A statistical model for QTL mapping in polysomic autotetraploids underlying double reduction

Fang Xu*, Yafei Lyu*, Chunfa Tong*, Weimiao Wu, Xuli Zhu, Danni Yin, Qin Yan, Jian Zhang, Xiaoming Pang, Christian M. Tobias and Rongling Wu

Submitted: 1st September 2013; Received (in revised form): 23rd September 2013

Abstract
As a group of economically important species, linkage mapping of polysomic autotetraploids, including potato, sugarcane and rose, is difficult to conduct due to their unique meiotic property of double reduction that allows sister chromatids to enter into the same gamete. We describe and assess a statistical model for mapping quantitative trait loci (QTLs) in polysomic autotetraploids. The model incorporates double reduction, built in the mixture model-based framework and implemented with the expectation–maximization algorithm. It allows the simultaneous estimation of QTL positions, QTL effects and the degree of double reduction as well as the assessment of the estimation precision of these parameters. We performed computer simulation to examine the statistical properties of the method and validate its use through analyzing real data in tetraploid switchgrass.

Keywords: quantitative trait loci; polysomic autotetraploid; EM algorithm; quantitative genetic model

Corresponding author. Rongling Wu, Center for Computational Biology, Beijing Forestry University, Beijing 100083, China. Tel.: +01186 10 6233 6283; Fax: +01186 10 6233 6283; E-mail: rwu@bjfu.edu.cn or Center for Statistical Genetics, Pennsylvania State University, Hershey, PA 17033, USA. Tel.: +1 717 531 2037; Fax: +1 717 531 0480; E-mail: rwu@phys.psu.edu

*These authors contributed to this work equally.

Fang Xu is a PhD student in forest genetics and tree breeding at the Center for Computational Biology at Beijing Forestry University. He studies the genetic architecture of quantitative traits using an integrative computational and experimental approach.

Yafei Lyu is a PhD student in bioinformatics at The Pennsylvania State University. His research focuses on the statistical genetics of quantitative traits.

Chunfa Tong is Professor of Forest Genetics and Genomics at Nanjing Forestry University. His research interest is in the methodological and algorithmic development of genetic mapping for complex traits.

Weimiao Wu is a fourth-year undergraduate student in the Liang Xi Talented Class at Beijing Forestry University. She develops statistical models for genetic mapping of complex traits.

Xuli Zhu is a Master’s student in quantitative molecular genetics at the Center for Computational Biology at Beijing Forestry University. His thesis focuses on the statistical modeling of genetic data from open-pollinated designs.

Danni Yin is a Master’s student in quantitative molecular genetics at the Center for Computational Biology at Beijing Forestry University. Her thesis focuses on the genetic dissection of apomixis in plants.

Qin Yan is a Master’s student in quantitative molecular genetics at the Center for Computational Biology at Beijing Forestry University. Her thesis focuses on the genetic dissection of allometry.

Jian Zhang is a PhD student in quantitative molecular genetics in the Center for Computational Biology at Beijing Forestry University. His research interest lies in the genetic analysis of saline-tolerant traits in forest tree.

Xiaoming Pang is Associate Professor of Tree Breeding at Beijing Forestry University. His research interest focuses on the utilization of molecular genetics and biotechnologies to study population genetic diversity and map quantitative trait loci in fruit trees.

Christian M. Tobias is Research Molecular Biologist in Genomics and Gene Discovery Research Unit, Western Regional Research Center, USDA. Part of Dr Tobias’ s ongoing projects 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 Changjiang Scholars Professor of Genetics and the Director of the Center for Computational Biology at Beijing Forestry University. His interest is to unravel the genetic roots for the outcome of a biological trait by dissecting the trait into its biochemical and developmental pathways.
INTRODUCTION
Because of their biological and economic importance, genetic analysis in polyploids has long intrigued geneticists and evolutionary biologists [1–8]. Linkage analysis with molecular markers has particular power to study the structure, organization and function of polyploid genomes [9–18] and found a basis for mapping quantitative trait loci (QTLs) that affect complex traits [19–23]. Statistical models for linkage mapping in polyploids are qualitatively different from those in diploids because of unique properties of the former. Several models that take into account these properties have been developed and play an increasingly important part in polyploid mapping [17,18,24–30].

Specifically, these models are classified into two categories based on the origin of polyploids. One focuses on linkage mapping of allopolyploids derived from the combination of chromosomes from different species. For this type of polyploids, Wu et al. [26,28,29] incorporated the so-called chromosomal pairing preference [4] into the linkage analysis framework, more precisely characterizing the process of their chromosomal pairing, and thereby marker segregation patterns. Based on this property, statistical models for QTL mapping in allopolyploids have been developed [20,28].

The second category concerns with linkage mapping of autopolyploids, with the origin mostly from the duplication of similar genomes [5,31]. In autopolyploids, chromosomes pair among more than two homologous copies as in allopolyploids, which thus leads to the occurrence of double reduction, i.e. two sister chromatids of a chromosome are sorted into the same gamete [2]. In the 1940s, Fisher [3] pioneered a conceptual model for identifying 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. This model found the basis of polysomic linkage analysis by Wu et al. [24,25] and other teams [16–18]. The models for linkage analysis in autotetraploids have been extended to include three markers and can use any type of markers, regardless of their informativeness and dominant or codominant nature [30].

