Statistical Analysis of Ordered Tetrads

Hongyu Zhao* and Terence P. Speed†

**Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, Connecticut 06520 and †Department of Statistics, University of California, Berkeley, California 94720*

Manuscript received January 12, 1998
Accepted for publication May 18, 1998

ABSTRACT

Ordered tetrad data yield information on chromatid interference, chiasma interference, and centromere locations. In this article, we show that the assumption of no chromatid interference imposes certain constraints on multilocus ordered tetrad probabilities. Assuming no chromatid interference, these constraints can be used to order markers under general chiasma processes. We also derive multilocus tetrad probabilities under a class of chiasma interference models, the chi-square models. Finally, we compare centromere map functions under the chi-square models with map functions proposed in the literature. Results in this article can be applied to order genetic markers and map centromeres using multilocus ordered tetrad data.

GENETIC studies using tetrad data are very valuable in studying the chance mechanisms in meiosis, including: (1) positions of crossovers along the four-strand bundle; (2) nonsister strand pairs involved in each crossover; (3) spindle-centromere attachment at the first meiotic division; and (4) spindle-centromere attachment at the second meiotic division. Deviation from random distributions of crossovers on the four-strand bundle is called chiasma interference. Deviation from random involvement of nonsister chromatid pairs in each crossover is called chromatid interference. Compared with single spore data, where the four products from a single meiosis can only be recovered separately, tetrad data, where four meiotic products can be recovered together, have several advantages. First, chromatid interference and chiasma interference can be distinguished using tetrad data. Second, when chromatid interference is absent, chiasma interference can be detected with only two markers, whereas at least three markers are needed for single spore data. Chiasma interference can even be detected with one marker in some studies. Third, the position of the centromere can be inferred. In some organisms, such as *Neurospora crassa*, the asci are produced in a linear order corresponding to the meiotic divisions and are called ordered tetrads. In other organisms, such as *Saccharomyces cerevisiae*, the asci are produced as a group without order and are called unordered tetrads.

Ordered tetrads have been used extensively to study the crossover process during meiosis since Lindegren (1932, 1933, 1936a,b). However, most studies on ordered tetrads and unordered tetrads, for example, Papazian (1952) and Perkins (1955), used only three loci for the detection of chromatid interference and one locus for mapping centromeres. Several studies have taken a multilocus approach: (1) under the assumption of no chromatid interference (NCI), Risch and Lange (1983) fitted one class of chiasma interference models, the count-location model, to one multilocus unordered tetrad data set; (2) Zhao *et al.* (1995b) fitted another class of chiasma interference models, the chi-square model, to several multilocus unordered tetrad data sets; and (3) Zhao *et al.* (1995a) derived a set of linear equality and inequality constraints on the probabilities of unordered tetrad patterns, with an arbitrary number of loci under the assumption of NCI, and tested these constraints on data sets from a variety of organisms reported in the literature.

In this article, ordered tetrad data are studied under different assumptions on the chance mechanisms. For each assumption, a detailed discussion is provided for single marker and two marker data. General results for multiple markers are then presented. Although the number of spores is four in a tetrad and eight in an octad, there is no loss of generality for discussing only tetrads when aberrant segregations can be ignored. Half-tetrad data are another type of genetic data that are closely related to ordered tetrad data and widely used in genetic studies. A detailed study of half-tetrad data is given in the accompanying article (Zhao and Speed 1998).

We adopt the following notation in this article. Markers are denoted by script letters; for example, we use $\mathcal{A}$ and $\mathcal{B}$ to denote markers. Alleles are denoted by italic letters. For example, A and a denote two alleles of marker $\mathcal{A}$. We use $[X, Y, Z, W]$ to denote the observed marker configuration for an ordered tetrad, where X and Y are attached to one centromere and Z and W are attached to the other centromere. For example, [*AB*,

Corresponding author: Hongyu Zhao, Department of Epidemiology and Public Health, Yale University School of Medicine, 60 College St., New Haven, CT 06520. E-mail: hongyu.zhao@yale.edu

TABLE 1

Six possible configurations at marker $\mathcal{A}$

Type	
1	[*A*, *A*, *a*, *a*]
2	[*a*, *a*, *A*, *A*]
3	[*A*, *a*, *A*, *a*]
4	[*a*, *A*, *a*, *A*]
5	[*A*, *a*, *a*, *A*]
6	[*a*, *A*, *A*, *a*]

TABLE 2

Eight ordered tetrad types with the same probability under RSCA

[1, 2, 3, 4]
[1, 2, 4, 3]
[2, 1, 3, 4]
[2, 1, 4, 3]
[3, 4, 1, 2]
[3, 4, 2, 1]
[4, 3, 1, 2]
[4, 3, 2, 1]

Ab, *aB*, *ab*] represents an ordered tetrad with two strands carrying *AB* and *Ab* attached to one centromere and with two strands carrying *aB* and *ab* attached to the other centromere. The centromere is denoted by *CEN*. For patterns between a pair of markers, we use *P* to denote parental ditype where all four strands retain the parental type, *T* to denote tetratype where two of the four strands show recombination, and *N* to denote nonparental ditype where all four strands are recombinants.

METHODS

Random spindle-centromere attachment assumption: Random spindle-centromere attachment (RSCA) assumes that two centromeres have the same chance to go to either pole at the first meiotic division, and the divided centromeres have the same chance to go to either pole at the second meiotic division (Griffiths *et al.* 1996).

For marker $\mathcal{A}$ with alleles *A* and *a* inherited from two parents, there are six distinguishable configuration types, as illustrated in Table 1. Under RSCA, types 1 and 2 ([*A*, *A*, *a*, *a*] and [*a*, *a*, *A*, *A*]) have the same probability because of random spindle-centromere attachment at the first meiotic division, whereas types 3 to 6 ([*A*, *a*, *A*, *a*], [*a*, *A*, *a*, *A*], [*A*, *a*, *a*, *A*], and [*a*, *A*, *A*, *a*]) have the same probability because of random spindle-centromere attachment at the second meiotic division. Types 1 and 2 are called first division segregation (FDS) pattern, and types 3 to 6 are called second division segregation (SDS) pattern (Griffiths *et al.* 1996). For a single marker, RSCA can be tested by examining whether types 1 and 2 occur with equal frequency and whether types 3, 4, 5, and 6 occur with equal frequency. RSCA is generally confirmed (Fincham *et al.* 1979).

For the two markers $\mathcal{A}$ and $\mathcal{B}$, there are six distinguishable configurations at $\mathcal{A}$ and six distinguishable configurations at $\mathcal{B}$. Therefore, there are 36 distinguishable patterns jointly at two markers. Under RSCA, if one pattern can be changed to another pattern through one of the eight permutations in Table 2, these two patterns should have the same probability. For example, the following eight types have the same probability: [*AB, Ab, aB, ab*], [*Ab, AB, aB, ab*], [*AB, Ab, ab, aB*], [*Ab, AB, ab, aB*], [*aB, ab, AB, Ab*], [*ab, aB, AB, Ab*], [*aB, ab, Ab, AB*], and [*ab, aB, Ab, AB*]. After examining all 36 distinguishable patterns, the number of distinct probabilities is seven under RSCA. These seven groups are shown in Table 3. When $\mathcal{A}$ shows FDS, there are three distinct groups corresponding to whether $\mathcal{A}$ and $\mathcal{B}$ have parental ditype, nonparental ditype, or tetratype, denoted by P_1, N_1, and T_{12}, respectively, in Table 3. When $\mathcal{A}$ shows SDS, there are four distinct groups. Two groups correspond to $\mathcal{A}$ and $\mathcal{B}$ having parental and nonparental ditypes, denoted by P_2 and N_2, respectively, in Table 3. When $\mathcal{A}$ and $\mathcal{B}$ have tetratype, $\mathcal{B}$ can show either FDS or SDS. Under RSCA, these two groups may have different probabilities, denoted by T_{21} and T_2. RSCA can be tested by examining whether all distinguishable patterns within each group occur with equal frequency. For example, the 8 distinguishable patterns listed above, which correspond to group T_{12} with $\mathcal{A}$ showing FDS and $\mathcal{A}$ and $\mathcal{B}$ having tetratype, should occur with equal frequency. These and more cases were first studied by Whitehouse (1942).

For *n* markers, there are 6^n distinguishable patterns. Under RSCA, these 6^n patterns reduce to $(6^n + 5 \times 2^n)/8$ distinct probabilities. This result was first derived by Papazian (1952). In the appendix (Proposition 1), we present a different derivation of this result through a more direct counting method. As for the one- and two-marker cases, RSCA can be tested by examining whether all the distinguishable tetrad patterns that should have the same probability (as discussed in the proof of Proposition 1 in the appendix) under RSCA do occur equally frequently.

No chromatid interference (one marker): For the case of one marker, $\mathcal{A}$, if NCI holds, the four configurations corresponding to the SDS pattern have the same probability before meiotic divisions. Therefore, these four types should occur with the same frequency even if RSCA fails. As a result, only when both NCI and RSCA fail can these four types occur with different probabilities.

As shown by Mather (1935), under NCI, the probabilities that $\mathcal{A}$ shows FDS and SDS patterns, given *k* chiasmata between *CEN* and $\mathcal{A}$, are

$$F_{\mathcal{A}}^{(k)} = P(\text{FDS} \mid k \text{ chiasmata}) = \tfrac{2}{3}(\tfrac{1}{2} + (-\tfrac{1}{2})^k), \quad (1)$$

and

$$S_{\mathcal{A}}^{(k)} = P(\text{SDS} \mid k \text{ chiasmata}) = \tfrac{2}{3}(1 - (-\tfrac{1}{2})^k).$$

TABLE 3

Seven distinct groups under RSCA

		$\mathcal{B}$				
		FDS		SDS		
		P_1	N_1	T_{12}		
$\mathcal{A}$	FDS	[*AB, AB, ab, ab*]	[*Ab, Ab, aB, aB*]	[*AB, Ab, aB, ab*]		
	SDS	[*AB, aB, Ab, ab*]		[*AB, ab, AB, ab*]	[*AB, ab, Ab, aB*]	[*Ab, aB, Ab, aB*]
		T_{21}		P_2	T_2	N_2

Suppose *A/a* and *B/b* are both segregating. The $6 \times 6 = 36$ combinations of ordered tetrads reduce to just 7 under RSCA. *P*, parental ditype; *T*, tetratype; *N*, nonparental ditype.

As k increases, the probability of SDS tends to $\frac{2}{3}$. For one marker, $\mathcal{A}$, NCI imposes no constraints on the probabilities of FDS and SDS, denoted by $F_{\mathcal{A}}$ and $S_{\mathcal{A}}$. For any observed FDS and SDS proportions, we can always construct a chiasma process model that gives rise to the observed FDS and SDS proportions. In fact, the process with probability $F_{\mathcal{A}}$ having zero chiasmata and with probability $S_{\mathcal{A}}$ having one chiasma is the simplest such model.

No chromatid interference (two markers): Two markers, $\mathcal{A}$ and $\mathcal{B}$, may be (1) on different chromosomes; (2) on the same chromosome but on different sides of the centromere (that is, the order is $\mathcal{A}$–*CEN*–$\mathcal{B}$); or (3) on the same chromosome and on the same side of the centromere (that is, the order is *CEN*–$\mathcal{A}$–$\mathcal{B}$ or *CEN*–$\mathcal{B}$–$\mathcal{A}$). We study these three cases separately. These three cases were first discussed in detail by Whitehouse (1942). We use the notation in Table 3 to denote seven distinct groups. For example, P_1 is the group with both markers showing FDS and no strand showing recombination between $\mathcal{A}$ and $\mathcal{B}$.

Two markers on different chromosomes: Let p and q denote the probability of SDS at $\mathcal{A}$ and $\mathcal{B}$. When both markers have FDS, tetrad types between $\mathcal{A}$ and $\mathcal{B}$ can be either parental ditype or nonparental ditype, depending on which pair of alleles are separated at the first meiotic division—*AB vs. ab* or *Ab vs. aB*. The probability of either outcome, group P_1 or N_1, is $(1 - p)(1 - q)/2$. Similar considerations lead to the probabilities of the seven groups in Table 4. These seven probabilities are determined by two independent parameters: p and q.

