Theoretical Population Biology **61**, 349–363 (2002)
doi:10.1006/tpbi.2002.1577, available online at http://www.idealibrary.com on IDEAL®

TPB

Simultaneous Maximum Likelihood Estimation of Linkage and Linkage Phases in Outcrossing Species

Rongling Wu and Chang-Xing Ma

Department of Statistics, University of Florida, Gainesville, Florida 32611
E-mail: rwu@stat.ufl.edu

and

Ian Painter and Zhao-Bang Zeng

Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695

Received March 16, 2001

With the advent of new molecular marker technologies, it is now feasible to initiate genome projects for outcrossing plant species, which have not received much attention in genetic research, despite their great agricultural and environmental value. Because outcrossing species typically have heterogeneous genomes, data structure for molecular markers representing an entire genome is complex: some markers may have more alleles than others, some markers are codominant whereas others are dominant, and some markers are heterozygous in one parent but fixed in the other parent whereas the opposite can be true for other markers. A major difficulty in analyzing these different types of marker at the same time arises from uncertainty about parental linkage phases over markers. In this paper, we present a general maximum-likelihood-based algorithm for simultaneously estimating linkage and linkage phases for a mixed set of different marker types containing fully informative markers (segregating 1:1:1:1) and partially informative markers (or missing markers, segregating 1:2:1, 3:1, and 1:1) in a full-sib family derived from two outbred parent plants. The characterization of linkage phases is based on the posterior probability distribution of the assignment of alternative alleles at given markers to two homologous chromosomes of each parent, conditional on the observed phenotypes of the markers. Two- and multi-point analyses are performed to estimate the recombination fraction and determine the most likely linkage phase between different types of markers. A numerical example is presented to demonstrate the statistical properties of the model for characterizing the linkage phase between markers.

Key Words: **EM algorithm; linkage phase; outcrossing species; partially informative marker; posterior probability; recombination fraction**

1. INTRODUCTION

A number of genome projects have been launched worldwide for agronomic crops in which homozygous inbred lines are available. These projects have been instrumental in unraveling genetic variation in economically important traits in crop species (Tanksley, 1993; Frary *et al*., 2000; Paterson *et al*., 2000). There is, however, a large group of plant species, including wild relatives of current crop cultivars and forest trees, in which inbred lines are not available. These plants are recalcitrant to genomic research because of their high genetic heterozygosity, resulting from outcrossing and long generation times (Wu *et al*., 2000). For this reason, despite extensive development of high capacity methods for generating genomic data, it appears that such development is not well matched by statistical methods that can provide effective analyses of and inferences from these data.

For crop plants whose genetic composition can be homogenized via successive inbreeding, linkage analysis

0040-5809/02 $35.00

of molecular markers is straightforward because linkage phases over different marker loci and, thus, the distinction of recombinants and non-recombinants in progeny is known (Lander *et al*., 1987). In progeny populations (e.g., F_2 or backcross populations) derived from two inbred lines, biallelic marker systems, such as restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), or amplified fragment length polymorphism (AFLP), are sufficient to characterize polymorphic variation throughout the entire genome. However, this may not be sufficient in outcrossing species in which biallelic markers characterize only a portion of their genomes. As observed in a number of molecular genetics studies, outcrossing plant populations have high but variable numbers of alleles at genetic loci (Hamrick and Godt, 1990; Weber and Wong, 1993; Pfeiffer *et al*., 1997). Also, in such populations, there frequently occurs a variable allelic relationship (dominance or codominance) from locus to locus (see Xiao *et al*., 1995 for a discussion of this issue).

Linkage analyses in outcrossing species should make full use of molecular markers with multiple alleles or dominant alleles. Simple sequence repeats, such as microsatellites, consisting of tandemly repeated multiple copies of mono-, bi-, tri-, or tetranucleotide motifs, provide an ideal tool to characterize polymorphic variation in outcrossing populations (Olson *et al*., 1989). To increase the resolution of genome characterization for a given outcrossing species, these highly polymorphic microsatellites should be combined with dominant marker systems to construct a single consensus map (Barreneche *et al*., 1998; Paglia *et al*., 1998). Previously, two independent linkage maps for outcrossing species were constructed based on a so-called double pseudo-testcross strategy using dominant markers, each map having markers segregating in one parent but fixed in the other (Grattapaglia and Sederoff, 1994; Hemmat *et al*., 1994). From a statistical perspective, the construction of linkage maps based on a mixed set of different marker types is complicated because many genetic assumptions made in traditional linkage analyses (e.g., known parental linkage phases throughout the genome) do not hold.

Methods for determining linkage phases between marker loci with a number of alleles have been proposed by Ritter *et al*. (1990), Arus *et al*. (1994), and Ritter and Salamini (1996). More recently, Maliepaard *et al*. (1997) gave a complete classification of all possible marker genotypes or phenotypes in terms of the number and types of marker alleles and presented equations for estimating recombination fractions under particular linkage phases in both parents. They also gave the function of the LOD scores, expressed as the logarithm to base 10 of the ratio of the likelihood under the estimated recombination fraction and the likelihood under the null hypothesis of no linkage, to test whether a pair of markers is linked. The most likely linkage phase should correspond to a minimum estimate of the recombination fraction and a significant LOD score used to acclaim the existence of linkage. Using a similar principle, Ridout *et al*. (1998) proposed a three-point linkage analysis for markers with four different alleles between two parents, providing a statistical foundation for ordering different markers on a genetic map. A major limitation of these analyses is that linkage analysis between two or three adjacent markers is performed independently rather than based on a multi-locus linkage phase inference, which may not be effective and efficient in many situations, especially when the adjacent markers have incomplete information. An additional limitation of these analyses is that a general formula has not been derived for simultaneously estimating linkage and linkage phase for a mixed dataset of markers including all possible segregation patterns. In some cases, two criteria for characterizing a linkage phase—the minimum estimate of recombination fraction and the highest LOD score—may not be consistent, as shown in Maliepaard *et al*. (1997). In this article, we develop a general statistical algorithm for simultaneous estimation of recombination fractions and linkage phase configurations over all linked markers in a full-sib family. We first derive a two-locus marker model for linkage analysis, followed by the inference of multi-locus marker relationships. A simulated example is then used to demonstrate the statistical properties of our linkage model.

2. IDENTIFICATION OF SEGREGATION PATTERNS

2.1. Marker Type

In a full-sib family derived from two parents of an outcrossing species, up to four marker alleles, besides a null allele, may be segregating at a single locus. Furthermore, the number of alleles may vary over loci. We assume that each of the marker alleles, symbolized by a, b, c, and d, is codominant with respect to the others but dominant to the null allele, symbolized by o. We assume that all markers undergo Mendelian segregation without distortion. Depending on how different alleles are combined in the two parents used for the cross, there exists a total of 18 possible cross types for a marker locus (Table I). Based on both parental and offspring marker

TABLE I

Possible Marker Genotype Cross Combinations and Observed Marker Band Patterns for Parents and Their Offspring

Cross type			Parent			Offspring		
			Cross	Observed band	Remark	Observed bands	Segregation	No. phenotypes
A		1	$ab \times cd$	$ab \times cd$	Asymmetry	ac, ad, bc, bd	1:1:1:1	4
		2	$ab \times ac$	$ab \times ac$	Asymmetry	a, ac, ba, bc	1:1:1:1	4
		3	$ab \times co$	$ab \times c$	Asymmetry	ac, a, bc, b	1:1:1:1	4
		4	$ao \times bo$	$a \times b$	Asymmetry	ab, a, b, o	1:1:1:1	4
B	B_1	5	$ab \times ao$	$ab \times a$	Asymmetry	$ab, 2a, b$	1:2:1	3
	B_2	6	$ao \times ab$	$a \times ab$	Asymmetry	$ab, 2a, b$	1:2:1	3
	B_3	7	$ab \times ab$	$ab \times ab$	Symmetry	$a, 2ab, b$	1:2:1	3
C		8	$ao \times ao$	$a \times a$	Symmetry	$3a, o$	3:1	2
D	D_1	9	$ab \times cc$	$ab \times c$	Asymmetry	ac, bc	1:1	2
		10	$ab \times aa$	$ab \times a$	Asymmetry	a, ab	1:1	2
		11	$ab \times oo$	$ab \times o$	Asymmetry	a, b	1:1	2
		12	$bo \times aa$	$b \times a$	Asymmetry	ab, a	1:1	2
		13	$ao \times oo$	$a \times o$	Asymmetry	a, o	1:1	2
	D_2	14	$cc \times ab$	$c \times ab$	Asymmetry	ac, bc	1:1	2
		15	$aa \times ab$	$a \times ab$	Asymmetry	a, ab	1:1	2
		16	$oo \times ab$	$o \times ab$	Asymmetry	a, b	1:1	2
		17	$aa \times bo$	$a \times b$	Asymmetry	ab, a	1:1	2
		18	$oo \times ao$	$o \times a$	Asymmetry	a, o	1:1	2

Note. Marker types B_3 and C have the same genotypes between the two parents and, therefore, are called symmetrical marker cross types. The other marker types have parent-specific marker genotypes and are called asymmetrical marker cross types. For the partially informative markers whose cross types are asymmetrical between the two parents, the reciprocals, e.g., B_1 vs B_2 and D_1 vs D_2, supply different information for the characterization of linkage phase when these markers are paired and, thus, are presented as two distinct groups.

band patterns, these cross types can be classified into seven groups (see also Maliepaard *et al*., 1997):

A. Loci that are heterozygous in both parents and segregate in a 1:1:1:1 ratio, involving either four alleles $ab \times cd$, three non-null alleles $ab \times ac$, three non-null alleles and a null allele $ab \times co$, or two null alleles and two non-null alleles $ao \times bo$;
B. Loci that are heterozygous in both parents and segregate in a 1:2:1 ratio, which include three groups:
   B_1. One parent has two different dominant alleles and the other has one dominant allele and one null allele, e.g., $ab \times ao$;
   B_2. The reciprocal of B_1;
   B_3. Both parents have the same genotype of two codominant alleles, i.e., $ab \times ab$;
C. Loci that are heterozygous in both parents and segregate in a 3:1 ratio, i.e., $ao \times ao$;
D. Loci that are in the testcross configuration between the parents and segregate in a 1:1 ratio, which include two groups:
   D_1. Heterozygous in one parent and homozygous in the other, including three alleles $ab \times cc$, two alleles $ab \times aa$, $ab \times oo$ and $bo \times aa$, and one allele (with three null alleles) $ao \times oo$;
   D_2. The reciprocals of D_1.