What is lacking in the literature is that statistical models for QTL mapping in autotetraploids have not been well investigated. A few authors explored the possibility of using model selection to map autotetraploid QTLs [21], but did not make full use of Fisher’s [3] 11 classifications of gamete formation. Li et al. [32] made an interesting attempt to incorporate these classifications based on a single marker–single QTL model. The merit of this approach lies in its capacity to estimate the extent of double reduction of the QTL. Its disadvantage is that it cannot map the location of a putative QTL. Interval mapping based on two flanking markers allows the estimation and test of not only the QTL effect but also of the QTL-marker linkage [33]. In this article, we highlight a two-stage hierarchical model for autotetraploid QTL mapping, by estimating the probabilities of gamete formation modes and therefore double reduction in the upper hierarchy and estimating the marker–QTL recombination fraction in the lower hierarchy. The model was formulated within the maximum likelihood context and implemented with the expectation–maximization (EM) algorithm. We used computer simulation to demonstrate statistical properties of the method and its analytical merits. Its biological merit is demonstrated by an analysis of real data in tetraploid switchgrass.

MODEL
Genetic design
As a commonly used mapping population, consider an F1 full-sib family derived from two heterozygous parental lines of autotetraploids. This family is genotyped by a set of molecular markers. It is possible that such a family contains two kinds of segregating markers: (1) testcross markers at which one parent is heterozygous and the other is homozygous and (2) intercross markers at which both parents are heterozygous. This article will focus on the first kink of markers in which the genotypes of progeny can be identified by examining the genotypes of gametes produced by the heterozygous parent. Thus, the segregation of gametes can find a basis of model development for QTL mapping in tetraploids. Consider the full-sib family of n individuals, each genotyped by codominant testcross markers and measured for a trait that is controlled by QTLs. It is supposed that a linkage map is constructed from these markers on which QTLs are scanned for their existence through the genome.

Considering a test cross QTL of four different alleles that segregates due to one heterozygous parent. This parent has a genotype $Q_iQ_jQ_kQ_l$, which generates four double reduction gametes $Q_iQ_j$, $Q_iQ_k$, $Q_iQ_l$, and $Q_jQ_l$ and six non-double reduction gametes $Q_iQ_j$, $Q_iQ_k$, $Q_iQ_l$, $Q_jQ_k$, $Q_jQ_l$, and $Q_kQ_l$. The genotypic
values of these diploid gametes, denoted as $\mu_{jk}$, $j < k = 1, 2, 3, 4$, can be partitioned in the following way:

\[
\begin{array}{c|c|c|c|c}
    Q_1 & Q_2 & Q_3 & Q_4 \\
    \left(\frac{1}{2}a_1\right) & \left(\frac{1}{2}a_2\right) & \left(\frac{1}{2}a_3\right) & \left(-\frac{1}{2}a_1 - \frac{1}{2}a_2 - \frac{1}{2}a_3\right) \\
    \hline
    Q_1 & \mu_{11} = \mu_{12} = & \mu_{13} = & \mu_{14} = \\
    \left(\frac{1}{2}a_1\right) & \mu + a_1 & \mu + \frac{1}{2}a_1 + \frac{1}{2}a_2 + \frac{1}{2}a_3 & \mu - \frac{1}{2}a_1 - \frac{1}{2}a_2 - \frac{1}{2}a_3 + \frac{1}{2}d_1 \\
    \hline
    Q_2 & \mu_{22} = & \mu_{23} = & \mu_{24} = \\
    \left(\frac{1}{2}a_2\right) & \mu + a_2 & \mu + \frac{1}{2}a_1 + \frac{1}{2}a_2 + \frac{1}{2}d_2 & \mu - \frac{1}{2}a_1 - \frac{1}{2}a_2 - \frac{1}{2}d_2 \\
    \hline
    Q_3 & \mu_{33} = & \mu_{34} = \\
    \left(\frac{1}{2}a_3\right) & \mu + a_3 & \mu - \frac{1}{2}a_1 - \frac{1}{2}a_2 - \frac{1}{2}a_3 + \frac{1}{2}d_2 \\
    \hline
    Q_4 & \mu_{44} = & & \\
    \left(-\frac{1}{2}a_1 - \frac{1}{2}a_2 - \frac{1}{2}a_3\right) & \mu - a_1 - a_2 - a_3 & \\
\end{array}
\]

where $\mu$ is the overall mean, $a_1$, $a_2$ and $a_3$ are double the additive effect of alleles $Q_1$, $Q_2$ and $Q_3$, respectively, and $d_4$ is the dominant effect due to the interaction of alleles $j$ and $k$ ($j < k = 1, 2, 3, 4$). All these genetic effects can be calculated in terms of genotypic values from the above equations.

### Characterizing diploid gamete modes

The detection of the QTL relies on the co-segregation of the markers that are linked with it. Suppose there are two fully informative markers $A$ (of alleles $A_1$--$A_4$) and $B$ (of alleles $B_1$--$B_3$) that bracket the QTL. Let $r$, $r_1$ and $r_2$ denote the recombination fractions between markers $A$ and $B$, marker $A$ and the QTL and the QTL and marker $B$, respectively.

