids that display bivalent associations at different markers based on Fisher’s model, Wu and colleagues [12,14] do not rely on any assumption about the occurrence of double reduction, showing greater robustness and flexibility to analyze any autotetraploid data. In this article, we review and assess a general model for handling multiple markers simultaneously by estimating the coefficients of double reduction at individual markers and their pairwise recombination fractions. As a universal feature in a majority of eukaryotic organisms, crossover interference is the non-random occurrence of crossovers along the length of individual chromosomes. The model allows the estimation of crossover interference at adjacent marker intervals. The usefulness of this model was validated by reanalyzing a published data for tetraploid switchgrass [25]. The new three-point model provides an analytical tool for studying the genome structure and organization of autotetraploids by linkage mapping.

THE MODEL

Consider a cross between two outcrossing autotetraploids as the parents. The progeny from this cross is genotyped for molecular markers including two types, testcross and intercross. The testcross markers are those for which one parent is heterozygous whereas the other is homozygous. For the intercross markers, both parents are heterozygous. The linkage analysis of the former type can be performed per the segregation of gametic genotypes, since gametic genotypes and progeny genotypes follow the same pattern of segregation in this case.

Gametic segregation

For one autotetraploid parent with four homologous sets of chromosomes, let $a_h$, $b_h$ and $c_h$ ($h = 1, 2, 3, 4$) denote different alleles for three ordered markers $A$, $B$ and $C$. Assume the allelic configuration or linkage phase of the autotetraploid as

$$a_1 \mid a_2 \mid a_3 \mid a_4 \mid b_1 \mid b_2 \mid b_3 \mid b_4 \mid c_1 \mid c_2 \mid c_3 \mid c_4,$$

where the lines indicate individual homologous chromosomes on which three markers are located.
It is assumed that this autotetraploid undergoes normal multivalent pairing during meiosis, producing diploid gametes.

For two linked markers, a total of 100 observable diploid gamete genotypes actually contain 136 diploid gamete configurations. Fisher [23] figured out that these configurations can be classified into 11 gamete configuration modes among which there are nine observable gamete genotype modes. We used Fisher’s approach to find 2080 diploid gamete configurations classified into 107 gamete configuration modes for three linked markers. These gamete configuration modes can be distinguished as 59 observable gamete genotypes each with a frequency denoted by \( g_i \) (i = 1, ..., 59) (Supplementary Table S1). The frequencies of some genotypes can be directly expressed as the frequencies of their gamete configuration modes because each genotype contains only one gamete mode. If different gamete configuration modes have the same genotype, the frequency of each gamete configuration mode can be expressed as a proportion of the genotype frequency.

This proportion is determined by the expected numbers of recombinants contained within a particular gamete configuration mode. Thus, the frequency of a typical gamete is expressed by the frequency of its observable genotype and the linkage information. Let \( r_1 \) and \( r_2 \) denote the recombination fractions between markers A and B, and between markers B and C, respectively [12]. We define

\[
\begin{align*}
\varphi_1 &= \frac{r_1^2}{9 - 18r_1 + 10r_1^2} \\
\varphi_2 &= \frac{r_2^2}{9 - 18r_2 + 10r_2^2} \\
\psi_1 &= \frac{r_1}{3 - 2r_1}
\end{align*}
\]

which are used to express the frequencies of all the gamete configuration modes in terms of \( g_i \), \( \varphi_i \)'s and \( \psi_i \)'s, as listed in the fourth column of Supplementary Table S1.

### Parameter estimates with fully informative markers

The heterozygous parent (1) produces 10 gametes at each marker, arrayed as \((a_1a_1, a_2a_2, a_3a_3, a_4a_4, a_1a_2, a_1a_3, a_1a_4, a_2a_3, a_2a_4, a_3a_4)\) for marker A, \((b_1b_1, b_2b_2, b_3b_3, b_4b_4, b_1b_2, b_1b_3, b_1b_4, b_2b_3, b_2b_4, b_3b_4)\) for marker B, and \((c_1c_1, c_2c_2, c_3c_3, c_4c_4, c_1c_2, c_1c_3, c_1c_4, c_2c_3, c_2c_4, c_3c_4)\) for marker C. The first four gametes in each case are created due to the double reduction, while the others are generated from the chromosome pairing. The frequencies of gametes at the three markers, determined by \( g_i \), constitute a \((10 \times 10 \times 10)\) cubic matrix \( G \) with the \( jk3 \)-th element denoted by \( G_{j,k,3} \), where the subscripts \( j, k \) and \( 3 \) stand for the genotype at marker A, B and C, respectively. We write the faces of \( G \) from genotypes \( c_1c_1 \) to \( c_3c_4 \) at marker C as

