A Bivalent Polyploid Model for Mapping Quantitative Trait Loci in Outcrossing Tetraploids

Rongling Wu,*,†,[1] Chang-Xing Ma* and George Casella*

**Department of Statistics, University of Florida, Gainesville, Florida 32611 and †Institute of Statistical Genetics, Zhejiang Forestry University, Lin'an, Zhejiang 311300, People's Republic of China*

Manuscript received November 5, 2002
Accepted for publication September 28, 2003

ABSTRACT

Two major aspects have made the genetic and genomic study of polyploids extremely difficult. First, increased allelic or nonallelic combinations due to multiple alleles result in complex gene actions and interactions for quantitative trait loci (QTL) in polyploids. Second, meiotic configurations in polyploids undergo a complex biological process including either bivalent or multivalent formation, or both. For bivalent polyploids, different degrees of preferential chromosome pairings may occur during meiosis. In this article, we develop a maximum-likelihood-based model for mapping QTL in tetraploids by considering the quantitative inheritance and meiotic mechanism of bivalent polyploids. This bivalent polyploid model is implemented with the EM algorithm to simultaneously estimate QTL position, QTL effects, and QTL-marker linkage phases by incorporating the impact of a cytological parameter determining bivalent chromosome pairings (the preferential pairing factor). Simulation studies are performed to investigate the performance and robustness of our statistical method for parameter estimation. The implication and extension of the bivalent polyploid model are discussed.

POLYPLOIDS represent a group of plant species that are of great importance to evolutionary studies and plant breeding (Zeven 1979; Bever and Felber 1992; Hilu 1993; Ramsey and Schemske 1998; Ott and Whitton 2000; Soltis and Soltis 2000). The genetic study of polyploids intrigued earlier pioneering geneticists (Haldane 1930; Mather 1935, 1936; Fisher 1947), who developed a series of theoretical models to study segregation and linkage in autotetraploids. Unfortunately, these seminal models have been limited in practical analysis, partly due to the fact that genetic information needed in the models could not be obtained with ease. Currently, the advent of molecular marker technologies has led to a resurgence of interest in the genetic analysis of polyploids (Leitch and Bennett 1997). Much theoretical and empirical emphasis has been made on marker inheritance and segregation and the construction of a genetic linkage map in polyploids (Wu *et al.* 1992; Da Silva *et al.* 1995; Grivet *et al.* 1996; Hackett *et al.* 1998; Ming *et al.* 1998; Brouwer and Osborn 1999; Ripol *et al.* 1999; Fjellstrom *et al.* 2001; Hoarau *et al.* 2001; Luo *et al.* 2001; Rajapakse *et al.* 2001; R. Wu *et al.* 2001, 2002a; S. Wu *et al.* 2001).

A significant gap that still remains in the current genetic study of polyploids is a serious lack of powerful statistical methods for mapping quantitative trait loci (QTL) on the basis of the genetic map of polymorphic markers. We know of only three articles that deal with the development of QTL-mapping methodologies (Doerge and Craig 2000; Xie and Xu 2000; Hackett *et al.* 2001). Considering the availability of marker and phenotype data in a variety of polyploid species ranging from tetraploids to octoploids (Yu and Pauls 1993; Grivet *et al.* 1996; Meyer *et al.* 1998; Ming *et al.* 1998; Brouwer and Osborn 1999; Fjellstrom *et al.* 2001; Hoarau *et al.* 2001; Rajapakse *et al.* 2001), this is a small number. One of these three articles did not use the appropriate biological process of meiosis in polyploids and its application is thus questionable (as noted by Hackett 2001). The other two articles were also based on limiting assumptions. Doerge and Craig assumed a completely preferential chromosome pairing mechanism for meiotic configurations and, therefore, their method can be appropriate only for extreme allopolyploids, in which chromosome pairings occur strictly between homologs. On the other hand, Hackett *et al.* treated bivalent pairings as a random event that occurs only when all chromosomes in the set are homologous (extreme autopolyploids). From a quantitative genetic perspective, none of the three articles have provided adequate estimations of allelic effects and dominant effects of different within-locus interaction levels for a putative QTL in polyploids. A major contribution of Hackett *et al.* (2001) is the implementation of Kempthorne's (1957) partitioning theory within a QTL-mapping framework to estimate additive and dominant effects of genes in polyploids. However, they did not explicitly show how the dominance effects were estimated from their model.

In this article, we have developed a new maximum-likelihood-based statistical infrastructure for mapping

[1]*Corresponding author:* Department of Statistics, 533 McCarty Hall C, University of Florida, Gainesville, FL 32611. E-mail: rwu@stat.ufl.edu

QTL in polyploids undergoing bivalent formation during meiosis. Beyond the existing statistical methods, our method integrates quantitative genetic knowledge about gene action and interaction and cytological mechanisms of chromosome pairing to gain better insights into the structure, organization, and function of polyploid genomes. It is observed that for many polyploids there is a higher probability of pairing between more similar chromosomes than between less similar chromosomes (Hickok 1978; Sybenga 1988, 1994, 1995; Allendorf and Danzmann 1997). By implementing powerful expectation-maximization (EM) algorithms, our method can provide simultaneous estimation of QTL position, QTL effects, linkage phase configuration, and cytological parameters. Moreover, results from our method will have potential implications for understanding the genetic architecture of a complex trait and evolutionary relatedness in polyploids. We present extensive simulation studies to investigate the statistical properties of our method built upon bivalent chromosome pairings.

MATHEMATICAL MODEL FOR LINKAGE ANALYSIS

Meiotic pairing: Consider a bivalent tetraploid, in which there are four sets of chromosomes. If chromosomes 1 and 2 are genetically more identical, as are chromosomes 3 and 4, there are three different combinations for the bivalent chromosome pairing. One of the three pairs is between more identical chromosomes 1 and 2 as well as 3 and 4 ($\mathcal{B}_1$) and the other two are between less identical chromosomes 1 and 3 as well as 2 and 4 ($\mathcal{B}_2$) or 1 and 4 as well as 2 and 3 ($\mathcal{B}_3$). In general, the probability of pairing between more identical chromosomes is higher than that between less identical chromosomes due to different evolutionary relatedness of chromosomes (Sybenga 1992, 1994, 1995) and such a difference is defined as the *preferential pairing factor*, denoted by p, that is bounded by $[0, \frac{2}{3}]$ (Sybenga 1988). Thus, the frequencies of the three bivalent pairings are expressed as $\frac{1}{3} + p$ for $\mathcal{B}_1$, $\frac{1}{3} - \frac{1}{2}p$ for $\mathcal{B}_2$, and $\frac{1}{3} - \frac{1}{2}p$ for $\mathcal{B}_3$. When $p = 0$, the four chromosomes in one group pair completely randomly. Extreme autopolyploids follow this pattern. When $p = \frac{2}{3}$, chromosome pairing occurs only between homologous ones and never occurs between homeologous ones. This pattern is characterized by extreme allopolyploids. Most polyploids are intermediate between these two extremes. Some polyploids that were originally classified as autotetraploids are found to belong to the intermediate types with $0 \le p \le \frac{2}{3}$ (reviewed in Sybenga 1996).

Tetraploid model for three-point linkage analysis: Linkage analysis in most diploid organisms is based on inbred line crosses, such as a backcross or F_2. However, for many other species including polyploids, inbred lines are not available and, thus, their linkage analysis should be based on a full-sib family derived from outbred parental lines. In such a full-sib family, numerous cross types of genes can be possible. To simplify our description of linkage analysis in polyploids, we first consider fully informative markers between the two parents. Our mapping model can be readily generalized to consider arbitrary polyploid cross types composed of any type of partially informative markers.

Suppose there is a full-sib family of size n derived from two heterozygous tetraploid parents P and Q. Consider two fully informative markers $\mathcal{M}^{\tau}$ and $\mathcal{M}^{\tau+1}$, which each have eight different alleles assigned to the four chromosomes of parents P and Q, respectively. The four alleles at marker $\mathcal{M}^{\tau}$ are labeled by M_1^{τ}, M_2^{τ}, M_3^{τ}, and M_4^{τ} for parent P and by N_1^{τ}, N_2^{τ}, N_3^{τ}, and N_4^{τ} for parent Q, and the four alleles at marker $\mathcal{M}^{\tau+1}$ are labeled by $M_1^{\tau+1}$, $M_2^{\tau+1}$, $M_3^{\tau+1}$, and $M_4^{\tau+1}$ for parent P and by $N_1^{\tau+1}$, $N_2^{\tau+1}$, $N_3^{\tau+1}$, and $N_4^{\tau+1}$, for parent Q. Between these two markers there is a putative QTL $\mathcal{Q}$ whose alleles are denoted by P_1, P_2, P_3, and P_4 for parent P and Q_1, Q_2, Q_3, and Q_4 for parent Q. The recombination fractions between marker $\mathcal{M}^{\tau}$ and QTL $\mathcal{Q}$, QTL $\mathcal{Q}$ and marker $\mathcal{M}^{\tau+1}$, and the two markers are denoted by θ_1, θ_2, and θ, respectively. For parents P and Q, these three loci (two markers and one QTL) have a total of $576 \times 576 = 331{,}776$ possible nonallelic configuration or linkage phase combinations, one of which can be schematically expressed as

$$\begin{array}{|c|c|c|c|} M_1^{\tau} & M_2^{\tau} & M_3^{\tau} & M_4^{\tau} \\ P_1 & P_2 & P_3 & P_4 \\ M_1^{\tau+1} & M_2^{\tau+1} & M_3^{\tau+1} & M_4^{\tau+1} \end{array} \otimes \begin{array}{|c|c|c|c|} N_1^{\tau} & N_2^{\tau} & N_3^{\tau} & N_4^{\tau} \\ Q_1 & Q_2 & Q_3 & Q_4 \\ N_1^{\tau+1} & N_2^{\tau+1} & N_3^{\tau+1} & N_4^{\tau+1} \end{array}, \quad (1)$$

where lines indicate the individual chromosomes on which the QTL is bracketed by the two markers and $\otimes$ is the Kronecker product. The specific linkage phase combination of parents P and Q, which is not known *a priori*, must be inferred from these possibilities for correct QTL mapping on the basis of marker and phenotype observations. In general, the linkage phase of the two markers is known before QTL mapping. Thus, we need to determine only the most likely linkage phase combination from $24 \times 24 = 576$ possibilities of the QTL relative to its two flanking markers.