Lu et al. extended Fisher's theory [3] to derive the polygenic inheritance of three linked loci in an autotetraploid [30]. This extension is the basis of interval QTL mapping. Totally, three loci have 2080 diploid gametes, which can be classified into 107 gamete modes. We rearrange these gamete modes for the two markers and QTL and find that there are only 59 typical genotypes each with a frequency denoted by $g_v$ ($v = 1, \ldots, 59$) (Table 1). Some of the typical genotypes contain two or four gamete modes with relative proportions determined by the recombination fractions $r_1$ and $r_2$. For example, the sum of frequencies of modes 15 and 16 (denoted by $g_{15a}$ and $g_{15b}$) constitutes the frequency of a typical genotype $A_1A_2Q_1Q_2B_1B_2$, i.e. $g_{15} = g_{15a}$ (for mode $A_1Q_1B_1|A_2Q_2B_2$) + $g_{15b}$ (for mode $A_1Q_1B_2|A_2Q_2B_1$). All the frequencies of genotypes at the two markers and QTL form a $10 \times 10 \times 10$ cubic matrix $G$ with its element denoted by $G_{ijkl}$. Let the subscripts $i$, $j$, $k$, $l$, $i_0$, $j_0$, $k_0$, and $l_0$ stand for the genotypes at markers $A$, $B$, and the QTL, respectively. By shrinking the matrix $G$, we obtain the expected frequencies of the observable genotypes at the two markers in matrix form expressed as:

\[
M = \begin{pmatrix}
\frac{1}{4} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} \\
\frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} \\
\frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} \\
\frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} \\
\frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} \\
\frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} \\
\frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{12} \\
\frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{12} \\
\frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{12} \\
\frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{12}
\end{pmatrix}
\]
where

\[
\begin{align*}
\pi_1 &= g_1 + g_2 + g_3 + g_4 \\
\pi_2 &= g_3 + g_6 + g_7 + g_8 + g_9 + g_{10} + g_{11} \\
\pi_3 &= g_{12} + g_{13} + g_{14} + g_{15} + g_{16} + g_{17} + g_{18} \\
\pi_4 &= g_{19} + g_{20} + g_{21} + g_{22} + g_{23} + g_{24} + g_{25} \\
\pi_5 &= g_{26} + g_{27} + g_{28} + g_{29} + g_{30} + g_{31} + g_{32} \\
\pi_6 &= g_{33} + g_{34} + g_{35} + g_{36} + g_{37} + g_{38} + g_{39} \\
\pi_7 &= g_{40} + g_{41} + g_{42} + g_{43} + g_{44} \\
\pi_8 &= g_{45} + g_{46} + g_{47} + g_{48} + g_{49} + g_{50} + g_{51} + g_{52} + g_{53} + g_{54} \\
\pi_9 &= g_{55} + g_{56} + g_{57} + g_{58} + g_{59}
\end{align*}
\]

and the \((l_A, l_B)\)th element of \(M\) is calculated as

\[
M_{l_A l_B} = \sum_{l_B=1}^{10} G_{(l_A l_B)}
\]

By defining

\[
\begin{align*}
n_1 &= n_{11} + n_{22} + n_{33} + n_{44} \\
n_2 &= n_{12} + n_{13} + n_{14} + n_{21} + n_{23} + n_{24} + n_{31} + n_{32} + n_{34} + n_{41} + n_{42} + n_{43} \\
n_3 &= n_{15} + n_{16} + n_{17} + n_{25} + n_{26} + n_{29} + n_{36} + n_{38} + n_{310} + n_{47} + n_{49} + n_{410} \\
n_4 &= n_{18} + n_{19} + n_{110} + n_{28} + n_{26} + n_{327} + n_{310} + n_{335} + n_{337} + n_{39} + n_{415} + n_{46} + n_{48} \\
n_5 &= n_{54} + n_{52} + n_{61} + n_{63} + n_{71} + n_{74} + n_{82} + n_{83} + n_{82} + n_{84} + n_{94} + n_{10,3} + n_{10,4} \\
n_6 &= n_{53} + n_{54} + n_{62} + n_{64} + n_{72} + n_{73} + n_{81} + n_{84} + n_{91} + n_{93} + n_{10,1} + n_{10,2} \\
n_7 &= n_{55} + n_{56} + n_{77} + n_{88} + n_{99} + n_{10,10} \\
n_8 &= n_{56} + n_{57} + n_{58} + n_{59} + n_{65} + n_{67} + n_{68} + n_{610} + n_{75} + n_{76} + n_{79} + n_{710} + n_{85} + n_{86} + n_{89} + n_{810} + n_{85} + n_{87} + n_{88} + n_{10,6} + n_{10,7} + n_{10,8} + n_{10,9} \\
n_9 &= n_{510} + n_{697} + n_{78} + n_{97} + n_{96} + n_{10,5}
\end{align*}
\]

where \(n_{l_A l_B}\) is the number of individuals with marker genotype \(l_A l_B\), we have

\[
\hat{\pi}_k = \frac{n_k}{\sum_{k=1}^{10} n_k}
\]

(4)

It can be seen that the coefficients of double reduction at marker \(A\) \((g)\), marker \(B\) \((\gamma)\) and the QTL \((\gamma)\) are calculated as:

\[
\alpha = \pi_1 + \pi_2 + \pi_3 + \pi_4
\]