\[
g_{c_1c_1} = \begin{pmatrix}
a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} & a_{19} & a_{110} \\
an_{11} & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} & a_{29} \\
an_{11} & a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & a_{37} & a_{38} & a_{39} \\
an_{11} & a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} & a_{47} & a_{48} & a_{49} \\
an_{11} & a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} & a_{57} & a_{58} & a_{59} \\
an_{11} & a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} & a_{67} & a_{68} & a_{69} \\
an_{11} & a_{71} & a_{72} & a_{73} & a_{74} & a_{75} & a_{76} & a_{77} & a_{78} & a_{79} \\
an_{11} & a_{81} & a_{82} & a_{83} & a_{84} & a_{85} & a_{86} & a_{87} & a_{88} & a_{89} \\
an_{11} & a_{91} & a_{92} & a_{93} & a_{94} & a_{95} & a_{96} & a_{97} & a_{98} & a_{99} \\
an_{11} & a_{101} & a_{102} & a_{103} & a_{104} & a_{105} & a_{106} & a_{107} & a_{108} & a_{109} \\
\end{pmatrix}
\]
on 15 August 2018

The maximum-likelihood estimates (MLEs) of $g_i$ can be obtained from observations of progeny genotypes by formulating a polynomial-based likelihood function. For fully informative markers, analytical MLEs of $g_i$ can be obtained.

Based on the definition of double reduction [22], we can easily give the MLE expression for estimating the coefficients of double reduction at markers A, B and C, i.e.

$$\alpha = \sum_{j_1=1}^{4} \sum_{j_2=1}^{4} \sum_{j_3=1}^{4} G_{j_1j_2j_3} = \sum_{l=1}^{25} g_i,$$

$$\gamma = \sum_{j_1=1}^{4} \sum_{j_2=1}^{4} \sum_{j_3=1}^{4} G_{j_1j_2j_3} = \sum_{l=1}^{45} g_i + \sum_{l=26}^{55} g_i,$$

$$\beta = \sum_{j_1=1}^{4} \sum_{j_2=1}^{4} \sum_{j_3=1}^{4} G_{j_1j_2j_3} = g_1 + g_2 + g_3 + g_7 + g_8 + g_9 + g_{10} + g_{11} + g_{12} + g_{13} + g_{14} + g_{15} + g_{16} + g_{17} + g_{18} + g_{19} + g_{20} + g_{21} + g_{22} + g_{23} + g_{24} + g_{25} + g_{26} + g_{27} + g_{28} + g_{29} + g_{30} + g_{31} + g_{32} + g_{33} + g_{34} + g_{35} + g_{36} + g_{37} + g_{38} + g_{39} + g_{40} + g_{41} + g_{42} + g_{43} + g_{44} + g_{45} + g_{46} + g_{47} + g_{48} + g_{49} + g_{50} + g_{51} + g_{52} + g_{53} + g_{54} + g_{55} + g_{56}.$$

By directly counting the numbers of recombinants contained within an observable diplod gamete genotype (Supplementary Table S1)[25], we can also provide the MLE expressions for calculating the recombination fractions between markers A and B ($r_1$), between markers B and C ($r_2$) and between markers A and C ($r_3$) as
The EM algorithm is implemented to estimate three coefficients of double reduction and three recombination fractions between the three markers. This algorithm includes a loop of iterations between the E step (aimed to calculate the expected proportions $\phi_1$, $\phi_2$, $\psi_1$ and $\psi_2$ expressed in Supplementary Table S1) and the M step expressed by Equations (3)–(8). The estimates at convergence are the MLEs of these parameters.

**Linkage phase inference**

In practice, we can only obtain the marker genotype of the heterozygous parent (1) and do not know its linkage phase of alleles among different markers *a priori*. This should be inferred from the marker data of parents and their progeny. Because of a huge number of linkage phase configurations given the three-marker genotype of a parent, we take two steps to infer the linkage phase as suggested by Luo *et al.*[17]. The first step is to determine the most likely alleles at each of the three linked markers. The second step is to choose the most likely linkage phase according to the likelihoods of progeny marker data under up to 24 possible linkage phases in the heterozygous parent.