Two markers on different sides of the centromere ($\mathcal{A}$–CEN–$\mathcal{B}$): We use $p(P_1^{(k,\ell)})$, $p(N_1^{(k,\ell)})$, $p(T_{12}^{(k,\ell)})$, $p(P_2^{(k,\ell)})$, $p(N_2^{(k,\ell)})$, $p(T_{21}^{(k,\ell)})$, and $p(T_2^{(k,\ell)})$ to denote the frequency of ordered tetrads of groups P_1, N_1, T_{12}, P_2, N_2, T_{21}, and T_2 among meioses with k chiasmata in $\mathcal{A}$–*CEN* and ℓ chiasmata in *CEN*–$\mathcal{B}$. We can easily check that

$$p(P_1^{(0,0)}) = 1, \quad p(T_{12}^{(0,1)}) = p(T_{21}^{(1,0)}) = 1,$$
$$p(P_1^{(0,2)}) = p(P_1^{(2,0)}) = p(N_1^{(0,2)}) = p(N_1^{(2,0)}) = \tfrac{1}{4},$$
$$p(T_{12}^{(0,2)}) = p(T_{21}^{(2,0)}) = \tfrac{1}{2},$$
$$p(P_2^{(1,1)}) = p(N_2^{(1,1)}) = \tfrac{1}{4},$$

and

$$p(T_2^{(1,1)}) = \tfrac{1}{2}.$$

For $k + \ell > 2$,

$$p(P_1^{(k,\ell)}) = p(N_1^{(k,\ell)}) = \tfrac{1}{2}F_{\mathcal{A}}^{(k)}F_{\mathcal{B}}^{(\ell)},$$
$$p(T_{12}^{(k,\ell)}) = F_{\mathcal{A}}^{(k)}S_{\mathcal{B}}^{(\ell)}, \quad p(T_{21}^{(k,\ell)}) = S_{\mathcal{A}}^{(k)}F_{\mathcal{B}}^{(\ell)},$$
$$p(P_2^{(k,\ell)}) = p(N_2^{(k,\ell)}) = \tfrac{1}{2}p(T_2^{(k,\ell)}) = \tfrac{1}{4}S_{\mathcal{A}}^{(k)}S_{\mathcal{B}}^{(\ell)},$$

where $F_{\mathcal{A}}^{(k)}$ and $S_{\mathcal{A}}^{(k)}$ were defined in (1). For a given chiasma process along the four-strand bundle, let $c_{k\ell}$ denote the joint probability of there being k chiasmata between *CEN* and $\mathcal{A}$ and ℓ chiasmata between *CEN* and $\mathcal{B}$. The above relations can be combined with expressions for $(c_{k\ell})$ and summed, to give our desired frequencies. For example,

$$\text{frequency of } P_1 = \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} c_{k\ell}\, p(P_1^{(k,\ell)})$$

and

$$\text{frequency of } N_1 = \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} c_{k\ell}\, p(N_1^{(k,\ell)}).$$

On the basis of the above results, it can be shown that, for any chiasma process, seven distinct groups can have at most five different probabilities: the probabilities of P_1, N_1, T_{12}, T_{21}, and $(P_2 + T_2 + N_2)$ can differ, and we denote them by α, β, γ, δ, and ε. The ratio of the probabilities of P_2:T_2:N_2 is 1:2:1. Therefore, the probabilities of P_2, T_2, and N_2 are $\varepsilon/4$, $\varepsilon/2$, and $\varepsilon/4$, respectively. These are summarized in Table 5.

The probabilities of these seven groups can be derived by another approach. If we treat the centromere as a marker, the results for unordered tetrads (Zhao *et al.* 1995a) can be applied in this context. If the two centromeres from the two parents could be distinguished, there would be three types, *P*, *T*, and *N*, between $\mathcal{A}$ and *CEN*, and three types, *P*, *T*, and *N*, between *CEN* and $\mathcal{B}$. This would lead to nine distinct probabilities. Let p_0^k, p_1^k, and p_2^k denote the conditional probabilities of

TABLE 4

The probabilities of seven groups when $\mathcal{A}$ and $\mathcal{B}$ are on different chromosomes

		$\mathcal{B}$				
		FDS		SDS		
		P_1	N_1	T_{12}		
$\mathcal{A}$	FDS	$\frac{1}{2}(1-p)(1-q)$	$\frac{1}{2}(1-p)(1-q)$	$(1-p)q$		
	SDS	$p(1-q)$		$\frac{1}{4}pq$	$\frac{1}{2}pq$	$\frac{1}{4}pq$
		T_{21}		P_2	T_2	N_2

p and q are SDS proportions at $\mathcal{A}$ and $\mathcal{B}$.

p_0^k, p_1^k, and p_2^k of P, T, and N, given k chiasmata between a pair of markers. Under NCI, Mather (1935) showed that, for $k \geq 1$,

$$\begin{aligned} p_0^k &= \tfrac{1}{3}(\tfrac{1}{2} + (-\tfrac{1}{2})^k), \\ p_1^k &= \tfrac{2}{3}(1 - (-\tfrac{1}{2})^k), \\ p_2^k &= \tfrac{1}{3}(\tfrac{1}{2} + (-\tfrac{1}{2})^k). \end{aligned} \tag{2}$$

When $k = 0$, $p_0^0 = 1$ and $p_1^0 = p_2^0 = 0$. Let p_{ij} denote the probability of joint tetrad pattern (ij), where $i, j = 0$, 1, or 2 corresponding to P, T, and N in each interval; then,

$$p_{ij} = \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} c_{k\ell} p_i^k p_j^\ell,$$

where p_i^k and p_j^ℓ were defined in (2). Because the two centromeres cannot be distinguished, some of these classes are not distinguishable. For example, (0, 0) and (2, 2) both give rise to FDS at two markers and no recombinations between these two markers. Using the notation in Table 5, we have $\alpha = p(P_1) = p_{00} + p_{22}$, $\beta = p(N_1) = p_{02} + p_{20}$, $\gamma = p(T_{12}) = p_{01} + p_{21}$, $\delta = p(T_{21}) = p_{10} + p_{12}$, and $\varepsilon = 4p(P_2) = 2p(T_2) = 4p(N_2) = p_{11}$.

It was shown in Zhao *et al.* (1995a) that NCI imposes equality and inequality constraints on the probabilities of distinguishable unordered tetrad patterns. Here we derive constraints on ordered tetrad probabilities under NCI. Following the definition in Zhao *et al.* (1995a), we say a chiasma process is *compatible* with a given set of joint tetrad probabilities $\mathbf{p}$ if, under NCI, this chiasma process gives rise to these joint probabilities. It was shown in Zhao *et al.* (1995a) that for any underlying chiasma process, if NCI holds, there is a chiasma process with at most two exhanges between each consecutive pair of markers, inducing the same tetrad probabilities. Using this property, Zhao *et al.* (1995a) showed that, for a given set of unordered tetrad probabilities $\mathbf{p} = (p_0, p_1, p_2)'$, the probabilities of P, T, and N between two markers, there is some chiasma process compatible with $\mathbf{p}$ if and only if $\mathbf{T}_1^{-1}\mathbf{p} \geq \mathbf{0}$, where

$$\mathbf{T}_1 = \begin{pmatrix} 1 & 0 & \frac{1}{4} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{4} \end{pmatrix}, \quad \mathbf{T}_1^{-1} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 4 \end{pmatrix},$$

and $\mathbf{0} = (0, 0, 0)'$. For two markers, write the p_{ij} in lexicographical order as $\mathbf{p}$; there is an underlying chiasma process satisfying NCI compatible with unordered tetrad probabilities $\mathbf{p}$ if and only if $\mathbf{T}_2^{-1}\mathbf{p} \geq \mathbf{0}$, where $\mathbf{T}_2 = \mathbf{T}_1 \otimes \mathbf{T}_1$, $\mathbf{T}_2^{-1} = \mathbf{T}_1^{-1} \otimes \mathbf{T}_1^{-1}$ and $\mathbf{0}$ is a column vector with nine 0's, plus equality constraints described in Zhao *et al.* (1995a). The operator $\otimes$ is the standard tensor product (see, *e.g.*, Bellman 1970). If the chiasma process has at most two chiasmata in each interval, the correspondence between $\mathbf{p}$ and $\mathbf{c} = (c_{k\ell})$ is simply $\mathbf{c} = \mathbf{T}_2^{-1}\mathbf{p}$. Using the property that, for any underlying chiasma process, there is a compatible chiasma process with at most two chiasmata in each interval, we may focus on the study of chiasma processes with at most two chiasmata in each interval. Using the notation in Table 5, we have the following proposition, whose proof is given in the appendix (Proposition 2): under NCI, for any joint ordered tetrad probabilities with two markers on different sides of the centromere, there is an underlying chiasma process that is compatible with these probabilities if and only if $\alpha \geq \beta$ and $\gamma + \delta \geq 2\beta$ and the ratios of $p(P_2):p(T_2):p(N_2)$ are 1:2:1.

Using α, β, γ, δ, and ε, we may express the probability

TABLE 5

The probabilities of seven groups when $\mathcal{A}$ and $\mathcal{B}$ are on different sides of the centromere

		$\mathcal{B}$				
		FDS		SDS		
		P_1	N_1	T_{12}		
$\mathcal{A}$	FDS	α	β	γ		
	SDS	δ		$\frac{1}{4}\varepsilon$	$\frac{1}{2}\varepsilon$	$\frac{1}{4}\varepsilon$
		T_{21}		P_2	T_2	N_2

In the table, α, β, γ, δ, ε all ≥ 0, $\alpha + \beta + \gamma + \delta + \varepsilon = 1$, with constraints $\alpha \geq \beta$, $\gamma + \delta \geq 2\beta$.

TABLE 6

The probabilities of seven groups when $\mathcal{A}$ and $\mathcal{B}$ are on one side of the centromere and in the order of *CEN–$\mathcal{A}$–$\mathcal{B}$*

		$\mathcal{B}$				
		FDS		SDS		
		P_1	N_1	T_{12}		
$\mathcal{A}$	FDS	α	β	γ		
	SDS	½ϕ		δ	½ϕ	ε
		T_{21}		P_2	T_2	N_2

In the table, α, β, γ, δ, ε, ϕ all ≥ 0, α + β + γ + δ + ε + ϕ = 1. The constraints are: α ≥ β, γ ≥ 2β, δ ≥ ε, ϕ ≥ 2ε.

that $\mathcal{A}$ and $\mathcal{B}$ show *P*, *T*, and *N*, as $p(P) = \alpha + \varepsilon/4$, $p(T) = \gamma + \delta + \varepsilon/2$, and $p(N) = \beta + \varepsilon/4$, respectively. In the unordered tetrad case, the constraints imposed by NCI are: $p(P) \geq P(N)$ and $p(T) \geq 2p(N)$ (Zhao *et al.* 1995a). Substituting α, β, γ, δ, and ε in these two inequalities, we get $\alpha \geq \beta$ and $\gamma + \delta \geq 2\beta$. Therefore, for markers on different sides of the centromere, the only extra constraints added by ordered tetrads are the 1:2:1 proportionality constriants among $p(P_2)$, $P(T_2)$, and $p(N_2)$.

