Statistics in Molecular Biology
IMS Lecture Notes – Monograph Series (1999) Volume 33

ON A MARKOV MODEL FOR CHROMATID INTERFERENCE

BY HONGYU ZHAO AND TERENCE P. SPEED

Yale University School of Medicine and University of California, Berkeley

Meiotic exchange between homologous chromosomes takes place after the formation of a bundle of four chromatids. Crossovers are precise breakage-and-reunion events. Random strand involvements (no chromatid interference) and random distribution of crossovers (no chiasma interference) are usually assumed in analyzing genetic data. In this paper, we discuss a Markov model for chromatid interference. Closed form expression for the probability of any multilocus recombination/tetrad pattern is derived. Both chromatid interference and chiasma interference can be studied together using this model. In particular, we discuss chi-square models for chiasma interference.

1. Introduction. In diploid cells, each chromosome is paired with its homologue during meiosis . Each member of a given homologous pair has two identical sister chromatids, so that each synapsed paired structure consists of four chromatids. Usually one or more crossovers occur among the four chromatids. A crossover is a precise breakage-and-reunion event occurring between two nonsister chromatids.

The types of genetic data considered here are single spore data, in which the products of a single meiosis are recovered separately, and tetrad data, in which all four meiotic products are recovered together. A tetrad consists of four spores, each of which is haploid, encased in a structure called an ascus. In some organisms, such as *Neurospora crassa* (red bread mold), tetrads consist of four spores in a linear order corresponding to the meiotic divisions; these are called ordered tetrads. In other organisms, such as *Saccharomyces cerevisiae* (baker's yeast), the four spores are produced as a group without order and are called unordered tetrads. Griffiths et al. (1996) covers relevant genetic background.

In this paper, genes (markers, loci) are denoted by script letters. For example, we use $\mathcal{A}$ and $\mathcal{B}$ to denote two genes. Alleles of $\mathcal{A}$ are denoted by A and a, while alleles of $\mathcal{B}$ are denoted by B and b. Suppose that $\mathcal{A}$ and $\mathcal{B}$ are on the same chromosome arm, and consider a diploid cell having AB and ab on homologous chromosomes. There are four possible products at these loci resulting from meiosis of this cell, namely, AB, ab, Ab, and aB. The first two are called

Supported in part by NIH grant HG-01093.
AMS 1991 *subject classifications.* Primary 60J20; secondary 60P10.
Key words and phrases. Chromatid interference, chiasma interference, Markov model, chi-square model, single spore data, tetrad data.

parental types or nonrecombinants, the other two types, Ab and aB, are called recombinants. If two markers are recombined by crossovers in a meiotic product, then during meiosis an odd number of crossovers must have occurred between the two markers on the strand carrying them. The proportion of recombinants, r_{AB}, is called the recombination fraction. Because recombination fractions are not additive, genetic (or map) distance is used as an additive measure of distance between loci. Genetic distance between two markers is defined as the average number of crossovers per strand per meiosis between these markers. The unit of genetic distance is Morgan (M). Two markers are 1M apart if on average there is one crossover occurring on a single strand per meiosis between these two markers. In practice, centiMorgan (cM = 0.01M) is more commonly used in genetic mapping.

The occurrence of crossovers cannot be observed directly and has to be inferred from observed recombination events. In the case of single spore data, a given meiotic product may be scored as recombinant or non-recombinant for each pair of markers. The map distance between two markers can be estimated from the observed recombination fraction. In the case of unordered tetrad data, there are three possible observed outcomes for each pair of markers: parental ditype when all four strands are non-recombinants, tetratype when exactly two of the four strands show recombination between the two markers, and nonparental ditype when all four strands are recombinants. The map distance between two markers can be estimated from the observed proportions of these three tetrad types. In the case of ordered tetrads bearing one marker $\mathcal{A}$ with alleles A and a, there are six distinguishable configurations:

1	2	3	4	5	6
A	a	a	A	A	a
A	a	A	a	a	A
a	A	a	A	a	A
a	A	A	a	A	a

Configurations 1 and 2 are called first division segregation (FDS) patterns and configurations 3 to 6 are called second division segregation (SDS) patterns. Because of random spindle to centromere attachments during meiosis, configurations 1 and 2 have the same probability, and the four configurations showing SDS pattern also have the same probability (Griffiths et al. 1996). The map distance between a marker and the centromere can then be estimated from the observed SDS proportion.

To estimate map distance from the observed data, a model is needed which connects the process of crossover to the observed outcomes. Any model has to

consider two aspects of crossover that are relevant to the observed recombination outcome: the distribution of crossover events along the bundle of four chromatids, and the pairs of nonsister chromatids involved in crossovers. To distinguish crossover events occurring on the four strand bundle and crossover events on single strands, we describe crossover events on the four strand bundle as chiasmata (singular: chiasma), and crossover events on single strands as crossovers. Chiasma interference refers to non-random distribution of chiasmata on the four strand bundle, whereas crossover interference refers to non-random distribution of crossovers along single strands. In this paper, we use the word *random* in the sense that all outcomes are equally likely. There is no chromatid interference (NCI) if any pair of non-sister chromatids are equally likely to be involved in any chiasma, independent of which pairs were involved in other chiasmata. It is possible that there is chiasma interference at the four strand bundle, but because of the presence of chromatid interference, there is no crossover interference on single strands, see Zhao and Speed (1996) for a model exhibiting this phenomenon.