Marker types B_3 and C have the same genotypes in both parents and, therefore, are called symmetrical marker cross types. The other marker types have parent-specific marker genotypes and are called asymmetrical marker cross types.

If we denote one of the parents by P, the other parent by Q, and the two chromosomes in each parent by P_1, P_2, Q_1 and Q_2, respectively, then, for a particular marker regardless of its marker type, there are four possible *parental chromosome pairings* (PCP) in the progeny, P_1Q_1, P_1Q_2, P_2Q_1, and P_2Q_2. Thus, the groups classified above can be defined in terms of the correspondence between the PCP genotypes and the observed phenotypes. Markers from group A have an exact correspondence between the PCPs and phenotypes

in the offspring (i.e., the four PCPs produce four different phenotypes) and, therefore, are fully informative. Markers from the other groups are regarded as partially informative. For the partially informative markers whose cross types are asymmetrical between the two parents, the reciprocals, e.g., B_1 vs B_2 and D_1 vs D_2, supply different information for the characterization of linkage phase when these markers are paired (see below), and, thus, are presented as two distinct groups.

To perform linkage analysis between two arbitrary markers, one should consider all possible combinations of these seven groups of markers. Theoretically, there are 28 such combinations. The combination of D_1 and D_2 does not, however, provide information to distinguish recombinant and non-recombinant offspring. Thus, linkage analysis cannot be performed between these two groups alone. However, their linkage relationships can be indirectly established through markers from other groups. To construct an integrated genetic map, therefore, a large number of markers from groups A–C should be available that are closely linked to markers from groups D_1 and D_2.

2.2. The Characterization of Marker Type

For a mapping project, both the parents and progeny are usually genotyped. Based on the segregation pattern of marker band data, we can determine the cross type to which a given marker belongs and re-code the marker data accordingly (Table I). We assume that the size of the progeny is large enough so that cross type can be determined unambiguously from observed marker genotype ratios.

Many marker types can be easily determined with no use of parental information. However, to distinguish groups D_1 vs D_2, the marker information of the two parents is needed. If the information of the two parents is missing or incomplete for these two marker cross types, a posterior probability is calculated for the cross types, assuming equal prior probabilities for each possible cross type. If the posterior probability for a particular cross type is large ($\geqslant 0.95$), that cross type is assumed to be the true cross type.

3. MODEL

3.1. Labeling of Parental Chromosomes

Linkage phase describes the configuration of alleles at a pair of heterozygous loci on homologous chromosomes in a single parent. The linkage phase between any two linked markers can be determined if we know what alternative allele one of the homologous chromosomes carries for each marker in a parent. Thus, the question of determining linkage phase becomes a question of labeling parental chromosomes using the alleles at given markers. The association of the marker alleles and homologous chromosomes can be anchored by calculating the probabilities of the PCPs of a marker conditional on the state at linked markers in the full-sib family.

Consider a linkage group of m markers, $\mathsf{M}^1, \mathsf{M}^2, \ldots, \mathsf{M}^m$, with a known order. For the first marker M^1 (at the top of the group in a vertical arrangement), the parental chromosomes can be arbitrarily labeled using its alleles. Assume that the parental chromosomes for the first marker are labeled as $P_1| \, |P_2$ and $Q_1| \, |Q_2$, where $| \, |$ stands for two homologous chromosomes on the left and right, respectively. The linkage phase between the alleles of the first marker and the second marker (M^2) can be determined by assigning the alternative alleles of the second marker to a different homologous chromosome given the defined label of the first marker. It is not difficult to imagine that there are a total of four legitimate assignments (A_ω, $\omega = 1, \ldots, 4$) for the two parents, as illustrated below:

$$
\begin{array}{lccc}
 & P & \times & Q \\
\mathsf{A}_1 & \begin{matrix} P_1^1 \,||\, P_2^1 \\ P_1^2 \,||\, P_2^2 \end{matrix} & & \begin{matrix} Q_1^1 \,||\, Q_2^1 \\ Q_1^2 \,||\, Q_2^2 \end{matrix} \\
\mathsf{A}_2 & \begin{matrix} P_1^1 \,||\, P_2^1 \\ P_1^2 \,||\, P_2^2 \end{matrix} & & \begin{matrix} Q_1^1 \,||\, Q_2^1 \\ Q_2^2 \,||\, Q_1^2 \end{matrix} \\
\mathsf{A}_3 & \begin{matrix} P_1^1 \,||\, P_2^1 \\ P_2^2 \,||\, P_1^2 \end{matrix} & & \begin{matrix} Q_1^1 \,||\, Q_2^1 \\ Q_1^2 \,||\, Q_2^2 \end{matrix} \\
\mathsf{A}_4 & \begin{matrix} P_1^1 \,||\, P_2^1 \\ P_2^2 \,||\, P_1^2 \end{matrix} & & \begin{matrix} Q_1^1 \,||\, Q_2^1 \\ Q_2^2 \,||\, Q_1^2 \end{matrix}
\end{array}
.
$$

In each assignment, the label of the first marker is kept unchanged. An appropriate assignment (and therefore the linkage phases for both parents) can be determined by calculating its posterior probability given the data.

3.2. Anchoring Algorithm

We assume that the two parents have the same recombination fractions for the corresponding markers.

Denote the recombination fraction between the first and second markers, M^1 and M^2, by ϑ^{12}. The conditional probability of the PCP genotype of M^2 in the progeny given the PCP genotype of M^1 is the transition probability of recombination events (h_ω^{12}) between the two markers at the PCP genotype level. The transition probability and the number (d_ω^{12}) of recombination events occurring between the two markers under the ωth assignment A_ω can be expressed in the matrix form:

$$\mathbf{H}_\omega^{12} \qquad\qquad \mathbf{D}_\omega^{12}$$

$$A_1:\quad \begin{array}{c} \\ P_1^1Q_1^1 \\ P_1^1Q_2^1 \\ P_2^1Q_1^1 \\ P_2^1Q_2^1 \end{array} \begin{array}{c} \begin{array}{cccc} P_1^2Q_1^2 & P_1^2Q_2^2 & P_2^2Q_1^2 & P_2^2Q_2^2 \end{array} \\ \begin{bmatrix} (1-\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) & \vartheta^{12}(1-\vartheta^{12}) & (\vartheta^{12})^2 \\ \vartheta^{12}(1-\vartheta^{12}) & (1-\vartheta^{12})^2 & (\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) \\ \vartheta^{12}(1-\vartheta^{12}) & (\vartheta^{12})^2 & (1-\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) \\ (\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) & \vartheta^{12}(1-\vartheta^{12}) & (1-\vartheta^{12})^2 \end{bmatrix} \end{array} \begin{array}{c} \begin{array}{cccc} P_1^2Q_1^2 & P_1^2Q_2^2 & P_2^2Q_1^2 & P_2^2Q_2^2 \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{bmatrix} \end{array},$$

$$A_2:\quad \begin{array}{c} \\ P_1^1Q_1^1 \\ P_1^1Q_2^1 \\ P_2^1Q_1^1 \\ P_2^1Q_2^1 \end{array} \begin{array}{c} \\ \begin{bmatrix} \vartheta^{12}(1-\vartheta^{12}) & (1-\vartheta^{12})^2 & (\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) \\ (1-\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) & \vartheta^{12}(1-\vartheta^{12}) & (\vartheta^{12})^2 \\ (\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) & \vartheta^{12}(1-\vartheta^{12}) & (1-\vartheta^{12})^2 \\ \vartheta^{12}(1-\vartheta^{12}) & (\vartheta^{12})^2 & (1-\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) \end{bmatrix} \end{array} \begin{array}{c} \\ \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 2 \\ 2 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{bmatrix} \end{array},$$

$$A_3:\quad \begin{array}{c} \\ P_1^1Q_1^1 \\ P_1^1Q_2^1 \\ P_2^1Q_1^1 \\ P_2^1Q_2^1 \end{array} \begin{array}{c} \\ \begin{bmatrix} \vartheta^{12}(1-\vartheta^{12}) & (\vartheta^{12})^2 & (1-\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) \\ (\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) & \vartheta^{12}(1-\vartheta^{12}) & (1-\vartheta^{12})^2 \\ (1-\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) & \vartheta^{12}(1-\vartheta^{12}) & (\vartheta^{12})^2 \\ \vartheta^{12}(1-\vartheta^{12}) & (1-\vartheta^{12})^2 & (\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) \end{bmatrix} \end{array} \begin{array}{c} \\ \begin{bmatrix} 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \\ 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \end{bmatrix} \end{array},$$

$$A_4:\quad \begin{array}{c} \\ P_1^1Q_1^1 \\ P_1^1Q_2^1 \\ P_2^1Q_1^1 \\ P_2^1Q_2^1 \end{array} \begin{array}{c} \\ \begin{bmatrix} (\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) & \vartheta^{12}(1-\vartheta^{12}) & (1-\vartheta^{12})^2 \\ \vartheta^{12}(1-\vartheta^{12}) & (\vartheta^{12})^2 & (1-\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) \\ \vartheta^{12}(1-\vartheta^{12}) & (1-\vartheta^{12})^2 & (\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) \\ (1-\vartheta^{12})^2 & \vartheta^{12}(1-\vartheta^{12}) & \vartheta^{12}(1-\vartheta^{12}) & (\vartheta^{12})^2 \end{bmatrix} \end{array} \begin{array}{c} \\ \begin{bmatrix} 2 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \\ 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix} \end{array},$$