Two markers on the same side of the centromere (CEN–$\mathcal{A}$–$\mathcal{B}$; the case of CEN–$\mathcal{B}$–$\mathcal{A}$ can be discussed similarly): As in the previous discussion, we use (i, j) to denote the nine distinct groups if the centromere can be treated as a marker. Because the centromere cannot be observed, (0, 0) and (2, 0) cannot be distinguished. Both (0, 0) and (2, 0) show FDS at $\mathcal{A}$ and parental ditype between $\mathcal{A}$ and $\mathcal{B}$. Therefore, we have $p(P_1) = p_{00} + p_{20}$. Similarly, $p(N_1) = p_{02} + p_{22}$, $p(T_{12}) = p_{01} + p_{21}$, $p(P_2) = p_{10}$, $p(N_2) = p_{12}$, and $p(T_{21}) = p(T_2) = p_{11}/2$. Therefore, these seven groups can have at most six different probabilities. Each of these 2 × 3 types can be represented as $(i_1 i_2)$, with $i_1 = 0$ or 1 corresponding to FDS or SDS at $\mathcal{A}$, and $i_2 = 0$, 1, or 2 corresponding to *P*, *T*, or *N* between $\mathcal{A}$ and $\mathcal{B}$. Denote these probabilities by $\alpha = p(P_1)$, $\beta = p(N_1)$, $\gamma = p(T_{12})$, $\delta = p(P_2)$, $\varepsilon = p(N_2)$, and $\phi = 2p(T_{21}) = 2p(T_2)$ (Table 6). It can be shown, as in Zhao *et al.* (1995a), that for joint tetrad probabilities with two markers on the same side of the centromere, there is a compatible chiasma process under NCI if and only if $\alpha \geq \beta$, $\gamma \geq 2\beta$, $\delta \geq \varepsilon$, $\phi \geq 2\varepsilon$, and $p(T_{21}) = p(T_2)$.

No chromatid interference (multiple markers): Here we consider only markers on the same chromosome. Markers on the same side of the centromere and markers on different sides of the centromere are discussed separately.

Markers on the same side of the centromere: Under the assumption of NCI, there are $2 \times 3^{n-1}$ distinct probabilities for *n* markers in the order of *CEN–*$\mathcal{A}_1$–$\mathcal{A}_2$– · · · –$\mathcal{A}_n$. Each of these $2 \times 3^{n-1}$ classes can be identified as follows: FDS and SDS are distinguished at $\mathcal{A}_1$, and for each pair of consecutive markers, there are three types: *P*, *T*, and *N*. Each of these $2 \times 3^{n-1}$ types can be represented as $(i_1 i_2 \cdots i_n)$, where $i_1 = 0$ or 1 corresponding to FDS or SDS at $\mathcal{A}_1$, and $i_r = 0$, 1, or 2 corresponding to *P*, *T*, or *N* between $\mathcal{A}_{r-1}$ and $\mathcal{A}_r$, for $r = 2, \ldots, n$. For unordered tetrads with *n* markers ($n - 1$ intervals), there are 3^{n-1} distinct probabilities because FDS and SDS cannot be differentiated at $\mathcal{A}_1$; that is, i_1 cannot be determined. Write the probabilities of the observable patterns $(i_2 \ldots i_n)$ from unordered tetrads, denoted by $p^t_{i_2 \ldots i_n}$, in lexicographical order as $\mathbf{p}^t$. It was shown that there is an underlying chiasma process satisfying NCI compatible with unordered tetrad probabilities if and only if $\mathbf{T}_{n-1}^{-1}\mathbf{p}^t \geq \mathbf{0}$, where $\mathbf{T}_{n-1} = \mathbf{T}_1 \otimes \ldots \otimes \mathbf{T}_1$ ($n - 1$ terms), plus equality constraints described in Zhao *et al.* (1995a).

In our discussion of multiple marker ordered tetrad data, the $p^t_{0i_2 \ldots i_n}$ and $p^t_{1i_2 \ldots i_n}$ are considered separately. Write the $p^t_{0i_2 \ldots i_n}$ in lexicographical order as $\mathbf{p}^t_0$, the $p^t_{1i_2 \ldots i_n}$ in lexicographical order as $\mathbf{p}^t_1$. If for a given $(0i_2 \ldots i_n)$, there are $k \geq 2$ tetratypes in the $n - 1$ intervals, these tetratype combinations may be subdivided further into 4^{k-1} subcells as follows. First, the strands can be labeled such that strands 1 and 3 always show recombination between two markers that have tetratype closest to the centromere. For the other $k - 1$ intervals showing tetratype, recombinations can occur on four possible pairs of strands. Therefore, there are 4^{k-1} subtypes. The probability of each subcell can be denoted by $p_{0i_2 \ldots i_n}(h_1 \ldots h_{k-1})$, where each h_j is 1, 2, 3, or 4. If for a given $(1i_2 \ldots i_n)$, there are $k \geq 1$ tetratypes in the $n - 1$ intervals, these tetratype combinations may be subdivided further into $2 \times 4^{k-1}$ subcells as follows. Suppose the first pair of markers showing tetratype from the centromere is $\mathcal{A}_{r-1}$ and $\mathcal{A}_r$. Marker $\mathcal{A}_{r-1}$ must show SDS, because otherwise there must be a tetratype before $\mathcal{A}_{r-1}$. Marker $\mathcal{A}_r$ can show either FDS or SDS. Two types can thus be distinguished, depending on whether $\mathcal{A}_r$ shows FDS or SDS. The strands can be labeled such that strands 1 and 3 always show recombination between $\mathcal{A}_{r-1}$ and $\mathcal{A}_r$. For the other $k - 1$ intervals showing tetratype, recombinations can occur on four possible pairs of strands. The probability of each subcell can be denoted by $p_{1i_2 \ldots i_n}(h_0, h_1 \ldots h_{k-1})$, where h_0 is 0 or 1 if $\mathcal{A}_r$ shows FDS or SDS, and each h_j is 1, 2, 3, or 4 for $j \geq 1$. Using arguments similar to those in Zhao *et al.* (1995a), it can be shown that there is an underlying chiasma process satisfying NCI compatible with $\mathbf{p}^t_0$ and $\mathbf{p}^t_1$ if and only if $\mathbf{T}_{n-1}^{-1}\mathbf{p}^t_0 \geq \mathbf{0}$, $\mathbf{T}_{n-1}^{-1}\mathbf{p}^t_1 \geq \mathbf{0}$, all the subcell probabilities $p^t_{0i_2 \ldots i_n}(h_1, \ldots, h_r)$ in a cell $i_2 \ldots i_n$ with $i_r = 1$ for more than one *r* are equal, and all the subcell probabilities $p^t_{1i_2 \ldots i_n}(h_0, h_1, \ldots, h_r)$ in a cell $i_2 \ldots i_n$ with $i_r = 1$ for one ore more *r* are equal.

Markers on different sides of the centromere: Consider markers on one side of the centromere in the order of *CEN–*$\mathcal{A}_1$–$\mathcal{A}_2$– · · · –$\mathcal{A}_{n_1}$ and markers on the other side in the order of *CEN–*$\mathcal{B}_1$–$\mathcal{B}_2$– · · · –$\mathcal{B}_{n_2}$. If the centromere

could be observed, any joint tetrad pattern can be represented by $(i_1 i_2 \ldots i_{n_1}; j_1 j_2 \ldots j_{n_2})$, where $i_r = 0$, 1, or 2 corresponding to *P*, *T*, or *N* between $\mathcal{A}_{r-1}$ and $\mathcal{A}_r$, $j_s =$ 0, 1, or 2 corresponding to *P*, *T*, or *N* between $\mathcal{B}_{s-1}$ and $\mathcal{B}_s$, and $\mathcal{A}_0$ and $\mathcal{B}_0$ both denote the same centromere. Because the centromere is not observable, both $(0i_2 \ldots i_{n_1}; 0j_2 \ldots j_{n_2})$ and $(2i_2 \ldots i_{n_1}; 2j_2 \ldots j_{n_2})$ show FDS at $\mathcal{A}_1$ and $\mathcal{B}_1$ and parental ditype between $\mathcal{A}_1$ and $\mathcal{B}_1$, they are not distinguishable. Similarly, $(0i_2 \ldots i_{n_1}; 1j_2 \ldots j_{n_2})$ is not distinguishable from $(2i_2 \ldots i_{n_1}; 1j_2 \ldots j_{n_2})$, $(1i_2 \ldots i_{n_1}; 0j_2 \ldots j_{n_2})$ is not distinguishable from $(1i_2 \ldots i_{n_1}; 2j_2 \ldots j_{n_2})$, and $(0i_2 \ldots i_{n_1}; 2j_2 \ldots j_{n_2})$ is not distinguishable from $(2i_2 \ldots i_{n_1}; 0j_2 \ldots j_{n_2})$. For SDS at both $\mathcal{A}_1$ and $\mathcal{B}_1$, that is, tetratype in the intervals $\mathcal{A}_1$–*CEN* and *CEN*–$\mathcal{B}_1$, there are three distinguishable types based on the configuration between $\mathcal{A}_1$ and $\mathcal{B}_1$: *P*, *T*, or *N*.

We combine tetrad types having *P* between $\mathcal{A}_1$ and $\mathcal{B}_1$, $(0i_2 \ldots i_{n_1}; 0j_2 \ldots j_{n_2})$, $(2i_2 \ldots i_{n_1}; 2j_2 \ldots j_{n_2})$, and one of the three types of $(1i_2 \ldots i_{n_1}; 1j_2 \ldots j_{n_2})$ showing *P* between $\mathcal{A}_1$ and $\mathcal{B}_1$, and denote the grouped type by $(P; i_2 \ldots i_{n_1}; j_2 \ldots j_{n_2})$. Similarly, we obtain new grouped types, $(T; i_2 \ldots i_{n_1}; j_2 \ldots j_{n_2})$ and $(N; i_2 \ldots i_{n_1}; j_2 \ldots j_{n_2})$, where the tetrad types between $\mathcal{A}_1$ and $\mathcal{B}_1$ are *T* and *P*. It can be shown that the inequality constraints imposed by NCI on ordered tetrads are the same as the inequality constraints imposed on unordered tetrads applied to the above new grouped types. In the new grouped types, FDS or SDS information is ignored at $\mathcal{A}_1$ and $\mathcal{B}_1$. The equality constraints can be established but are more complex; we omit the details here.

Genetic mapping (one marker): The probabilities of FDS and SDS at a marker $\mathcal{A}$ can be related to the map distance between *CEN* and $\mathcal{A}$ if a chiasma process model is specified. We study several chiasma models and compare various map functions derived from these models and map functions proposed in the literature. Note that centromeres can be mapped using other types of data. When markers at the centromere are available, the centromere can be treated as a marker and standard mapping procedures can be used to map centromeres (Ferguson-Smith *et al.* 1975). For unordered tetrads, centromeres can be mapped with three markers on three chromosomes (Perkins 1949).

Complete interference model: If there is at most one chiasma between *CEN* and $\mathcal{A}$, let c_0 and c_1 denote the probabilities of having 0 and 1 chiasma; then, $F_{\mathcal{A}} = p(\text{FDS}) = c_0$ and $S_{\mathcal{A}} = p(\text{SDS}) = c_1$. The map distance *d* between *CEN* and $\mathcal{A}$ is $c_1/2$. Therefore, $F_{\mathcal{A}} = 1 - 2d$ and $S_{\mathcal{A}} = 2d$.

If more than one chiasma is allowed, map distance *d* cannot be estimated from $F_{\mathcal{A}}$ and $S_{\mathcal{A}}$ unless the chiasma process is fully specified with the map distance as the only unknown parameter.

Poisson model: The most widely used chiasma process model is the Poisson process, which imposes no chiasma interference. In this model, the probability of *k* chiasmata between *CEN* and $\mathcal{A}$ is $e^{-2d}(2d)^k/k!$. Therefore, from (1),

$$F_{\mathcal{A}}(d) = \sum_{k=0}^{\infty}\left(e^{-2d}\frac{(2d)^k}{k!}\right)\left[\frac{2}{3}\left(\frac{1}{2} + \left(-\frac{1}{2}\right)^k\right)\right] = \frac{1}{3}(1 + 2e^{-3d}), \tag{3}$$

and

$$S_{\mathcal{A}}(d) = 1 - F_{\mathcal{A}}(d) = \frac{2}{3}(1 - e^{-3d}).$$

Under the complete interference model, the SDS proportion is twice the map distance. Under the Poisson model, which imposes no chiasma interference, the SDS proportion will never exceed $\frac{2}{3}$. Therefore, for ordered tetrad data, the presence of chiasma interference can be shown with just a single marker if NCI is assumed and the observed SDS proportion is significantly above $\frac{2}{3}$. In many organisms, the SDS proportion was observed to be larger than $\frac{2}{3}$ for some markers (Weinstein 1936; Barratt *et al.* 1954; Perkins 1962; Deka *et al.* 1990). On the other hand, for many markers, the observed SDS proportion is less than twice the map distance, especially for markers far from the centromere, indicating less than complete interference.