With virtually no restrictions on the chiasma process, Speed, McPeek and Evans (1992) derived the constraints on multilocus recombination probabilities for single spore data under the assumption of NCI. It was shown by Zhao, McPeek and Speed (1995) that the assumption of NCI also imposes constraints on multilocus tetrad probabilities. Based on these constraints, a statistical testing procedure for NCI was proposed in the last mentioned paper and applied to data from several organisms. Though there was an excess of two-strand double recombinations in some organisms, no strong evidence was found for chromatid interference.

Crossover interference has been observed in almost all organisms. The presence of one crossover usually inhibits the formation of crossovers in a nearby region. Fisher, Lyon and Owen (1947) modeled crossovers as a renewal process; that is, crossovers along a single strand were assumed to be formed as a regular sequence starting from the centromere, with the length between two adjacent crossovers always following the same distribution. Centromeres are constricted regions of nuclear chromosomes, to which the spindle fibers attach during division. Mather (1938) appeared to be the first one to propose the above sequential model for the chiasma process. Other mathematical models have also been proposed to model crossover interference (McPeek and Speed 1995). However, the biological nature of crossover interference is still not well understood.

In this paper, we discuss a Markov model for chromatid interference. There exist closed form expressions for joint recombination and tetrad probabilities under this model. Closed form expressions still exist when this model is combined with a class of chiasma interference models, the chi-square model (Zhao, Speed

and McPeek 1995), thus allowing joint modeling of both types of interference.

2. A Markov model for chromatid interference. The Markov chromatid interference model discussed here was first introduced and studied by Weinstein (1938). Later studies on chromatid interference (Carter and Robertson 1952, Sturt and Smith 1976 and Stam 1979) essentially used Weinstein's model.

Weinstein's model assumes that chiasmata occur according to a point process originating from the centromere, and that the choice of nonsister chromatids involved in one chiasma only depends on the pair involved in the previous chiasma. Thus it is a Markov model. The two pairs of sister chromatids are labeled as w^1, w^2 and m^1, m^2. If w^1 and m^1 are involved in the kth chiasma, the chances of the four nonsister pairs (w^1, m^1), (w^1, m^2), (w^2, m^1), and (w^2, m^2) being involved in the $(k+1)$th chiasma are denoted by α, β, β, and γ, respectively. That is, the conditional probabilities of two, three, and four-strand double chiasmata are α, 2β, and γ for any two consecutive chiasmata. Under this model, if a strand is involved in one chiasma, the chance that it will be involved in the next chiasma is $\eta = \alpha + \beta$, and the chance that it will not be involved is $\beta + \gamma = 1 - \eta$. So for a single strand, the degree of chromatid interference is determined by η. Different values of $(\alpha, 2\beta, \gamma)$ can correspond to the same η. When there is no chromatid interference, $(\alpha, 2\beta, \gamma) = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ and $\eta = \frac{1}{2}$, but any chromatid interference model with parameters $(\alpha, 2\beta, \gamma)$ satisfying $\alpha + \beta = \frac{1}{2}$ will be indistinguishable from a no chromatid interference model if only single spore data are available. On the other hand, different sets of $(\alpha, 2\beta, \gamma)$ that correspond to the same η value can be distinguished using tetrad data.

Previous work on this chromatid interference model has been confined to the two-locus case. Our aim is to derive general expressions for the probabilities of recombination or not, across a set of loci, and the analogous multilocus tetrad probabilities. We consider single spore data and tetrad data separately.

2.1 *Single spore data.* We assign a state "y" or "n" to any locus on any one of the four strands in a bundle as follows. If there has been at least one chiasma between the centromere and that locus, then a (locus, strand) pair is assigned state "y" if the strand was involved in the last chiasma before the locus; otherwise the (locus, strand) pair is assigned "n". If no chiasmata have occurred between a locus and the centromere, it can be shown that assigning the state "y" or "n" with probability $\frac{1}{2}$ fits in well with our development.