Let A_ω^2 be the ωth assignment of the alleles of the second marker M^2 given the fixed chromosomal labeling of the first marker M^1 and $\mathbf{M} = (M_1, M_2)^T$ be the data for the two markers. According to Bayes's theorem, the posterior probability of A_ω^2 given $\mathbf{M}$ is expressed as

$$P(A_\omega^2 \mid \mathbf{M}) = \frac{P(A_\omega^2)P(\mathbf{M} \mid A_\omega^2)}{\sum_{\omega=1}^{4} P(A_\omega^2)P(\mathbf{M} \mid A_\omega^2)} = \frac{P(\mathbf{M} \mid A_\omega^2)}{\sum_{\omega=1}^{4} P(\mathbf{M} \mid A_\omega^2)}, \tag{1}$$

where $P(A_\omega^2)$ is the unconditional or prior probability of A_ω^2 and is assumed to be uniform, and $P(\mathbf{M} \mid A_\omega^2)$ is the likelihood of $\mathbf{M}$ given A_ω^2. Assuming that the marker data are independent among N offspring, we have

$$\begin{aligned} P(\mathbf{M} \mid A_\omega^2) &= \prod_{j=1}^{N} P(M_j \mid A_\omega^2) \\ &= \prod_{j=1}^{N} \sum_{i_1=1}^{p_1} \sum_{i_2=1}^{p_2} P(M_{1j} \mid (PQ)_{i_1}) P((PQ)_{i_2} \mid (PQ)_{i_1}) \\ &\quad P(M_{2j} \mid (PQ)_{i_2}), \end{aligned} \tag{2}$$

where p_k is the number of distinguishable genotypes (phenotypes) in the offspring at marker M^k, which is 4, 3, 3, 3, 2, 2, and 2 for marker types A, B_1, B_2, B_3, C, D_1, and D_2 (Table I), respectively; $P(M_{kj} \mid (PQ)_{i_k})$ is the indicator variable describing the i_kth PCP phenotype of marker M^k for offspring j, which is one if the phenotype of offspring j at marker M^k (M_{kj}) is consistent with $(PQ)_{i_k}$, zero otherwise; and $P((PQ)_{i_2} \mid (PQ)_{i_1})$ is the conditional (transition) probability of the i_2th PCP phenotype of marker M^2 given the i_1th PCP of marker M^1, as described above for different assignments.

Equation (2) can be written in matrix form as

$$P(\mathbf{M} \mid \mathsf{A}_\omega^2) = \prod_{j=1}^{N} \mathbf{m}_{i_1 j}^{\mathrm{T}} \mathbf{I}_{p_1}^{\mathrm{T}} \mathbf{H}_\omega^{12} \mathbf{I}_{p_2} \mathbf{m}_{i_2 j}$$
$$= \prod_{j=1}^{N} \mathbf{m}_{i_1 j}^{\mathrm{T}} \mathbf{P}_\omega^{12} \mathbf{m}_{i_2 j}, \qquad (3)$$

where $\mathbf{m}_{i_k j}$ is the p_k-dimensional vector of the indicator variable $P(M_{kj} \mid (PQ)_{i_k})$ for marker M^k; $\mathbf{I}_{p_k}$ is a $(4 \times p_k)$ incidence matrix relating the PCP genotypes to phenotypes, which can be designed for each of the four allelic assignments as

$$\mathbf{I}_{p_{\mathrm{A}}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

for the marker cross type A,

$$\mathbf{I}_{p_{\mathrm{B}_1}}^{\mathrm{T}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

for the marker cross type B_1,

$$\mathbf{I}_{p_{\mathrm{B}_2}}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

for the marker cross type B_2,

$$\mathbf{I}_{p_{\mathrm{B}_3}}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

for the marker cross type B_3,

$$\mathbf{I}_{p_{\mathrm{C}}}^{\mathrm{T}} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

for the marker cross type C,

$$\mathbf{I}_{p_{\mathrm{D}_1}}^{\mathrm{T}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

for the marker cross type D_1, and

$$\mathbf{I}_{p_{\mathrm{D}_2}}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

for the marker cross type D_2, $\mathbf{H}_\omega^{12}$ is the (4×4) matrix of the transition probability of the PCP genotype from markers M^1 to M^2, $\mathbf{P}_\omega^{12} = \mathbf{I}_{p_1}^{\mathrm{T}} \mathbf{H}_\omega^{12} \mathbf{I}_{p_2}$ is a $(p_1 \times p_2)$ matrix of the transition probability from markers M^1 to M^2 at the phenotype level, and T denotes the transpose of a vector or matrix.

For a particular assignment for the second marker M^2, the recombination fraction between markers M^1 and M^2 can be estimated by maximizing the log-likelihood of Eq. (3) for ϑ^{12} and solving the likelihood equation

$$\frac{\partial}{\partial \vartheta^{12}} \ln[P(\mathbf{M} \mid \mathsf{A}_\omega^2)] = 0.$$

To obtain the MLE for ϑ^{12}, an iterative procedure based on expectation (E) and maximization (M) algorithms can be used for parameter estimation (Dempster *et al.*, 1977; Lander and Green, 1987). The general equations underlying the $\{\tau + 1\}$th EM step are given as follows:

E Step. At step τ, using the matrix $\mathbf{H}_\omega^{12(\tau)}$ based on the current estimate $\vartheta_\omega^{12(\tau)}$, calculate the expected number of recombination events between markers M^1 and M^2 for offspring j under assignment A_u^2,

$$c_{i_1 i_2 j}^{\omega\{\tau+1\}} = \frac{\mathbf{m}_{i_1 j}^{\mathrm{T}} [\mathbf{I}_{p_1}^{\mathrm{T}} (\mathbf{D}_\omega^{12} \circ \mathbf{H}_\omega^{12\{\tau\}}) \mathbf{I}_{p_2}] \mathbf{m}_{i_2 j}}{\mathbf{m}_{i_1 j}^{\mathrm{T}} \mathbf{P}_\omega^{12\{\tau\}} \mathbf{m}_{i_2 j}}, \qquad (4)$$

where $\mathbf{D}_\omega^{12}$ is a (4×4) matrix with elements given above and $\circ$ denotes an elementwise product of two matrices.

M Step. Calculate $\vartheta_\omega^{12\{\tau+1\}}$ under assignment A_ω^2 using the equation

$$\vartheta_\omega^{12\{\tau+1\}} = \frac{1}{2N} \sum_{j=1}^{N} \sum_{i_1=1}^{p_1} \sum_{i_2=1}^{p_2} c_{i_1 i_2 j}^{\omega\{\tau+1\}}. \qquad (5)$$

These iterative procedures are repeated between Eqs. (4) and (5) until ϑ_ω^{12} converges to a stable value. This stable value represents the MLE of the recombination fraction between markers M^1 and M^2 for assignment A_ω^2.

The most likely linkage phase between markers M^1 and M^2 can be inferred by calculating the posterior probabilities of the four different assignments A_ω^2 based on Eq. (1). The assignment corresponding to the highest posterior probability represents the most likely inference for the linkage phase between the two markers. It should be noted that there is some uncertainty in characterizing the assignments between two markers from a symmetrical cross type (B_3 or C; see Table 1), although such uncertainty can be removed by pairing them with a marker of non-symmetrical cross type except D_1 or D_2 (see below). When two markers of symmetrical cross type are paired, one cannot distinguish between assignments A_2 and A_3 because those two assignments have the same matrix for the transition probability at the phenotypic level (see Eq. (3)). If these two assignments have the highest probability the reasonable solution is to arbitrarily choose one of them for inclusion in the next two-point analysis with the following marker. In

addition, assignments A_1 and A_2, as well as A_3 and A_4, are not distinguishable for the linkage analysis of marker type D_1 to any other types. The same is also true between assignments A_1 and A_3 as well as A_2 and A_4 for marker type D_2. However, these uncertainties for marker types D_1 and D_2 do not affect our inferences because the same allele at such a marker is contained in one of the two parents. The least informative inference for linkage analysis exists between marker types D_1 and D_2, between which recombinants cannot be identified.

After the linkage phase between the first and second markers M^1 and M^2 is determined, one should move to analyze the linkage phase between the second and third markers M^2 and M^3 based on the data of these three markers. However, if there is no interference between the marker intervals, the posterior probability of an assignment of this marker is calculated just based on the estimated PCP of the second marker according to the memoryless property of Markov chains. This chain proceeds forward across a chromosome until the last two markers. Strictly speaking, an effective application of the Markov property relies upon whether or not the assignment of the previous adjacent marker has been determined precisely. If the previous marker is a symmetrical cross type (i.e., marker groups B_3 or C), the use of this property may be frustrated due to missing information. In this case, as pointed out by Jiang and Zeng (1997), the determination of the assignment of a given marker should include both the previous adjacent marker and the nearest fully informative marker on the left of the marker under consideration.