The first step involves estimating the coefficient of double reduction at a marker and calculating the likelihood of the progeny marker data. Four alleles at a marker form a total of 16 diploid gamete configurations which are collapsed into 10 observable gamete genotypes because some of four alleles at the marker are identical or recessive. The frequencies of gamete genotypes can be derived by collapsing the cubic matrix of fully informative markers $G$. This procedure is quite tedious, but the MLEs of $g_i$’s can be obtained by using the EM algorithm derived in Wu and Ma [14]. As an example, we consider three partially informative markers: $A$, partially informative codominant marker with alleles $a_1$, $a_2$ and $a_3$; and $B$ and $C$, partially informative dominant markers, with alleles $b_1$, $b_2$ and $b_3$, and $c_1$ and $c_2$, respectively. Consider a full-sib family derived from a pseudo-test backcross design generated by a heterozygous autotetraploid with a particular linkage phase and a homozygous autotetraploid, expressed as

$$
\begin{array}{ccccccc}
  a_1 & a_2 & a_3 & a & a & a \\
  b_3 & b_1 & b_2 & b & b & b \\
  c_2 & c_1 & c_1 & c & c & c \\
\end{array}
$$

The symbol $\varnothing$ in the heterozygous parent denotes a null or recessive allele. We provide a two-stage hierarchic EM algorithm to estimate double reductions and recombination fractions for partially informative markers. Since this derivation is quite lengthy, we show it in Supplementary Text S1.

**RESULT**

**Worked example**

The three-point linkage analysis model was used to reanalyze a tetraploid mapping data. Okada *et al.*
[26] reported a linkage map for tetraploid switchgrass (Panicum virgatum) with a full-sib family of 238 progeny derived from two heterozygous parents. The mapping population was genotyped for 1509 dominant markers (each with a dominant allele 1 and recessive allele 0), of which 606 are the testcross markers segregating for the female parent (K5), 667 are the testcross markers for the male parent (A4) and 126 are intercross markers for both parents. All these markers are single-dose amplicons (simplex), double-dose amplicons (duplex), or triple-dose amplicons (triplex). For three ordered testcross markers, each of which can be simplex (S), duplex (D), or triplex (T), there are a total of 3 × 3 × 3 = 27 possible genotype combinations. Let us first consider three of the combinations: S × S × S, D × D × D and T × T × T. Combination S × S × S may have four possible linkage phases as follows:

\[
\begin{align*}
\text{SSS1} & : 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \\
\text{SSS2} & : 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \\
\text{SSS3} & : 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \\
\text{SSS4} & : 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \\
\end{align*}
\]

Combination D × D × D has eight possible linkage phases, i.e.

\[
\begin{align*}
\text{DDD1} & : 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 1 \ 1 \ 1 \ 0 \\
\text{DDD2} & : 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \\
\text{DDD3} & : 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \\
\text{DDD4} & : 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \\
\text{DDD5} & : 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 1 \ 0 \ 0 \ 1 \\
\text{DDD6} & : 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \\
\text{DDD7} & : 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \\
\text{DDD8} & : 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \\
\end{align*}
\]

Combination T × T × T has four possible linkage phases, i.e.

\[
\begin{align*}
\text{TTT1} & : 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 1 \ 1 \ 1 \ 0 \\
\text{TTT2} & : 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \\
\text{TTT3} & : 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 1 \\
\text{TTT4} & : 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 1 \\
\end{align*}
\]

The other 24 combinations each have different linkage phases (Supplementary Text S2). The model analyzes all possible genotype combinations and all possible linkage phases for each combination. An optimal genotype combination and optimal phase are determined as one that gives the largest likelihood, in which the MLEs of the recombination fractions and double reductions at three markers are obtained. In Supplementary Text S3, we provide a procedure for parameter estimation implemented with a two-stage hierarchic EM algorithm.

Yang et al. [19] reported high-density linkage maps for the full-sib family of switchgrass using Okada et al.'s [26] marker data. As a worked example, we chose the first nine markers from the first linkage group constructed from the female heterozygous parent K5, whose linkage was reanalyzed using our three-point autotetraploid model. In Supplementary Table S2, we provide the results used to find the best genotype combination and the most likely linkage phase for these nine markers, under which the MLEs of the recombination fractions and double reductions for each set of three markers were obtained simultaneously. Since our model does not assume that a particular pattern occurs for the genomic distribution of double reduction, the MLEs of double reduction from our three-point linkage model and a single-marker segregation model [13] should be similar. In Supplementary Text S4, we provided a simple EM algorithm for estimating the double reduction for individual dominant markers. Both linkage and segregation models gave similar results, i.e. the coefficients of double reduction at all the markers analyzed are close to zero, suggesting that this portion of the switchgrass genome does not undergo a multivalent process.