Suppose that $\mathcal{A}$ and $\mathcal{B}$ are two loci on the same chromosome in the order $CEN - \mathcal{A} - \mathcal{B}$. For any given strand, there are four possible joint states, (y, y), (y, n), (n, y) and (n, n). Consider the case in which $\mathcal{A}$ is in state y and that $k \geq 1$ chiasmata have occurred between $\mathcal{A}$ and $\mathcal{B}$. A given strand could have

been involved in an odd (o) or an even (e) number of the chiasmata occurring between $\mathcal{A}$ and $\mathcal{B}$, i.e., $\mathcal{A}$ and $\mathcal{B}$ may have recombined or not, on that strand. Given $\mathcal{A}$ is in state y on a strand, let $p^o_{y,y}(k)$ be the conditional probability that $\mathcal{B}$ is in the state y and that an odd number of the $k \geq 1$ chiasmata between $\mathcal{A}$ and $\mathcal{B}$ involve this strand. Define $p^o_{y,n}(k)$, $p^e_{y,y}(k)$, and $p^e_{y,n}(k)$ in similar fashion, where e denotes an even number. If $\mathcal{A}$ is in state n, $p^o_{n,y}(k)$, $p^o_{n,n}(k)$, $p^e_{n,y}(k)$, and $p^e_{n,n}(k)$ can be similarly defined. Group $p^o_{y,y}(k)$, $p^o_{y,n}(k)$, $p^o_{n,y}(k)$, and $p^o_{n,n}(k)$ into a 2×2 matrix, define

$$\mathbf{T}^1_k = \begin{pmatrix} p^o_{y,y}(k) & p^o_{y,n}(k) \\ p^o_{n,y}(k) & p^o_{n,n}(k) \end{pmatrix},$$

and similarly,

$$\mathbf{T}^0_k = \begin{pmatrix} p^e_{y,y}(k) & p^e_{y,n}(k) \\ p^e_{n,y}(k) & p^e_{n,n}(k) \end{pmatrix}.$$

Recursive relationships among these quantities can be established as follows. Suppose that $\mathcal{B}$ is in state y on one strand, and that $\mathcal{A}$ and $\mathcal{B}$ are recombined on that strand after k chiasmata have taken place on the bundle between $\mathcal{A}$ and $\mathcal{B}$. If a $(k+1)$th chiasma on the bundle occurs between $\mathcal{A}$ and $\mathcal{B}$, then the strand has a chance η of being involved in it, and $1-\eta$ of not being involved. In the first case, $\mathcal{B}$ will remain in state y and $\mathcal{A}$ and $\mathcal{B}$ will not be recombined on that strand; in the second case, $\mathcal{B}$ will change to state n and $\mathcal{A}$ and $\mathcal{B}$ will still be recombined on that strand. Other cases can be considered similarly, and we can thus derive the following relationship:

$$\begin{pmatrix} p^o_{y,y}(k+1) \\ p^o_{y,n}(k+1) \\ p^e_{y,y}(k+1) \\ p^e_{y,n}(k+1) \end{pmatrix} = \begin{pmatrix} 0 & 0 & \eta & 1-\eta \\ 1-\eta & \eta & 0 & 0 \\ \eta & 1-\eta & 0 & 0 \\ 0 & 0 & 1-\eta & \eta \end{pmatrix} \begin{pmatrix} p^o_{y,y}(k) \\ p^o_{y,n}(k) \\ p^e_{y,y}(k) \\ p^e_{y,n}(k) \end{pmatrix}.$$

Similarly, if $\mathcal{A}$ is in state n, with $p^o_{n,y}(k)$, $p^o_{n,n}(k)$, $p^e_{n,y}(k)$, and $p^e_{n,n}(k)$ defined as above, similar recursive relationship can be established.

Let $p^y_{\mathcal{A}}$ (or $p^n_{\mathcal{A}}$) be the probability that $\mathcal{A}$ is in state y (or n) on a given strand and $s_{\mathcal{A}}$ and $s_{\mathcal{B}}$ be the states at $\mathcal{A}$ and $\mathcal{B}$. From the definitions of the $p^r_{s_{\mathcal{A}},s_{\mathcal{B}}}(k)$, after $k \geq 1$ chiasmata have occurred along the four-strand bundle between $\mathcal{A}$ and $\mathcal{B}$, the chance that $\mathcal{A}$ and $\mathcal{B}$ are recombined on that strand is

$$(p^y_{\mathcal{A}}, p^n_{\mathcal{A}})\mathbf{T}^1_k \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

and the chance that $\mathcal{A}$ and $\mathcal{B}$ are not recombined on that strand is:

$$(p^y_{\mathcal{A}}, p^n_{\mathcal{A}})\mathbf{T}^0_k \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Let $\delta = 2\eta - 1$, define

$$\mathbf{M} = \frac{1}{2}\begin{pmatrix} 1+\delta^k & 1-\delta^k \\ 1-\delta^k & 1+\delta^k \end{pmatrix},$$

$$\mathbf{N}_0 = \begin{pmatrix} -\delta^{\frac{k}{2}} & 0 \\ 0 & -\delta^{\frac{k}{2}} \end{pmatrix} \qquad \text{and} \qquad \mathbf{N}_1 = \begin{pmatrix} \eta\delta^{\frac{k-1}{2}} & -(1-\eta)\delta^{\frac{k-1}{2}} \\ (1-\eta)\delta^{\frac{k-1}{2}} & -\eta\delta^{\frac{k-1}{2}} \end{pmatrix}.$$

General expressions of $\mathbf{T}^1_k$ and $\mathbf{T}^0_k$ can be obtained through the above recursive relationships among the $p^r_{s_{\mathcal{A}},s_{\mathcal{B}}}(k)$ as summarized in the following theorem. The proof is given in section 6.