(5)

\[
\beta = \pi_1 + \pi_2 + \pi_5 + \pi_6
\]

(6)

\[
\gamma = g_1 + g_2 + g_3 + g_6 + g_7 + g_8 + g_9 + g_{10} + g_{11} + g_{12} + g_{13} + g_{14} + g_{19} + g_{20} + g_{21} + g_{22} + g_{23} + g_{24} + g_{25} + g_{29} + g_{30} + g_{31} + g_{32} + g_{34} + g_{35} + g_{36} + g_{38} + g_{39} + g_{41} + g_{42} + g_{43} + g_{44} + g_{45} + g_{46} + g_{48} + g_{53} + g_{55} + g_{57} + g_{59} + g_{61} + g_{62} + g_{63} + g_{64} + g_{65} + g_{67} + g_{68} + g_{69} + g_{70} + g_{71} + g_{73} + g_{74} + g_{75} + g_{76} + g_{79} + g_{810} + g_{85} + g_{86} + g_{89} + g_{810} + g_{85} + g_{87} + g_{88} + g_{10,6} + g_{10,7} + g_{10,8} + g_{10,9}
\]

(7)

The expected number of recombination events in each gamete mode should be the weighted average of the number of recombination events for each typical genotype. If the frequencies of all the 107 gamete modes are known, the recombination fractions \(r_1, r_2\) and \(r\) can be directly counted as:

\[
r_1 = g_2 + g_4 + g_6 + g_7 + g_{10} + g_{11} + g_{12} + g_{13} + g_{14} + g_{19} + g_{20} + g_{21} + g_{22} + g_{23} + g_{24} + g_{25} + g_{29} + g_{30} + g_{31} + g_{32} + g_{34} + g_{35} + g_{36} + g_{38} + g_{39} + g_{41} + g_{42} + g_{43} + g_{44} + g_{45} + g_{46} + g_{48} + g_{53} + g_{55} + g_{57} + g_{59} + g_{61} + g_{62} + g_{63} + g_{64} + g_{65} + g_{67} + g_{68} + g_{69} + g_{70} + g_{71} + g_{73} + g_{74} + g_{75} + g_{76} + g_{79} + g_{810} + g_{85} + g_{86} + g_{89} + g_{810} + g_{85} + g_{87} + g_{88} + g_{10,6} + g_{10,7} + g_{10,8} + g_{10,9}
\]

(8)

\[
r_2 = g_2 + g_4 + g_5 + g_7 + g_9 + g_{11} + g_{14} + g_{18} + g_{19} + g_{21} + g_{23} + g_{27} + g_{28} + g_{31} + g_{32} + g_{34} + g_{35} + g_{38} + g_{39} + g_{41} + g_{44} + g_{46} + g_{48} + g_{53} + g_{55} + g_{57} + g_{59} + g_{61} + g_{62} + g_{63} + g_{64} + g_{65} + g_{67} + g_{68} + g_{69} + g_{70} + g_{71} + g_{73} + g_{74} + g_{75} + g_{76} + g_{79} + g_{810} + g_{85} + g_{86} + g_{89} + g_{810} + g_{85} + g_{87} + g_{88} + g_{10,6} + g_{10,7} + g_{10,8} + g_{10,9}
\]

(9)

\[
r = \pi_2 + \pi_4 + \pi_6 + \pi_0 + g_{52} + \frac{g_{53}}{2} + \pi_5 + g_{60} + g_{41} + g_{44} + g_{49} + g_{50} + g_{51} + \frac{g_{45}}{2} + g_{46} + g_{47} + g_{48} + g_{53} + g_{54} + \frac{g_{42} + g_{40}}{2} + \pi_1 + \frac{g_{42} + g_{50}}{2} + \Phi_2 - \frac{g_{42} + g_{50}}{2} \Phi_2 + \frac{g_{42} + g_{50}}{2} + \frac{g_{43} + g_{50}}{2} + g_{51} - g_{52} \Phi_1 + g_{50} \Phi_2 \Phi_1 + \frac{g_{43} + g_{49}}{2} + g_{51} - g_{52} - g_{49} \Phi_1 \Phi_2 - (g_{43} + g_{51} - g_{52}) \Phi_1 \Phi_2
\]

(10)
Table 1: Modes of gamete formation for a QTL flanked by two markers in a polysomic autotetraploid