3.3. Joint Two-Point Analysis Based on a Hidden Markov Model

For the m markers in the linkage group, a more robust result for linkage phase and recombination fraction estimates can be expected when all these markers are considered simultaneously. By arbitrarily labeling the parental chromosomes for the first marker, one will have 4^{m-1} combinations of the assignments for the rest of the $(m-1)$ markers. The most likely assignment for each of the $(m-1)$ markers is determined by comparing the posterior probabilities for all these combinations calculated by

$$P(\mathsf{A}_{\omega_2\omega_3\ldots\omega_m} \mid \mathbf{M}) = \frac{P(\mathbf{M} \mid \mathsf{A}_{\omega_2\omega_3\ldots\omega_m})}{\sum_{\omega_2}^{4}\sum_{\omega_3}^{4}\cdots\sum_{\omega_m}^{4} P(\mathbf{M} \mid \mathsf{A}_{\omega_2\omega_3\ldots\omega_m})},$$

where $\mathsf{A}_{\omega_2\omega_3\ldots\omega_m}$ is the combination of the ω_2th assignment for marker M^2, the ω_3th assignment for marker $M^3, \ldots,$ and the ω_mth assignment for marker M^m, $\mathbf{M} = (M_1, M_2, \ldots, M_m)^{\mathrm{T}}$ is the data for all m markers, and $P(\mathbf{M} \mid \mathsf{A}_{\omega_2\omega_3\ldots\omega_m})$ is the likelihood of the data given $\mathsf{A}_{\omega_2\omega_3\ldots\omega_m}$, which is calculated through a hidden Markov chain model in a recursive manner as outlined below. Assuming that the marker data are independent among N offspring and that there is no interference in recombination between the marker intervals, we have

$$\begin{aligned} P(\mathbf{M} \mid \mathsf{A}_{\omega_2\omega_3\ldots\omega_m}) &= \prod_{j=1}^{N}\sum_{i_1=1}^{4}\cdots\sum_{i_m=1}^{4}\prod_{k=1}^{m} P(M_{kj} \mid (PQ)_{i_k}) \\ &\quad \prod_{k=2}^{m} P((PQ)_{i_k} \mid (PQ)_{i_{k-1}}) \\ &= \prod_{j=1}^{N}\prod_{k=2}^{m} [\mathbf{m}_{i_{k-1}j}^{\mathrm{T}}(\mathbf{I}_{p_{k-1}}^{\mathrm{T}}\mathbf{H}_{\omega_k}^{(k-1)k}\mathbf{I}_{p_k})\mathbf{m}_{i_kj}] \\ &= \prod_{j=1}^{N}\prod_{k=2}^{m} [\mathbf{m}_{i_{k-1}j}^{\mathrm{T}}\mathbf{P}_{\omega_k}^{(k-1)k}\mathbf{m}_{i_kj}], \end{aligned}$$

where the indicator variables, transition probabilities, vectors and matrices are defined as above. The recombination fraction between markers M^{k-1} and M^k ($\vartheta^{(k-1)k}$) can be estimated through iterative EM-algorithms similar to Eqs. (4) and (5) using the most likely allelic assignment.

3.4. Three-Point Analysis

Statistical algorithms for estimating the recombination fraction based on a two-point analysis may not be powerful, especially in the case where partially informative markers are involved. Ridout *et al.* (1998) demonstrated an example in which three-point analysis can detect more linkage relationships between three loci than two-point analysis.

Consider three markers in order, M^1, M^2, and M^3. Assuming that the allele assignment in the two parents is fixed for marker M^1, there are $4 \times 4 = 16$ possible assignments ($\mathsf{A}^{23}_{\omega_2\omega_3}$) for markers M^2 and M^3. Let ϑ^{12} denote the recombination fraction between markers M^1 and M^2, with ϑ^{23} and ϑ^{13} defined similarly. These recombination fractions are associated with the probabilities with which a crossover occurs between markers M^1 and M^2 and between markers M^2 and M^3. The event that a crossover or no crossover occurs in each interval is denoted by G^{11} and G^{00}, respectively, whereas the event that a crossover occurs only in the first interval or in the second interval is denoted by G^{10} and G^{01}, respectively. The probabilities of these events are denoted by g^{00}, g^{01}, g^{10}, and g^{11}, respectively, whose sum equals 1. According to the definition of recombination fraction as the probability of a crossover between a pair of loci, it is clear that $\vartheta^{12} = g_{10} + g_{11}$,

$\vartheta^{23} = g_{01} + g_{11}$, and $\vartheta^{13} = g_{01} + g_{10}$, and

$$g^{11} = \tfrac{1}{2}(\vartheta^{12} + \vartheta^{23} - \vartheta^{13}),$$
$$g^{10} = \tfrac{1}{2}(\vartheta^{12} + \vartheta^{13} - \vartheta^{23}),$$
$$g^{01} = \tfrac{1}{2}(\vartheta^{23} + \vartheta^{13} - \vartheta^{12}),$$
$$g^{00} = 1 - \tfrac{1}{2}(\vartheta^{12} + \vartheta^{13} + \vartheta^{23}).$$

For three-point analysis, there are a total of 16 (16×4)-transition probability matrices from markers M^1 to M^2 and M^3, denoted by $\mathbf{H}^{123}_{\omega_2\omega_3}$, each corresponding to a particular assignment $\mathsf{A}^{23}_{\omega_2\omega_3}$ conditional on fixing the parental status of M^1. Similarly, there are 16 (16×4) matrices for the numbers of crossovers that have occurred for G^{00}, G^{01}, G^{10}, and G^{11} among the three markers M^1, M^2, and M^3 under an assumed assignment, denoted by $\mathbf{G}^{00}_{\omega_2\omega_3}$, $\mathbf{G}^{01}_{\omega_2\omega_3}$, $\mathbf{G}^{10}_{\omega_2\omega_3}$, and $\mathbf{G}^{11}_{\omega_2\omega_3}$, respectively.

For a given assignment $\mathsf{A}^{23}_{\omega_2\omega_3}$, the probability of the marker data is expressed as

$$P(\mathbf{M} \mid \mathsf{A}^{23}_{\omega_2\omega_3}) = \prod_{j=1}^{N} [\mathbf{m}^{\mathrm{T}}_{i_1 j} \otimes \mathbf{m}^{\mathrm{T}}_{i_2 j}][\mathbf{I}^{\mathrm{T}}_{p_1} \otimes \mathbf{I}^{\mathrm{T}}_{p_2}]\mathbf{H}^{123}_{\omega_2\omega_3}\mathbf{I}_{p_3}\mathbf{m}_{i_3 j}, \quad (6)$$

where $\otimes$ denotes the Kronecker product and the vectors and matrices are as defined for the two-point analysis. If the assignment for the three markers is

$$\begin{matrix} P^1_1 \\ P^2_1 \\ P^3_1 \end{matrix} \left\| \begin{matrix} P^1_2 \\ P^2_2 \\ P^3_2 \end{matrix} \right. \times \left\| \begin{matrix} Q^1_1 \\ Q^2_1 \\ Q^3_1 \end{matrix} \right\| \begin{matrix} Q^1_2 \\ Q^2_2 \\ Q^3_2 \end{matrix},$$

with the allelic assignment for markers M^2 and M^3 denoted by A^{23}_{11} according to the definition for two-point analysis, the transition probability matrix ($\mathbf{H}^{123}_{11}$) and the matrices for the numbers of crossovers ($\mathbf{G}^{00}_{11}$, $\mathbf{G}^{01}_{11}$, $\mathbf{G}^{10}_{11}$, and $\mathbf{G}^{11}_{11}$,) under this assignment are given in Table II.

The EM algorithm is used to obtain the MLE for the recombination fractions between the three markers. The general equations formulating the iteration of the $\{\tau + 1\}$th EM step are given as follows:

E Step. Calculate the expected number of recombination events associated with $G^{00}(\alpha)$, $G^{01}(\beta)$, $G^{01}(\gamma)$ and $G^{11}(\delta)$ for offspring j under a given assignment, respectively:

$$\alpha^{\omega_2\omega_3\{\tau+1\}}_{i_1 i_2 i_3 j} = \frac{[\mathbf{m}^{\mathrm{T}}_{i_1 j} \otimes \mathbf{m}^{\mathrm{T}}_{i_2 j}][\mathbf{I}^{\mathrm{T}}_{p_1} \otimes \mathbf{I}^{\mathrm{T}}_{p_2}][(\mathbf{G}^{00}_{\omega_2\omega_3} \circ \mathbf{H}^{123\{\tau\}}_{\omega_2\omega_3})\mathbf{I}_{p_3}]\mathbf{m}_{i_3 j}}{[\mathbf{m}^{\mathrm{T}}_{i_1 j} \otimes \mathbf{m}^{\mathrm{T}}_{i_2 j}][\mathbf{I}^{\mathrm{T}}_{p_1} \otimes \mathbf{I}^{\mathrm{T}}_{p_2}][\mathbf{H}^{123\{\tau\}}_{\omega_2\omega_3}\mathbf{I}_{p_3}]\mathbf{m}_{i_3 j}} \quad (7a)$$

$$\beta^{\omega_2\omega_3\{\tau+1\}}_{i_1 i_2 i_3 j} = \frac{[\mathbf{m}^{\mathrm{T}}_{i_1 j} \otimes \mathbf{m}^{\mathrm{T}}_{i_2 j}][\mathbf{I}^{\mathrm{T}}_{p_1} \otimes \mathbf{I}^{\mathrm{T}}_{p_2}][(\mathbf{G}^{01}_{\omega_2\omega_3} \circ \mathbf{H}^{123\{\tau\}}_{\omega_2\omega_3})\mathbf{I}_{p_3}]\mathbf{m}_{i_3 j}}{[\mathbf{m}^{\mathrm{T}}_{i_1 j} \otimes \mathbf{m}^{\mathrm{T}}_{i_2 j}][\mathbf{I}^{\mathrm{T}}_{p_1} \otimes \mathbf{I}^{\mathrm{T}}_{p_2}][\mathbf{H}^{123\{\tau\}}_{\omega_2\omega_3}\mathbf{I}_{p_3}]\mathbf{m}_{i_3 j}} \quad (7b)$$

$$\gamma^{\omega_2\omega_3\{\tau+1\}}_{i_1 i_2 i_3 j} = \frac{[\mathbf{m}^{\mathrm{T}}_{i_1 j} \otimes \mathbf{m}^{\mathrm{T}}_{i_2 j}][\mathbf{I}^{\mathrm{T}}_{p_1} \otimes \mathbf{I}^{\mathrm{T}}_{p_2}][(\mathbf{G}^{10}_{\omega_2\omega_3} \circ \mathbf{H}^{123\{\tau\}}_{\omega_2\omega_3})\mathbf{I}_{p_3}]\mathbf{m}_{i_3 j}}{[\mathbf{m}^{\mathrm{T}}_{i_1 j} \otimes \mathbf{m}^{\mathrm{T}}_{i_2 j}][\mathbf{I}^{\mathrm{T}}_{p_1} \otimes \mathbf{I}^{\mathrm{T}}_{p_2}][\mathbf{H}^{123\{\tau\}}_{\omega_2\omega_3}\mathbf{I}_{p_3}]\mathbf{m}_{i_3 j}} \quad (7c)$$

$$\delta^{\omega_2\omega_3\{\tau+1\}}_{i_1 i_2 i_3 j} = \frac{[\mathbf{m}^{\mathrm{T}}_{i_1 j} \otimes \mathbf{m}^{\mathrm{T}}_{i_2 j}][\mathbf{I}^{\mathrm{T}}_{p_1} \otimes \mathbf{I}^{\mathrm{T}}_{p_2}][(\mathbf{G}^{11}_{\omega_2\omega_3} \circ \mathbf{H}^{123\{\tau\}}_{\omega_2\omega_3})\mathbf{I}_{p_3}]\mathbf{m}_{i_3 j}}{[\mathbf{m}^{\mathrm{T}}_{i_1 j} \otimes \mathbf{m}^{\mathrm{T}}_{i_2 j}][\mathbf{I}^{\mathrm{T}}_{p_1} \otimes \mathbf{I}^{\mathrm{T}}_{p_2}][\mathbf{H}^{123\{\tau\}}_{\omega_2\omega_3}\mathbf{I}_{p_3}]\mathbf{m}_{i_3 j}}. \quad (7d)$$