Theorem 1. *When k is even,*

$$\mathbf{T}^1_k = \frac{1}{2}(\mathbf{M} + \mathbf{N}_0) \qquad \textit{and} \qquad \mathbf{T}^0_k = \frac{1}{2}(\mathbf{M} - \mathbf{N}_0),$$

and when k is odd,

$$\mathbf{T}^1_k = \frac{1}{2}(\mathbf{M} + \mathbf{N}_1) \qquad \textit{and} \qquad \mathbf{T}^0_k = \frac{1}{2}(\mathbf{M} - \mathbf{N}_1).$$

For a strand chosen at random, $\mathcal{A}$ has an equal chance to be in state y or n. The probability that $\mathcal{A}$ and $\mathcal{B}$ are recombined is $\frac{1}{2}(1-\delta^{\frac{k}{2}})$ for k even and $\frac{1}{2}$ for k ood.

This result was proved by Weinstein (1938) and Sturt and Smith (1976). Our approach, however, is different, and easily generalizes to the multilocus case, as illustrated later. We now consider three special cases: (a) $\eta = \frac{1}{2}$, (b) $\eta = 0$, and (c) $\eta = 1$. Case (a) includes no chromatid interference. Case (b) implies that a strand is never involved in two consecutive chiasmata, i.e., that only four-strand double chiasmata occur. In case (c), all chiasmata involve the same two strands, i.e., only two-strand double chiasmata occur. We calculate $\mathbf{T}^1_k$ for each case. Case (a):

$$\mathbf{T}^1_k = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

Case (b): when k is even,

$$\mathbf{T}_k^1 = \begin{pmatrix} \frac{1}{2}(1-(-1)^{\frac{k}{2}}) & 0 \\ 0 & \frac{1}{2}(1-(-1)^{\frac{k}{2}}) \end{pmatrix},$$

and when k is odd,

$$\mathbf{T}_k^1 = \begin{pmatrix} 0 & \frac{1}{2}(1-(-1)^{\frac{k-1}{2}}) \\ \frac{1}{2}(1+(-1)^{\frac{k-1}{2}}) & 0 \end{pmatrix}.$$

Case (c): when k is even,

$$\mathbf{T}_k^1 = \begin{pmatrix} 0\,0 \\ 0\,0 \end{pmatrix},$$

and when k is odd,

$$\mathbf{T}_k^1 = \begin{pmatrix} 1\,0 \\ 0\,0 \end{pmatrix}.$$

The recombination fraction is: case (a) $r = \frac{1}{2}$ for $k \geq 1$; case (b) $r = \frac{1}{2}(1-(-1)^{\frac{k}{2}})$ for k even and $\frac{1}{2}(1-(-1)^{\frac{k-1}{2}})$ for k odd; and case (c) $r = 0$ for k even and 1 for k odd.

Suppose q_k is the probability that there are k chiasmata between $\mathcal{A}$ and $\mathcal{B}$, and let

$$\mathbf{T}_{\mathcal{AB}}^1 = \sum q_k \mathbf{T}_k^1 = \begin{pmatrix} p_{y,y}^o & p_{y,n}^o \\ p_{n,y}^o & p_{n,n}^o \end{pmatrix}.$$

Then given that $\mathcal{A}$ is in state $s_{\mathcal{A}}$, the value of $p_{s_{\mathcal{A}},s_{\mathcal{B}}}^o$ is the conditional probability that $\mathcal{B}$ is in state $s_{\mathcal{B}}$ and that $\mathcal{A}$ and $\mathcal{B}$ are recombined. We can similarly define

$$\mathbf{T}_{\mathcal{AB}}^0 = \sum q_k \mathbf{T}_k^0 = \begin{pmatrix} p_{y,y}^e & p_{y,n}^e \\ p_{n,y}^e & p_{n,n}^e \end{pmatrix}.$$

Define $p_{\mathcal{A}}^y$ and $p_{\mathcal{A}}^n$ as above, the chance that $\mathcal{A}$ and $\mathcal{B}$ are recombined is

$$(p_{\mathcal{A}}^y, p_{\mathcal{A}}^n)\mathbf{T}_{\mathcal{AB}}^1 \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

and the chance that they are not recombined is

$$(p^y_{\mathcal{A}}, p^n_{\mathcal{A}})\mathbf{T}^0_{\mathcal{AB}}\begin{pmatrix}1\\1\end{pmatrix}.$$