Apart from the effect of different linkage phases on gamete formation frequencies, as a case in diploid organisms (Wu *et al.* 2002b), different chromosome pairings ($\mathcal{B}_1$, $\mathcal{B}_2$, and $\mathcal{B}_3$) in bivalent polyploids also affect the patterns of gene segregation and, thus, gamete frequencies. However, these two factors have different influences. For a particular parent, there can be only one linkage phase, whereas different bivalent pairings may occur simultaneously with different frequencies. Hence, overall frequencies of gametes from three possible bivalent pairings should be expressed in terms of the preferential pairing factor p for parents P and Q (Wu *et al.*

2002a). Given the linkage phase of display (1), three possible bivalent pairings and their frequencies are expressed as

$$\begin{bmatrix} M_1^{\tau} & M_2^{\tau} \\ P_1 & P_2 \\ M_1^{\tau+1} & M_2^{\tau+1} \end{bmatrix} \begin{bmatrix} M_3^{\tau} & M_4^{\tau} \\ P_3 & P_4 \\ M_3^{\tau+1} & M_4^{\tau+1} \end{bmatrix} f(\mathcal{B}_1) = \tfrac{1}{3} + p$$
$$\begin{bmatrix} M_1^{\tau} & M_3^{\tau} \\ P_1 & P_3 \\ M_1^{\tau+1} & M_3^{\tau+1} \end{bmatrix} \begin{bmatrix} M_2^{\tau} & M_4^{\tau} \\ P_2 & P_4 \\ M_2^{\tau+1} & M_4^{\tau+1} \end{bmatrix} f(\mathcal{B}_2) = \tfrac{1}{3} - \tfrac{1}{2}p \otimes$$
$$\begin{bmatrix} M_1^{\tau} & M_4^{\tau} \\ P_1 & P_4 \\ M_1^{\tau+1} & M_4^{\tau+1} \end{bmatrix} \begin{bmatrix} M_2^{\tau} & M_3^{\tau} \\ P_2 & P_3 \\ M_2^{\tau+1} & M_3^{\tau+1} \end{bmatrix} f(\mathcal{B}_3) = \tfrac{1}{3} - \tfrac{1}{2}p$$
$$\begin{bmatrix} N_1^{\tau} & N_2^{\tau} \\ Q_1 & Q_2 \\ N_1^{\tau+1} & N_2^{\tau+1} \end{bmatrix} \begin{bmatrix} N_3^{\tau} & N_4^{\tau} \\ Q_3 & Q_4 \\ N_3^{\tau+1} & N_4^{\tau+1} \end{bmatrix} f(\mathcal{B}_1) = \tfrac{1}{3} + p$$
$$\begin{bmatrix} N_1^{\tau} & N_3^{\tau} \\ Q_1 & Q_3 \\ N_1^{\tau+1} & N_3^{\tau+1} \end{bmatrix} \begin{bmatrix} N_2^{\tau} & N_4^{\tau} \\ Q_2 & Q_4 \\ N_2^{\tau+1} & N_4^{\tau+1} \end{bmatrix} f(\mathcal{B}_2) = \tfrac{1}{3} - \tfrac{1}{2}p,$$
$$\begin{bmatrix} N_1^{\tau} & N_4^{\tau} \\ Q_1 & Q_4 \\ N_1^{\tau+1} & N_4^{\tau+1} \end{bmatrix} \begin{bmatrix} N_2^{\tau} & N_3^{\tau} \\ Q_2 & Q_3 \\ N_2^{\tau+1} & N_3^{\tau+1} \end{bmatrix} f(\mathcal{B}_3) = \tfrac{1}{3} - \tfrac{1}{2}p \qquad (2)$$

where double lines are used to distinguish the two sets of paired chromosomes. For one parent, each of these three different bivalent pairings produces four diploid gamete types at a single locus. When the gametes are mixed from these pairings, a total of six gamete types will be produced for a locus. Thus, under bivalent pairings, parent P generates 36 diploid gametes at the two markers, whose genotypes are arrayed by

$$\mathbf{G}_{\mathcal{M}}^{\mathrm{P}} = (M_{k_1}^{\tau} M_{k_2}^{\tau})_{6\times 1} (M_{r_1}^{\tau+1} M_{r_2}^{\tau+1})_{1\times 6}$$
$$= (M_{k_1}^{\tau} M_{k_2}^{\tau} M_{r_1}^{\tau+1} M_{r_2}^{\tau+1})_{6\times 6}, \quad 1 \le k_1 < k_2 \le 4,\ 1 \le r_1 < r_2 \le 4.$$

The probabilities of these marker gametes, $\mathbf{p}_{\mathcal{M}}^{\mathrm{P}} = (\mathrm{Prob}(M_{k_1}^{\tau} M_{k_2}^{\tau} M_{r_1}^{\tau+1} M_{r_2}^{\tau+1}))$, can be derived in terms of the preferential pairing factor and the recombination fraction between the two markers (Wu *et al.* 2002a; Table 1. Each of these 36 two-marker gametes corresponds to one of six possible QTL genotypes arrayed by $\mathbf{G}_{\mathcal{Q}}^{\mathrm{P}} = (P_{u_1} P_{u_2})_{1\times 6}$, $1 \le u_1 < u_2 \le 4$, which are produced in the same way as the generation of the marker gametes [expression (2)]. Table 1 lists the joint probabilities of the two-marker and one-QTL gamete genotypes, $\mathbf{p}_{\mathcal{M}\mathcal{Q}}^{\mathrm{P}} = (\mathrm{Prob}(M_{k_1}^{\tau} M_{k_2}^{\tau} P_{u_1} P_{u_2} M_{r_1}^{\tau+1} M_{r_2}^{\tau+1}))$, in parent P when three possible bivalent pairings occur at meiosis given a particular linkage phase of expression (1).

Similarly, for parent Q, we can write the array of the two-marker gamete genotypes, $\mathbf{G}_{\mathcal{M}}^{\mathrm{Q}} = ((N_{l_1}^{\tau} N_{l_2}^{\tau})_{6\times 1} (N_{s_1}^{\tau+1} \times N_{s_2}^{\tau+1})_{1\times 6}) = (N_{l_1}^{\tau} N_{l_2}^{\tau} N_{s_1}^{\tau+1} N_{s_2}^{\tau+1})_{6\times 6}$, $1 \le l_1 < l_2 \le 4$, $1 \le s_1 < s_2 \le 4$, and the array of one-QTL gamete genotypes, $\mathbf{G}_{\mathcal{Q}}^{\mathrm{Q}} = (Q_{v_1} Q_{v_2})_{1\times 6}$, $1 \le v_1 < v_2 \le 4$. The probabilities of two-marker gamete genotypes, $\mathbf{p}_{\mathcal{M}}^{\mathrm{Q}} = (\mathrm{Prob}(N_{l_1}^{\tau} N_{l_2}^{\tau} N_{s_1}^{\tau+1} \times N_{s_2}^{\tau+1}))$, and of joint marker and QTL gamete genotypes, $\mathbf{p}_{\mathcal{M}\mathcal{Q}}^{\mathrm{Q}} = (\mathrm{Prob}(N_{l_1}^{\tau} N_{l_2}^{\tau} Q_{v_1} Q_{v_2} N_{s_1}^{\tau+1} N_{s_2}^{\tau+1}))$, can also be written.

With the information of the two parents, we can express the arrays of zygote genotypes for the markers and the QTL, respectively, as

$$\mathbf{G}_{\mathcal{M}} = \mathbf{G}_{\mathcal{M}}^{\mathrm{P}} \otimes \mathbf{G}_{\mathcal{M}}^{\mathrm{Q}},$$
$$\mathbf{G}_{\mathcal{Q}} = \mathbf{G}_{\mathcal{Q}}^{\mathrm{P}} \otimes \mathbf{G}_{\mathcal{Q}}^{\mathrm{Q}},$$

and the probabilities of two-marker zygote genotypes and of joint marker and QTL zygote genotypes, respectively, as

$$\mathbf{p}_{\mathcal{M}} = \mathbf{p}_{\mathcal{M}}^{\mathrm{P}} \otimes \mathbf{p}_{\mathcal{M}}^{\mathrm{Q}}, \qquad (3)$$
$$\mathbf{p}_{\mathcal{M}\mathcal{Q}} = \mathbf{p}_{\mathcal{M}\mathcal{Q}}^{\mathrm{P}} \otimes \mathbf{p}_{\mathcal{M}\mathcal{Q}}^{\mathrm{Q}}. \qquad (4)$$

The conditional probabilities of the QTL zygote genotypes upon the marker zygote genotypes can be derived as

$$\mathbf{p}_{\mathcal{Q}|\mathcal{M}} = (\mathrm{Prob}(P_{u_1} P_{u_2} Q_{v_1} Q_{v_2} | M_{k_1}^{\tau} M_{k_2}^{\tau} N_{l_1}^{\tau} N_{l_2}^{\tau} M_{r_1}^{\tau+1} M_{r_2}^{\tau+1} N_{s_1}^{\tau+1} N_{s_2}^{\tau+1}))$$
$$= \mathbf{p}_{\mathcal{M}\mathcal{Q}} \varnothing \mathbf{p}_{\mathcal{M}}, \qquad (5)$$

which forms a (1296 × 36) matrix, where ∅ is the elementwise division of the two matrices. These conditional probabilities are used for QTL mapping as described in the next section.

STATISTICAL METHOD FOR QTL MAPPING

The mixture model: A fundamental statistical model for QTL mapping is the mixture model (Lander and Botstein 1989). In such a mixture model, each observation y is assumed to have arisen from one of n (n possibly unknown but finite) components, each component being modeled by a density from the parametric family f,

$$p(y|\pi, \phi, \eta) = \pi_1 f(y; \phi_1, \eta) + \ldots + \pi_n f(y; \phi_n, \eta), \qquad (6)$$

where $\pi = (\pi_1, \ldots, \pi_n)$ are the mixture proportions that are constrained to be nonnegative and sum to unity; $\phi = (\phi_1, \ldots, \phi_n)$ are the component-specific parameters, with ϕ_j being specific to component j; and η is a parameter that is common to all components.

For the mixture model used in genetic mapping, each component represents a class of QTL genotypes and, thus, the mixture model provides a framework by which observations may be clustered together into different classes of QTL genotypes. The mixture proportions represent the relative frequency of occurrence of each QTL genotype in the population. For a particular two-marker genotype, $M_{k_1}^{\tau} M_{k_2}^{\tau} N_{l_1}^{\tau} N_{l_2}^{\tau} M_{r_1}^{\tau+1} M_{r_2}^{\tau+1} N_{s_1}^{\tau+1} N_{s_2}^{\tau+1}$, the frequency of the QTL genotype $P_{u_1} P_{u_2} Q_{v_1} Q_{v_2}$ is the corresponding conditional probability described by Equation 5 and given in Table 1.

Linear model of a quantitative trait: The mixture components in the mixture model of Equation 6 follow a normal distribution, with the mean equal to the expected genotypic value ($\mu_{u_1 u_2 v_1 v_2}$) of a QTL genotype and the variance equal to the residual variance (σ^2) within the QTL genotype. The phenotype of a quantitative trait observed for individual i can be described by a linear model,

$$y_i = \mu + \sum_{u_1=1}^{3} \sum_{u_2=u_1+1}^{4} \sum_{v_1=1}^{3} \sum_{v_2=v_1+1}^{4} \mu_{u_1 u_2 v_1 v_2} X_{i(u_1 u_2 v_1 v_2)} + e_i, \qquad (7)$$

where $X_{i(u_1 u_2 v_1 v_2)}$ is the indicator variable defined as 1 if individual i has the QTL genotype $P_{u_1} P_{u_2} Q_{v_1} Q_{v_2}$ and 0 otherwise, and e_i is the residual effect, distributed as

TABLE 1

Joint probabilities of marker-QTL gamete genotypes for two fully informative markers $\mathcal{M}^{\tau}$ and $\mathcal{M}^{\tau+1}$ under an assumed linkage phase as given in expression (1) for parent P