M Step. Calculate $g^{00\{\tau+1\}}$, $g^{01\{\tau+1\}}$, $g^{10\{\tau+1\}}$, and $g^{11\{\tau+1\}}$ for this assignment by solving the linear equations

$$\begin{bmatrix} \mathrm{A}^{12} & \mathrm{B}^{12} & \Gamma^{12} & \Delta^{12} \\ \mathrm{A}^{13} & \mathrm{B}^{13} & \Gamma^{13} & \Delta^{13} \\ \mathrm{A}^{23} & \mathrm{B}^{23} & \Gamma^{23} & \Delta^{23} \end{bmatrix} \begin{bmatrix} g^{00} \\ g^{01} \\ g^{10} \\ g^{11} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (8)$$

where

$$\mathrm{A}^{12} = \sum_{j=1}^{N} \sum_{\substack{i_1=1 \\ i_1=i_2}}^{p_1} \sum_{i_2=1}^{p_2} \sum_{i_3=1}^{p_3} \alpha^{\{\tau+1\}}_{i_1 i_2 i_3 j} - \sum_{j=1}^{N} \sum_{\substack{i_1=1 \\ i_1 \neq i_2}}^{p_1} \sum_{i_2=1}^{p_2} \sum_{i_3=1}^{p_3} \alpha^{\{\tau+1\}}_{i_1 i_2 i_3 j},$$

$$\mathrm{A}^{13} = \sum_{j=1}^{N} \sum_{\substack{i_1=1 \\ i_1=i_3}}^{p_1} \sum_{i_2=1}^{p_2} \sum_{i_3=1}^{p_3} \alpha^{\{\tau+1\}}_{i_1 i_2 i_3 j} - \sum_{j=1}^{N} \sum_{\substack{i_1=1 \\ i_1 \neq i_3}}^{p_1} \sum_{i_2=1}^{p_2} \sum_{i_3=1}^{p_3} \alpha^{\{\tau+1\}}_{i_1 i_2 i_3 j},$$

$$\mathrm{A}^{23} = \sum_{j=1}^{N} \sum_{\substack{i_1=1 \\ i_2=i_3}}^{p_1} \sum_{i_2=1}^{p_2} \sum_{i_3=1}^{p_3} \alpha^{\{\tau+1\}}_{i_1 i_2 i_3 j} - \sum_{j=1}^{N} \sum_{\substack{i_1=1 \\ i_2 \neq i_3}}^{p_1} \sum_{i_2=1}^{p_2} \sum_{i_3=1}^{p_3} \alpha^{\{\tau+1\}}_{i_1 i_2 i_3 j},$$

with $\mathrm{B}^{kk'}$, $\Gamma^{kk'}$ and $\Delta^{kk'}$ defined by replacing α by β, γ and δ, respectively, in the above. Because $g^{11} = 1 - g^{00} - g^{01} - g^{10}$, Eq. (8) can be reduced to

$$\begin{bmatrix} \mathrm{A}^{12} - \Delta^{12} & \mathrm{B}^{12} - \Delta^{12} & \Gamma^{12} - \Delta^{12} \\ \mathrm{A}^{13} - \Delta^{13} & \mathrm{B}^{13} - \Delta^{13} & \Gamma^{13} - \Delta^{13} \\ \mathrm{A}^{23} - \Delta^{23} & \mathrm{B}^{23} - \Delta^{23} & \Gamma^{23} - \Delta^{23} \end{bmatrix} \begin{bmatrix} g^{00} \\ g^{01} \\ g^{10} \end{bmatrix} = \begin{bmatrix} -\Delta^{12} \\ -\Delta^{13} \\ -\Delta^{23} \end{bmatrix}.$$

Using $\mathbf{F}$ to denote the matrix, $\mathbf{g}$ the vector on the left, and Δ the vector on the right, we have

$$\mathbf{g} = \mathbf{F}^{-1}\Delta. \quad (9)$$

The E and M steps are repeated until g_{00}, g_{01}, g_{10}, and g_{11} converge to stable values, which are the MLEs of these g's for the corresponding assignment. The MLEs of the g's can be transformed to give the MLEs of recombination fractions ϑ^{12}, ϑ^{13} and ϑ^{23}, because MLEs are invariant under parameter transformation. Because all possible recombination fractions between the three markers are estimated, the three-point analysis provides important information about marker ordering. For example, if either $\hat{\vartheta}^{12}$ or $\hat{\vartheta}^{23}$ estimated under an assumed marker order M^1–M^2–M^3 is greater than $\hat{\vartheta}^{13}$, then the

TABLE II

Transition Probability Matrix and Matrices for the Number of Crossovers among Different PCP Genotypes of Three Markers with a Known Order M^1–M^3 **for a Particular Allelic Assignment** $\begin{array}{c\|\|c} P_1^1 & P_2^1 \\ P_1^2 & P_2^2 \\ P_1^3 & P_2^3 \end{array} \times \begin{array}{c\|\|c} Q_1^1 & Q_2^1 \\ Q_1^2 & Q_2^2 \\ Q_1^3 & Q_2^3 \end{array}$

		M^3																			
						Matrix for the number of crossovers															
		Transition probability matrix				$\mathbf{G}_{11}^{00}$				$\mathbf{G}_{11}^{01}$				$\mathbf{G}_{11}^{10}$				$\mathbf{G}_{11}^{11}$			
M^1	M^2	$P_1^3Q_1^3$	$P_1^3Q_2^3$	$P_2^3Q_1^3$	$P_2^3Q_2^3$	$P_1^3Q_1^3$	$P_1^3Q_2^3$	$P_2^3Q_1^3$	$P_2^3Q_2^3$	$P_1^3Q_1^3$	$P_1^3Q_2^3$	$P_2^3Q_1^3$	$P_2^3Q_2^3$	$P_1^3Q_1^3$	$P_1^3Q_2^3$	$P_2^3Q_1^3$	$P_2^3Q_2^3$	$P_1^3Q_1^3$	$P_1^3Q_2^3$	$P_2^3Q_1^3$	$P_2^3Q_2^3$
$P_1^1Q_1^1$	$P_1^2Q_1^2$	$(g^{00})^2$	$g^{00}g^{01}$	$g^{00}g^{01}$	$(g^{01})^2$	2	1	1	0	0	1	1	2	0	0	0	0	0	0	0	0
	$P_1^2Q_2^2$	$g^{00}g^{11}$	$g^{00}g^{10}$	$g^{01}g^{11}$	$g^{01}g^{10}$	1	1	0	0	0	0	1	1	0	1	0	1	1	0	1	0
	$P_2^2Q_1^2$	$g^{00}g^{11}$	$g^{01}g^{11}$	$g^{00}g^{10}$	$g^{01}g^{10}$	1	0	1	0	0	1	0	1	0	0	1	1	1	1	0	0
	$P_2^2Q_2^2$	$(g^{11})^2$	$g^{10}g^{11}$	$g^{10}g^{11}$	$(g^{10})^2$	0	0	0	0	0	0	0	0	0	1	1	2	2	1	1	0
$P_1^1Q_2^1$	$P_1^2Q_1^2$	$g^{00}g^{10}$	$g^{00}g^{11}$	$g^{01}g^{10}$	$g^{01}g^{11}$	1	1	0	0	0	0	1	1	1	0	1	0	0	1	0	1
	$P_1^2Q_2^2$	$g^{00}g^{01}$	$(g^{00})^2$	$(g^{01})^2$	$g^{00}g^{01}$	1	2	0	1	1	0	2	1	0	0	0	0	0	0	0	0
	$P_2^2Q_1^2$	$g^{10}g^{11}$	$(g^{11})^2$	$(g^{10})^2$	$g^{10}g^{11}$	0	0	0	0	0	0	0	0	1	0	2	1	1	2	0	1
	$P_2^2Q_2^2$	$g^{01}g^{11}$	$g^{00}g^{11}$	$g^{01}g^{10}$	$g^{00}g^{10}$	0	1	0	1	1	0	1	0	0	0	1	1	1	1	0	0
$P_2^1Q_1^1$	$P_1^2Q_1^2$	$g^{00}g^{10}$	$g^{01}g^{10}$	$g^{00}g^{11}$	$g^{01}g^{11}$	1	0	1	0	0	1	0	1	1	1	0	0	0	0	1	1
	$P_1^2Q_2^2$	$g^{10}g^{11}$	$(g^{10})^2$	$(g^{11})^2$	$g^{10}g^{11}$	0	0	0	0	0	0	0	0	1	2	0	1	1	0	2	1
	$P_2^2Q_1^2$	$g^{00}g^{01}$	$(g^{01})^2$	$(g^{00})^2$	$g^{00}g^{01}$	1	0	2	1	1	2	0	1	0	0	0	0	0	0	0	0
	$P_2^2Q_2^2$	$g^{01}g^{11}$	$g^{01}g^{10}$	$g^{00}g^{11}$	$g^{00}g^{10}$	0	0	1	1	1	1	0	0	0	1	0	1	1	0	1	0
$P_2^1Q_2^1$	$P_1^2Q_1^2$	$(g^{10})^2$	$g^{10}g^{11}$	$g^{10}g^{11}$	$(g^{11})^2$	0	0	0	0	0	0	0	0	2	1	1	0	0	1	1	2
	$P_1^2Q_2^2$	$g^{01}g^{10}$	$g^{00}g^{10}$	$g^{01}g^{11}$	$g^{00}g^{11}$	0	1	0	1	1	0	1	0	1	1	0	0	0	0	1	1
	$P_2^2Q_1^2$	$g^{01}g^{10}$	$g^{01}g^{11}$	$g^{00}g^{10}$	$g^{00}g^{11}$	0	0	1	1	1	1	0	0	1	0	1	0	0	1	0	1
	$P_2^2Q_2^2$	$(g^{01})^2$	$g^{00}g^{01}$	$g^{00}g^{01}$	$(g^{00})^2$	0	1	1	2	2	1	1	0	0	0	0	0	0	0	0	0