<table>
<thead>
<tr>
<th>Mode</th>
<th>Typical gamete</th>
<th>Number</th>
<th>Frequency</th>
<th>Mode</th>
<th>Typical gamete</th>
<th>Number</th>
<th>Frequency</th>
<th>Mode</th>
<th>Typical gamete</th>
<th>Number</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AQ1B/AQ1B</td>
<td>16</td>
<td>62%</td>
<td>17</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>83%</td>
<td>18</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>84%</td>
</tr>
<tr>
<td>2</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>84%</td>
<td>25</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>86%</td>
<td>26</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>87%</td>
</tr>
<tr>
<td>3</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>86%</td>
<td>27</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>88%</td>
<td>28</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>89%</td>
</tr>
<tr>
<td>4</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>88%</td>
<td>30</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>90%</td>
<td>31</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>91%</td>
</tr>
<tr>
<td>5</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>90%</td>
<td>33</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>92%</td>
<td>34</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>93%</td>
</tr>
<tr>
<td>6</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>92%</td>
<td>36</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>94%</td>
<td>37</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>95%</td>
</tr>
<tr>
<td>7</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>94%</td>
<td>39</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>96%</td>
<td>40</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>97%</td>
</tr>
<tr>
<td>8</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>96%</td>
<td>42</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>98%</td>
<td>43</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>99%</td>
</tr>
<tr>
<td>9</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>98%</td>
<td>45</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>99%</td>
<td>46</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
</tr>
<tr>
<td>10</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>99%</td>
<td>48</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>49</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
</tr>
<tr>
<td>11</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>51</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>52</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
</tr>
<tr>
<td>12</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>54</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>55</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
</tr>
<tr>
<td>13</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>57</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>58</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
</tr>
<tr>
<td>14</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>60</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>61</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
</tr>
<tr>
<td>15</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>63</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>64</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
</tr>
<tr>
<td>16</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>66</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>67</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
</tr>
<tr>
<td>17</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>69</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>70</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
</tr>
<tr>
<td>18</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>72</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
<td>73</td>
<td>AQ1B/AQ1B</td>
<td>24</td>
<td>100%</td>
</tr>
</tbody>
</table>

where $\phi_1$, $\phi_2$, $\psi_1$, and $\psi_2$ are the proportions defined as: $\phi_1 = \frac{n}{3-2n}$, $\phi_2 = \frac{n}{3-2n}$, $\psi_1 = \frac{n}{3-2n}$, and $\psi_2 = \frac{n}{3-2n}$ (see also Table 2).

Equations (5–10) establish the procedure for estimating the double reductions and recombination fractions between the two markers and QTL. This will be incorporated into the statistical estimation of these parameters as well as the genetic effects of QTL.

Likelihood and estimation

For a mapping population of $n$ individuals that are genotyped (M) and phenotyped (y), its likelihood in terms of the putative QTL can be formulated as:

$$L(\Omega|y,M) = \prod_{l=1}^{10} \prod_{q=1}^{10} \prod_{i=1}^{10} \prod_{j=1}^{10} \left( \frac{G_l | l_{ij} | g_{ij}}{M_{l_{ij}}} \right) f_{l_{ij}}(y_{l_{ij}}, h_{l_{ij}}; \sigma^2)$$

where $\Omega = (g_1, \ldots, g_{59}; \mu_1, \ldots, \mu_{10}; \sigma^2)$ is the vector of unknown parameters to be estimated; $G_l | l_{ij} | g_{ij}$ is the conditional (prior) probability of the $l_{ij}$th QTL genotype given the $l_{ij}$th marker genotype, with the frequency of joint marker-QTL genotype $G_l | l_{ij} | g_{ij}$ expressed in terms of 59 $g$-probabilities; and $f_{l_{ij}}(y_{l_{ij}}, h_{l_{ij}}; \sigma^2)$ is the probability density function of normal distribution with mean $\mu_{l_{ij}}$ and variance $\sigma^2$.

Note that the parameters of $g$ in Equation (11) are required to satisfy the linear restrictions of expression (2) and Equations (8–10) for the recombination fractions $r_1$, $r_2$, and $r$.

We implement the EM algorithm to find the maximum likelihood estimates (MLEs) of $\Omega$ through specific linear restrictions on the parameters. By assuming a QTL at a particular position and scanning it at every 1 cM along the genome, we determine an
Table 2: Frequencies of the mixed gamete modes in term of $g_v$

<table>
<thead>
<tr>
<th>$G_{1a}$</th>
<th>$G_{1b}$</th>
<th>$G_{1c}$</th>
<th>$G_{1d}$</th>
<th>$G_{1e}$</th>
<th>$G_{1f}$</th>
<th>$G_{1g}$</th>
<th>$G_{1h}$</th>
<th>$G_{1i}$</th>
<th>$G_{1j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(1-\psi_1)\phi_2\psi_3$</td>
<td>$(1-\phi_1)\psi_2\psi_3$</td>
<td>$(1-\phi_1)\phi_2\psi_3$</td>
<td>$(1-\psi_1)\psi_2\psi_3$</td>
<td>$(1-\psi_1)\phi_2\psi_3$</td>
<td>$(1-\phi_1)\psi_2\psi_3$</td>
<td>$(1-\phi_1)\phi_2\psi_3$</td>
<td>$(1-\psi_1)\psi_2\psi_3$</td>
<td>$(1-\psi_1)\phi_2\psi_3$</td>
<td>$(1-\phi_1)\psi_2\psi_3$</td>
</tr>
</tbody>
</table>

Note: $\phi_1 = \frac{\gamma_1}{\gamma_1+\phi_2}$, $\phi_2 = \frac{\gamma_2}{\gamma_2+\phi_2}$, $\psi_1 = \frac{\gamma_1}{\gamma_1+\psi_2}$, and $\psi_2 = \frac{\gamma_2}{\gamma_2+\psi_2}$.

optimal position of the QTL. Given an assumed QTL position, we develop an iterative procedure to calculate the parameters as follows:

In the E step, the posterior probability with which an individual (with marker genotype $l_{Alb}$) carries a QTL genotype $l_Q$ is calculated using

$$P_{l_Q/l_{Alb}} = \frac{G_{l_Q/l_{Alb}}(y_{l_{Alb}}, \mu_l, \sigma_l^2)}{\sum_{l_{Alb}=1}^{10} G_{l_Q/l_{Alb}}(y_{l_{Alb}}, \mu_l, \sigma_l^2)}.$$  \hspace{1cm} (12)