There are several proposals in the literature to incorporate chiasma interference in relating the map distance and the SDS proportion. The earliest one appears to be the model proposed by Barratt *et al.* (1954). In their model, the probability of having $r \geq 1$ chiasmata is

$$e^{-2x}\frac{(2x)^r}{r!}\alpha^{r-1}\frac{S_{\alpha=1}}{S_\alpha}, \tag{4}$$

where

$$S_\alpha = \sum_{r=1}^{\infty} e^{-2x}\frac{(2x)^r}{r!}\alpha^{r-1}.$$

Map distances and SDS proportions can be expressed in terms of *x* and α. Barratt *et al.* (1954) used *k* instead of α in (4). To avoid confusion with other notation in this article, α is used in the following discussion. Barratt *et al.* (1954) found that α between 0.2 and 0.3 provided good fit to Drosophila and Neurospora data.

After trying out many candidates for simple map functions for SDS proportions, Ott *et al.* (1976) found that, for SDS proportions between 0 and 0.6, the function $S_{\mathcal{A}} = \frac{2}{3}\sin(3d)$ was in excellent agreement with the empirical data in Perkins (1962).

On the basis of a map function relating the map distance *d* and the recombination fraction θ between two markers proposed by Rao *et al.* (1977),

$$\begin{aligned} d = \{&p(2p-1)(1-4p)\ln(1-2\theta) \\ &+ 16p(p-1)(2p-1)\tan^{-1}(2\theta) \\ &+ 2p(1-p)(8p+2)\tanh^{-1}(2\theta) \\ &+ 6(1-p)(1-2p)(1-4p)\theta\}/6, \end{aligned}$$

Morton *et al.* (1990) proposed the map function $S_{\mathcal{A}} = 3\theta - d$.

Here we will compare these map functions with map functions derived from the chi-square chiasma interference models. The chi-square model, first introduced by Fisher *et al.* (1947), was suggested as a plausible biological model by Foss *et al.* (1993), although there are now doubts concerning the appropriateness of this motivation (Foss and Stahl 1995). In the Foss *et al.* (1993) study, the model is represented in the form of $Cx(Co)^m$, as follows: assume the crossover intermediates are randomly distributed along the four strand bundle, and every intermediate resolves either as a crossover (*Cx*) or not (*Co*). When an intermediate resolves as a *Cx*, the next m intermediates must resolve as a *Co*, and after mCo's the next intermediate must resolve as a *Cx*. The process is made stationary by allowing the leftmost crossover intermediate an equal chance to be one of $Cx(Co)^m$. The chi-square model was found to provide good fit to data from different organisms (Zhao *et al.* 1995b). One nice property of the chi-square model is that the probability of any ordered tetrad pattern has a closed-form expression, thus facilitating genetic data analysis under this model.

Let $p = m + 1$, and define $\mathbf{D}_k(y)$ to be the matrix whose (i, j)th entry is $d_k(ij) = e^{-y}y^{pk+j-i}/(pk + j - i)!$ if $pk + j - i \geq 0$, and $d_k(ij) = 0$ otherwise. Let $\mathbf{1} = (1, 1, \ldots, 1)'$ and $\alpha = (1/p)\mathbf{1}'$. For an interval defined by parameter y, the map distance d is $y/2p$ because (1) the average number of crossover intermediates between these two markers is y; (2) one out of every $p = m + 1$ intermediates resolves as a crossover; and (3) each strand has a chance of ½ of being involved in each crossover. Therefore, a given strand is involved in a crossover for every $2p$ crossover intermediates. The probability of having k chiasmata between two markers is $c_k = \boldsymbol{\alpha}\mathbf{D}_k\mathbf{1}$ (Zhao *et al.* 1995b). For the simple case of $m = 1$,

$$\mathbf{D}_k(y) = e^{-y}\begin{pmatrix} y^{2k}/(2k)! & y^{2k+1}/(2k+1)! \\ y^{2k-1}/(2k-1)! & y^{2k}/(2k)! \end{pmatrix},$$

$$\text{for } k \geq 1.$$

Therefore, from (1),

$$\begin{aligned} F_{\mathcal{A}}(y) &= \sum_{k=0}^{\infty} c_k\left[\frac{2}{3}\left(\frac{1}{2} + \left(-\frac{1}{2}\right)^k\right)\right] \\ &= \boldsymbol{\alpha}\left(\sum_{k=0}^{\infty}\left[\frac{2}{3}\left(\frac{1}{2} + \left(-\frac{1}{2}\right)^k\right)\right]\mathbf{D}_k(y)\right)\mathbf{1} \\ &= \frac{1}{2}(1,1)e^{-y}\frac{1}{6} \\ &\quad \times \begin{pmatrix} e^y + e^{+y} + 4\cos\left(\frac{y}{\sqrt{2}}\right) & e^y - e^{-y} + 4\sqrt{2}\sin\left(\frac{y}{\sqrt{2}}\right) \\ e^y - e^{-y} - 2\sqrt{2}\sin\left(\frac{y}{\sqrt{2}}\right) & e^y + e^{-y} + 4\cos\left(\frac{y}{\sqrt{2}}\right) \end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} \\ &= \frac{e^{-y}}{6}\left(2e^y + 4\cos\left(\frac{y}{\sqrt{2}}\right) + \sqrt{2}\sin\left(\frac{y}{\sqrt{2}}\right)\right), \end{aligned}$$

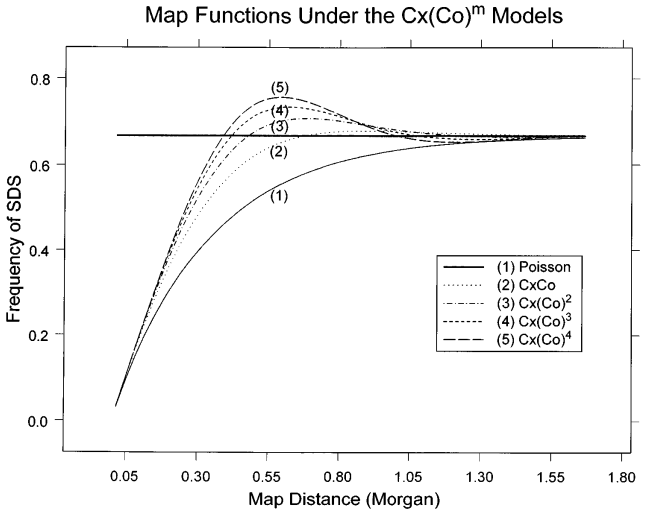

Figure 1.—Map function relating the map distance between the centromere and the marker to the second division segregation proportion under different chi-square models. The upper limit for the frequency of SDS under the Poisson model, $y = ⅔$, is also plotted in the figure.

where d is related to y by $d = y/4$ from the above discussion. The expressions of $F_{\mathcal{A}}(y)$ and $S_{\mathcal{A}}(y)$ are more complicated for $m > 1$. Map functions relating the SDS proportion and the map distance for different m's are plotted in Figure 1. Note that $m = 0$ corresponds to the no-interference model, that is, the Poisson model. Under the no interference model, the SDS proportion never goes above ⅔. For $m > 0$, the SDS proportion rises above ⅔. As m increases, the maximal value of $S_{\mathcal{A}}$ increases, and it is achieved at smaller d. For $m > 0$, there is no one-to-one correspondence between $S_{\mathcal{A}}$ and d. Therefore, the centromere cannot be uniquely mapped when the SDS proportion is larger than 0.6, and chiasma interference cannot be ruled out.

To compare map functions proposed in the literature, we plot different map functions in Figure 2. The map functions presented are: (1) the map function under the complete-interference model, (2) the map function under the no-interference model, (3) the map function proposed by Barratt *et al.* (1954) with $\alpha = 0.3$, (4) the map function proposed by Ott *et al.* (1976), (5) the map function proposed by Morton *et al.* (1990) with $p = 0.40$, and (6) the map function under the $Cx(Co)^2$ model. It is clear from Figure 2 that all map functions, except those under the complete-interference model and the no-interference model, agree with each other fairly well for $S_{\mathcal{A}}$ up to ⅔. Therefore, the map functions proposed in the literature can be well approximated by the map functions under the $Cx(Co)^2$ model. In the context of single-spore data, it was also found that map functions under the chi-square model can approximate most map functions in the literature

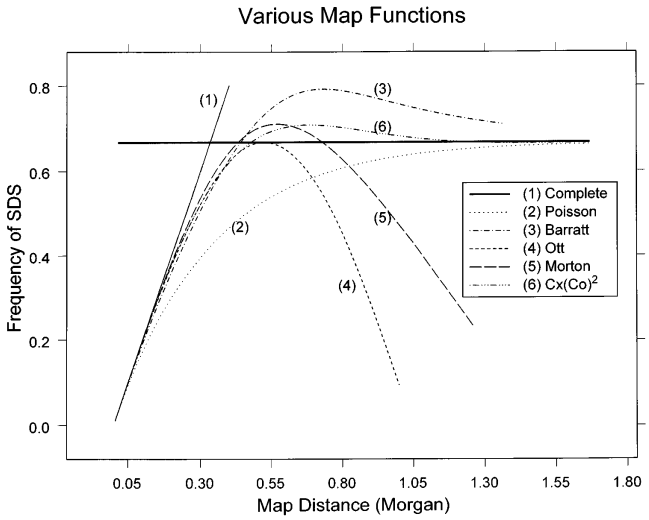

Figure 2.—Comparison of different map functions proposed in the literature. The upper limit for the frequency of SDS under the Poisson model, $y = \frac{2}{3}$, is also plotted in the figure.

(Zhao and Speed 1996). Therefore, the chi-square model is a good candidate for multilocus analysis of ordered tetrad data.

Genetic mapping (two markers): From two-marker ordered tetrad data, the map distances among two markers and the centromere can be estimated for a given chiasma process model. Here we derive joint ordered tetrad probabilities under the chi-square model. A special case of the chi-square model, the Poisson model, is studied separately, because joint tetrad probabilities can be expressed rather easily under this model. We consider markers on different sides of the centromere and markers on the same side of the centromere in turn.

Markers on different sides of the centromere (Poisson model): For a Poisson chiasma process, if the map distance between *CEN* and $\mathcal{A}$ is d_1, and if the centromere could be observed, $p_0(d_1) = \frac{1}{6}(1 + 2e^{-3d_1} + 3e^{-2d_1})$, $p_1(d_1) = \frac{1}{3}(2 - 2e^{-3d_1})$, and $p_2(d_1) = \frac{1}{6}(1 + 2e^{-3d_1} - 3e^{-2d_1})$, where $p_0(d_1)$, $p_1(d_1)$, and $p_2(d_1)$ are the probabilities of *P*, *T*, and *N* between *CEN* and $\mathcal{A}$, respectively (Haldane 1931). If the map distance between *CEN* and $\mathcal{B}$ is d_2, similarly we obtain $p_0(d_2)$, $p_1(d_2)$, and $p_2(d_2)$, the probabilities of *P*, *T*, and *N* between *CEN* and $\mathcal{B}$. The joint tetrad probability p_{ij} for type (ij), where $i, j = 0$, 1, or 2, is $p_i(d_1)p_j(d_2)$. Therefore, the five probabilities in Table 5 are:

$$\alpha = p_{00} + p_{22} = p_0(d_1)p_0(d_2) + p_2(d_1)p_2(d_2),$$
$$\beta = p_{02} + p_{20} = p_0(d_1)p_2(d_2) + p_2(d_1)p_0(d_2),$$
$$\gamma = p_{01} + p_{21} = p_0(d_1)p_1(d_2) + p_2(d_1)p_1(d_2),$$
$$\delta = p_{10} + p_{12} = p_1(d_1)p_0(d_2) + p_1(d_1)p_2(d_2),$$
$$\varepsilon = p_{11} = p_1(d_1)p_1(d_2).$$

Markers on different sides of the centromere (chi-square model): The chi-square model $Cx(Co)^m$ assumes that the chiasma process is stationary. This model has been applied mostly to markers on the same side of the centromere. Because the chiasma interference pattern may be different across the centromere, two chi-square models starting from the centromere toward two different telomeres may be necessary to model the chiasma process. In general, we may model interference across the centromere by relating the two most proximal crossover intermediates on two sides of the centromere. For example, we may assign the most proximal crossover intermediates on both sides of the centromere as the first *Co* after a *Cx*, thus inducing a higher chiasma interference in the centromere region than those in other regions. Or we may assign these crossover intermediates as the *m*th *Co* after a *Cx*. This will induce a lower chiasma interference in the centromere region than those in other regions. For simplicity, in this discussion we assume that starting from the centromere, there are two stationary chiasma processes on the two arms of the chromosome. In this case, there is no chiasma interference between the two arms.