assumed order is likely to be wrong. Based on the mutual relationships of $\hat{\vartheta}^{12}$, $\hat{\vartheta}^{23}$, and $\hat{\vartheta}^{13}$, a new order can be determined on which the three-point analysis is performed again to make new estimates for the three recombination rates. When every three adjacent markers from a linkage group are subject to three-point analysis and the marker ordering has been confirmed, we can obtain two different estimates of the recombination fraction for the same marker interval (except for the intervals at the two ends of a linkage group). The best way to combine these estimates is to take a weighted mean, with the weights being the reciprocals of the variances of the two separate estimates (Ridout *et al.*, 1998).

4. NUMERICAL EXAMPLE

We describe a hypothetical example to demonstrate the statistical procedures for calculating the posterior probability characterizing linkage phases between different markers in a full-sib family. Consider a cross of two outbred parents involving five markers on the same chromosome, with a known order M^1–M^2–M^3–M^4–M^5. These five loci are from different marker groups, with parental genotypes being $a^1b^1 \times c^1d^1$ (type A segregating 1:1:1:1), $a^2b^2 \times a^2b^2$ (type B_3 segregating 1:2:1), $a^3o^3 \times a^3o^3$ (type C segregating 3:1), $a^4b^4 \times b^4b^4$ (type D_1 segregating 1:1), and $a^5b^5 \times c^5d^5$ (type A segregating 1:1:1:1), respectively. The linkage phases for the two parents between the five markers are

$$\begin{array}{c\|\|c} a^1 & b^1 \\ a^2 & b^2 \\ a^3 & o^3 \\ a^4 & b^4 \\ a^5 & b^5 \end{array} \times \begin{array}{c\|\|c} c^1 & d^1 \\ a^2 & b^2 \\ o^3 & a^3 \\ b^4 & b^4 \\ c^5 & d^5 \end{array}$$

The full-sib family is assumed to include 100 progeny, which are allocated to different combinations of the

TABLE III

Results from Linkage Analysis for the Five Hypothesized Markers in a Full-Sib Family of Size 100

Marker order		M^1	——	M^2	——	M^3	——	M^4	——	M^5
		1:1:1:1		1:2:1		3:1		1:1		1:1:1:1
Hypothesized recom. fraction			0.14		0.20		0.10		0.22	
Two-point analysis: *single test*										
Posterior probability	$P(A_1)$		0.52		0.08		0.39		0.30	
	$P(A_2)$		0.25		0.35		0.39		0.30	
	$P(A_3)$		0.11		0.35		0.11		0.20	
	$P(A_4)$		0.12		0.22		0.11		0.20	
Estimate for recombination fraction[a]			0.12 ± 0.027		0.16 ± 0.045		0.09 ± 0.042		0.20 ± 0.041	
Two-point analysis: *joint test*										
Estimate for recombination fraction[a]			0.12 ± 0.026		0.16 ± 0.043		0.09 ± 0.040		0.20 ± 0.042	
Three-point analysis										
				M^1–M^2–M^3		M^2–M^3–M^4		M^3–M^4–M^5		
Posterior probability	$P(A_{11})$[b]			0.08		0.09		0.20		
	$P(A_{12})$			0.24		0.09		0.08		
	$P(A_{13})$			0.06		0.02		0.04		
	$P(A_{14})$			0.07		0.02		0.03		
	$P(A_{21})$			0.05		0.23		0.20		
	$P(A_{22})$			0.06		0.23		0.08		
	$P(A_{23})$			0.05		0.01		0.04		
	$P(A_{24})$			0.08		0.01		0.03		
	$P(A_{31})$			0.04		0.05		0.05		
	$P(A_{32})$			0.02		0.05		0.04		
	$P(A_{33})$			0.03		0.01		0.02		
	$P(A_{34})$			0.05		0.01		0.04		
	$P(A_{41})$			0.03		0.05		0.05		
	$P(A_{42})$			0.04		0.05		0.04		
	$P(A_{43})$			0.06		0.04		0.02		
	$P(A_{44})$			0.04		0.04		0.04		
Estimate for recombination fraction[a]			0.14 ± 0.021		0.19 ± 0.031		0.10 ± 0.025			
					0.18 ± 0.034		0.10 ± 0.023		0.21 ± 0.035	

[a] The estimates ($\pm$SE) for recombination fraction presented here are those under the most likely assignment (see text).
[b] The first and second numbers of the subscript denote the assignment of the second marker given the first marker and that of the third marker given the second marker, respectively, for the three markers considered.

phenotypes at the five markers according to Mendelian inheritance given the recombination fraction and phase (Table III). For this cross, there are 192 distinct offspring phenotypes.

4.1. Single Two-Point Analysis

Two-point linkage analysis is used to estimate the recombination fraction and linkage phase separately for each adjacent marker pair from the simulated marker dataset including both fully informative and partially informative markers (Table III). The linkage phases between two adjacent markers are characterized by comparing posterior probabilities of the four different allele assignments between the homologous chromosomes of the two parents calculated using Eq. (1). Given the fixed assignment of marker M^1, the posterior probability of the first assignment of marker M^2 is the highest (0.52), suggesting that the parental phase configurations for these two markers are

$$\begin{matrix} a^1 \\ a^2 \end{matrix} \Big\| \begin{matrix} b^1 \\ b^2 \end{matrix} \times \begin{matrix} c^1 \\ a^2 \end{matrix} \Big\| \begin{matrix} d^1 \\ b^2 \end{matrix}$$

Based on the estimated allele assignment for marker M^2, the same procedure is used to determine the posterior probability of marker M^3 and, therefore, the linkage phase between markers M^2 and M^3. It is found that both

the second and third assignments of marker M^3, given M^2, have the highest posterior probability (0.35), which cannot provide a precise result for the next assignment analysis due to incomplete information of both markers, M^2 and M^3. We arbitrarily choose one assignment (say the third one) from these two to characterize the assignment of marker M^4 and continue the procedures until marker M^5. It is noted that marker M^4 cannot be discriminated between its first and second assignments as well as its third and fourth assignments (Table III). But this may not pose a problem for the subsequent quantitative trait locus analysis because these indistinguishable assignments result from the homozygous nature of this marker in one parent. Finally, from the calculated posterior probabilities (Table III), one can conclude that the most likely parental phases for all these five markers are

$$\begin{matrix} a^1 \\ a^2 \\ o^3 \\ a^4 \\ a^5 \end{matrix} \left\| \begin{matrix} b^1 \\ b^2 \\ a^3 \\ b^4 \\ b^5 \end{matrix} \right. \begin{matrix} c^1 \\ a^2 \\ \times\ a^3 \\ b^4 \\ d^5 \end{matrix} \left\| \begin{matrix} d^1 \\ b^2 \\ o^3 \\ b^4 \\ c^5 \end{matrix} \right. ,$$

which does not exactly match the assumed pattern. Note, however, that the correct pattern is among the equally probable ones, had the right choices been made for each ambiguous assignment.

Table III also presents estimates for the recombination rates under the most likely assignment for each marker, which are close to the values used for the simulation. But the estimation bias is higher for the two markers of symmetrical parental cross type, M^2 and M^3, than other marker combinations. The accuracy of the estimates for the recombination rates is reasonably high, as assessed by the sampling errors estimated from the Fisher information index, i.e., minus the second derivative of the log-likelihood function (Ritter and Salamini, 1996).

4.2. Joint Analysis

A joint test of two-point analyses via the Markov chain is further performed to examine its performance relative to separate two-point analyses as shown above and also used by other authors (Ritter *et al.*, 1990; Ritter and Salamini, 1996; Maliepaard *et al.*, 1997). In this case, five markers have $4^4 = 256$ possible combinations of assignments from which a most likely combination is inferred based on a joint analysis. However, most of these combinations can be ignored based on the results from the separate analysis above. We keep those combinations for which the posterior probability between two adjacent markers calculated from separate tests is equal to, or greater than, 0.20. It is found from Table III that there are a total of 48 assignment combinations meeting this criterion. The posterior probabilities for the various combinations of assignments differ sharply when the joint analysis is conducted (Fig. 1), thus suggesting that the joint analysis has greater power to characterize linkage phases. A major advantage of the joint analysis is that the ambiguity in determining the most likely assignment of marker M^3 given M^2 can be clarified because of a multi-locus linkage phase inference including fully informative markers M^1 and M^5. Such clarification is not possible with the separate analyses as shown above. The most likely parental phase configurations from the joint analysis (Fig. 1) are

$$\begin{matrix} a^1 \\ a^2 \\ a^3 \\ a^4 \\ a^5 \end{matrix} \left\| \begin{matrix} b^1 \\ b^2 \\ o^3 \\ b^4 \\ b^5 \end{matrix} \right. \begin{matrix} c^1 \\ a^2 \\ \times\ o^3 \\ b^4 \\ c^5 \end{matrix} \left\| \begin{matrix} d^1 \\ b^2 \\ a^3 \\ b^4 \\ d^5 \end{matrix} \right. ,$$

which are correct.