Marker genotype	Frequency	QTL genotype					
		P_1P_2	P_1P_3	P_1P_4	P_2P_3	P_2P_4	P_3P_4
$\mathbf{M_1^{\tau}M_2^{\tau}M_1^{\tau+1}M_2^{\tau+1}}$	$\frac{1}{4}(1-r)^2(\frac{2}{3}-p)$	$\frac{1}{4}(1-r_1)^2(1-r_2)^2(\frac{2}{3}-p)$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1^2r_2^2(\frac{2}{3}-p)$
$M_1^{\tau}M_2^{\tau}M_1^{\tau+1}M_3^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$	0	0	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$M_1^{\tau}M_2^{\tau}M_1^{\tau+1}M_3^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$M_1^{\tau}M_2^{\tau}M_2^{\tau+1}M_3^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$M_1^{\tau}M_2^{\tau}M_2^{\tau+1}M_4^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$	0	0	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$\mathbf{M_1^{\tau}M_2^{\tau}M_3^{\tau+1}M_4^{\tau+1}}$	$\frac{1}{4}r^2(\frac{2}{3}-p)$	$\frac{1}{4}(1-r_1)^2r_2^2(\frac{2}{3}-p)$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1^2(1-r_2)^2(\frac{2}{3}-p)$
$M_1^{\tau}M_3^{\tau}M_1^{\tau+1}M_2^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	0	0	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$
$\mathbf{M_1^{\tau}M_3^{\tau}M_1^{\tau+1}M_3^{\tau+1}}$	$\frac{1}{4}(1-r)^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2(1-r_2)^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1^2r_2^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$M_1^{\tau}M_3^{\tau}M_1^{\tau+1}M_4^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}+p)$	0	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}+p)$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}+p)$	0
$M_1^{\tau}M_3^{\tau}M_2^{\tau+1}M_3^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}+p)$	0	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}+p)$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}+p)$	0
$\mathbf{M_1^{\tau}M_3^{\tau}M_2^{\tau+1}M_4^{\tau+1}}$	$\frac{1}{4}r^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2r_2^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1^2(1-r_2)^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$M_1^{\tau}M_3^{\tau}M_3^{\tau+1}M_4^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	0	0	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$
$M_1^{\tau}M_4^{\tau}M_1^{\tau+1}M_2^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$
$M_1^{\tau}M_4^{\tau}M_1^{\tau+1}M_3^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}+p)$	0	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}+p)$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}+p)$	0
$\mathbf{M_1^{\tau}M_4^{\tau}M_1^{\tau+1}M_4^{\tau+1}}$	$\frac{1}{4}(1-r)^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}(1-r_1)^2(1-r_2)^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1^2r_2^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$\mathbf{M_1^{\tau}M_4^{\tau}M_2^{\tau+1}M_3^{\tau+1}}$	$\frac{1}{4}r^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}(1-r_1)^2r_2^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1^2(1-r_2)^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$M_1^{\tau}M_4^{\tau}M_2^{\tau+1}M_4^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}+p)$	0	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}+p)$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}+p)$	0
$M_1^{\tau}M_4^{\tau}M_3^{\tau+1}M_4^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$

(*continued*)

TABLE 1

(Continued)

Marker genotype	Frequency	QTL genotype					
		P_1P_2	P_1P_3	P_1P_4	P_2P_3	P_2P_4	P_3P_4
$M_2^{\tau}M_3^{\tau}M_1^{\tau+1}M_2^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$
$M_2^{\tau}M_3^{\tau}M_1^{\tau+1}M_3^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}+p)$	0	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}+p)$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}+p)$	0
$\mathbf{M_2^{\tau}M_3^{\tau}M_1^{\tau+1}M_4^{\tau+1}}$	$\frac{1}{4}r^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)(1-r_2)r_2(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1^2(1-r_2)^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2r_2^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$\mathbf{M_2^{\tau}M_3^{\tau}M_2^{\tau+1}M_3^{\tau+1}}$	$\frac{1}{4}(1-r)^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1^2r_2^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2(1-r_2)^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$M_2^{\tau}M_3^{\tau}M_2^{\tau+1}M_4^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}+p)$	0	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}+p)$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}+p)$	0
$M_2^{\tau}M_3^{\tau}M_3^{\tau+1}M_4^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$
$M_2^{\tau}M_4^{\tau}M_1^{\tau+1}M_2^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	0	0	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$
$\mathbf{M_2^{\tau}M_4^{\tau}M_1^{\tau+1}M_3^{\tau+1}}$	$\frac{1}{4}r^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1^2(1-r_2)^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}(1-r_1)^2r_2^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$M_2^{\tau}M_4^{\tau}M_1^{\tau+1}M_4^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}+p)$	0	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}+p)$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}+p)$	0
$M_2^{\tau}M_4^{\tau}M_2^{\tau+1}M_3^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}+p)$	0	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}+p)$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}+p)$	0
$\mathbf{M_2^{\tau}M_4^{\tau}M_2^{\tau+1}M_4^{\tau+1}}$	$\frac{1}{4}(1-r)^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1^2r_2^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}+p)$	$\frac{1}{4}(1-r_1)^2(1-r_2)^2(\frac{2}{3}+\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$M_2^{\tau}M_4^{\tau}M_3^{\tau+1}M_4^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	0	0	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$
$\mathbf{M_3^{\tau}M_4^{\tau}M_1^{\tau+1}M_2^{\tau+1}}$	$\frac{1}{4}r^2(\frac{2}{3}-p)$	$\frac{1}{4}r_1^2(1-r_2)^2(\frac{2}{3}-p)$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2r_2^2(\frac{2}{3}-p)$
$M_3^{\tau}M_4^{\tau}M_1^{\tau+1}M_3^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$	0	0	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$M_3^{\tau}M_4^{\tau}M_1^{\tau+1}M_4^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$M_3^{\tau}M_4^{\tau}M_2^{\tau+1}M_3^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$	0	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$M_3^{\tau}M_4^{\tau}M_2^{\tau+1}M_4^{\tau+1}$	$\frac{1}{4}r(1-r)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2^2(\frac{1}{3}-\frac{p}{2})$	0	0	$\frac{1}{4}r_1(1-r_1)(1-r_2)^2(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$
$\mathbf{M_3^{\tau}M_4^{\tau}M_3^{\tau+1}M_4^{\tau+1}}$	$\frac{1}{4}(1-r)^2(\frac{2}{3}-p)$	$\frac{1}{4}r_1^2r_2^2(\frac{2}{3}-p)$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}r_1(1-r_1)r_2(1-r_2)(\frac{1}{3}-\frac{p}{2})$	$\frac{1}{4}(1-r_1)^2(1-r_2)^2(\frac{2}{3}-p)$

The conditional probabilities of the QTL gamete genotypes upon the marker gamete genotypes, as used in QTL mapping, are derived according to Bayes' theorem. No double crossovers between the two markers are assumed. The recombination fractions between the two markers, between marker M^{τ} and the QTL, and between the QTL and marker $M^{\tau+1}$, are denoted by r, r_1, and r_2, respectively. p is the preferential pairing factor describing the propensity of pairing between more similar rather than less similar chromosomes.

$N(0, \sigma^2)$. The genotypic value of QTL genotype $P_{u_1}P_{u_2}Q_{v_1}Q_{v_2}$ is partitioned into additive and dominant (interaction) effects of different orders:

$$\begin{aligned}\mu_{u_1u_2v_1v_2} &= \alpha^{P}_{u_1} + \alpha^{P}_{u_2} + \alpha^{Q}_{v_1} + \alpha^{Q}_{v_2} && \text{the main (additive) effects}\\ &+ \beta^{PP}_{u_1u_2} + \beta^{PQ}_{u_1v_1} + \beta^{PQ}_{u_1v_2} + \beta^{PQ}_{u_2v_1} + \beta^{PQ}_{u_2v_2} + \beta^{QQ}_{v_1v_2} && \text{the diallelic interactions}\\ &+ \gamma^{PPQ}_{u_1u_2v_1} + \gamma^{PPQ}_{u_1u_2v_2} + \gamma^{PQQ}_{u_1v_1v_2} + \gamma^{PQQ}_{u_2v_1v_2} && \text{the triallelic interactions}\\ &+ \delta_{u_1u_2v_1v_2} && \text{the tetraallelic interaction.}\end{aligned} \tag{8}$$

In a full-sib family, an individual will inherit two QTL alleles, $P_{u_1}P_{u_2}$, from parent P and two QTL alleles, $Q_{v_1}Q_{v_2}$, from parent Q. Because both parents P and Q have a total of eight different alleles, the above genetic model includes eight main effects, 28 diallelic interactions [6 due to two alleles from the same parent P ($\beta^{PP}_{u_1u_2}$), 6 from the same parent Q ($\beta^{QQ}_{v_1v_2}$), and 16 from different parents ($\beta^{PQ}_{u_1v_1}$, $\beta^{PQ}_{u_1v_2}$, $\beta^{PQ}_{u_2v_1}$, $\beta^{PQ}_{u_2v_2}$,)], 48 triallelic interactions [24 due to two alleles from parent P and one from parent Q ($\beta^{PPQ}_{u_1u_2v_1}$, $\beta^{PPQ}_{u_1u_2v_2}$) and 24 due to one allele from parent P and two from parent Q ($\beta^{PQQ}_{u_1v_1v_2}$, $\beta^{PQQ}_{u_2v_1v_2}$)], and 36 tetraallelic interactions.

Because some of the main and interaction effects are not independent, a parameterization process based on effect partitioning is needed to obtain a smaller number of estimable independent parameters (APPENDIX A). After this, estimable parameters include 6 for the main effects, 13 for the diallelic interactions (2 for interactions between alleles from parent P, 2 for parent Q, and 9 for interactions between alleles from different parents), 12 triallelic interactions, and 4 tetraallelic interactions (see also HACKETT *et al.* 2001). These 35 independent effect parameters, plus the overall mean, are denoted by the vector **a**.

We also used orthogonal polynomials to parameterize the main and interaction effects into linear contrasts, quadratic contrasts, and, if any, cubic contrasts (C.-X. MA and R. L. WU, unpublished results). Yet, we do not report the results from this parameterization approach here because of space limitation.

Computational algorithm: A maximum-likelihood approach is used to fit a single QTL affecting a quantitative trait in tetraploids. The likelihood of the phenotypes (**y**) for n offspring in a full-sib family of two outcrossing tetraploids is expressed as

$$L(\mathbf{y}|\mathbf{\Omega}) = \prod_{i=1}^{n}\left[\sum_{u_1=1}^{3}\sum_{u_2=u_1+1}^{4}\sum_{v_1=1}^{3}\sum_{v_2=v_1+1}^{4} p_{u_1u_2v_1v_2i} f_{u_1u_2v_1v_2}(y_i)\right], \tag{9}$$

where $\mathbf{\Omega} = (\mathbf{a}, r_1 \text{ or } r_2, \sigma^2, p)$ is the vector of unknown parameters containing the overall mean, QTL effects, QTL position, residual variance, and the preferential pairing factor; $p_{u_1u_2v_1v_2i}$ is the probability of progeny i to have QTL genotype $P_{u_1}P_{u_2}Q_{v_1}Q_{v_2}$, which is the probability of the QTL genotype conditional upon marker genotypes (Table 1) when the marker information is combined. Last, $f_{u_1u_2v_1v_2}(y_i)$ is a normal distribution density for QTL genotype $P_{u_1}P_{u_2}Q_{v_1}Q_{v_2}$, with the mean equal to the expected genotypic value from Equation 7 and the variance equal to the residual variance (σ^2) within this genotype.

As seen from above, the total number of QTL effects equals the number of the QTL genotypes in bivalent tetraploids. This permits us to estimate the overall mean and QTL effect parameters from the estimated values ($\hat{\mu}_{u_1u_2v_1v_2}$) of the QTL genotypes by solving a group of regular equations. From a computational perspective, it is more efficient to estimate the expected genotypic values ($\mu_{u_1u_2v_1v_2}$) from the mixture model of Equation 7 than to estimate the overall mean μ and QTL effect parameters that comprise vector **a**. We use the two parameterization approaches, as mentioned above, to estimate vector **a** from the 36 normal mixtures of QTL genotypes for bivalent tetraploids. The process of estimating $\hat{\mu}_{u_1u_2v_1v_2}$ and σ^2 on the basis of the EM algorithm is given in APPENDIX B. The maximum-likelihood estimates (MLEs) of r_1 or r_2 and p can be obtained using the grid approach because these two parameters each have a particular bound, $0 \le r_1$ or $r_2 \le 1$ and $0 \le p \le \frac{2}{3}$.