In the M step, the genotypic values of QTL and residual variance are estimated by using

$$\mu_l = \frac{1}{n} \sum_{l_{Alb}=1}^{10} \sum_{y_{l_{Alb}}} \sum_{l_{Q}} P_{l_Q/l_{Alb}}(y_{l_{Alb}}, \mu_l, \sigma_l^2)$$  \hspace{1cm} (13)

$$\sigma_l^2 = \frac{1}{n} \sum_{l_{Alb}=1}^{10} \sum_{y_{l_{Alb}}} \sum_{l_{Q}} P_{l_Q/l_{Alb}}(y_{l_{Alb}}, \mu_l, \sigma_l^2)$$  \hspace{1cm} (14)

Different from the traditional EM algorithm for QTL mapping, we here need an additional step to estimate $g_v$ values that satisfy the linear constraints described by Equations (2) and (8–10). By assuming a fixed position of QTL, we obtain the values of $r_1$ and $r_2$ to generate $g_v$ under constraints (8–10), from which to calculate $\pi_1-\pi_9$ using Equation (2) and then $G_{l_Q/l_{Alb}}$. This procedure is repeated in the E and M steps until the MLEs of $\Omega$ are obtained.

**Hypothesis testing**

The first hypothesis is to test whether a QTL exists to affect the trait in the autotetraploid population. This can be done by testing the hypotheses:

$$H_0: \mu_l = \mu \cup I_Q = 1, \ldots, 10$$

$$H_1: \text{at least one of the equalities above does not hold}$$

where the reduced model $H_0$ assumes no QTLs, whereas the full model $H_1$ is associated with a QTL. The log-likelihood ratio of the full over-reduced model is calculated and compared against the critical threshold for asserting the existence of a QTL determined from permutation tests [34].

After a significant QTL is detected, the next step is to test its mode of action. The additive genetic effect of the QTL is tested using the null hypothesis:

$$H_0: a_1 = a_2 = a_3 = 0$$ \hspace{1cm} (15)

under which the parameter MLEs can be obtained with the algorithm as described above by shrinking
the number of QTL genotypic values with $\mu_{11} = \mu_{22} = \mu_{33} = \mu_{44}$. Similarly, we can also test the dominant genetic effect by the null hypothesis,

$$H_0: d_{12} = d_{13} = d_{14} = d_{23} = d_{24} = d_{34} = 0$$

(16)

under which only four genotypic values need to be estimated because all the genotypic values can be uniquely expressed in terms of $\mu_{11}$, $\mu_{22}$, $\mu_{33}$ and $\mu_{44}$.

The occurrence of double reduction is a product of selection and evolution in species [31,35,36]. Thus, it is interesting to test its significance. This can be done by testing:

$$H_0: \gamma = 0$$

(17)

under which parameter estimates are obtained under the restraints of Equation (7). The significance of double reduction at the markers can also be tested, although this has been done in linkage analysis [24,25,30].

**Partially informative markers**

In tetraploid linkage analysis, it is common to use partially informative markers, i.e. those whose gamete genotypes may be derived from multiple gamete configurations. Because of this, the amount of information for linkage analysis is reduced [30], but it is still feasible to map QTLs using these markers. For partially informative markers, the matrix of marker genotypes $M$ will be shrunk by collapsing the corresponding gamete configurations into the same cell of gamete genotype. Lu et al. [30] provided a detailed procedure for collapsing the gamete configurations and using the collapsed genotypes to estimate the recombination fractions. This procedure can be used to derive the algorithm for QTL mapping.

**Monte Carlo simulation**

We performed computer simulation to examine the statistical properties of the QTL mapping method for a polysomic tetraploid population. Assuming a QTL at the middle of a marker interval with a genetic length of 20 cM, marker and QTL genotypes were simulated under a set of double reduction coefficients at the two markers and QTL. The phenotypic values of mapping individuals were assumed to be normally distributed, simulated for different QTL heritabilities ($H^2 = 0.05$, $0.1$, $0.2$). Different population sizes were considered ($n = 200$, $400$, $800$).

Table 3 gives the means of the MLEs of the parameters and their standard errors based on 1000 simulation replicates. It is seen that the QTL position and double reduction coefficients at the two markers can accurately be estimated even with a small sample size (200) and a low heritability (0.05). The estimation of the coefficient of double reduction at the QTL strongly relies on sample size and, also, on heritability level, but to a lesser extent. When a larger sample size, i.e. 400, is used, it can be accurately estimated even with a low heritability (<0.1). When the sample size is small (200), the precise estimation of the QTL double reduction needs a high heritability (0.2). Generally, a sample size of 400 can guarantee the precision for estimating the double reduction of a QTL that explains >10% of the phenotypic variation.

It is not surprising that the precision of QTL effect estimation is affected largely by the heritability and sample size. As the heritability or the sample size increases, the QTL genotypic values can be more accurately estimated, which led to the improvement of estimating the additive and dominant effects. The genotypic values of QTL genotypes derived from chromosome pairing ($\mu_{12}$, $\mu_{13}$, $\mu_{14}$, $\mu_{23}$, $\mu_{24}$, $\mu_{34}$) can be reasonably estimated even with a modest sample size and heritability. However, estimating the genotypic values of QTL genotypes generated due to double reduction ($\mu_{11}$, $\mu_{22}$, $\mu_{33}$, $\mu_{44}$) is influenced heavily by the sample size when the heritability is small, which directly affects the estimation precision of the additive and dominant effects.