For marker $\mathcal{A}$, if the centromeres from the two parents could be distinguished, the probabilities p_0, p_1, and p_2 of *P*, *T*, or *N* between *CEN* and $\mathcal{A}$ can be evaluated as follows. Let $\mathbf{D}_k(y)$ be as defined above; the probability of having *k* chiasmata between *CEN* and $\mathcal{A}$ is $c_k = \boldsymbol{\alpha}\mathbf{D}_k\mathbf{1}$. Define $\mathbf{P}(y) = \Sigma_{k=0}^{\infty} p_0^k \mathbf{D}_k(y)$, $\mathbf{T}(y) = \Sigma_{k=0}^{\infty} p_1^k \mathbf{D}_k(y)$, and $\mathbf{N}(y) = \Sigma_{k=0}^{\infty} p_2^k \mathbf{D}_k(y)$, where p_0^k, p_1^k, and p_2^k were defined in (2). Then $p_0 = \boldsymbol{\alpha}\mathbf{P}(y)\mathbf{1}$, $p_1 = \boldsymbol{\alpha}\mathbf{T}(y)\mathbf{1}$, and $p_2 = \boldsymbol{\alpha}\mathbf{N}(y)\mathbf{1}$. The relation between the map distance *d* and the parameter *y* is $d = y/2(m + 1)$. Using these results, $p_0(d_1)$, $p_1(d_1)$, and $p_2(d_1)$ can be obtained. Similarly, the probability of *P*, *T*, or *N* between *CEN* and $\mathcal{B}$, $p_0(d_2)$, $p_1(d_2)$, and $p_2(d_2)$ can be evaluated. The joint tetrad probability p_{ij} is $p_i(d_1)p_j(d_2)$. Therefore, the five probabilities can be obtained as in the Poisson model.

When $m = 1$, it can be shown that

$$\mathbf{P}(y) = e^{-y}\frac{1}{12} \times \begin{pmatrix} 6 + e^{y} + e^{-y} + 4\cos\left(\frac{y}{\sqrt{2}}\right) & 6y + e^{y} - e^{-y} + 4\sqrt{2}\sin\left(\frac{y}{\sqrt{2}}\right) \\ e^{y} - e^{-y} - 2\sqrt{2}\sin\left(\frac{y}{\sqrt{2}}\right) & 6 + e^{y} + e^{-y} + 4\cos\left(\frac{y}{\sqrt{2}}\right) \end{pmatrix},$$

$$\mathbf{T}(y) = e^{-y}\frac{1}{3} \times \begin{pmatrix} e^{y} + e^{-y} - 2\cos\left(\frac{y}{\sqrt{2}}\right) & e^{y} - e^{-y} - 2\sqrt{2}\sin\left(\frac{y}{\sqrt{2}}\right) \\ e^{y} - e^{-y} + \sqrt{2}\sin\left(\frac{y}{\sqrt{2}}\right) & e^{y} + e^{-y} - 2\cos\left(\frac{y}{\sqrt{2}}\right) \end{pmatrix},$$

and

$$\mathbf{N}(y) = e^{-y}\frac{1}{12} \times \begin{pmatrix} -6 + e^{y} + e^{-y} + 4\cos\left(\frac{y}{\sqrt{2}}\right) - 6y + e^{y} - e^{-y} + 4\sqrt{2}\sin\left(\frac{y}{\sqrt{2}}\right) \\ e^{y} - e^{-y} - 2\sqrt{2}\sin\left(\frac{y}{\sqrt{2}}\right) - 6 + e^{y} + e^{-y} + 4\cos\left(\frac{y}{\sqrt{2}}\right) \end{pmatrix},$$

where $d = y/4$. Even for this simple model, the analytical forms are not so simple. No general results for arbitrary m are presented in this article.

Markers on the same side of the centromere (Poisson model): For a Poisson chiasma process, if the map distance between *CEN* and $\mathcal{A}$ is d_1, as shown in Equation 3, then $F_{\mathcal{A}}(d_1) = \frac{1}{3}(1 + 2e^{-3d_1})$ and $S_{\mathcal{A}}(d_1) = \frac{2}{3}(1 - e^{-3d_1})$. If the map distance between $\mathcal{A}$ and $\mathcal{B}$ is d_2, $p_0(d_2) = \frac{1}{6}(1 + 2e^{-3d_2} + 3e^{-2d_2})$, $p_1(d_2) = \frac{2}{3}(1 - e^{-3d_2})$, and $p_2(d_2) = \frac{1}{6}(1 + 2e^{-3d_2} - 3e^{-2d_2})$, where $p_0(d_2)$, $p_1(d_2)$, and $p_2(d_2)$ are the probabilities of *P*, *T*, and *N* between $\mathcal{A}$ and $\mathcal{B}$, respectively. The six probabilities in Table 6 are

$$\begin{aligned} \alpha &= F_{\mathcal{A}}(d_1)p_0(d_2), \\ \beta &= F_{\mathcal{A}}(d_1)p_1(d_2), \\ \gamma &= F_{\mathcal{A}}(d_1)p_2(d_2), \\ \delta &= S_{\mathcal{A}}(d_1)p_0(d_2), \\ \phi &= S_{\mathcal{A}}(d_1)p_1(d_2), \\ \varepsilon &= S_{\mathcal{A}}(d_1)p_2(d_2). \end{aligned}$$

Markers on the same side of the centromere (chi-square model): Under the chi-square model, the joint tetrad probability $c_{j\ell}$ of having k and ℓ chiasmata in the intervals (*CEN*, $\mathcal{A}$) and ($\mathcal{A}$, $\mathcal{B}$) is $\boldsymbol{\alpha}\mathbf{D}_k(y_1)\mathbf{D}_\ell(y_2)\mathbf{1}$, where $\boldsymbol{\alpha}$, $\mathbf{D}_k(y)$, and $\mathbf{1}$ were defined above (Zhao *et al.* 1995b). For joint tetrad type $(i_1 i_2)$,

$$\begin{aligned} p_{i_1 i_2} &= \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} c_{k\ell} p_{i_1}^{k} p_{i_2}^{\ell} \\ &= \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} [\boldsymbol{\alpha}\mathbf{D}_k(y_1)\mathbf{D}_\ell(y_2)\mathbf{1}] p_{i_1}^{k} p_{i_2}^{\ell} \\ &= \boldsymbol{\alpha}\left[\sum_{k=0}^{\infty} p_{i_1}^{k}\mathbf{D}_k(y_1)\right]\left[\sum_{\ell=0}^{\infty} p_{i_2}^{\ell}\mathbf{D}_\ell(y_2)\right]\mathbf{1}, \end{aligned}$$

where $p_{i_1}^{k}$ is the conditional probability for FDS ($i_1 = 0$) or SDS ($i_1 = 1$) defined in Equation 1, and $p_{i_2}^{\ell}$ is the conditional tetrad type probability defined in Equation 2. Define $\mathbf{F}(y) = \Sigma_{k=0}^{\infty}[\frac{2}{3}(\frac{1}{2} + (-\frac{1}{2})^k)]\mathbf{D}_k(y)$ and $\mathbf{S}(y) = \Sigma_{k=0}^{\infty}[\frac{2}{3}(1 - (-\frac{1}{2})^k)]\mathbf{D}_k(y)$. For any joint tetrad pattern $(i_1 i_2)$, $p_{i_1 i_2} = \boldsymbol{\alpha}\mathbf{M}_1(y_1)\mathbf{M}_2(y_2)\mathbf{1}$, where $\mathbf{M}_1(y_1) = \mathbf{F}(y_1)$ or $\mathbf{S}(y_1)$ when $i_1 = 0$ or 1, and $\mathbf{M}_2(y_2) = \mathbf{P}(y_2)$, $\mathbf{T}(y_2)$, or $\mathbf{N}(y_2)$ when $i_2 = 0$, 1, or 2. The matrices $\mathbf{P}(y_2)$, $\mathbf{T}(y_2)$, and $\mathbf{N}(y_2)$ were defined above. Explicit expressions for $\mathbf{F}(y)$, $\mathbf{S}(y)$, $\mathbf{P}(y)$, $\mathbf{T}(y)$, and $\mathbf{N}(y)$ were obtained in previous discussion under the *CxCo* model.

Genetic mapping (multiple markers): As before, markers on the same side of the centromere and on different sides of the centromere are considered separately.

Markers on the same side of the centromere: Consider n markers $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n$ in the order of *CEN*–$\mathcal{A}_1$–$\mathcal{A}_2$–$\cdots$–$\mathcal{A}_n$. Under NCI, there are $2 \times 3^{n-1}$ different probabilities corresponding to patterns $(i_1 i_2 \ldots i_n)$. These $2 \times 3^{n-1}$ types were mentioned in the discussion of the NCI assumption for the multiple marker case. Denote the map distance between $\mathcal{A}_{r-1}$ and $\mathcal{A}_r$ by d_r, where $\mathcal{A}_0$ is the centromere.

For a Poisson chiasma process, from the previous discussion, $F_{\mathcal{A}_1}(d_1) = \frac{1}{3}(1 + 2e^{-3d_1})$, $S_{\mathcal{A}_1}(d_1) = \frac{2}{3}(1 - e^{-3d_1})$, $p_0(d_r) = \frac{1}{6}(1 + 2e^{-3d_r} + 3e^{-2d_r})$, $p_1(d_r) = \frac{2}{3}(1 - e^{-3d_r})$, and $p_2(d_r) = \frac{1}{6}(1 + 2e^{-3d_r} - 3e^{-2d_r})$. The probability of tetrad pattern $(i_1 i_2 \ldots i_n)$ is $f \times \Pi_{r=2}^{n} p_{i_r}(d_r)$, where f is $F_{\mathcal{A}_1}(d_1)$ or $S_{\mathcal{A}_1}(d_1)$ when $i_1 = 0$ or 1.

Under the chi-square model, define $\mathbf{F}(y)$, $\mathbf{S}(y)$, $\mathbf{P}(y)$, $\mathbf{T}(y)$, and $\mathbf{N}(y)$ as above. The probability of tetrad pattern $(i_1 i_2 \ldots i_n)$ is $\boldsymbol{\alpha}(\Pi_{r=1}^{n}\mathbf{M}_r)\mathbf{1}$, where $\mathbf{M}_1 = \mathbf{F}(y_1)$ or $\mathbf{S}(y_1)$ for $i_1 = 0$ or 1, and $\mathbf{M}_r = \mathbf{P}(y_r)$, $\mathbf{T}(y_r)$, or $\mathbf{N}(y_r)$ for $i_r = 0$, 1, or 2 when $r \geq 2$. The parameter y_r and the map distance d_r are related by $d_r = y_r/2(m + 1)$.