The characterization of linkage phase: Above, we have derived a statistical procedure for estimating the recombination fraction and the preferential pairing factor in polyploids when their chromosome pairings at meiosis follow the bivalent model. The procedure assumes the linkage phase combination of the two markers and QTL as indicated by display (1). However, this represents only one of the 576 possible combinations for the two phase-known flanking markers and the QTL. Optimal estimates of all parameters should be based on a most likely linkage phase combination. Different linkage phases of the QTL relative to its flanking markers can be assigned on the basis of the permutation of four QTL alleles on four different chromosomes for each parent. A most likely linkage phase combination should correspond to the largest likelihood value calculated from Equation 9.

However, a new question arises about the comparisons of the likelihood values among different phase combinations. If we change different linkage phases, we may obtain different estimates for a QTL effect parameter, but we will obtain the same likelihood value. We therefore should pose constraints on allelic effects of the two parents to obtain comparable likelihood values. In fact, the occurrence of a particular linkage phase implies that alleles should be different for both loci under consideration. A total of 576 phase combinations between the QTL and its flanking fully informative markers are based on the condition that four QTL alleles are different for each parent. The direct description of such differences can be provided by allelic effects. Thus, we can pose the inequality constraints of three allelic effects from each parent. Without loss of generality, such constraints can be taken as

$$\alpha_1^P > \alpha_2^P > \alpha_3^P > \alpha_4^P \quad (10)$$

for parent P and

$$\alpha_1^Q > \alpha_2^Q > \alpha_3^Q > \alpha_4^Q \quad (11)$$

for parent Q. Under these constraints, we will obtain different likelihood values from different linkage phase combinations and, therefore, it will be possible to select a most likely linkage phase combination.

Hypothesis tests: After the optimal estimates for the linkage and linkage phase are obtained on the basis of the largest likelihood value, we test for the significance of linkage by calculating the likelihood-ratio test (LRT) statistic,

$$\text{LRT} = -2 \log \left[\frac{L(\mathbf{y}|A_o, \tilde{\mu}, \mathbf{a} = 0, \tilde{r}_1, \tilde{p}, \tilde{\sigma}^2)}{L(\mathbf{y}|A_o, \hat{\mu}, \hat{\mathbf{a}}, \hat{r}_1, \hat{p}, \hat{\sigma}^2)} \right], \quad (12)$$

where A_o stands for the most likely linkage phase combination between the QTL and its flanking markers under which the likelihood value is highest, calculated from Equation 9 with the above-mentioned constraints. Here, $\hat{}$ and $\tilde{}$ stand for the MLEs for unknown parameters under the full model (at least one element in $\mathbf{a}$ is not equal to zero) and reduced model ($\mathbf{a} = 0$), respectively. By formulating similar reduced models, we can also test for the significance of additive effects or dominance effects at different interaction levels.

As in diploid mapping, simulation studies can be used to determine critical threshold values. We can declare the existence of a significant QTL located between two markers $\mathcal{M}^\tau$ and $\mathcal{M}^{\tau+1}$ if the LRT is greater than the critical threshold for an appropriate choice of the type I error rate α. Similarly, we can formulate a hypothesis for testing whether or not the preferential pairing factor is equal to zero (a set of four chromosomes are all homologous; the autopolyploid model) or $\frac{2}{3}$ (homeologous chromosomes do not pair; the allopolyploid model). Results from such a test are useful for examining the level of relatedness between different genomes.

RESULTS

Simulation studies are performed to examine the statistical behavior of our bivalent polyploid model. We first focus our simulation to quantify the effects of trait heritability and sample size on the estimation of QTL parameters and of the bivalent chromosome pairing parameter. Then, we compare the differences of parameter estimates between our method and Doerge and Craig's (2000) method, in which completely preferential bivalent chromosome pairings are assumed, and Hackett *et al.*'s (2001) method, in which random chromosome pairings are assumed.

Experimental design: Two outcrossing tetraploid parents are simulated for two fully informative markers and a QTL with an assumed linkage phase configuration shown in display (1). The recombination fractions between the two markers and between the first marker and the QTL are given as 0.20 and 0.10, respectively. The preferential pairing factor $p = 0.30$ is assumed. These two parents are crossed to generate a full-sib family of 200, 400, and 800 offspring. Given a sample size, the observations of each of $36 \times 36 = 1296$ offspring genotypes at these two markers are simulated on the basis of their respective frequencies (Equation 4).

The numbers of offspring within each marker genotype carrying each of 36 QTL genotypes are simulated on the basis of the conditional probability matrices of Equation 5. Because of the QTL effects, offspring with different QTL genotypes will be different for a quantitative trait. The genotypic values of the offspring carrying different QTL genotypes are calculated on the basis of their structures, as given in $\mathbf{D}^{-1}\mathbf{a}$ (appendix a), using the hypothesized values of the overall mean and 35 effects in the vector $\mathbf{a}$ (Table 2). The variance among these genotypic values is the genetic variance explained by this QTL. The phenotypic values of the offspring are calculated as an overall mean of 10, plus the genotypic values and the residual effects distributed as $N(0, \sigma^2)$. Different σ^2 values are assigned by assuming different heritability levels 0.20 and 0.40. The heritability is defined as the proportion of the genetic variance to the total phenotypic variance.

For the simulated marker and phenotypic data, we use the bivalent polyploid model to estimate unknown parameters contained in the vector $\mathbf{\Theta}$ and further obtain the MLEs of $\mathbf{\Omega}$ using a procedure described in appendix a. By permutating the arrangements of four QTL alleles among the four chromosomes for each parent, we obtain the MLEs of $\mathbf{\Omega}$ with the constraints, as given in displays (10) and (11), under a total of 576 linkage phase combinations. The phase combination that has the largest likelihood value is regarded as a most likely one, under which the MLEs of $\mathbf{\Omega}$ are given in Table 2. The simulations are repeated 100 times to calculate the means and standard errors of the MLEs from our model.

The effects of trait heritability and sample size: Using the computational algorithms described in appendix b, we obtain the MLEs of $\mathbf{\Theta}$. The recombination fraction between the first marker and the QTL can be accurately estimated for different sample sizes (n) and heritability (H^2) levels considered, although its estimation precision increases with sample sizes and heritability levels. The estimate of residual variance (σ^2) is considerably downward biased, especially for a trait with low heritability, if the sample size used is <400.

The real genotypic values of the 36 QTL genotypes are determined from $\mathbf{a} = \mathbf{D}^{-1}\mathbf{m}$ (see appendix a). The EM algorithm provides accurate estimates for these genotypic values, even when sample size or heritability is low (results not shown). If the genotypic values can be well estimated, the QTL gene effects ($\mathbf{a}$) can also be well estimated because, according to our parameterization, the sampling variances of $\hat{\mathbf{a}}$ will be reduced relative

TABLE 2

MLEs of allelic action and interaction effects for a QTL bracketed by two flanking markers $\mathcal{M}^{\tau}$ and $\mathcal{M}^{\tau+1}$ under different sample sizes (n) and heritability (H^2) levels

	$n = 200$		$n = 400$		$n = 800$	
Parameter	$H^2 = 0.2$	$H^2 = 0.4$	$H^2 = 0.2$	$H^2 = 0.4$	$H^2 = 0.2$	$H^2 = 0.4$
$r_1 = 0.10$	0.10(0.04)	0.10(0.03)	0.11(0.04)	0.10(0.02)	0.10(0.02)	0.10(0.01)
$\sigma^2 = 3.45/1.29$	2.37(1.16)	0.88(0.45)	2.97(0.53)	1.11(0.21)	3.20(0.32)	1.19(0.12)
$\alpha_1^{\mathrm{P}} = 0.95$	0.87(0.34)	0.87(0.20)	0.92(0.22)	0.92(0.13)	0.93(0.12)	0.93(0.09)
$\alpha_2^{\mathrm{P}} = 0.35$	0.34(0.29)	0.33(0.21)	0.36(0.18)	0.30(0.13)	0.31(0.13)	0.33(0.08)
$\alpha_3^{\mathrm{P}} = -0.40$	−0.30(0.32)	−0.35(0.18)	−0.39(0.23)	−0.35(0.13)	−0.37(0.13)	−0.35(0.09)
$\alpha_1^{\mathrm{Q}} = 0.84$	0.76(0.32)	0.77(0.21)	0.79(0.20)	0.79(0.12)	0.84(0.12)	0.85(0.07)
$\alpha_2^{\mathrm{Q}} = 0.31$	0.23(0.34)	0.26(0.21)	0.27(0.20)	0.29(0.15)	0.30(0.15)	0.32(0.09)
$\alpha_3^{\mathrm{Q}} = -0.29$	−0.18(0.39)	−0.24(0.20)	−0.27(0.19)	−0.28(0.12)	−0.29(0.11)	−0.31(0.08)
$\beta_{12}^{\mathrm{PP}} = -0.08$	−0.08(0.31)	−0.05(0.21)	−0.08(0.19)	−0.07(0.22)	−0.08(0.15)	−0.09(0.08)
$\beta_{13}^{\mathrm{PP}} = -0.13$	−0.14(0.26)	−0.11(0.18)	−0.11(0.16)	−0.15(0.15)	−0.12(0.12)	−0.12(0.07)
$\beta_{12}^{\mathrm{QQ}} = 0.20$	0.20(0.33)	0.20(0.21)	0.20(0.21)	0.22(0.18)	0.21(0.15)	0.19(0.08)
$\beta_{13}^{\mathrm{QQ}} = -0.00$	−0.00(0.28)	−0.02(0.16)	−0.02(0.17)	0.03(0.13)	−0.00(0.11)	−0.01(0.07)
$\beta_{11}^{\mathrm{PQ}} = 0.08$	0.15(0.58)	0.14(0.29)	0.14(0.36)	0.05(0.21)	0.09(0.18)	0.10(0.15)
$\beta_{12}^{\mathrm{PQ}} = -0.04$	0.04(0.54)	−0.00(0.29)	−0.08(0.33)	−0.05(0.21)	−0.07(0.22)	−0.01(0.15)
$\beta_{13}^{\mathrm{PQ}} = 0.08$	0.02(0.55)	−0.01(0.31)	0.13(0.35)	0.10(0.23)	0.16(0.26)	0.05(0.14)
$\beta_{21}^{\mathrm{PQ}} = 0.00$	−0.05(0.34)	0.02(0.29)	−0.03(0.30)	0.01(0.22)	−0.03(0.26)	0.00(0.14)
$\beta_{22}^{\mathrm{PQ}} = -0.02$	−0.02(0.39)	−0.03(0.31)	−0.14(0.39)	0.01(0.22)	−0.10(0.22)	−0.03(0.14)
$\beta_{23}^{\mathrm{PQ}} = -0.02$	−0.18(0.40)	−0.02(0.33)	0.08(0.36)	−0.02(0.21)	0.03(0.24)	−0.02(0.15)
$\beta_{31}^{\mathrm{PQ}} = 0.20$	0.18(0.46)	0.15(0.32)	0.16(0.35)	0.20(0.23)	0.18(0.20)	0.19(0.15)
$\beta_{32}^{\mathrm{PQ}} = 0.00$	0.00(0.53)	−0.04(0.33)	0.05(0.38)	−0.01(0.23)	0.07(0.23)	−0.01(0.14)
$\beta_{33}^{\mathrm{PQ}} = -0.10$	−0.04(0.51)	−0.02(0.36)	−0.15(0.35)	−0.12(0.23)	−0.15(0.23)	−0.09(0.16)
$\gamma_{121}^{\mathrm{PPQ}} = 0.18$	0.09(0.47)	0.04(0.38)	0.14(0.39)	0.07(0.25)	0.20(0.27)	0.18(0.14)
$\gamma_{122}^{\mathrm{PPQ}} = 0.16$	0.19(0.57)	0.07(0.32)	0.07(0.39)	0.10(0.25)	0.17(0.25)	0.16(0.16)
$\gamma_{123}^{\mathrm{PPQ}} = 0.04$	0.05(0.56)	0.14(0.33)	0.09(0.38)	0.12(0.26)	−0.01(0.26)	0.03(0.16)
$\gamma_{131}^{\mathrm{PPQ}} = -0.21$	−0.17(0.40)	−0.11(0.34)	−0.21(0.30)	−0.14(0.22)	−0.21(0.20)	−0.24(0.13)
$\gamma_{132}^{\mathrm{PPQ}} = -0.15$	−0.14(0.36)	−0.09(0.28)	−0.13(0.33)	−0.12(0.20)	−0.13(0.22)	−0.14(0.14)
$\gamma_{133}^{\mathrm{PPQ}} = -0.03$	−0.08(0.54)	−0.09(0.32)	−0.04(0.38)	−0.07(0.21)	−0.02(0.22)	−0.01(0.15)
$\gamma_{112}^{\mathrm{PQQ}} = -0.00$	0.05(0.37)	−0.11(0.35)	0.08(0.39)	−0.04(0.25)	−0.01(0.26)	−0.01(0.15)
$\gamma_{212}^{\mathrm{PQQ}} = -0.01$	−0.08(0.51)	−0.01(0.37)	0.08(0.38)	−0.08(0.23)	−0.04(0.29)	−0.01(0.15)
$\gamma_{312}^{\mathrm{PQQ}} = -0.03$	0.09(0.59)	−0.02(0.33)	−0.10(0.41)	0.06(0.24)	0.01(0.25)	−0.01(0.16)
$\gamma_{113}^{\mathrm{PQQ}} = -0.02$	−0.06(0.52)	0.04(0.31)	−0.05(0.29)	0.04(0.18)	0.01(0.20)	−0.00(0.14)
$\gamma_{213}^{\mathrm{PQQ}} = -0.01$	−0.09(0.48)	0.02(0.31)	−0.05(0.25)	0.04(0.20)	0.03(0.22)	0.01(0.13)
$\gamma_{313}^{\mathrm{PQQ}} = 0.00$	−0.04(0.47)	−0.05(0.35)	−0.01(0.35)	−0.05(0.16)	−0.04(0.21)	−0.01(0.13)
$\delta_{1212} = -0.09$	−0.07(0.48)	−0.06(0.38)	−0.04(0.33)	−0.07(0.35)	−0.06(0.30)	−0.10(0.16)
$\delta_{1213} = -0.01$	0.08(0.48)	−0.06(0.37)	−0.09(0.37)	0.01(0.25)	0.01(0.23)	−0.01(0.15)
$\delta_{1312} = 0.03$	0.05(0.48)	0.08(0.33)	−0.05(0.33)	0.02(0.22)	0.05(0.24)	0.03(0.12)
$\delta_{1313} = 0.03$	−0.08(0.43)	−0.00(0.27)	0.07(0.30)	0.03(0.18)	−0.01(0.17)	0.02(0.11)