The model was further tested for its properties to map a QTL using partially informative markers. Assume that a polysomic tetraploid parent has an allelic configuration

\[
\begin{array}{cccc}
A_1 & A_2 & A_3 & A_3 \\
Q_1 & Q_2 & Q_3 & Q_4 \\
B_1 & B_2 & B_2 & B_2 \\
\end{array}
\]

at two partially informative markers A and B and a fully informative QTL. For these two markers, the dimension of matrix $M$ reduces from $10 \times 10$ to $6 \times 3$. This reduced matrix is used to specify the conditional probabilities of QTL genotypes given the marker genotypes. Using the same genetic values as given in Table 3, we simulated a new mapping population for these two markers. Table 4 provides the results of parameter estimation from partially informative markers in a polysomic tetraploid population. In general, the estimation precision of all parameters reduces when markers are partially informative, but the estimation can still achieve sufficient precision with a large sample size and for a high heritability.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>$H^2 = 0.05$</th>
<th>$H^2 = 0.1$</th>
<th>$H^2 = 0.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
</tr>
</tbody>
</table>

Table 3: The MLEs of double reduction coefficients, QTL location and QTL effects and the standard errors of these estimates based on fully informative markers by a polysomic QTL model in an autotetraploid population from 1000 simulation replicates under different heritabilities ($H^2$) and sample sizes (n).
Table 4: The MLEs of double reduction coefficients, QTL location and QTL effects and the standard errors of these estimates based on partially informative markers by a polysomic QTL model in an autotetraploid population from 1000 simulation replicates under different heritabilities ($H^2$) and sample sizes ($n$).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>$H^2 = 0.05$</th>
<th>$H^2 = 0.1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 800$</td>
<td>$n = 200$</td>
</tr>
<tr>
<td>Position</td>
<td>32</td>
<td>32.419 (0.023)</td>
<td>32.181 (0.7307)</td>
</tr>
<tr>
<td>$x$</td>
<td>0.2</td>
<td>0.198 (0.033)</td>
<td>0.200 (0.024)</td>
</tr>
<tr>
<td>$y$</td>
<td>0.2</td>
<td>0.195 (0.037)</td>
<td>0.197 (0.025)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.2</td>
<td>0.208 (0.047)</td>
<td>0.208 (0.043)</td>
</tr>
<tr>
<td>$\beta_{0}$</td>
<td>1.4</td>
<td>1.477 (1.939)</td>
<td>1.539 (2.226)</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>1.5</td>
<td>1.453 (2.201)</td>
<td>1.500 (2.150)</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>1.6</td>
<td>1.880 (1.922)</td>
<td>1.596 (1.314)</td>
</tr>
<tr>
<td>$\beta_{14}$</td>
<td>0.5</td>
<td>0.544 (2.317)</td>
<td>0.796 (1.387)</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>2.2</td>
<td>2.306 (0.910)</td>
<td>2.256 (0.671)</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>2.4</td>
<td>2.327 (1.016)</td>
<td>2.467 (0.692)</td>
</tr>
<tr>
<td>$\beta_{14}$</td>
<td>0.4</td>
<td>0.396 (0.955)</td>
<td>0.406 (0.628)</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>2.5</td>
<td>2.538 (1.043)</td>
<td>2.999 (0.664)</td>
</tr>
<tr>
<td>$\beta_{34}$</td>
<td>0.5</td>
<td>0.499 (0.887)</td>
<td>0.467 (0.641)</td>
</tr>
<tr>
<td>$\beta_{34}$</td>
<td>0.6</td>
<td>0.659 (0.981)</td>
<td>0.526 (0.647)</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>17.364</td>
<td>16.394 (2.066)</td>
<td>17.133 (1.358)</td>
</tr>
</tbody>
</table>
It has been found that the statistical power for QTL detection is sensitive to sample size and heritability (Table 5). If the trait is highly heritable, a small sample size can produce good power. If the trait is thought to have a low heritability, a large sample size is required. In general, for a heritability of 0.05, 800 samples are needed. When the heritability increases to 0.2, only 200 samples are adequate.

### Application

Okada et al. [17] reported a linkage map for tetraploid switchgrass (*Panicum virgatum*) with a full-sib family of 238 progenies 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 can be single-dose amplicons (simplex), double-dose amplicons (duplex) or triple-dose amplicons (triplex). Our previous work for tetraploid linkage analysis has determined optimal numbers of doses for each marker and the linkage phase of a pair of adjacent markers [29,30].