The recombination probabilities can be computed explicitly for some simple chiasma processes. Consider the Poisson chiasma process model. Under this model, the number of chiasmata between $\mathcal{A}$ and $\mathcal{B}$ is Poisson distributed, $q_k = e^{-2d}(2d)^k/k!$, where d is the genetic distance between $\mathcal{A}$ and $\mathcal{B}$. There is a factor 2 because each chiasma involves two of the four chromatids, thus the average number of crossovers on a single strand is half the number of chiasmata on the four strand bundle. The matrix $\mathbf{T}^1_{\mathcal{AB}}$ is

$$\frac{1}{4}\begin{pmatrix}x & y-z\\ y+z & x\end{pmatrix},$$

where

$$x = 1 + e^{-4d(1-\eta)} - \frac{(\eta+\phi)e^{-2d(1+\phi)} - (\eta-\phi)e^{-2d(1-\phi)}}{\phi},$$

$$y = 1 - e^{-4d(1-\eta)},$$

$$z = \frac{(1-\eta)(e^{-2d(1-\phi)} - e^{-2d(1+\phi)})}{\phi},$$

and $\phi = \sqrt{\delta}$.

Therefore, for any ϕ, the recombination probability r is $\frac{1}{4}\{2 - (e^{-2d(1+\phi)} + e^{-2d(1-\phi)})\}$. Consider the above-mentioned three special cases. For case (a), r is $\frac{1}{2}(1-e^{-2d})$; for case (b), r is $\frac{1}{2}\{1-\cos(2d)e^{-2d}\}$; and for case (c), r is $\frac{1}{4}(1-e^{-4d})$.

Now consider three markers, $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}_3$. Given $\mathcal{A}_1$ is in state s_1, and there are k_1 and k_2 chiasmata in the two intervals, denote the probability that $\mathcal{A}_3$ is in state s_3 and that the strand is involved in an odd number of chiasmata in both intervals (both intervals show recombination) as $p^{o,o}_{s_1,s_3}(k_1,k_2)$. Define the matrix $\mathbf{T}^{1,1}_{\mathcal{A}_1\mathcal{A}_3}(k_1,k_2)$ as

$$\mathbf{T}^{1,1}_{\mathcal{A}_1\mathcal{A}_3}(k_1,k_2) = \begin{pmatrix}p^{o,o}_{y,y}(k_1,k_2) & p^{o,o}_{y,n}(k_1,k_2)\\ p^{o,o}_{n,y}(k_1,k_2) & p^{o,o}_{n,n}(k_1,k_2)\end{pmatrix}.$$

Considering the state of $\mathcal{A}_2$, denoted by s_2, and using the Markovian property of the model (that the choice of strands involved in the next chiasma depends only on the pair involved in the previous one), we obtain

$$p^{o,o}_{y,y}(k_1,k_2) = p^o_{y,s_2=y}(k_1)p^o_{s_2=y,y}(k_2) + p^o_{y,s_2=n}(k_1)p^o_{s_2=n,y}(k_2).$$

Similar relationships hold for the other $p^{o,o}_{s_1,s_3}(k_1,k_2)$. Writing this in matrix form, we have

$$\mathbf{T}^{1,1}_{\mathcal{A}_1\mathcal{A}_3}(k_1,k_2) = \mathbf{T}^{1}_{\mathcal{A}_1\mathcal{A}_2}(k_1)\mathbf{T}^{1}_{\mathcal{A}_2\mathcal{A}_3}(k_2).$$

Other $\mathbf{T}^{i_1,i_2}_{\mathcal{A}_1\mathcal{A}_3}(k_1,k_2)$ can be defined similarly, where i_1, i_2 is either 1 or 0 corresponding to whether two genes are recombined or not over that interval. We have

$$\mathbf{T}^{i_1,i_2}_{\mathcal{A}_1\mathcal{A}_3}(k_1,k_2) = \mathbf{T}^{i_1}_{\mathcal{A}_1\mathcal{A}_2}(k_1)\mathbf{T}^{i_2}_{\mathcal{A}_2\mathcal{A}_3}(k_2).$$

Let $p^{i_1,i_2}_{\mathcal{A}_1\mathcal{A}_3}(k_1,k_2)$ be the conditional probability that the recombination pattern for $\mathcal{A}_1\mathcal{A}_2\mathcal{A}_3$ is (i_1,i_2) given that there are k_1 and k_2 chiasmata in the two intervals. Then $p^{i_1,i_2}_{\mathcal{A}_1\mathcal{A}_3}(k_1,k_2)$ can be calculated as:

$$p^{i_1,i_2}_{\mathcal{A}_1\mathcal{A}_3}(k_1,k_2) = (p^{y}_{\mathcal{A}_1}, p^{n}_{\mathcal{A}_1})\mathbf{T}^{i_1}_{\mathcal{A}_1\mathcal{A}_2}(k_1)\mathbf{T}^{i_2}_{\mathcal{A}_2\mathcal{A}_3}(k_2)\begin{pmatrix}1\\1\end{pmatrix}.$$