The numbers given in the first column (parameter) are the hypothesized values for these QTL parameters. The symbols for QTL effects are given in the text and in APPENDIX A. The standard errors of the MLEs are estimated from 100 simulations and are given in parentheses.

to those of $\hat{\mathbf{m}}$ [see the structure of $\mathbf{D}^{-1}(\mathbf{D}^{-1})^{\mathrm{T}}$ in APPENDIX A]. It is shown that the estimators of additive effects of alleles for each parent have only one-sixteenth of the sampling variance of the estimated residual variances. The estimates of dominant effects vary depending upon the type and degree of interactions. If dominant effects are derived from the two alleles of one same parent, their estimators will be even more precise than those of the allelic effects. The estimators of dominant effects are derived from two alleles of different parents having the lowest precision, whose sampling variances are $\frac{9}{64}$ of those of the estimated residual variance. It is interesting to note that the estimators of tetraallelic effects have better precision than those of the diallelic dominance effect with two alleles from different parents. From the structure of $\mathbf{D}^{-1}(\mathbf{D}^{-1})^{\mathrm{T}}$, the estimators of different QTL effect parameters are basically independent. Their dependence occurs only within the QTL effects of the same type. The structure analysis of $\mathbf{D}^{-1}(\mathbf{D}^{-1})^{\mathrm{T}}$ suggests that the parameterization process of QTL effects will produce favorable effects on their estimates from the EM algorithm as described in APPENDIX B.

As expected, the allelic (or additive) effects can be estimated both more accurately and more precisely than

TABLE 3

The probabilities of detecting a correct linkage phase combination (Pr1), multiple linkage phase combinations including the correct one (Pr2), and an incorrect linkage phase combination (Pr3) from a total of 576 possible phase combinations between two phase-known fully informative markers and a putative QTL for two tetraploid plants

	$n = 200$		$n = 400$		$n = 800$	
	$H^2 = 0.2$	$H^2 = 0.4$	$H^2 = 0.2$	$H^2 = 0.4$	$H^2 = 0.2$	$H^2 = 0.4$
Pr1	30	49	39	48	46	53
Pr2	23	37	49	51	53	47
Pr3	47	14	12	1	1	0

the dominant effects, and the dominant effects of lower-order interactions can be estimated more precisely than the dominant effects of higher-order interactions (Table 2). It is interesting to note that the diallelic dominance effects between two alleles from the same parent can be estimated better than those between two alleles from different parents.

For all kinds of gene effects in bivalent tetraploids, the estimation accuracy and precision are increased when sample sizes and heritability levels are increased (Table 2). In general, a sample size of 200 can provide reasonably precise estimates of the allelic additive effects for a quantitative trait with a heritability of 0.20. But the estimation precision can be significantly improved if n is increased to 400 or for a quantitative trait with an increased H^2 level. There is not much improvement if n is further increased from 400 to 800, even for a less inheritable trait.

For the diallelic dominance effects between two alleles from the same parent, it seems that for a lower heritability (0.20) a sample size of at least 400 is needed to achieve reasonable estimation precision, whereas for a heritability of at least 0.40 a smaller sample size (200) may be adequate, compared to the magnitudes of the actual values of these effects that are hypothesized (Table 2). For the diallelic dominance effects between two alleles from different parents, reasonable estimates need a sample size of at least 400 for a trait with a heritability of at least 0.40. In general, it is difficult to estimate triallelic dominance effects unless a sample size is extremely large (say 800). To obtain reasonable estimates for tetraallelic effects, an extremely large sample size should accompany a highly inheritable quantitative trait (see Table 2).

The estimates of all parameters listed in Table 2 were based on an optimal linkage phase combination selected from all possibilities in terms of the estimated likelihood values. The probabilities of detecting a correct linkage phase combination were estimated for different sample sizes and heritability levels (Table 3). When $N = 200$ and $H^2 = 0.20$, we have only about one-third probability to detect a correct linkage phase combination. Other probabilities include about one-quarter to detect two linkage phase combinations and about one-half to detect an incorrect linkage phase combination. When a sample size or heritability is doubled, the probability of detecting an incorrect linkage phase combination is reduced. If a sample size of 400 is used for a trait of $H^2 = 0.40$, no incorrect linkage phase combination will be detected.

The log-likelihood ratios (LRT) of Equation 12 were used to test for the significance of QTL effects under different sample sizes and heritability levels. Except for a few cases where $N = 200$ and $H^2 = 0.20$, QTL can be detected at a significance level of $P = 0.05$ in all 100 repeated simulations. The critical threshold value was calculated by simulating data sets with QTL effects set to zero and examining the distribution of the LRT (see also Hackett *et al.* 2001). Using the 95% point of the distribution of the LRT gives a test of significance at a 5% level for the presence of a QTL.

The effects of completely preferential pairings and random pairings: Doerge and Craig (2000) assumed that chromosomes pair strictly between homologs during polyploid meiosis. If this assumption is true, we will have only one bivalent pairing pattern, as opposed to three patterns when incompletely preferential pairings are considered [see expression (2)]. Thus, under this assumption there will be only 16 gamete genotypes at two informative markers and 4 gamete genotypes at one QTL for each parent. Such a (16×4) matrix of conditional probabilities with the completely preferential pairing assumption represents an allopolyploid model and is a subset of the (36×6) matrix used in our method.

Hackett *et al.*'s assumption of random bivalent pairings (the autopolyploid model) leads to the same structure of the conditional probability matrix that we have in our bivalent polyploid model. Because our model covers the allo- and autopolyploid model, it can be regarded as the general polyploid model. Here, we make a comparison between Hackett *et al.*'s method and our method by first looking at the conditional probability matrix derived for the general polyploid model listed in Table 1. From the table it is found that only the conditional probabilities of QTL genotypes of 12 bold-

faced marker genotypes contain p and the conditional probabilities of the rest of the 24 marker genotypes do not contain p because p is canceled out. This means that p may have a relatively small influence on the conditional probability matrix and therefore on parameter estimates under the general polyploid model when two markers considered are fully informative. In other words, for fully informative markers, results from the autopolyploid model will be similar to those from the general polyploid model. A small simulation study has confirmed this inference (results not shown).

However, for partially informative markers (R. Wu *et al.* 2001), some of the QTL genotypes will be collapsed into one so that the corresponding joint genotypic probabilities will be summed up. For example, for a single-dose restriction fragment (simplex) (*Pppp*), six QTL gamete genotypes (P_1P_2, P_1P_3, P_1P_4, P_2P_3, P_2P_4, and P_3P_4) will be reduced to two (*Pp* and *pp*) with each summed from three gamete genotypes. Similar reductions are also true for two flanking simplex markers. In this case, p would not be canceled out in the conditional probability matrix and, therefore, will play an important role in affecting the estimates of QTL position and effect parameters.

DISCUSSION

The development of statistical methods for mapping QTL in polyploids is one of the most difficult tasks in genetic and genomic study. Although quite a few studies of linkage analysis have used polymorphic markers in polyploids (Hackett *et al.* 1998; Ripol *et al.* 1999; Luo *et al.* 2001; R. Wu *et al.* 2001, 2002a; S. Wu *et al.* 2001), we know of only three articles published about the statistical developments of QTL mapping in this recalcitrant group of species (Doerge and Craig 2000; Xie and Xu 2000; Hackett *et al.* 2001), with one, unfortunately, based on an improper biological process of polyploid meiosis (as noted by Hackett 2001). The other two articles require simplifying assumptions, which are not likely to hold in real life. Doerge and Craig's (2000) method can be appropriate only for extreme allopolyploids, in which chromosome pairings occur strictly between two homologs. On the contrary, the assumption used in Hackett *et al.* (2001) is random bivalent pairings during meiosis and, thus, that method can fit only extreme bivalent autopolyploids having identical chromosomes in the set.