In this study, we aim to identify the possible existence of significant QTLs for whole-plant fresh weight in this population. Because only dominant markers were used, we assume that a putative QTL has only two alleles Q1 and Q2, forming three different genotypes Q1Q1, Q1Q2 and Q2Q2. As a demonstration of the mapping model, we only scanned part of the linkage group (Figure 1). In scanning the first four linkage groups of the female parent (K5) map, several adjacent peaks of the log-likelihood ratio (LR) curve were detected on group 4, at which the LR values are

### Table 5: The power of QTL detection using different types of markers

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Heritability</th>
<th>0.05</th>
<th>0.1</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fully informative markers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.13</td>
<td>0.46</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.51</td>
<td>0.78</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>800</td>
<td>0.85</td>
<td>0.91</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Partially informative markers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>0.48</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.48</td>
<td>0.79</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>800</td>
<td>0.81</td>
<td>0.94</td>
<td>0.94</td>
<td></td>
</tr>
</tbody>
</table>
beyond the critical threshold determined from permutation tests. Because these peaks are so close, only one single QTL may be considered in this area. This QTL is bracketed by two markers sww389_247 and nfsg025_121 (Table 6). The linkage phase of the heterozygous parent at these markers and QTL was found to be

\[
\begin{align*}
A & | O & O & O & O \\
Q_1 & | Q_2 & Q_2 & Q_2 & Q_2.
\end{align*}
\]

An optimal linkage phase is determined by calculating the information criteria (Akaike information criterion) for all possible phases and then choosing one corresponding to the minimum Akaike information criterion value. There was a large rate of occurrence (0.76) of double reduction at the QTL, implicating an influence of natural selection on this QTL for plant weight. This QTL was found to act in a partial dominant manner.

DISCUSSION

In this article, we described a statistical method for genetic mapping of QTLs in a full-sib family derived from polysomic autotetraploids undergoing double reduction. Double reduction is a ubiquitous phenomenon during the meiosis of autotetraploids and has been thought to play a pivotal role in plant evolution and maintenance of genetic polymorphism in natural populations [5,31]. Therefore, the estimation of double reduction, its occurrence, chromosomal distribution and association with a quantitative trait can shed light on the evolution and speciation of autopolyploids [35,36]. Given its influence on the segregation pattern of gametes in meiosis [24,25], double reduction should be incorporated into a QTL mapping framework.

When autotetraploids undergo double reduction, traditional mapping approaches fail to handle the pattern of marker-QTL segregation and cannot be used for their QTL identification. The method described capitalizes on 107 different classifications of three-locus gamete formations, derived by Fisher [3], and can provide the simultaneous estimation of frequencies of double reduction and the recombination fraction between different loci. Although statistical development for QTL mapping in polyploids is not new [21,23], this method is among the first that integrates Fisher’s tetrasomic inheritance into the mapping framework and, thus, facilitates the cytological relevance of QTL detection. Results from simulation studies showed that the method is applicable to map QTLs in a controlled cross of polysomic autotetraploids when the mapping population is adequately large (say 400). The method can estimate the double reduction of a QTL using fully informative, partially informative or dominant markers. The method has been validated by analyzing real data in tetraploid switchgrass, leading to the identification of a significant QTL for plant weight.

This article provides a comprehensive framework to map polysomic QTLs using any kind of molecular markers in an autotetraploid population derived from two heterozygous parents. The framework focuses on the inheritance of testcross QTLs to describe the analytical method clearly. It is, however, crucial to take into account intercross QTLs that are heterozygous for both parents. Assume an intercross QTL that have four alleles \(Q_1-\) \(Q_4\) for one parent and four alleles \(Q_5-\) \(Q_8\) for the second. Together, this QTL forms \(10 \times 10 = 100\) tetraploid genotypes whose values can be partitioned into the overall mean, six allelic effects and 93 diallelic, triallelic and tetraallelic interaction effects. To accurately estimate these parameters, a considerably large sample size is expected, although a precise power calculation needs further investigation. Based on Wu et al.’s [28] simulation studies, at least 800 individuals are needed to estimate six allelic effects, 13 diallelic interaction effects, 12 triallelic interaction effects and 4 tetraallelic interaction effects for an intercross QTL in a full-sib family of two allotetraploid lines.

The method is based on interval mapping, an approach that localizes a QTL with two flanking

<table>
<thead>
<tr>
<th>Table 6: QTL position and effect on fresh weight in switchgrass data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>
markers and has proven to be useful for QTL mapping [33]. A more powerful approach based on genome-wide association studies, through incorporating double reduction, should hold potential to reveal the genetic architecture of autopolyploids. This approach includes thousands of single nucleotide polymorphisms on a much smaller number of samples. Variable selection approaches that handle high-dimensional data analysis have been implemented for genome-wide association studies [38] and should play a key role in genetic dissection of complex traits in autotetraploids when the idea presented in this article is integrated. It is extremely important to characterize the contributions of genetic interactions and genotype–environment interactions to quantitative variation in an autotetraploid population [39]. Although the idea for considering these interactions is straightforward, their estimation requires sophisticated computing algorithms and capacity. In developing a method to map the genes involved in autotetraploid populations, we open a window to understanding the possible genetic mechanisms that characterize natural and artificial variation in polyploids.

**Key Points**

- The complexity of genetic mapping in polyploid autopolyploids is derived from their process of double reduction, which drives sister chromatids to migrate into the same gametes.
- We describe a statistical model for QTL mapping in these species by incorporating double reduction, allowing different genetic configurations to be separated from genotype data.
- The model is examined and validated through computer simulation and real data analysis, providing a useful tool for QTL mapping in polyploid autopolyploids.

**Acknowledgements**

This work is supported by Changjiang Scholars Award and ‘Thousand-person Plan’ Award.

**References**