Estimates for the recombination rates are similar under the most likely assignment combination from the joint analysis and from the separate analyses in terms of biases and sampling errors (Table III). This is because these estimates are independent in the different marker intervals.

4.3. Three-Point Analysis

We further carry out a three-point analysis on the same marker dataset. The first three-point analysis for markers M^1, M^2, and M^3 simultaneously infers the assignments of markers M^2 and M^3 by assuming a fixed assignment of marker M^1. The second three-point analysis for markers M^2, M^3, and M^4 simultaneously infers the assignments of markers M^3 and M^4 based on the estimated assignment of marker M^2. This process continues until the last marker. There are two major advantages of three-point analysis over either separate or joint two-point analyses. First, three-point analysis has much greater power to determine the linkage phases between adjacent markers, because, as indicated by Table III, its posterior probabilities differ most between the correct and alternate assignments. Second, the estimates of recombination fraction are improved, exhibiting less bias and lower standard errors estimated from the Fisher information index, under the most likely

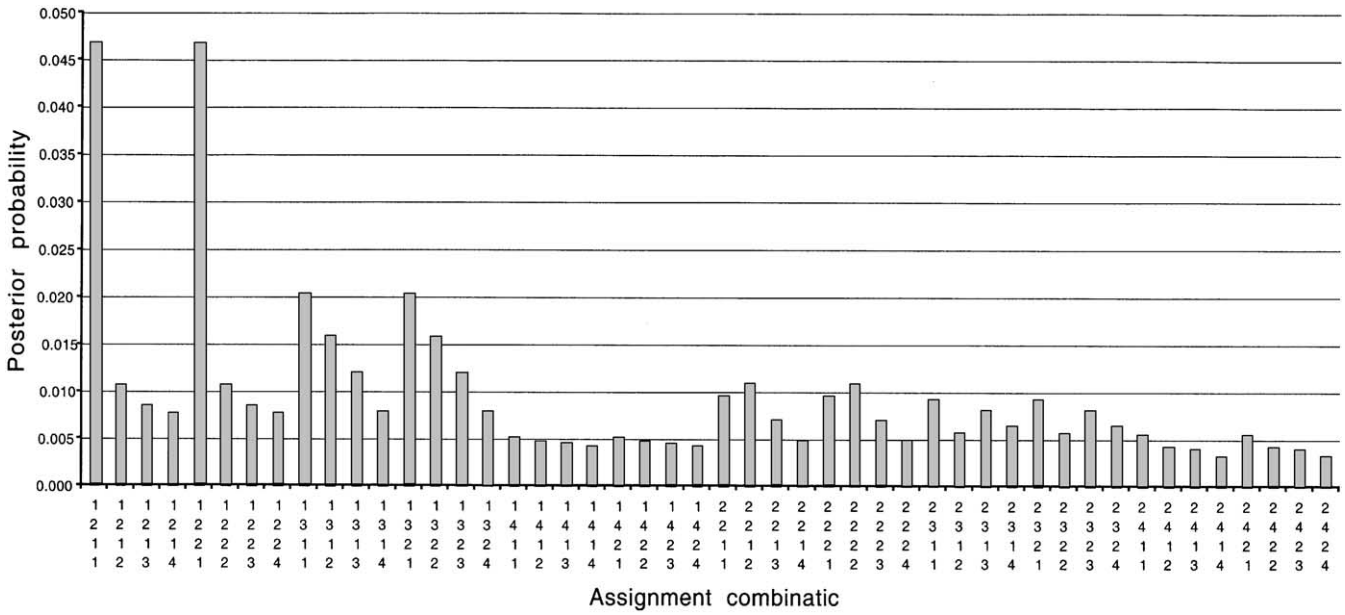

FIG. 1. Posterior probabilities of the 48 selected combinations of assignments among the five assumed markers calculated from a joint analysis. The numbers below the *x*-axis indicate the assignment combinations where the first, second, third, and fourth numbers denote the assignments of markers M^2, M^3, M^3, and M^5 conditional on their previous markers, respectively. Because the fourth marker belongs to type D_1, its phase assignment relative to the third marker is identical between A_1 and A_2 as well as between A_3 and A_4. For the assignments given in this figure, thus, we have 1211 = 1221, 1212 = 1222, and so on. Note that the second and the third assignment of marker M^3 conditional on marker M^2 can be well specified through the joint analysis.

assignment for three-point analysis compared with two-point analysis (Table III). These desirable properties of three-point analysis may result from the inclusion of partially informative markers, e.g., M^3, which cannot be well used in two-point analyses (see also Ridout *et al.*, 1998). A problem arising with three-point analysis is that there are usually two different estimates for the recombination fraction between two adjacent markers not from the two ends. In this example, three-point analysis of markers M^1–M^2–M^3 and M^2–M^3–M^4 generate two different estimates of recombination fraction ϑ^{23}. The same is true for the estimates of ϑ^{34}. The best way of combining these estimates is to take a weighted mean, with the weights being the reciprocals of the variances of the two separate estimates.

5. DISCUSSION

The molecular process of recombination during meiosis enables the relative positions of genetic loci to be inferred from the patterns of allele transmission from parents to offspring. The construction of high-density linkage maps based on meiotic recombination is aided by the increased availability of easily measured genetic markers and also by the development of powerful statistical methods. The statistical techniques for linkage analysis in model systems with well-designed pedigrees derived from two inbred lines have been well established and have led to the construction of hundreds of genetic maps in a number of organisms. However, when performing a linkage analysis in a full-sib family of outbred parents, one would likely encounter two major uncertainties, one being a variable number of segregating alleles and the other being that parental linkage phases between different markers are unknown. The current strategy for overcoming these problems is to account for all the possible combinations of marker types between the two parents and estimate the recombination fraction and linkage phase for these combinations individually using particular formulae (Ritter *et al.*, 1990; Ritter and Salamini, 1996; Maliepaard *et al.*, 1997; Ridout *et al.*, 1998). However, these analyses are not statistically effective because the inference from the minimum estimate of the recombination fraction and from the significant LOD score may be inconsistent. In this study, we have derived a general algorithm for simultaneously estimating linkage and

parental linkage phases over all linked molecular markers of any kind in a full-sib family derived from two outbred parents. Compared to previous procedures, our algorithm can analyze all possible segregating marker types at the same time, thus increasing both computational effectiveness and efficiency.

We started a single two-point analysis to show the procedures for deriving the general statistical algorithm for linkage mapping, followed by a joint two-point analysis combining all markers based on a hidden Markov model. The joint two-point analysis is not a simple extension of a single two-point analysis because the former provides estimates of parental linkage phases over all markers of any kind under consideration. We used a numerical example to demonstrate the statistical properties of our new algorithm and examine the advantages of the joint over single two-point analyses. The effect of sample sizes on the detection and estimation of linkage has been discussed in detail by Maliepaard *et al*. (1997). For an infinite sample size, it can be proven analytically or through simulations that the MLE of the recombination fraction for any marker combination is unbiased. But for a finite sample size, say 50, the estimates of recombination are downward biased because large estimates correspond to small test statistics (e.g., LOD scores) and, therefore, are ignored. Maliepaard *et al*. (1997) found that the accuracy and power of the estimate for recombination rates smaller than 0.25 are reasonably good when the number of progeny is increased to 100, which is a sample size widely used in practical studies of linkage map construction (e.g., Grattapaglia and Sederoff, 1994). We thus used a sample size of 100 in our example. Linkage analyses for markers with different levels of genetic information differ in accuracy and power. The linkage phase between markers of symmetrical parental cross types, e.g., B_3 or C (Table I), cannot be well characterized through a simple two-point test, and estimates of recombination for these markers have large biases and sampling errors (Table III; Maliepaard *et al*., 1997). These problems, however, can be solved when a two-point analysis is extended to make simultaneous inferences concerning multi-locus phases including at least one marker of asymmetrical cross type using the Markov chain. As shown by the example, such a joint analysis has enabled clear separation of the two ambiguous assignments associated with a dominant marker. The idea of multipoint analysis for linkage mapping is not new in human genetics (Lander and Green, 1987; Kruglyak *et al*., 1996; Kruglyak and Lander, 1998). But multipoint analysis in humans is mostly based on fully informative markers obtained from a pedigree with a clear structure in which the inference of linkage phase is not an issue. The method presented in this paper has generalized the multipoint analysis for any kind of marker in a single full-sib family. The method can increase the accuracy of linkage estimates for partially informative markers by correctly specifying the linkage phases between these markers and more informative markers.

For clarity of expression, we assume that the first marker at the top of a linkage group has a fixed assignment. This is no problem if the parental origin of all markers is known. However, when not all markers have parental marker information, an efficient analysis is to choose a marker of parental information from an asymmetrical cross type (Table I) as a starting point, simultaneously extending upward and downward to include all markers in the group (see Lander and Green, 1987; Jiang and Zeng, 1997). Using this method, one can effectively characterize linkage phases between any markers regardless of the number of markers as well as their distances in the linkage group.