In this article we report on the development of a novel statistical methodology for QTL mapping in bivalent polyploids that represent an important group of polyploids including alfalfa, potato, and wheat. Using extensive simulations, we examined the robustness and performance of this bivalent polyploid method in estimating QTL effects, QTL position, and QTL linkage phase relative to known-phase markers under different sample sizes and heritability levels. We also compared the results from our method and the current methods on the basis of the allo- and autopolyploid model. Our method has four significant improvements over these current statistical methods for QTL mapping in bivalent polyploids. First, our method incorporates a general bivalent pairing mechanism of meiotic configuration by defining a cytological parameter called the preferential pairing factor. The preferential pairing factor (p) is defined as the propensity of bivalent pairings between more similar rather than less similar chromosomes (Sybenga 1988, 1994, 1995, 1996). Different values of this parameter, ranging from 0 to $\frac{2}{3}$, describe different degrees of relatedness between the chromosomes in the set. When $p = 0$, it means that chromosomes pair randomly and that our method is automatically reduced to the autopolyploid model. When $p = \frac{2}{3}$, only identical chromosomes can pair and our method is reduced to the allopolyploid model. Our method therefore represents a general model for QTL mapping in bivalent polyploids. It can, in particular, be applied for those polyploids whose chromosome origins (auto- *vs.* allopolyploids) are unknown *a priori*. In a recent review by Soltis and Soltis (2000), such origin-unknown polyploids commonly occur in nature. On the basis of the estimate of p, we will be in a better position to study the origin and relatedness of the genomes contained in a polyploid (Sybenga 1996).

The second improvement of our method is a thorough exploration of QTL action and interaction effects on phenotypes in polyploids. As with diploids, the inheritance mode of QTL in polyploids can be additive or dominant. But compared with diploids, these gene actions and interactions are much more complicated because of an increased number of alleles and allele combinations. Kempthorne (1957) extended the diploid theory of quantitative genetics to partition genetic effects of a QTL into additive and dominant components of different within-locus interaction levels in polyploids. For a bivalent tetraploid having four different alleles at a QTL, we are confronted with 4 allelic or additive effects, 28 diallelic dominant interaction effects, 48 triallelic dominant interaction effects, and 36 tetraallelic interaction effects. Because these 120 parameters are not completely independent, their dependence needs to be removed to obtain estimable parameters. We used a parameterization process to reduce these parameters to 36 independent ones. Such a reduced space of unknown parameters was also embedded in Hackett *et al.*'s (2001) QTL-mapping framework, but those authors have not provided a tractable estimation of all these components. In fact, it is impossible to obtain accurate and precise estimates of these 36 independent parameters on the basis of a sample size we can have in practice, using a traditional treatment for QTL mapping in diploids.

The efficient estimation of these 36 quantitative genetic parameters in tetraploid mapping, therefore, offers the third improvement of our method over the current methods. In this article, we incorporate the EM algorithm (Lander and Botstein 1989; Meng and Rubin 1993) and techniques of experimental design to

estimate QTL effects at different levels. In a statistical mixture model for QTL mapping, the EM algorithm can provide robust estimates for the expected means of QTL genotypes. This advantage is combined with a parameterization process to provide robust estimates of QTL effects that constitute the QTL genotypic means. Through a parameterization process, the sampling variance of the estimator of each QTL effect is only a small portion of the sampling variance of the estimated residual variance (see APPENDIX A). Also, the influences of the estimator of one QTL effect by other effects are limited within the QTL effects of similar nature [see the structure of $\mathbf{D}^{-1}(\mathbf{D}^{-1})^{T}$]. These two favorable properties of the parameterized QTL effects assure the estimation precision of QTL actions and interactions, as demonstrated in the simulation study.

The correct characterization of linkage phases is a prerequisite for genome mapping in species like polyploids, in which homozygous inbred lines cannot be obtained. In this article, we used a modified EM algorithm to simultaneously estimate linkage and linkage phases. LUO *et al.* (2001) determined a most likely linkage phase of different markers on the basis of the largest maximum log-likelihood ratio and the lowest estimate of the recombination fraction. However, when this procedure is used for characterizing a most likely linkage phase of a QTL relative to its flanking markers, there is a technical difficulty. When permuting different linkage phases, the same likelihood value will be detected, despite different estimates of QTL effects obtained, and thus gives no way of selecting a most likely linkage phase. In this article, the EM algorithm was performed by posing some constraints on additive effects of different QTL alleles. This modified approach greatly increases the probability of correctly selecting a most likely linkage phase, as shown in the simulation study. The more efficient characterization of linkage phase presents a fourth technical merit of our method.

Our method can be extended to several more general situations. First, we should consider statistical properties of using markers with lower degrees of informativeness to map QTL in polyploids. HACKETT *et al.* (2001) found that less informative markers, *e.g.*, dominant markers, displayed reduced power of QTL detection compared to fully informative markers. But it is unclear how much the estimation precision and power will be reduced for our bivalent polyploid model, incorporating the preferential pairing factor when partially informative markers are used. It appears that partially informative markers, *e.g.*, single-, double-, or multidose restriction fragments, occupy a larger portion of polyploid genomes (DA SILVA *et al.* 1995; BROUWER *et al.* 2000; MING *et al.* 2001). In this study, we assume that the recombination fractions and the preferential pairing factor are identical between the two parents. Because these two parameters are often population or sex specific (*e.g.*, ALLENDORF and DANZMANN 1997), we should study the effects of parent-dependent linkages and preferential pairing factors on the estimates of QTL parameters. In the real world, many polyploids undergo multivalent formations. Our bivalent polyploids cannot solve the issues arising from multivalent formation, which leads to be typical genetic phenomenon of double reduction (BUTRUILLE and BOITEUX 2000). Last but not least, our bivalent tetraploid model should be extended to study polyploids at a higher polyploidy level. The model reported in this article represents a platform on which complicated problems related to polyploid mapping can be solved within our framework, integrating statistics, genetics, computer science, and cytology.

We thank two anonymous reviewers for their constructive comments. This work is partially supported by an Outstanding Young Investigator Award of the National Natural Science Foundation of China (30128017), a University of Florida Research Opportunity Fund (02050259), and a University of South Florida Biodefense grant (7222061-12) to R.W. The publication of this manuscript is approved as a journal series no. R-08795 by the Florida Agricultural Experiment Station.

LITERATURE CITED

ALLENDORF, F. W., and R. G. DANZMANN, 1997 Secondary tetrasomic segregation of *MDH-B* and preferential pairing of homeologues in rainbow trout. Genetics **145:** 1083–1092.

BEVER, J. D., and F. FELBER, 1992 The theoretical population genetics of autopolyploidy. Oxf. Surv. Evol. Biol. **8:** 185–217.

BROUWER, D. J., and T. C. OSBORN, 1999 A molecular marker linkage map of tetraploid alfalfa (*Medicago sativa* L.). Theor. Appl. Genet. **99:** 1194–1200.

BROUWER, D. J., S. H. DUKE and T. C. OSBORN, 2000 Mapping genetic factors associated with winter hardiness, fall growth, and freezing injury in autotetraploid alfalfa. Crop Sci. **40:** 1387–1396.

BUTRUILLE, D. V., and L. S. BOITEUX, 2000 Selection-mutation balance in polysomic tetraploids: impact of double reduction and gametophytic selection on the frequency and subchromosomal localization of deleterious mutations. Proc. Natl. Acad. Sci. USA **97:** 6608–6613.

DA SILVA, J., M. E. SORRELLS, W. L. BURNQUIST and S. D. TANKSLEY, 1995 RFLP linkage map and genome analysis of *Saccharum spontaneum*. Genome **36:** 782–791.

DEMPSTER, A. P., N. M. LAIRD and D. B. RUBIN, 1977 Maximum likelihood from incomplete data via EM algorithm. J. R. Stat. Soc. Ser. B **39:** 1–38.

DOERGE, R. W., and B. A. CRAIG, 2000 Model selection for quantitative trait locus analysis in polyploids. Proc. Natl. Acad. Sci. USA **97:** 7951–7956.

FISHER, R. A., 1947 The theory of linkage in polysomic inheritance. Philos. Trans. R. Soc. Ser. B **233:** 55–87.

FJELLSTROM, R. G., P. R. BEUSELINCK and J. J. STEINER, 2001 RFLP marker analysis supports tetrasomic inheritance in *Lotus corniculatus* L. Theor. Appl. Genet. **102:** 718–725.

GRIVET, L., A. D'HONT, D. ROQUES, P. FELDMANN, C. LANAUD *et al.*, 1996 RFLP mapping in cultivated sugarcane (*Saccharum* spp): genome organization in a highly polyploid and aneuploid interspecific hybrid. Genetics **142:** 987–1000.

HACKETT, C. A., 2001 A comment on Xie and Xu: 'Mapping quantitative trait loci in tetraploid species'. Genet. Res. **78:** 187–189.

HACKETT, C. A., J. E. BRADSHAW, R. C. MEYER, J. W. MCNICOL, D. MILBOURNE *et al.*, 1998 Linkage analysis in tetraploid species: a simulation study. Genet. Res. **71:** 143–154.

HACKETT, C. A., J. E. BRADSHAW and J. W. MCNICOL, 2001 Interval mapping of quantitative trait loci in autotetraploid species. Genetics **159:** 1819–1832.

HALDANE, J. B. S., 1930 Theoretical genetics of autopolyploids. Genetics **22:** 359–372.

HICKOK, L. G., 1978 Homoeologous chromosome pairing: frequency difference in inbred and intraspecific hybrid polyploid ferns. Science **202:** 982–984.

Hilu, K. W., 1993 Polyploidy and the evolution of domesticated plants. Am. J. Bot. **80:** 1491–1499.

Hoarau, J. Y., B. Offmann, A. D'Hont, A. M. Risterucci, D. Roques *et al.*, 2001 Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). I. Genome mapping with AFLP markers. Theor. Appl. Genet. **103:** 84–97.

Kempthorne, O., 1957 *An Introduction to Genetic Statistics*. John Wiley & Sons, New York.

Lander, E. S., and D. Botstein, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

Leitch, I. J., and M. D. Bennett, 1997 Polyploidy in angiosperms. Trends Plant Sci. **2:** 470–476.

Luo, Z. W, C. A. Hackett, J. E. Bradshaw, J. W. McNicol and D. Milbourne, 2001 Construction of a genetic linkage map in tetraploid species using molecular markers. Genetics **157:** 1369–1385.

Mather, K., 1935 Reductional and equational separation of the chromosomes in bivalents and multivalents. J. Genet. **30:** 53–78.

Mather, K., 1936 Segregation and linkage in autotetraploids. J. Genet. **32:** 287–314.

Meng, X. L., and D. B. Rubin, 1993 Maximum likelihood estimation via the ECM algorithm—a general framework. Biometrika **80:** 267–278.

Meyer, R. C., D. Milbourne, C. A. Hackett, J. E. Bradshaw, J. W. McNichol *et al.*, 1998 Linkage analysis in tetraploid potato and association of markers with quantitative resistance to late blight (*Phytophthora infestans*). Mol. Gen. Genet. **259:** 150–160.

Ming, R., S. C. Liu, Y. R. Lin, J. da Silva, W. Wilson *et al.*, 1998 Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics **150:** 1663–1682.

Ming, R., S. C. Liu, P. H. Moore, J. E. Irvine and A. H. Paterson, 2001 QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. Genome Res. **11:** 2075–2084.

Ott, S. P., and J. Whitton, 2000 Polyploid incidence and evolution. Annu. Rev. Genet. **34:** 401–437.

Rajapakse, S., D. H. Byrne, L. Zhang, N. Anderson, K. Arumuganathan *et al.*, 2001 Two genetic linkage maps of tetraploid roses. Theor. Appl. Genet. **103:** 575–583.