Markers on different sides of the centromere: Consider markers in the order of $\mathcal{B}_{n_2}$– $\cdots$ –$\mathcal{B}_1$–*CEN*–$\mathcal{A}_1$–$\mathcal{A}_2$– $\cdots$ –$\mathcal{A}_{n_1}$. If the two chiasma processes on different sides of the centromere are independent, we may first consider the case in which the centromere could be observed. For tetrad pattern $(i_1 i_2 \ldots i_n)$ on markers *CEN*, $\mathcal{A}_1$, $\mathcal{A}_2$, $\cdots$, and $\mathcal{A}_n$, $p_{(i_1 i_2 \ldots i_{n_1})} = \boldsymbol{\alpha}(\Pi_{r=1}^{n_1}\mathbf{M}_r)\mathbf{1}$, where $\mathbf{M}_r = \mathbf{P}(y_r)$, $\mathbf{T}(y_r)$, and $\mathbf{N}(y_r)$ for $i_r = 0$, 1, and 2. The map distance between $\mathcal{A}_{r-1}$ and $\mathcal{A}_r$ is $d_r = y_r/2(m + 1)$. For tetrad pattern $(j_1 j_2 \ldots j_{n_2})$ on markers *CEN*, $\mathcal{B}_1$, $\mathcal{B}_2$, $\cdots$, and $\mathcal{B}_{n_2}$, $p_{(j_1 j_2 \ldots j_{n_2})} = \boldsymbol{\alpha}(\Pi_{s=1}^{n_2}\mathbf{M}_s)\mathbf{1}$, where $\mathbf{M}_s = \mathbf{P}(y_s)$, $\mathbf{T}(y_s)$, and $\mathbf{N}(y_s)$ *for* $i_s = 0$, 1, and 2. The map distance between $\mathcal{B}_{s-1}$ and $\mathcal{B}_s$ is $d_s = y_s/2(m + 1)$. Because the centromere is not observable, instead of $3^{n_1+n_2}$ probabilities, there are $5 \times 3^{n_1+n_2-2}$ distinct probabilities. These $5 \times 3^{n_1+n_2-2}$ distinct probabilities can be denoted by $(o; i_2 \ldots i_{n_1}; j_2 \ldots j_{n_2})$, where $o = 0$ corresponds to FDS at both $\mathcal{A}_1$ and $\mathcal{B}_1$ and the tetrad type between $\mathcal{A}_1$ and $\mathcal{B}_1$ being *P*, $o = 1$ corresponds to FDS at both $\mathcal{A}_1$ and $\mathcal{B}_1$ and the tetrad type between $\mathcal{A}_1$ and $\mathcal{B}_1$ being *N*, $o = 2$ corresponds to FDS at $\mathcal{A}_1$ and SDS at $\mathcal{B}_1$, $o = 3$ corresponds to SDS at $\mathcal{A}_1$ and FDS at $\mathcal{B}_1$, and $o = 4$ corresponds to SDS at both $\mathcal{A}_1$ and $\mathcal{B}_1$. The probability of type $(o; i_2 \ldots i_{n_1}; j_2 \ldots j_{n_2})$ is

$$\begin{aligned} p_{(0;i_2\ldots i_{n_1};j_2\ldots j_{n_2})} &= p_{(0i_2\ldots i_{n_1})}p_{(0j_2\ldots j_{n_2})} + p_{(2i_2\ldots i_{n_1})}p_{(2j_2\ldots j_{n_2})}, \\ p_{(1;i_2\ldots i_{n_1};j_2\ldots j_{n_2})} &= p_{(0i_2\ldots i_{n_1})}p_{(2j_2\ldots j_{n_2})} + p_{(2i_2\ldots i_{n_1})}p_{(0j_2\ldots j_{n_2})}, \\ p_{(2;i_2\ldots i_{n_1};j_2\ldots j_{n_2})} &= p_{(0i_2\ldots i_{n_1})}p_{(1j_2\ldots j_{n_2})} + p_{(2i_2\ldots i_{n_1})}p_{(1j_2\ldots j_{n_2})}, \\ p_{(3;i_2\ldots i_{n_1};j_2\ldots j_{n_2})} &= p_{(1i_2\ldots i_{n_1})}p_{(0j_2\ldots j_{n_2})} + p_{(1i_2\ldots i_{n_1})}p_{(2j_2\ldots j_{n_2})}, \\ p_{(4;i_2\ldots i_{n_1};j_2\ldots j_{n_2})} &= p_{(1i_2\ldots i_{n_1})}p_{(1j_2\ldots j_{n_2})}. \end{aligned}$$

TABLE 7

Observed and expected counts of seven groups for markers $\mathcal{MT}$ and $\mathcal{AD}$ under four possible orders for $\mathcal{MT}$ and $\mathcal{AD}$

Type	Observed	(1)	(2)	(3)	(4)
[*AB*, *AB*, *ab*, *ab*]	1014	486	1014	1014	1014
[*Ab*, *Ab*, *aB*, *aB*]	0	486	0	0	0
[*AB*, *Ab*, *aB*, *ab*]	97	140	97	97	48
[*AB*, *aB*, *Ab*, *ab*]	1	44	1	0	1
[*AB*, *ab*, *AB*, *ab*]	49	2	12	49	49
[*AB*, *ab*, *Ab*, *aB*]	0	3	25	0	48
[*Ab*, *aB*, *Ab*, *aB*]	0	2	12	0	0
log-likelihood		−1416	−609	−541	−607

(1) On different chromosomes (CEN_1–$\mathcal{MT}$ and CEN_2–$\mathcal{AD}$); (2) on different sides of the centromere ($\mathcal{MT}$–*CEN*–$\mathcal{AD}$); (3) on the same side of the centromere in the order of *CEN*–$\mathcal{AD}$–$\mathcal{MT}$; and (4) on the same side of the centromere in the order of *CEN*–$\mathcal{MT}$–$\mathcal{AD}$. The two alleles at $\mathcal{AD}$ are represented by *A* and *a*, and the two alleles at $\mathcal{MT}$ are represented by *B* and *b*.

RESULTS

In this section, the methods developed and described in methods are used to find the order of a set of markers and to estimate map distances between the centromere and genetic markers.

Order markers under NCI (two markers): As discussed above and summarized in Tables 4–6, different orders of two markers impose different constraints among the probabilities of seven groups in Table 3. These constraints can be used to order two markers. Data from Howe (1956) are analyzed here to illustrate the procedure. The first pair of markers analyzed are $\mathcal{MT}$ and $\mathcal{AD}$. The observed numbers of tetrads for the seven groups are shown in Table 7. It is clear that the data satisfy the constraints under the order *CEN*–$\mathcal{AD}$–$\mathcal{MT}$ but not the constraints under other orders. Thus, the order *CEN*–$\mathcal{AD}$–$\mathcal{MT}$ can be established. To make the inference more rigorous, the maximum likelihood estimates of the probabilities for the seven groups and the corresponding maximum likelihoods were calculated under the four possible orders: (1) CEN_1–$\mathcal{AD}$, CEN_2–$\mathcal{MT}$, (2) $\mathcal{AD}$–*CEN*–$\mathcal{MT}$, (3) *CEN*–$\mathcal{AD}$–$\mathcal{MT}$, and (4) *CEN*–$\mathcal{MT}$–$\mathcal{AD}$. It is straightforward to obtain the maximum likelihood estimates under order (1). To find the maximum likelihood estimates under the linear inequality constraints among the seven probabilities for orders (2), (3), and (4), an expectation maximization (EM) algorithm (Dempster *et al.* 1977) was implemented to find the maximum likelihood estimates under each order. This algorithm is similar to the EM algorithm used in Zhao *et al.* (1995a, p. 1061) in that it treats the unobserved chiasma frequencies as constituting the complete data. The details of this algorithm are provided in the appendix (EM algorithm). The expected number of observations for each group and the maximized log-likelihood under each order are given in Table 7. Among the four orders, the order *CEN*–$\mathcal{AD}$–$\mathcal{MT}$ yielded the largest maximized log-likelihood, thus establishing the order *CEN*–$\mathcal{AD}$–$\mathcal{MT}$. The second pair of markers analyzed are $\mathcal{MT}$ and $\mathcal{VIS}$. The observations as well as the expected values and maximized log-likelihoods under the four orders are summarized in Table 8. Comparing the maximized log-likelihoods under the four orders leads to the order $\mathcal{MT}$–*CEN*–$\mathcal{VIS}$. The last pair of markers studied are $\mathcal{MT}$ and $\mathcal{RIB}$ (Table 9). The data are consistent with $\mathcal{MT}$ and $\mathcal{RIB}$ being on different chromosomes. Perkins (1953) discussed the detection of linkage using unordered tetrads. With ordered tetrads, as can be seen from these examples, it is not only possible to detect linkage, but it is also possible to order the two markers relative to the centromere. This results from the extra information in ordered tetrads.

Order markers under NCI (three markers): Consider three markers $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$. Under RSCA, there are 32 distinct probabilities (appendix: Proposition 1). When $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are on the same chromosome, there are a total of 12 possible orders among them: (1) *CEN*–$\mathcal{A}$–$\mathcal{B}$–$\mathcal{C}$, (2) *CEN*–$\mathcal{A}$–$\mathcal{C}$–$\mathcal{B}$, (3) *CEN*–$\mathcal{B}$–$\mathcal{A}$–$\mathcal{C}$, (4) *CEN*–$\mathcal{B}$–$\mathcal{C}$–$\mathcal{A}$, (5) *CEN*–$\mathcal{C}$–$\mathcal{A}$–$\mathcal{B}$, (6) *CEN*–$\mathcal{C}$–$\mathcal{B}$–$\mathcal{A}$, (7) $\mathcal{A}$–*CEN*–$\mathcal{B}$–$\mathcal{C}$, (8) $\mathcal{A}$–*CEN*–$\mathcal{C}$–$\mathcal{B}$, (9) $\mathcal{B}$–*CEN*–$\mathcal{A}$–$\mathcal{C}$, (10) $\mathcal{B}$–*CEN*–$\mathcal{C}$–$\mathcal{A}$, (11) $\mathcal{C}$–*CEN*–$\mathcal{A}$–$\mathcal{B}$, and (12) $\mathcal{C}$–*CEN*–$\mathcal{B}$–$\mathcal{A}$. For a given order, NCI imposes linear equality and inequality constraints among these 32 probabilities. The maximum likelihood estimates of the 32 probabilities under these constraints can be derived using the EM algorithm (appendix, EM algorithm). Here we study the same data set analyzed for the two-marker case except that we are simultaneously analyzing three markers, $\mathcal{MT}$, $\mathcal{AD}$, and $\mathcal{VIS}$. Although there are 32 groups with distinct probabilities under RSCA, only 12 groups were observed for this data set. The number of observations for each group is given in Table 10. Because of limited space, we summarize only the maximized log-likelihoods under 12 different orders in Table 11 without giving the expected number of observations for each group under each order. Comparing the log-likelihoods, the order of $\mathcal{VIS}$–*CEN*–

TABLE 8

Observed and expected counts of seven groups for markers $\mathcal{MT}$ and $\mathcal{VIS}$ under four possible orders for $\mathcal{MT}$ and $\mathcal{VIS}$

Type	Observed	(1)	(2)	(3)	(4)
[*AB*, *AB*, *ab*, *ab*]	888	445	888	888	888
[*Ab*, *Ab*, *aB*, *aB*]	1	445	1	1	1
[*AB*, *Ab*, *aB*, *ab*]	128	128	128	128	69
[*AB*, *aB*, *Ab*, *ab*]	126	126	126	68	126
[*AB*, *ab*, *AB*, *ab*]	5	5	5	5	5
[*AB*, *ab*, *Ab*, *aB*]	10	9	9	68	69
[*Ab*, *aB*, *Ab*, *aB*]	3	5	5	3	3
log-likelihood		−1509	−900	−958	−960

(1) On different chromosomes (CEN_1–$\mathcal{MT}$ and CEN_2–$\mathcal{VIS}$); (2) on different sides of the centromere ($\mathcal{MT}$–*CEN*–$\mathcal{VIS}$); (3) on the same side of the centromere in the order of *CEN*–$\mathcal{VIS}$–$\mathcal{MT}$; and (4) on the same side of the centromere in the order of *CEN*–$\mathcal{MT}$–$\mathcal{VIS}$. The two alleles at $\mathcal{VIS}$ are represented by *A* and *a*, and the two alleles at $\mathcal{MT}$ are represented by *B* and *b*.