Let q_{k_1,k_2} be the chance of k_1 and k_2 chiasmata in the two intervals. Denote $p^{i_1,i_2}_{\mathcal{A}_1\mathcal{A}_3}$ as the probability that the recombination pattern is (i_1,i_2), we have

$$p^{i_1,i_2}_{\mathcal{A}_1\mathcal{A}_3} = (p^{y}_{\mathcal{A}_1}, p^{n}_{\mathcal{A}_1})\left(\sum q_{k_1,k_2}\mathbf{T}^{i_1}_{\mathcal{A}_1\mathcal{A}_2}(k_1)\mathbf{T}^{i_2}_{\mathcal{A}_2\mathcal{A}_3}(k_2)\right)\begin{pmatrix}1\\1\end{pmatrix}.$$

For any $l+1$ markers, $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_{l+1}$, an explicit expression exists for the probability of any recombination pattern $\mathbf{i} = (i_1, i_2, \ldots, i_l)$, where $i_j = 0$, or 1 according to whether there is no recombination or recombination in the jth interval. Define $\mathbf{k} = (k_1,k_2,\ldots,k_l)$ and let $q_{\mathbf{k}}$ be the probability of there being $k_1, k_2, \ldots, k_l$ chiasmata in these l intervals. The probability of recombination type $\mathbf{i}$ is denoted $p_{\mathbf{i}}$. The general result for single spore data involving $l+1$ markers is:

Theorem 2.

$$p_{\mathbf{i}} = (p^{y}_{\mathcal{A}_1}, p^{n}_{\mathcal{A}_1})\left(\sum q_{\mathbf{k}}\mathbf{T}^{i_1}_{\mathcal{A}_1\mathcal{A}_2}(k_1)\mathbf{T}^{i_2}_{\mathcal{A}_2\mathcal{A}_3}(k_2)\cdots\mathbf{T}^{i_l}_{\mathcal{A}_l\mathcal{A}_{l+1}}(k_l)\right)\begin{pmatrix}1\\1\end{pmatrix},$$

where the $\mathbf{T}$ *matrices are defined as for the two marker case, and the sum is over all* $\mathbf{k} = (k_1,k_2,\ldots,k_l)$.

Given the joint chiasma probabilities on the four strand bundle across studied intervals, $q_{\mathbf{k}}$, Theorem 2 allows us to obtain a closed form expression for any joint recombination probability.

2.2. *Ordered tetrad data.* For the simplicity of our discussion, we assume for the moment that the tetrads are ordered. Unordered tetrads will be discussed in the next subsection. For a tetrad ordered from top to bottom, marker $\mathcal{A}$ with two alleles A and a can have six distinguishable types as illustrated in the introduction section.

For the moment, assume that there is at least one chiasma between the centromere and $\mathcal{A}$. Then for each type, tetrads can be further divided into subclasses according to the two strands involved in the last chiasma before $\mathcal{A}$. Each possible pair can be represented by (l_w, l_m), where $l_w = 0$ if the top one of the two strands bearing A was involved in the last chiasma, l_w=1 otherwise, and, l_m=0 if the top strand bearing a was involved in the last chiasma, l_m=1 otherwise. Thus, we can classify each locus into 6×4 states according to the order type of the strands, and according to the nonsister pair involved in the previous crossover. Each state is written as $[h, (l_w, l_m)]$, where h is the order type of the four strands 1, 2, 3, 4, 5, or 6, and (l_w, l_m) defines the strands involved in the last chiasma before $\mathcal{A}$.

Consider two markers $\mathcal{A}$ and $\mathcal{B}$. Suppose $\mathcal{A}$ is in some state $s_{\mathcal{A}}$, say $[1, (0, 0)]$, and let $t^{s_{\mathcal{A}}, s_{\mathcal{B}}}(k)$ be the conditional probability that $\mathcal{B}$ is in state $s_{\mathcal{B}}$ after k chiasmata between these two markers. For each $s_{\mathcal{B}}$=$[h, (l_w, l_m)]$, $t^{s_{\mathcal{A}}, s_{\mathcal{B}}}(k+1)$ is a linear function of the $t^{s_{\mathcal{A}}, s_{\mathcal{B}}}(k)$. For example, in order for $\mathcal{B}$ to be in state $s_{\mathcal{B}}$=$[1, (0, 0)]$ after $k+1$ chiasmata, $\mathcal{B}$ must be in one of the four states $[6, (0, 0)]$, $[6, (0, 1)]$, $[6, (1, 0)]$, or $[6, (1, 1)]$ after k chiasmata; these states have chance β, γ, α, and β, respectively, to change to $[1, (0, 0)]$ after the $(k+1)$th chiasma. So $t^{s_{\mathcal{A}}, [1,(0,0)]}(k+1)$ is equal to

$$\beta t^{s_{\mathcal{A}}, [6,(0,0)]}(k) + \gamma t^{s_{\mathcal{A}}, [6,(0,1)]}(k) + \alpha t^{s_{\mathcal{A}}, [6,(1,0)]}(k) + \beta t^{s_{\mathcal{A}}, [6,(1,1)]}(k).$$