Ramsey, J., and D. W. Schemske, 1998 Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu. Rev. Ecol. Syst. **29:** 467–501.

Ripol, M. I., G. A. Churchill, J. A. G. da Silva and M. Sorrells, 1999 Statistical aspects of genetic mapping in autopolyploids. Gene **235:** 31–41.

Soltis, P. S., and D. E. Soltis, 2000 The role of genetic and genomic attributes in the success of polyploids. Proc. Natl. Acad. Sci. USA **97:** 7051–7057.

Sybenga, J., 1988 Mathematical models for estimating preferential pairing and recombination in triploid hybrids. Genome **30:** 745–757.

Sybenga, J., 1992 *Cytogenetics in Plant Breeding*. Springer-Verlag, Berlin/Heidelberg/New York.

Sybenga, J., 1994 Preferential pairing estimates from multivalent frequencies in tetraploids. Genome **37:** 1045–1055.

Sybenga, J., 1995 Meiotic pairing in autohexaploid Lathyrus: a mathematical model. Heredity **75:** 343–350.

Sybenga, J., 1996 Chromosome pairing affinity and quadrivalent formation in polyploids: Do segmental allopolyploids exist? Genome **39:** 1176–1184.

Wu, K. K., W. Burnquist, M. E. Sorrells, T. L. Tew, P. H. Moore *et al.*, 1992 The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor. Appl. Genet. **83:** 294–300.

Wu, R. L., M. Gallo-Meagher, R. C. Littell and Z-B. Zeng, 2001 A general polyploid model for analyzing gene segregation in outcrossing tetraploid species. Genetics **159:** 869–882.

Wu, R. L., C.-X. Ma and G. Casella, 2002a A preferential pairing model of linkage analysis in outcrossing tetraploids. Theor. Popul. Biol. **62:** 129–151.

Wu, R. L., C.-X. Ma, I. Painter and Z-B. Zeng, 2002b Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. Theor. Popul. Biol. **61:** 349–363.

Wu, S. S., R. L. Wu, C.-X. Ma, Z-B. Zeng, M. C. K. Yang *et al.*, 2001 A multivalent pairing model of linkage analysis in autotetraploids. Genetics **159:** 2000–2012.

Xie, C. G., and S. H. Xu, 2000 Mapping quantitative trait loci in tetraploid populations. Genet. Res. **76:** 105–115.

Yu, K. F., and K. P. Pauls, 1993 Segregation of random amplified polymorphic DNA markers and strategies for molecular mapping in tetraploid alfalfa. Genome **36:** 844–851.

Zeven, A. C., 1979 Polyploidy and domestication: the origin and survival of polyploids in cytotype mixtures, pp. 385–408 in *Polyploidy, Biological Relevance*, edited by W. H. Lewis. Plenum Press, New York.

Communicating editor: M. A. Asmussen

APPENDIX A: PARAMETERIZATION OF GENE EFFECTS

All the main and interaction effects in bivalent tetraploids should be parameterized to obtain a group of estimable parameters. In this article, the parameterization of these gene effects is based on different constraints posed on them. The constraints on the allelic (main) effects are expressed as

$$\sum_{u_1=1}^{4} \alpha_{u_1}^{\mathrm{P}} = \sum_{v_1=1}^{4} \alpha_{v_1}^{\mathrm{Q}} = 0,$$

which lead to six estimable independent parameters. The constraints on the diallelic interaction effects are

$$\sum_{u_1=1}^{4} \beta_{u_1u_2}^{\mathrm{PP}} = \sum_{v_1=1}^{4} \beta_{v_1v_2}^{\mathrm{QQ}} = 0,\ u_1 \neq u_2,\ v_1 \neq v_2,$$

$$\sum_{u_1=1}^{4} \beta_{u_1v_1}^{\mathrm{PQ}} = \sum_{v_1=1}^{4} \beta_{u_1v_1}^{\mathrm{PQ}} = 0,$$

which lead to two independent parameters for interactions between two alleles from parents P and Q, respectively, and nine independent parameters for interactions between two alleles each from a different parent. The constraints on the triallelic interaction effects are

$$\sum_{v_1=1}^{4} \gamma_{u_1u_2v_1}^{\mathrm{PPQ}} = \sum_{u_1=1}^{3} \sum_{u_2=u_1+1}^{4} \gamma_{u_1u_2v_1}^{\mathrm{PPQ}} = \sum_{u_1=1}^{4} \gamma_{u_1v_1v_2}^{\mathrm{PQQ}} = \sum_{v_1=1}^{3} \sum_{v_2=v_1+1}^{4} \gamma_{u_1v_1v_2}^{\mathrm{PPQ}} = 0,$$

which lead to 12 independent parameters. The constraints on the tetraallelic interaction effects are

$$\begin{aligned}
\sum_{v_2=1}^{4}\delta_{u_1u_2v_1v_2} &= \sum_{v_1=1}^{4}\delta_{u_1u_2v_1v_2} = \sum_{u_2=1}^{4}\delta_{u_1u_2v_1v_2} = \sum_{u_1=1}^{4}\delta_{u_1u_2v_1v_2}\\
&= \sum_{v_1=1}^{3}\sum_{v_2=v_1+1}^{4}\delta_{u_1u_2v_1v_2} = \sum_{u_1=1}^{3}\sum_{u_2=u_1+1}^{4}\delta_{u_1u_2v_2v_2} = \sum_{u_2=1}^{4}\sum_{v_2=1}^{4}\delta_{u_1u_2v_2v_2}\\
&= \sum_{u_1=1}^{4}\sum_{v_1=1}^{4}\delta_{u_1u_2v_2v_2} = \sum_{u_1=1}^{4}\sum_{v_2=1}^{4}\delta_{u_1u_2v_2v_2} = \sum_{u_1=1}^{4}\sum_{v_1=1}^{4}\delta_{u_1u_2v_2v_2}\\
&= \sum_{u_2=1}^{4}\sum_{v_1=1}^{3}\sum_{v_2=v_1+1}^{4}\delta_{u_1u_2v_2v_2} = \sum_{u_1=1}^{4}\sum_{v_1=1}^{3}\sum_{v_2=v_1+1}^{4}\delta_{u_1u_2v_2v_2}\\
&= \sum_{u_1=1}^{3}\sum_{u_2=u_1+1}^{4}\sum_{v_2=1}^{4}\delta_{u_1u_2v_2v_2} = \sum_{u_1=1}^{3}\sum_{u_2=u_1+1}^{4}\sum_{v_1=1}^{4}\delta_{u_1u_2v_2v_2}\\
&= 0,
\end{aligned}$$

which lead to four independent parameters.

By these parameterization constraints, a total of 120 QTL effect parameters contained in the genotypic values $P_{u_1}P_{u_2}Q_{v_1}Q_{v_2}$ ($1 \le u_1 < u_2 \le 4$, $1 \le v_1 < v_2 \le 4$) in Equation 8 can be reduced to 35 independent estimable parameters. Without loss of generality, 6 independent allelic (main) effect parameters are assigned as α_1^{P}, α_2^{P}, and α_3^{P} for parent P and α_1^{Q}, α_2^{Q}, and α_3^{Q} for parent Q. Thirteen diallelic interaction parameters are assigned as β_{12}^{PP} and β_{13}^{PP} for two alleles from parent P; β_{12}^{QQ} and β_{13}^{QQ} for two alleles from parent Q; and β_{11}^{PQ}, β_{12}^{PQ}, β_{13}^{PQ}, β_{21}^{PQ}, β_{22}^{PQ}, β_{23}^{PQ}, β_{31}^{PQ}, β_{32}^{PQ}, and β_{33}^{PQ} for two alleles, one from parent P and the other from parent Q. Twelve triallelic interaction parameters are assigned as $\gamma_{121}^{\mathrm{PPQ}}$, $\gamma_{122}^{\mathrm{PPQ}}$, $\gamma_{123}^{\mathrm{PPQ}}$, $\gamma_{131}^{\mathrm{PPQ}}$, $\gamma_{132}^{\mathrm{PPQ}}$, and $\gamma_{133}^{\mathrm{PPQ}}$ for three alleles, two from parent P and one from parent Q; and $\gamma_{112}^{\mathrm{PQQ}}$, $\gamma_{212}^{\mathrm{PQQ}}$, $\gamma_{312}^{\mathrm{PQQ}}$, $\gamma_{113}^{\mathrm{PQQ}}$, $\gamma_{213}^{\mathrm{PQQ}}$, and $\gamma_{313}^{\mathrm{PQQ}}$ for three alleles, one from parent P and two from parent Q. Four tetraallelic interaction parameters are assigned as δ_{1212}, δ_{1213}, δ_{1312}, and δ_{1313}.

The vector ($\mathbf{m} = (\mu_{u_1u_2v_1v_2})_{36\times 1}$, $1 \le u_1 < u_2 \le 4$, $1 \le v_1 < v_2 \le 4$) of the 36 QTL genotypic values can be expressed in terms of these assigned effect parameters. We have

$$\mathbf{m} = \mathbf{Da},$$

where **D** is a design matrix expressed as

$$\begin{bmatrix}
1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\
1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0\\
1 & 1 & 1 & 0 & 0 & -1 & -1 & 1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & 0\\
1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & -1 & -1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & 0\\
1 & 1 & 1 & 0 & -1 & 0 & -1 & 1 & 0 & 0 & 1 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0\\
1 & 1 & 1 & 0 & -1 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\
1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0\\
1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1\\
1 & 1 & 0 & 1 & 0 & -1 & -1 & 0 & 1 & -1 & -1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & 0 & -1 & -1\\
1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & -1 & -1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & 0 & -1 & -1\\
1 & 1 & 0 & 1 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & -1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1\\
1 & 1 & 0 & 1 & -1 & -1 & 0 & 0 & 1 & 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0\\
1 & 0 & -1 & -1 & 1 & 1 & 0 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & 0 & -1\\
1 & 0 & -1 & -1 & 1 & 0 & 1 & -1 & -1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & 0 & -1\\
1 & 0 & -1 & -1 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1\\
1 & 0 & -1 & -1 & 0 & 1 & 1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1\\
1 & 0 & -1 & -1 & -1 & 0 & -1 & -1 & -1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & 0 & -1\\
1 & 0 & -1 & -1 & -1 & -1 & 0 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & -1 & 0 & -1 & 0\\
1 & 0 & 1 & 1 & 1 & 1 & 0 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & -1 & 0\\
1 & 0 & 1 & 1 & 1 & 0 & 1 & -1 & -1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & -1 & 0 & -1\\
1 & 0 & 1 & 1 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & -1 & -1 & 0 & -1 & -1 & 1 & 1 & 1 & 1\\
1 & 0 & 1 & 1 & 0 & 1 & 1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & -1 & 1 & 1 & 1 & 1\\
1 & 0 & 1 & 1 & -1 & 0 & -1 & -1 & -1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & -1 & -1 & 0 & -1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & -1 & 0 & -1\\
1 & 0 & 1 & 1 & -1 & -1 & 0 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & -1 & 0\\
1 & -1 & 0 & -1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0\\
1 & -1 & 0 & -1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & -1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 1\\
1 & -1 & 0 & -1 & 0 & -1 & -1 & 0 & 1 & -1 & -1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & -1 & -1\\
1 & -1 & 0 & -1 & 0 & 1 & 1 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & 0\\
1 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\
1 & -1 & -1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\
1 & -1 & -1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 0 & 0\\
1 & -1 & -1 & 0 & 0 & -1 & -1 & 1 & 0 & -1 & -1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & -1 & -1 & 0 & 0\\
1 & -1 & -1 & 0 & 0 & 1 & 1 & 1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & -1 & -1 & 0 & 0\\
1 & -1 & -1 & 0 & -1 & 0 & -1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 0 & 0\\
1 & -1 & -1 & 0 & -1 & -1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & -1 & -1 & 0 & 0
\end{bmatrix},$$

and **a** is the vector of gene effects, expressed as

$$[\mu\alpha_1^P\alpha_2^P\alpha_3^P\alpha_1^Q\alpha_2^Q\alpha_3^Q\beta_{12}^{PP}\beta_{13}^{PP}\beta_{12}^{QQ}\beta_{13}^{QQ}\beta_{11}^{PQ}\beta_{12}^{PQ}\beta_{13}^{P}\beta_{21}^{PQ}\beta_{22}^{PQ}\beta_{23}^{PQ}\beta_{31}^{PQ}\beta_{32}^{PQ}\beta_{33}^{PQ}\gamma_{121}^{PPQ}\gamma_{122}^{PPQ}\gamma_{123}^{PPQ}\gamma_{131}^{PPQ}\gamma_{132}^{PPQ}\gamma_{133}^{PPQ}\gamma_{112}^{PQQ}\gamma_{212}^{PQQ}\gamma_{312}^{PQQ}\gamma_{113}^{PQQ}\delta_{1212}\delta_{1213}\delta_{1312}\delta_{1313}]^T.$$