TABLE 9

Observed and expected counts of seven groups for markers $\mathcal{MT}$ and $\mathcal{RIB}$ under four possible orders for $\mathcal{MT}$ and $\mathcal{RIB}$

Type	Observed	(1)	(2)	(3)	(4)
[*AB*, *AB*, *ab*, *ab*]	473	497	473	473	473
[*Ab*, *Ab*, *aB*, *aB*]	521	497	228	181	221
[*AB*, *Ab*, *aB*, *ab*]	21	21	58	361	11
[*AB*, *aB*, *Ab*, *ab*]	143	143	398	72	443
[*AB*, *ab*, *AB*, *ab*]	2	1	1	2	2
[*AB*, *ab*, *Ab*, *aB*]	1	2	2	72	11
[*Ab*, *aB*, *Ab*, *aB*]	0	1	1	0	0
log-likelihood		−1248	−1509	−1832	−1541

(1) On different chromosomes (CEN_1–$\mathcal{MT}$ and CEN_2–$\mathcal{RIB}$); (2) on different sides of the centromere ($\mathcal{MT}$–*CEN*–$\mathcal{RIB}$); (3) on the same side of the centromere in the order of *CEN*–$\mathcal{MT}$–$\mathcal{RIB}$; and (4) on the same side of the centromere in the order of *CEN*–$\mathcal{RIB}$–$\mathcal{MT}$. The two alleles at $\mathcal{MT}$ are represented by *A* and *a*, and the two alleles at $\mathcal{RIB}$ are represented by *B* and *b*.

TABLE 10

Observed counts for markers $\mathcal{MT}$, $\mathcal{AD}$, and $\mathcal{VIS}$

Type	Observed counts
[*ABD*, *ABD*, *ABD*, *ABD*]	888
[*ABD*, *aBD*, *Abd*, *abd*]	85
[*ABD*, *abD*, *ABd*, *abd*]	43
[*ABD*, *ABd*, *abD*, *abd*]	126
[*ABD*, *aBd*, *AbD*, *abd*]	3
[*ABD*, *abd*, *ABD*, *abd*]	2
[*ABD*, *aBd*, *abD*, *abd*]	5
[*aBD*, *ABd*, *AbD*, *abd*]	3
[*ABD*, *abd*, *ABd*, *abD*]	2
[*aBD*, *abD*, *ABd*, *Abd*]	1
[*aBD*, *ABd*, *abD*, *Abd*]	1
[*ABd*, *abD*, *ABd*, *abD*]	2

The two alleles at $\mathcal{MT}$ are represented by *A* and *a*, the two alleles at $\mathcal{AD}$ are represented by *B* and *b*, and the two alleles at $\mathcal{VIS}$ are represented by *D* and *d*.

TABLE 11

The maximum log-likelihoods for the 12 possible orders among $\mathcal{MT}$, $\mathcal{AD}$, and $\mathcal{VIS}$ under the assumption of no chromatid interference

Order	log-likelihood
CEN–$\mathcal{MT}$–$\mathcal{AD}$–$\mathcal{VIS}$	−1090
CEN–$\mathcal{MT}$–$\mathcal{VIS}$–$\mathcal{AD}$	−1311
CEN–$\mathcal{AD}$–$\mathcal{MT}$–$\mathcal{VIS}$	−1119
CEN–$\mathcal{AD}$–$\mathcal{VIS}$–$\mathcal{MT}$	−1238
CEN–$\mathcal{VIS}$–$\mathcal{MT}$–$\mathcal{AD}$	−1208
CEN–$\mathcal{VIS}$–$\mathcal{AD}$–$\mathcal{MT}$	−1087
$\mathcal{MT}$–*CEN*–$\mathcal{AD}$–$\mathcal{VIS}$	−1091
$\mathcal{MT}$–*CEN*–$\mathcal{VIS}$–$\mathcal{AD}$	−1246
$\mathcal{AD}$–*CEN*–$\mathcal{MT}$–$\mathcal{VIS}$	−1183
$\mathcal{AD}$–*CEN*–$\mathcal{VIS}$–$\mathcal{MT}$	−1218
$\mathcal{VIS}$–*CEN*–$\mathcal{MT}$–$\mathcal{AD}$	−1126
$\mathcal{VIS}$–*CEN*–$\mathcal{AD}$–$\mathcal{MT}$	−1000

TABLE 12

Observed and expected number of individuals under different chi-square models for Neurospora

Type	Observed	Expected		
		Cx	*CxCo*	$Cx(Co)^2$
000	104	115	100	89
100	34	34	40	44
010	26	22	31	38
001	76	61	70	75
110	11	7	5	3
101	15	18	18	18
011	7	12	10	8
002	2	3	1	1
111	3	3	1	1
log-likelihood		−471	−469	−476

Data from Whitehouse (1942). Three markers on the second chromosome were studied in the order of *CEN–Pe–Tu–F*. The observed types are represented by $i_1i_2i_3$, where $i_1 = 0$ or 1 corresponding to FDS or SDS pattern at *Pe*, $i_2 = 0$, 1, or 2 corresponding to parental ditype, tetratype, or nonparental ditype between *Pe* and *Tu*, and $i_3 = 0$, 1, or 2 corresponding to parental ditype, tetratype, or nonparental ditype between *Tu* and *F*.

$\mathcal{AD}$–$\mathcal{MT}$ can be clearly established. The difference between the maximized log-likelihood under this order and the largest maximized log-likelihood under the other orders is 90. Note that when two markers on the same chromosome were analyzed, the differences between the maximized log-likelihood under the best order and the largest maximized log-likelihood under other orders were 66 (Table 7) and 58 (Table 8). This increased difference between the maximized log-likelihoods using three markers, 90 *vs.* 66 and 58, demonstrates the power gained in discriminating different marker orders by simultaneously considering multiple markers.

Gene-centromere mapping: To illustrate the application of the chi-square model to analyzing ordered tetrad data, Neurospora data discussed by Whitehouse (1942, Table 12) are presented in Table 12, along with the expected number of observations for each ordered tetrad type under the Poisson model, the *CxCo* model, and the $Cx(Co)^2$ model. Three genes were studied: *Pe*, *Tu*, and *F* on the second chromosome of Neurospora. The order of these genes is *CEN–Pe–Tu–f*. A total of 278 asci were genotyped at these three genes. The largest maximized likelihood was obtained under the *CxCo* model in the chi-square model class. Although this indicates the presence of chiasma interference, the difference between the maximum likelihoods under the Poisson model and the *CxCo* model is too small to make any definite conclusions. Under the *CxCo* model, the map distances were estimated to be 12, 9, and 20 cM in the three intervals: *CEN–Pe*, *Pe–Tu*, and *Tu–F*. We estimated the standard errors of the parameter estimates using the parametric bootstrap method by: (1) simulating data sets with the same sample size under the chi-square model, assuming the estimated parameter values; (2) estimating model parameters for each simulated data set; and (3) approximating the standard errors of the parameter estimates using the standard errors of the estimated parameter values from these simulated data sets. Using this method, the standard errors were estimated to be 1, 1, and 2 cM, respectively. Note that the $Cx(Co)^2$ model was found to be the best model using different sets of Neurospora data (Zhao *et al.* 1995b). The data analyzed in Zhao *et al.* (1995b) were of unordered type. The difference between the estimated best chi-square model could be the result of sampling errors or real difference in chiasma interference among the Neurospora strains used in different genetic experiments.

DISCUSSION

Ordered tetrads studied in this article offer information both on chromatid interference and on chiasma interference. In addition, they can be used to map centromeres. Although the number of distinguishable patterns for ordered tetrad is 6^n, under RSCA these distinguishable types reduce to $(6^n + 5 \times 2^n)/8$ distinct groups. The NCI assumption further reduces the number of distinct probabilities.

The equality and inequality constraints among ordered tetrad probabilities imposed by NCI were derived. We showed how these general constraints can be used to order markers.

Because it makes use of all the data available, multilocus analysis is advantageous over the analysis in which only a limited number of markers can be analyzed. For example, for four markers in the order of $\mathcal{A}$–$\mathcal{B}$–$\mathcal{C}$–$\mathcal{D}$, using the multilocus approach described above, if markers $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{D}$ are studied in one experiment and markers $\mathcal{A}$, $\mathcal{C}$, and $\mathcal{D}$ are studied in another experiment, data from these two experiments can be analyzed together to infer the genetic distances among these markers. However, if only one or two markers can be studied in a single analysis, data from the second experiment cannot be used to infer the map distances between $\mathcal{B}$ and other markers.

We have assumed that genotypes are available at all markers for all tetrads in our discussion. If genotypes for some markers are missing, the EM algorithm (Dempster *et al.* 1977) can be employed to carry out the likelihood analysis efficiently.

Chiasma interference has been observed in many organisms. One difficulty in multilocus analysis has been to find a good model that both incorporates chiasma interference and permits tractable analysis. Interference was considered for the analysis of unordered tetrad data (Risch and Lange 1983; Zhao *et al.* 1995b). Risch and Lange's tetrad analysis used the count-location model for the chiasma process. The count-location model was

shown to be inconsistent with experimental observations (McPeek and Speed 1995). On the other hand, the chi-square model is both consistent with biological observations and tractable for multilocus analysis. This class of models fits data from a variety of organisms well.

It was shown by Zhao and Speed (1996) that for most map functions, there is an underlying stationary renewal process that can give rise to these map functions. Furthermore, the inter-event distributions of these renewal processes can be closely approximated by gamma distributions. In this article, it was shown that three map functions proposed in the literature that relate the SDS proportion to the map distance are almost identical to the map function under the $Cx(Co)^2$ model when the SDS proportion is less than $\frac{2}{3}$. Therefore, multilocus ordered tetrad analysis using the chi-square model should provide a tractable and flexible approach to studying multilocus ordered tetrad data. The chi-square model has recently been extended to allow a more general class of interevent distributions (Lange *et al.* 1997). This generalized class of models is still tractable.

Half-tetrad data are closely related to ordered tetrad data. A detailed study of half-tetrad data is given in the accompanying article (Zhao and Speed 1998).

This work was supported by National Institutes of Health grant HG-01093. We thank two referees for their constructive comments.

LITERATURE CITED

Barratt, R. W., D. Newmeyer, D. D. Perkins and L. Garnjobst, 1954 Map construction in *Neurospora crassa*. Advan. Genet. **6:** 1–93.

Bellman, R. E., 1970 *Introduction to Matrix Analysis*, Ed. 2, McGraw-Hill, New York.

Deka, R., A. Chakravarti, U. Surti, E. Hauselman, J. Reefer *et al.*, 1990 Genetics and biology of human ovarian teratomas. II. Molecular analysis of origin of nondisjunction and gene-centromere mapping of chromosome I markers. Am. J. Hum. Genet. **47:** 644–655.

Dempster, A. P., N. M. Laird and D. B. Rubin, 1977 Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Statis. Soc. Ser. B **39:** 1–22.

Ferguson-Smith, M. A., P. M. Ellis, O. Mutchinick, K. P. Glen, G. B. Cote *et al.*, 1975 Centromeric linkage. Hum. Genet. Map. Workshop. **2:** 300–307.

Fincham, J. R. S., P. R. Day and A. Radford, 1979 *Fungal Genetics*. University of California Press, Berkeley/Los Angeles.

Fisher, R. A., M. F. Lyon and A. R. G. Owen, 1947 The sex chromosome in the house mouse. Heredity **1:** 335–365.

Foss, E., and F. W. Stahl, 1995 A test of a counting model for chiasma interference. Genetics **139:** 1201–1209.

Foss, E., R. Lande, F. W. Stahl and C. M. Steinberg, 1993 Chiasma interference as a function of genetic distance. Genetics **133:** 681–691.

Griffiths, A. J. F., J. H. Miller, D. T. Suzuki, R. C. Lewontin and W. M. Gelbart, 1996 *An Introduction to Genetic Analysis*, Ed. 6, W. H. Freeman and Company, New York.