As shown by the estimated standard errors by bootstrapping approaches, the MLEs of the recombination fractions from our three-point linkage model are reasonably precise (Table 1) given the sample size used (238). The nine markers were also subject to a conventional two-point linkage analysis, which gave much larger estimates of the standard errors of the MLEs (20–40% larger than those in Table 1 from the three-point model). Our three-point analysis allows the estimates of crossover interference. It is interesting to see that some genomic intervals have strong crossover interference by which recombinant events between two adjacent marker intervals are affected by one another. The occurrence of crossover interference suggests that adjacent crossovers may influence the segregation of a chromosome pair during meiosis so as to offer a selective advantage for the organism to adapt to adverse environments [27].

**Simulation**

To evaluate the accuracy of the estimates of our model, we simulated a group of pseudo-test backcross markers in a full-sib family derived from an autotetraploid parent with linkage phase and another homozygous parent (10). Assume double reductions
for markers $A$, $B$ and $C$, $(\alpha, \beta, \gamma) = (0.15, 0.05, 0.10)$ or $(0.10, 0.15, 0.08)$ and recombination fractions $(r_1, r_2, r_3) = (0.10, 0.15, 0.20)$ and $(0.15, 0.30, 0.39)$, with the coefficient of coincidence $c = 2.5$ and $c = 1.0$, presenting strong crossover interference and no crossover interference, respectively. With these parameters, we simulated the tetraploid family of size 200, 400, or 800.

In Supplementary Text S1, the EM algorithm is given to estimate the recombination fractions and double reductions at different markers. Table 2 tabulates the means and standard errors of the MLEs

Table 1: The MLEs of the recombination fractions and the standard deviations of the MLEs (in parentheses) for nine ordered testcross markers in a full-sib family of switchgrass

<table>
<thead>
<tr>
<th>Marker</th>
<th>$r_1/r_2$</th>
<th>$r_3$</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sww2654 308</td>
<td>0.0673(0.0051)</td>
<td>0.0673(0.0060)</td>
<td>1.19</td>
</tr>
<tr>
<td>sww2262 237</td>
<td>0.0000(0.0000)</td>
<td>0.0003(0.0000)</td>
<td></td>
</tr>
<tr>
<td>nfg50 139</td>
<td>0.0266(0.0021)</td>
<td>0.0263(0.0012)</td>
<td>36.47</td>
</tr>
<tr>
<td>nfg50 137</td>
<td>0.0577(0.0042)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sww2524 187</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sww2022 417</td>
<td>0.1019(0.0098)</td>
<td>0.1586(0.0124)</td>
<td>0.08</td>
</tr>
<tr>
<td>sww737 258</td>
<td>0.0616(0.0053)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sww50 204</td>
<td>0.0061(0.0060)</td>
<td>5.97</td>
<td></td>
</tr>
<tr>
<td>sww50 194</td>
<td>0.0000(0.0000)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: $r_1$ and $r_2$ are the recombination fractions between two adjacent markers, $r_3$ is the recombination fraction between the two markers separated by a middle marker, and $c$ is the coefficient of crossover coincidence.

Table 2: The MLEs of the parameters and their standard deviations (in parentheses) from a simulated full-sib family derived from a heterozygous parent (10) and a homozygous parent at three partially informative markers

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$\hat{a}$</th>
<th>$\hat{b}$</th>
<th>$\hat{c}$</th>
<th>$\hat{r}_1$</th>
<th>$\hat{r}_2$</th>
<th>$\hat{r}_3/c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 200$</td>
<td>0.15</td>
<td>0.05</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.20/2.5</td>
</tr>
<tr>
<td>$n = 400$</td>
<td>0.15</td>
<td>0.05</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.20/2.5</td>
</tr>
<tr>
<td>$n = 800$</td>
<td>0.15</td>
<td>0.05</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.20/2.5</td>
</tr>
</tbody>
</table>