Number the 24 possible states in the following way,

$$(1, 2, \ldots, 24) = ([1, (0, 0)], [1, (0, 1)], \ldots, [6, (1, 0)], [6, (1, 1)])$$

and let $\mathbf{T}(k)$=$(t^{i,j}(k))$, where $t^{i,j}(k), i, j = 1, \ldots, 24$, is as defined above. We have $\mathbf{T}(k+1) = \mathbf{U}\mathbf{T}(k)$, where $\mathbf{U}$ is the following 24×24 matrix.

$$\left(\begin{array}{cccc|cccc|cccc|cccc|cccc|cccc}
0&0&0&0&0&0&0&0&0&0&\beta&0&0&0&\beta&0&0&0&\gamma&0&0&0&\alpha&0\\
0&0&0&0&0&0&0&0&0&0&\alpha&0&0&0&\gamma&0&0&0&\beta&0&0&0&\beta&0\\
0&0&0&0&0&0&0&0&0&0&\gamma&0&0&0&\alpha&0&0&0&\beta&0&0&0&\beta&0\\
0&0&0&0&0&0&0&0&0&0&\beta&0&0&0&\beta&0&0&0&\alpha&0&0&0&\gamma&0\\
\hline
0&0&0&0&0&0&0&0&0&\beta&0&0&0&\beta&0&0&0&\alpha&0&0&0&\gamma&0&0\\
0&0&0&0&0&0&0&0&0&\alpha&0&0&0&\gamma&0&0&0&\beta&0&0&0&\beta&0&0\\
0&0&0&0&0&0&0&0&0&\gamma&0&0&0&\alpha&0&0&0&\beta&0&0&0&\beta&0&0\\
0&0&0&0&0&0&0&0&0&\beta&0&0&0&\beta&0&0&0&\gamma&0&0&0&\alpha&0&0\\
\hline
0&\beta&0&0&0&\beta&0&0&0&0&0&0&0&0&0&0&\alpha&0&0&0&0&0&0&\gamma\\
0&\gamma&0&0&0&\alpha&0&0&0&0&0&0&0&0&0&0&\beta&0&0&0&0&0&0&\beta\\
0&\alpha&0&0&0&\gamma&0&0&0&0&0&0&0&0&0&0&\beta&0&0&0&0&0&0&\beta\\
0&\beta&0&0&0&\beta&0&0&0&0&0&0&0&0&0&0&\gamma&0&0&0&0&0&0&\alpha\\
\hline
0&0&\beta&0&0&0&\beta&0&0&0&0&0&0&0&0&0&0&\gamma&\alpha&0&0&0\\
0&0&\gamma&0&0&0&\alpha&0&0&0&0&0&0&0&0&0&0&\beta&\beta&0&0&0\\
0&0&\alpha&0&0&0&\gamma&0&0&0&0&0&0&0&0&0&0&\beta&\beta&0&0&0\\
0&0&\beta&0&0&0&\beta&0&0&0&0&0&0&0&0&0&0&\alpha&\gamma&0&0&0\\
\hline
0&0&0&\beta&\beta&0&0&0&\alpha&0&0&0&0&0&0&\gamma&0&0&0&0&0&0&0&0\\
0&0&0&\gamma&\alpha&0&0&0&\beta&0&0&0&0&0&0&\beta&0&0&0&0&0&0&0&0\\
0&0&0&\alpha&\gamma&0&0&0&\beta&0&0&0&0&0&0&\beta&0&0&0&0&0&0&0&0\\
0&0&0&\beta&\beta&0&0&0&\gamma&0&0&0&0&0&0&\alpha&0&0&0&0&0&0&0&0\\
\hline
\beta&0&0&0&0&0&0&\beta&0&0&0&\gamma&\alpha&0&0&0&0&0&0&0&0&0&0&0\\
\gamma&0&0&0&0&0&0&\alpha&0&0&0&\beta&\beta&0&0&0&0&0&0&0&0&0&0&0\\
\alpha&0&0&0&0&0&0&\gamma&0&0&0&\beta&\beta&0&0&0&0&0&0&0&0&0&0&0\\
\beta&0&0&0&0&0&0&\beta&0&0&0&\alpha&\gamma&0&0&0&0&0&0&0&0&0&0&0
\end{array}\right)$$