We also derived a general algorithm based on three-point analysis to study linkage relationships for any possible marker combinations. Three-point analysis has also been described by Ridout *et al*. (1998) to analyze marker data in a full-sib family. However, their analysis was not based on the EM algorithm. Also, they did not convert likelihoods into posterior probabilities. Compared to two-point analysis, three-point analysis takes longer computing times because more extensive computations are needed. On average, a single three-point analysis in our example including five markers of different cross types can take a few seconds of computing time to detect a most likely linkage phase combination and estimate the corresponding recombination rates, whereas a joint two-point analysis combining all the five markers takes about one second. However, three-point analysis has two advantages over two-point analysis in terms of precise estimates of recombination and marker ordering. Our example indicates that both linkage phases and recombination fractions can be estimated more accurately from a three- than from a two-point analysis, though such improvements may be modest when many informative markers are included (sec also Thompson, 1984; Lathrop *et al*., 1985). It should be pointed out that, like two-point analyses, three-point analysis has reduced power to specify correct linkage phases when the three markers are each from a symmetrical parental cross type. But this problem can be overcome by the inclusion of a marker of asymmetrical cross type. In addition, three-point analysis can be used to estimate the locus order

among three given markers because it estimates all possible recombination fractions between the three markers. If these estimates are not consistent with the assumed order, the ordering should be revised, and new estimates of the recombination rates may be calculated. This proceeds until these estimates and the order achieve consistency. Although three-point analysis can well estimate the locus order among three given markers, it remains unsolved how to order more than three markers on a linkage group using this approach. Because tremendous computational demands are expected for marker ordering, more powerful computational technologies, such as the Bayesian method implemented with Markov chain Monte Carlo (MCMC) algorithms (Robert and Casella, 1999), may be needed.

Finally, a few complexities that may arise from linkage analysis in a full-sib family deserve consideration. First, markers may exhibit distorted segregation because of inbreeding or outbreeding depression; for example, many distorted markers were observed among progeny derived from two individuals from the same population of pedunculate oak (Barreneche *et al.*, 1998). Second, the recombination fraction may not be the same for all meioses because of genetic differences in the rate of crossovers within a certain marker interval between two parents. Divergence in recombination fraction between female- and male-parental meioses has been well documented for many species (e.g., Plomion and O'Malley, 1996; Maliepaard *et al.*, 1998). The deviation from Mendelian segregation can be incorporated into our current model by considering its biological causes. And sex differences in recombination rates can be included using our algorithm by defining sex-specific recombination fractions.

Last, and most difficult, we should extend our current full-sib linkage analysis to include multiple families. For plant species, this is not a problem because a virtually unlimited number of offspring can be produced in a single family. A single family, though commonly used in many current molecular experiments, does not, however, represent a random sample of a natural or experimental population, and thus the analysis of a single family can be affected by genetic drift errors. Also, due to some unpredictable factors in practical pollination, it is possible that one does not produce enough offspring in a single family for genome mapping, which would thereby reduce power to detect target genes of small effects in plant genome projects. For these reasons, a linkage analysis should be based on the combination of multiple families derived from a random sample of parents with particular population genetic properties. Theoretically, such a combined analysis can also shed light on the genetic architecture of a population by estimating population genetic parameters, such as allele frequencies, the departure from Hardy–Weinberg equilibrium, and linkage disequilibria between polymorphic markers. In a companion paper, we have developed a new strategy for estimating these genetic parameters for an outcrossing population based on fully informative markers (Wu and Zeng, 2001), thus providing a useful framework for us to extend the current work to study natural populations. Because partially informative markers are involved in such an extension, it is worthwhile incorporating the Bayesian method implemented with MCMC. The power of the Bayesian method to perform joint linkage mapping with multiple families and the comparison of this method with maximum likelihood method deserves further exploration.

We will write window friendly software for the statistical methods proposed in this study for public use. The program that implements the proposed method can be requested from the senior author.

ACKNOWLEDGMENTS

The senior author thanks Dr. George Casella for his support for this and other studies. We are grateful to Dr. S. P. Otto and two anonymous reviewers for helpful comments on this manuscript. This research is supported from Grant NIH GM 45344 to Z.-B. Z. This manuscript was approved for publication as Journal Series Paper R-08224 by the Florida Agricultural Experiment Station.

REFERENCES

Arus, P., Olarte, C., Romero, M., and Vargas, F. 1994. Linkage analysis of 10 isozyme genes in F_1 segregating almond progenies, *J. Am. Soc. Hort. Sci.* **119**, 339–344.

Barreneche, T., Bodenes, C., Lexer, C., Trontin, J. F., Fluch, S., Streiff, R., Plomion, C., Roussel, G., Steinkellner, H., Burg, K., Favre, J. M., Glossl, J., and Kremer, A. 1998. A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers, *Theor. Appl. Genet.* **97**, 1090–1103.

Basten, C. J., Weir, B. S., and Zeng, Z.-B. 1999. "Zmap—A, QTL Cartographer," North Carolina State University, Raleigh, NC.

Bradshaw, H. D., Jr., Villar, M., Watson, B. D., Otto, K. G., Stewart, S., and Stettler, R. F. 1994. Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers, *Theor. Appl. Genet.* **89**, 551–558.

Dempster, A. P., Laird, N. M., and Rubin, D. B. 1977. Maximum likelihood from incomplete data via KM algorithm, *J. Roy. Stat. Soc. Ser. B* **39**, 1–38.

Devey, M. E., Fiddler, T. A., Liu, B. H., Knapp, S. J., and Neale, D. B. 1994. An RFLP linkage map for loblolly pine based on a three generation outbred pedigree, *Theor. Appl. Genet.* **88**, 273–278.

Frary, A., Nesbitt, C., Frary, A., Grandillo, S., van der Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K. B., and Tanksley, S. D. 2000. *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size, *Science* **289**, 85–88.

Grattapaglia, D., and Sederoff, R. 1994. Genetic linkage maps of *Eucalyptus grandis and Eucalyptus urophylla* using a pseudo-testcross: Mapping strategy and RAPD markers, *Genetics* **137**, 1121–1137.

Hamrick, J. L., and Godt, M. J. W. 1990. Allozyme diversity in plant species. *in* "Plant Population Genetics, Breeding and Genetic Resources," (H. D. Brown, T. Clegg, A. L. Kahler, and B. S. Weir, Eds.), pp. 43–63, Sinauer, Sunderland, MA.

Hemmat, M., Weeden, N. F., Manganaris, A. G., and Lawson, D. M. 1994. Molecular marker linkage map for apple, *J. Hered.* **85**, 4–11.

Jiang, C. J., and Zeng, Z.-B. 1997. Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines, *Genetica* **101**, 47–58.

Kruglyak, L., Daly, M. J., RecveDaly, M. P., and Lander, E. S. 1996. Parametric and nonparametric linkage analysis: A unified multipoint approach, *Am. J. Hum. Genet.* **58**, 1347–1363.

Kruglyak, L., and Lander, E. S. 1998. Faster multipoint linkage analysis using Fourier transforms, *J. Comput. Biol.* **5**, 1–7.

Lander, E. S., and Green, P. 1987. Construction of multilocus genetic linkage maps in human, *Proc. Natl. Acad. Sci.* **84**, 2363–2367.

Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E., and Newburg, L. 1987. Mapmaker: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations, *Genomics* **1**, 174–181.

Lathrop, G. M., Lalouel, J. M., Julier, C., and Ott, J. 1985. Multilocus linkage analysis in humans: Detection of linkage and estimation of recombination, *Am. J. Hum. Genet.* **37**, 482–498.

Maliepaard, C., Alston, F. H., and van Arkel, G., *et al.* 1998. Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers, *Theor. Appl. Genet.* **97**, 60–73.

Maliepaard, C., Jansen, J., and van Ooijen, J. W. 1997. Linkage analysis in a full-sib family of an outbreeding plant species: Overview and consequences for applications, *Genet. Res.* **70**, 237–250.

Olson, M., Hood, L., Cantor, C., and Botstein, D. 1989. A common language for physical mapping of the human genome, *Science* **245**, 1434–1435.

Paglia, G. P., Olivieri, A. M., and Morgante, M. 1998. Towards second-generation STS (sequence-tagged sites) linkage maps in conifers: A genetic map of Norway spruce (*Picea abies* K.), *Mol. Gen. Genet.* **258**, 466–478.

Paterson, A. H., Bowers, J. E., Burow, M. D., Draye, X., Elsik, C. G., Jiang, C. X., Katsar, C. S., Lan, T. H., Lin, Y. R., Ming, R. G., and Wright, R. J. 2000. Comparative genomics of plant chromosomes, *Plant Cell* **12**, 1523–1539.

Pfeiffer, A., Olivieri, A. M., and Morgante, M. 1997. Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.), *Genome* **40**, 411–419.

Plomion, C., and O'Malley, D. M. 1996. Recombination rate differences for pollen parents and seed parents in *Pinus pinaster*, *Heredity* **77**, 341–350.

Ridout, M. S., Tong, S., Vowden, C. J., and Tobutt, K. R. 1998. Three-point linkage analysis in crosses of allogamous plant species, *Genet. Res.* **72**, 111–121.

Ritter, E., Gebhardt, C., and Salamini, F. 1990. Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents, *Genetics* **125**, 645–654.

Ritter, E., and Salamini, F. 1996. The calculation of recombination frequencies in crosses of allogamous plant species with applications to linkage mapping, *Genet. Res.* **67**, 55–65.

Robert, C. P., and Casella, G. 1999. "Monte Carlo Statistical Methods," Springer-Verlag, New York.

Tanksley, S. D. 1993. Mapping genes, *Annu. Rev. Genet.* **27**, 205–233.

Thompson, E. A. 1984. Information gain in joint linkage analysis, *IMA J. Math. Appl. Med. Biol.* **1**, 31–49.

Weber, J. L., and Wong, C. 1993. Mutation of human short tandem repeats, *Hum. Mol. Genet.* **2**, 1123–1128.

Wu, R. L., and Zeng, Z.-B. 2001. Joint linkage and linkage disequilibrium mapping in natural populations, *Genetics* **157**, 899–909.

Wu, R. L., Zeng, Z.-B., McKeand, S. E., and O'Malley, D. M. 2000. The case for molecular mapping in forest tree breeding. *Plant Breed. Rev.* **19**, 41–68.

Xiao, J., Li, J., Yuan, L., and Tanksley, S. D. 1995. Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers, *Genetics* **140**, 745–754.