Note: $n, a, b, c, r_1, r_2, r_3$ and $c$ are the sample size, the coefficients of double reduction at three markers, recombination fractions between markers $A$ and $B$, between markers $B$ and $C$ and between markers $A$ and $C$ and the coefficient of coincident, respectively.
DISCUSSION

A set of statistical models has been developed for linkage analysis in autotetraploids by integrating their cytological behavior of double reduction [12,14]. By producing more gametes, double reduction may play a significant role in the maintenance of genetic diversity that favors evolution under natural populations. Thus, the estimation of double reduction becomes essential for shedding light on the genetic variation and organization of natural populations in autopolyploids [28]. Based on our previous computer simulation, the estimation of the linkage among markers would be biased for a progeny of autotetraploids that undergo double reduction if double reduction was neglected during the estimation procedure [12,14].

In this article, we assess and investigate a general model for three-point analysis of the linkage, crossover interference, and double reduction for codominant markers and less informative dominant makers. The merits of the model are 3-fold: First, three-point analysis is more powerful and more precise for linkage detection and estimation, when marker genotypes are less informative, distortedly segregating, and misclassified and/or contains missing values [19,29], compared to commonly used two-point analysis. Second, the model estimates the double reduction, an important cytological parameter of evolutionary significance [20,21,24], without needing the assumption of its genomic distribution. Third, the model is equipped with a capacity to estimate and test the degree of crossover interference. By influencing the segregation of adjacent chromosome pairs, crossover interference may be an evolutionary force for the organism to better adapt to environmental fluctuations [27].

The statistical behavior of the model was well validated through simulation studies. The usefulness and utilization of the model for the autotetraploid multilocus linkage analysis of dominant markers were demonstrated by analyzing a published data set collected from a full-sib family derived from tetraploid switchgrass. Previous studies suggest that switchgrass is a disomic polyploid that undergoes bivalent pairing [26,30,31]. It appears that the evidence of absence of double reduction observed from our autotetraploid linkage model supports the bivalent pairing feature of switchgrass, although more informative codominant markers are needed to address this question.

The model can be extended to study the linkage of molecular markers for higher-level polyploids, hexaploids (such as wheat and kiwifruit), octaploids (such as dahlias), decaploids (such as strawberries) and dodecaploids (such as plumed cockscomb), by considering more allelic and linkage phase combinations. Given the exponential increase of these combinations with polyploid level, an efficient computing algorithm should be developed, allowing the selection among the most likely combinations. The development of statistical models for mapping quantitative trait loci (QTLs) that control complex traits in polyploids are still in its infant stage (but see [32,33]). Li et al. [34] has for the first time incorporated Fisher’s [23] classification theory into a QTL mapping model for autotetraploids undergoing double reduction. The uniqueness of this model lies in its estimation of the double reduction at a putative QTL, thereby providing an additional dimension to study the diversification and evolution of complex traits in polyploids. Our three-point analysis model provides a general framework for interval mapping of QTLs [25,32] and, therefore, the precise understanding of the genetic control of complex phenotypes in a multivalent tetraploid. The computer code for the model can be freely downloaded from http://statgen.psu.edu/software.

SUPPLEMENTARY DATA

Supplementary data are available online at http://bib.oxfordjournals.org/.

Table 3: The power of correctly finding the linkage phase for a simulated full-sib family derived from a heterozygous parent (10) and a homozygous parent at three partially informative markers

<table>
<thead>
<tr>
<th>Design</th>
<th>n</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200</td>
<td>0.942</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.986</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.998</td>
</tr>
<tr>
<td>Design2</td>
<td>200</td>
<td>0.983</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note: Designs 1 and 2 have parameters \((\alpha, \beta, \gamma, r_1, r_2, r_3) = (0.15, 0.05, 0.10, 0.10, 0.15, 0.20/2.5)\) and \((0.10, 0.15, 0.08, 0.15, 0.30, 0.39/1.0)\), respectively.
Key Points
- Autotetraploids, a group of important species to agriculture and biology, are increasingly studied because of their pivotal role in agriculture, evolution, and speciation.
- We review and assess a three-point linkage analysis model for a controlled cross population of autotetraploids by considering the unique meiotic behavior of autopolyploids that undergo double reduction through multivalent pairing.
- The model allows the linkage, double reduction and crossover interference to be estimated at the same time, facilitating our study of the genome structure and organization of autotetraploids.

FUNDING
This work was supported by Changjiang Scholars Award and “Thousand-person Plan” Award.

References