In APPENDIX B, we provide the EM algorithm for estimating $\hat{\mathbf{m}}$, from which the MLEs of **a** are solved:

$$\hat{\mathbf{a}} = \mathbf{D}^{-1}\hat{\mathbf{m}}.$$

The sampling variance of $\hat{\mathbf{a}}$ is thus $\mathbf{V}_{\hat{\mathbf{a}}} = \mathbf{D}^{-1}\mathbf{V}_{\hat{\mathbf{m}}}(\mathbf{D}^{-1})^T = \mathbf{D}^{-1}\,(\mathbf{D}^{-1})^T\hat{\sigma}^2$. We have $\mathbf{D}^{-1}(\mathbf{D}^{-1})^T =$

$$\begin{bmatrix}
\frac{1}{36} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & \frac{1}{16} & \frac{-1}{48} & \frac{-1}{48} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & \frac{-1}{48} & \frac{1}{16} & \frac{-1}{48} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & \frac{-1}{48} & \frac{-1}{48} & \frac{1}{16} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & \frac{1}{16} & \frac{-1}{48} & \frac{-1}{48} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & \frac{-1}{48} & \frac{1}{16} & \frac{-1}{48} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & \frac{-1}{48} & \frac{-1}{48} & \frac{1}{16} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{18} & \frac{-1}{36} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-1}{36} & \frac{1}{18} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{18} & \frac{-1}{36} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-1}{36} & \frac{1}{18} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{9}{64} & \frac{-3}{64} & \frac{-3}{64} & \frac{-3}{64} & \frac{1}{64} & \frac{1}{64} & \frac{-3}{64} & \frac{1}{64} & \frac{1}{64} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-3}{64} & \frac{9}{64} & \frac{-3}{64} & \frac{1}{64} & \frac{-3}{64} & \frac{1}{64} & \frac{1}{64} & \frac{-3}{64} & \frac{1}{64} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-3}{64} & \frac{-3}{64} & \frac{9}{64} & \frac{-3}{64} & \frac{1}{64} & \frac{-3}{64} & \frac{1}{64} & \frac{1}{64} & \frac{-3}{64} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-3}{64} & \frac{1}{64} & \frac{1}{64} & \frac{9}{64} & \frac{-3}{64} & \frac{1}{64} & \frac{-3}{64} & \frac{1}{64} & \frac{1}{64} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{64} & \frac{-3}{64} & \frac{1}{64} & \frac{-3}{64} & \frac{9}{64} & \frac{1}{64} & \frac{1}{64} & \frac{-3}{64} & \frac{1}{64} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{64} & \frac{1}{64} & \frac{-3}{64} & \frac{1}{64} & \frac{1}{64} & \frac{9}{64} & \frac{-3}{64} & \frac{1}{64} & \frac{-3}{64} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-3}{64} & \frac{1}{64} & \frac{1}{64} & \frac{-3}{64} & \frac{1}{64} & \frac{1}{64} & \frac{9}{64} & \frac{-3}{64} & \frac{1}{64} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{64} & \frac{-3}{64} & \frac{1}{64} & \frac{1}{64} & \frac{-3}{64} & \frac{1}{64} & \frac{1}{64} & \frac{9}{64} & \frac{-3}{64} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{64} & \frac{1}{64} & \frac{-3}{64} & \frac{1}{64} & \frac{1}{64} & \frac{-3}{64} & \frac{-3}{64} & \frac{-3}{64} & \frac{9}{64} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{8} & \frac{-1}{24} & \frac{-1}{24} & \frac{-1}{16} & \frac{1}{48} & \frac{1}{48} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-1}{24} & \frac{1}{8} & \frac{-1}{24} & \frac{1}{48} & \frac{-1}{16} & \frac{1}{48} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-1}{24} & \frac{-1}{24} & \frac{1}{8} & \frac{1}{48} & \frac{1}{48} & \frac{-1}{16} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-1}{16} & \frac{1}{48} & \frac{1}{48} & \frac{1}{8} & \frac{-1}{24} & \frac{-1}{24} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{48} & \frac{-1}{16} & \frac{-1}{48} & \frac{-1}{24} & \frac{1}{8} & \frac{-1}{24} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{48} & \frac{1}{48} & \frac{-1}{16} & \frac{-1}{24} & \frac{-1}{24} & \frac{1}{8} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{9} & \frac{-1}{18} & \frac{-1}{18} & \frac{1}{36} & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-1}{18} & \frac{1}{9} & \frac{1}{36} & \frac{-1}{18} & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-1}{18} & \frac{1}{36} & \frac{1}{9} & \frac{-1}{18} & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{36} & \frac{-1}{18} & \frac{-1}{18} & \frac{1}{9} & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix},$$

which has two desirable properties: (1) the elements on its diagonal are much smaller than one, ranging from $\frac{9}{64}$ to $\frac{1}{36}$, and (2) most elements off its diagonal are zero. The first property implies that the sampling variance of each estimator in the vector **a** from our parameterization approach is always smaller than the estimated residual variance. The second property suggests that different estimators in vector **a** are independent of each other.

APPENDIX B: IMPLEMENTATION OF THE EM ALGORITHM

The parameter vector in which we are interested is denoted by $\boldsymbol{\Omega}$. But the estimation of this vector is not most efficient from a computational standpoint. As explained in the text, we define a new vector $\boldsymbol{\Theta} = (\mathbf{m}, r_1 \text{ or } r_2, \sigma^2, p)$, which can be more easily estimated by implementing the EM algorithm (DEMPSTER *et al.* 1977; MENG and RUBIN 1993). The log-likelihood of the new vector is given by

$$\log L(\mathbf{y}|\boldsymbol{\Theta}) = \sum_{i=1}^{n}\log\left[\sum_{u_1=1}^{3}\sum_{u_2}^{4}\sum_{v_1=1}^{3}\sum_{v_2=v_1}^{4} p_{u_1u_2v_1v_2i}f_{u_1u_2v_1v_2}(y_i)\right], \tag{B1}$$

with derivatives

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\Theta}_m}\log L(\mathbf{y}|\Theta) &= \sum_{i=1}^{n}\sum_{u_1=1}^{3}\sum_{u_2=u_1+1}^{4}\sum_{v_1=1}^{3}\sum_{v_2=v_1+1}^{4}\frac{p_{u_1u_2v_1v_2i}\,\frac{\partial}{\partial\Theta_m}f_{u_1u_2v_1v_2}(y_i)}{\sum_{u_1=1}^{3}\sum_{u_2=u_1+1}^{4}\sum_{v_1=1}^{3}\sum_{v_2=v_1+1}^{4}p_{u_1u_2v_1v_2i}f_{u_1u_2v_1v_2}(y_i)}\\ &= \sum_{i=1}^{n}\sum_{u_1=1}^{3}\sum_{u_2=u_1+1}^{4}\sum_{v_1=1}^{3}\sum_{v_2=v_1+1}^{4}\frac{p_{u_1u_2v_1v_2i}f_{u_1u_2v_1v_2}(y_i)}{\sum_{u_1=1}^{3}\sum_{u_2=u_1+1}^{4}\sum_{v_1=1}^{3}\sum_{v_2=v_1+1}^{4}p_{u_1u_2v_1v_2i}f_{u_1u_2v_1v_2}(y_i)}\frac{\partial}{\partial\Theta_m}\log f_{u_1u_2v_1v_2}(y_i)\\ &= \sum_{i=1}^{n}\sum_{u_1=1}^{3}\sum_{u_2=u_1+1}^{4}\sum_{v_1=1}^{3}\sum_{v_2=v_1+1}^{4}P_{u_1u_2v_1v_2i}\frac{\partial}{\partial\Theta_m}\log f_{u_1u_2v_1v_2}(y_i),\end{aligned}$$

where we define

$$P_{u_1u_2v_1v_2i}=\frac{P_{u_1u_2v_1v_2i}f_{u_1u_2v_1v_2}(y_i)}{\sum_{u_1=1}^{3}\sum_{u_2=u_1+1}^{4}\sum_{v_1=1}^{3}\sum_{v_2=v_1+1}^{4}p_{u_1u_2v_1v_2i}f_{u_1u_2v_1v_2}(y_i)},\tag{B2}$$

which could be thought of as a posterior probability that progeny i ($i = 1, \ldots, N$) has QTL genotype $u_1u_2v_1v_2$ ($1 \le u_1 < u_2 \le 4$, $1 \le v_1 < v_2 \le 4$). We then implement the EM algorithm with the expanded parameter set {$\boldsymbol{\Theta}$, $\mathbf{P}$}, where P = $\{P_{u_1u_2v_1v_2}\}$. Conditional on $\mathbf{P}$, we solve for the zeros of $\partial/\partial\boldsymbol{\Theta}_m \log L(\mathbf{y}|\Theta)$ to get our estimates of $\boldsymbol{\Theta}$ under the constraints of displays (10) and (11) (the M step). The estimates are then used to update $\mathbf{P}$ (the E step), and the process is repeated until convergence. The values at convergence are the MLEs.

The log-likelihood equations for the MLEs of $\mathbf{m}$ and σ^2 are given as

$$\mu_{u_1u_2v_1v_2i}=\frac{\sum_{i=1}^{n}\sum_{u_1=1}^{3}\sum_{u_2=u_1+1}^{4}\sum_{v_1=1}^{3}\sum_{v_2=v_1+1}^{4}P_{u_1u_2v_1v_2i}\,y_i}{\sum_{i=1}^{n}\sum_{u_1=1}^{3}\sum_{u_2=u_1+1}^{4}\sum_{v_1=1}^{3}\sum_{v_2=v_1+1}^{4}P_{u_1u_2v_1v_2i}},$$

$$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}\sum_{u_1=1}^{3}\sum_{u_2=u_1+1}^{4}\sum_{v_1=1}^{3}\sum_{v_2=v_1+1}^{4}P_{u_1u_2v_1v_2i}(y_i-\mu_{u_1u_2v_1v_2i})^2.$$

The QTL position and the preferential pairing factor are generally estimated using the grid approach by fixing them at particular values in their space. The values of these two parameters, at which the maximum-likelihood value is obtained, are their MLEs.