Haldane, J. B. S., 1931 The cytological basis of genetical interference. Cytologia **3:** 54–65.

Howe, H. B., Jr., 1956 Crossing-over and nuclear passing in *Neurospora crassa*. Genetics **41:** 610–622.

Lange, K., H. Zhao and T. P. Speed, 1997 The Poisson-skip model of crossing-over. Ann. Appl. Prob. **7:** 299–313.

Lindegren, C. C., 1932 The genetics of Neurospora. II. Segregation of the sex factors in the asci of *N. crassa*, *N. sitophila* and *N. tetrasperma*. Bull. Torrey Bot. Cl. **59:** 119–138.

Lindegren, C. C., 1933 The genetics of Neurospora. III. Pure bred stocks and crossing-over in *N. crassa*. Bull. Torrey Bot. Cl. **60:** 133–154.

Lindegren, C. C., 1936a A six-point map of the sex-chromosome of *Neurospora crassa*. J. Genet. **32:** 243–256.

Lindegren, C. C., 1936b The structure of the sex-chromosomes of *Neurospora crassa* suggested by genetic analysis. J. Hered. **27:** 251–259.

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

McPeek, M. S., and T. P. Speed, 1995 Modeling interference in genetic recombination. Genetics **139:** 1031–1044.

Morton, N. E., B. J. Keats, P. A. Jacobs, T. Hassold, D. Pettay *et al.*, 1990 A centromere map of the X chromosome from trisomies of maternal origin. Ann. Hum. Genet. **54:** 39–47.

Ott, J., D. Linder, B. K. McCaw, E. W. Lovrien and F. Hecht, 1976 Estimating distances from the centromere by means of benign ovarian teratomas in man. Ann. Hum. Genet. **49:** 191–196.

Papazian, H. P., 1952 The analysis of tetrad data. Genetics **37:** 175–188.

Perkins, D. D., 1949 Biochemical mutants in the smut fungus *Ustilago maydis*. Genetics **34:** 607–626.

Perkins, D. D., 1953 The detection of linkage in tetrad analysis, 1953 Genetics **38:** 187–197.

Perkins, D. D., 1955 Tetrads and crossing over. J. Cell. Comp. Physiol. **45**(Supp 2): 119–149.

Perkins, D. D., 1962 Crossing-over and interference in a multiply marked chromosome arm of Neurospora. Genetics **47:** 1253–1274.

Rao, D. C., N. E. Morton, J. Lindsten, M. Hulten and S. Yee, 1977 A mapping function for man. Hum. Hered. **27:** 99–104.

Risch, N., and K. Lange, 1983 Statistical analysis of multilocus recombination. Biometrics **39:** 949–963.

Weinstein, A., 1936 The theory of multiple-strand crossing over. Genetics **21:** 155–199.

Whitehouse, H. L. K., 1942 Crossing-over in Neurospora. New Phytologist **41:** 23–62.

Zhao, H., and T. P. Speed, 1996 On genetic map functions. Genetics **142:** 1369–1377.

Zhao, H., and T. P. Speed, 1998 Statistical analysis of half-tetrads. Genetics **150:** 473–485.

Zhao, H., M. S. McPeek and T. P. Speed, 1995a Statistical analysis of chromatid interference. Genetics **139:** 1057–1065.

Zhao, H., T. P. Speed and M. S. McPeek, 1995b Statistical analysis of crossover interference using the chi-square model. Genetics **139:** 1045–1056.

Communicating editor: D. Botstein

APPENDIX

Proposition 1. *Under RSCA, there are* $(6^n + 5 \times 2^n)/8$ *distinct probabilities for ordered tetrads with n markers.*

Proof. Two distinguishable patterns have the same probability under RSCA if one pattern can be changed to the other one through one of the eight permutations of the four strands in an order tetrad, as listed in Table 2. Each marker can have six distinguishable types, as summarized in Table 1.

(a) When configurations at all markers are type 1 or 2, for each pattern, there is another one with the same probability. For example, [*Ab*, *Ab*, *aB*, *aB*] and [*aB*, *aB*, *Ab*, *Ab*] have the same probability. This gives $2^n/2 = 2^{n-1}$ distinct probabilities.

(b) When configurations at all markers are either type 3 or 4, each pattern has three other patterns with the same probability. For example, [*Ab*, *aB*, *Ab*, *aB*],

[*Ab, aB, aB, Ab*], [*aB, Ab, Ab, aB*], and [*aB, Ab, aB, Ab*] have the same probability. Likewise when all markers are either type 5 or 6, each pattern has three other patterns with the same probability. This gives $(2^n + 2^n)/4 = 2^{n-1}$ distinct probabilities.

(c) When all markers have configurations 3, 4, 5, or 6, and with at least one being 3 or 4 and at least one being 5 or 6, each pattern has seven other patterns with the same probability. For example, [*AB, ab, Ab, aB*], [*AB, ab, aB, Ab*], [*ab, AB, Ab, aB*], [*ab, AB, aB, Ab*], [*Ab, aB, AB, ab*], [*Ab, aB, ab, AB*], [*aB, Ab, AB, ab*], and [*aB, Ab, ab, AB*] have the same probability. This gives $(4^n - 2^n - 2^n)/8$ distinct probabilities.

(d) When at least one marker has type 1 or 2 and at least one marker has type 3, 4, 5, or 6, there are seven other patterns with the same probability. For example, [*AB, Ab,* aB, ab], [*AB, Ab, ab, aB*], [*Ab, AB, aB, ab*], [*Ab, AB, ab, aB*], [*aB, ab, AB, Ab*], [*aB, ab, Ab, AB*], [*ab, aB, AB, Ab*], and [*ab, aB, Ab, AB*] all have the same probability. This gives $(6^n - 2^n - 4^n)/8$ distinct probabilities.

Adding (a) to (d), there are at most $(6^n + 5 \times 2^n)/8$ distinct probabilities.

Proposition 2. *Under NCI, for any joint ordered tetrad probabilities with two markers on different sides of the centromere, there is an underlying chiasma process that is compatible with these probabilities if and only if* $\alpha \geq \beta$ *and* $\gamma + \delta \geq 2\beta$ *and the ratio of* $p(P_2):p(T_2):p(N_2)$ *is 1:2:1.*

Proof. It is easy to see that, under NCI, the ratio of $p(P_2):p(T_2):p(N_2)$ being 1:2:1 is a necessary and sufficient condition. Let $c_{k\ell}$ denote the joint probability of there being k chiasmata between *CEN* and $\mathcal{A}$ and ℓ chiasmata between *CEN* and $\mathcal{B}$. If $c_{k\ell} = 0$ for either $k > 2$ *or* $\ell > 2$, the relations between $c_{k\ell}$ and $\alpha, \beta, \gamma, \delta$, and ε defined in Table 5 are

$$\begin{aligned}
\alpha &= c_{00} + \tfrac{1}{4}c_{02} + \tfrac{1}{4}c_{20} + \tfrac{1}{8}c_{22}, \\
\beta &= \tfrac{1}{4}c_{02} + \tfrac{1}{4}c_{20} + \tfrac{1}{8}c_{22}, \\
\gamma &= c_{01} + \tfrac{1}{2}c_{02} + \tfrac{1}{2}c_{21} + \tfrac{1}{4}c_{22}, \\
\delta &= c_{10} + \tfrac{1}{2}c_{12} + \tfrac{1}{2}c_{20} + \tfrac{1}{4}c_{22}, \\
\varepsilon &= c_{11} + \tfrac{1}{2}c_{12} + \tfrac{1}{2}c_{21} + \tfrac{1}{4}c_{22}.
\end{aligned}$$

Suppose $\alpha \geq \beta$ and $\gamma + \delta \geq 2\beta$.

(a) If $\gamma - \beta \geq 0$ and $\delta - \beta \geq 0$, let $c_{00} = \alpha - \beta$, $c_{01} = \gamma - \beta$, $c_{10} = \delta - \beta$, $c_{02} = c_{20} = 2\beta$, $c_{11} = \varepsilon$, and all other $c_{k\ell} = 0$. All the $c_{k\ell}$ are nonnegative, and they are compatible with $(\alpha, \beta, \gamma, \delta, \varepsilon)$.

(b) If $\gamma - \beta \leq 0$, $4\beta - 2\gamma$ must be ≥ 0 because $4\beta - 2\gamma \geq 2\beta - 2\gamma \geq 0$. Let $c_{00} = \alpha - \beta$, $c_{01} = 0$, $c_{10} = \gamma + \delta - 2\beta$, $c_{02} = 2\gamma$, $c_{20} = 4\beta - 2\gamma$, $c_{11} = \varepsilon$, and all other $c_{k\ell} = 0$. All the $c_{k\ell}$ are nonnegative, and they are compatible with $(\alpha, \beta, \gamma, \delta, \varepsilon)$.

(c) The case of $\delta - \beta < 0$ is similar to the case $\gamma - \beta < 0$.

Combining (a) to (c), this proves the necessity.

For any underlying chiasma process, if NCI holds, there is a chiasma process with at most two exhanges between each consecutive pair of markers, inducing the same tetrad probabilities (Zhao *et al.* 1995a). Therefore, we may consider only processes with no more than two chiasmata in each interval, that is, the $c_{k\ell}$ with $c_{k\ell} = 0$ for either $k > 2$ or $\ell > 2$. From the relations in Equation 5, it is easy to check that $\alpha \geq \beta$ and $\gamma + \delta \geq 2\beta$. This proves the sufficiency.

EM Algorithm. Here we sketch the EM algorithm used to obtain the maximum likelihood estimates under the linear constraints imposed by a given marker order. First, consider two markers in the order of $\mathcal{A}$–*CEN*–$\mathcal{B}$, which was discussed in the proof of Proposition 2. The "complete" data constitute $c_{k\ell}$, the joint probability of there being k chiasmata between *CEN* and $\mathcal{A}$ and ℓ chiasmata between *CEN* and $\mathcal{B}$. We consider only the $c_{k\ell}$ with $k, \ell = 0, 1$, and 2. This is because for any chiasma process, under NCI, there is a chiasma process with at most two exchanges between each consecutive pair of markers inducing the same tetrad probabilities (Zhao *et al.* 1995a). Let p_i denote the probability of group i, where $i = 1, \ldots, 7$ is one of the seven distinct groups in Table 3. Each p_i is a linear function of the $c_{k\ell}$, that is, $p_i = \Sigma t_{i,k\ell} c_{k\ell}$. Let x_i denote the number of observations for group i in the sample and n denote the sample size. Then the EM algorithm proceeds as follows:

1. Start with some initial estimates $c_{k\ell}^0$ of the $c_{k\ell}$.
2. Update the $c_{k\ell}$ using

$$c_{k\ell}^{\text{new}} = \frac{1}{n}\left\{\sum_{i=1}^{7}\left[\frac{t_{i,k\ell}c_{k\ell}^{\text{old}}}{\sum t_{i,k\ell}c_{k\ell}^{\text{old}}}\right]x_i\right\}$$

until the convergence criterion is satisfied.

If the two markers are in the order of *CEN*–$\mathcal{A}$–$\mathcal{B}$, we only consider the $c_{k\ell}$ with $k = 0$ and 1 and $\ell = 0$, 1, and 2. This is because under this order, for any chiasma process under NCI, there is a chiasma process with $c_{k\ell} = 0$ for $k > 1$ or $\ell > 2$ inducing the same tetrad probabilities.

The same procedure can be applied to more than two markers. For example, when there are three markers in the order of $\mathcal{A}$–*CEN*–$\mathcal{B}$–$\mathcal{C}$, there are 32 distinct groups under RSCA, and the "complete" data constitute 27 joint chiasma probabilities $c_{k\ell m}$ across three intervals ($\mathcal{A}$–*CEN*, *CEN*–$\mathcal{B}$, and $\mathcal{B}$–$\mathcal{C}$) with $k, \ell, m = 0$, 1, and 2. After establishing the relationships between the p_i and the $c_{k\ell m}$, the EM algorithm proceeds exactly the same as above.