Therefore, $\mathbf{T}(k) = \mathbf{U}^k\mathbf{T}(0) = \mathbf{U}^k$ because $\mathbf{T}(0) = I_{24\times 24}$. Matrix $\mathbf{T}(k)$ can be divided into 36 4×4 submatrices as follows:

$$\left(\begin{array}{c|c|c|c|c|c}
\mathbf{T}^{1,1}(k) & \mathbf{T}^{1,2}(k) & \mathbf{T}^{1,3}(k) & \mathbf{T}^{1,4}(k) & \mathbf{T}^{1,5}(k) & \mathbf{T}^{1,6}(k)\\
\hline
\mathbf{T}^{2,1}(k) & \mathbf{T}^{2,2}(k) & \mathbf{T}^{2,3}(k) & \mathbf{T}^{2,4}(k) & \mathbf{T}^{2,5}(k) & \mathbf{T}^{2,6}(k)\\
\hline
\mathbf{T}^{3,1}(k) & \mathbf{T}^{3,2}(k) & \mathbf{T}^{3,3}(k) & \mathbf{T}^{3,4}(k) & \mathbf{T}^{3,5}(k) & \mathbf{T}^{3,6}(k)\\
\hline
\mathbf{T}^{4,1}(k) & \mathbf{T}^{4,2}(k) & \mathbf{T}^{4,3}(k) & \mathbf{T}^{4,4}(k) & \mathbf{T}^{4,5}(k) & \mathbf{T}^{4,6}(k)\\
\hline
\mathbf{T}^{5,1}(k) & \mathbf{T}^{5,2}(k) & \mathbf{T}^{5,3}(k) & \mathbf{T}^{5,4}(k) & \mathbf{T}^{5,5}(k) & \mathbf{T}^{5,6}(k)\\
\hline
\mathbf{T}^{6,1}(k) & \mathbf{T}^{6,2}(k) & \mathbf{T}^{6,3}(k) & \mathbf{T}^{6,4}(k) & \mathbf{T}^{6,5}(k) & \mathbf{T}^{6,6}(k)
\end{array}\right).$$

The submatrix $\mathbf{T}^{h_{\mathcal{A}},h_{\mathcal{B}}}(k)$, $h_{\mathcal{A}}, h_{\mathcal{B}} = 1,\ldots,6$, is the transition matrix from type $h_{\mathcal{A}}$ at $\mathcal{A}$ to type $h_{\mathcal{B}}$ at $\mathcal{B}$ given k chiasmata between them. For example, $\mathbf{T}^{1,1}(k)$ is a 4×4 matrix with each entry being the conditional probability that, given $\mathcal{A}$ is in state $[1,(l^1_w,l^1_m)]$, $\mathcal{B}$ is in state $[1,(l^2_w,l^2_m)]$ after k chiasmata.

Let $\mathbf{p}_{\mathcal{A}} = (p^1_{\mathcal{A}}, \ldots, p^{24}_{\mathcal{A}})'$ be the initial distribution of states at $\mathcal{A}$, and let $\mathbf{S} = (\mathbf{p}_{\mathcal{A}}, \ldots, \mathbf{p}_{\mathcal{A}})$ and $\mathbf{P}(k) = \mathbf{ST}(k) = (p^{i,j}(k))$. Then $p^{i,j}(k)$ is the joint probability that $\mathcal{A}$ is in state i and $\mathcal{B}$ in state j given the occurrence of k chiasmata between them. The matrix $\mathbf{P}(k)$ can be also divided into 36 4×4 submatrices, and labeled as $P^{h_{\mathcal{A}},h_{\mathcal{B}}}(k)$. It is straightforward to obtain from $\mathbf{P}(k)$ the probability that $\mathcal{A}$ and $\mathcal{B}$ show parental ditype, tetratype and nonparental ditype with k chiasmata between them. For example, the chance of parental ditype with k chiasmata between the markers is the sum of all entries in the following matrices:

$$\mathbf{P}^{1,1}(k), \mathbf{P}^{2,2}(k), \mathbf{P}^{3,3}(k), \mathbf{P}^{4,4}(k), \mathbf{P}^{5,5}(k), \mathbf{P}^{6,6}(k).$$

Suppose there is chance q_k of there being k chiasmata between $\mathcal{A}$ and $\mathcal{B}$ and define $\mathbf{T}=\sum q_k\mathbf{T}(k)$ and $\mathbf{P}=\sum q_k\mathbf{P}(k)$. Then $t^{i,j}$ is the conditional probability that $\mathcal{B}$ is in state j given $\mathcal{A}$ is in state i, whereas $p^{i,j}$ is the joint probability that $\mathcal{A}$ is in state i and $\mathcal{B}$ in state j.

Using arguments similar to the single spore data case, we may obtain general results for multilocus ordered tetrads with $l+1$ markers, $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_{l+1}$, ordered starting from the centromere. Let $\mathbf{k} = (k_1, k_2, \ldots, k_l)$ and denote by $q_{\mathbf{k}}$ the joint probability of having k_i chiasmata between $\mathcal{A}_i$ and $\mathcal{A}_{i+1}$, $i = 1, \ldots, l$. Similarly, write $\mathbf{h}=(h_1, h_2, \ldots, h_{l+1})$, where h_i is the order type of the ith marker. If $\mathbf{p}^{h_1}_{\mathcal{A}_1}$ be the initial distribution of the state at $\mathcal{A}_1$, then we have

Theorem 3. *The multilocus probability of tetrad type* $\mathbf{h}$ *is*

$$\mathbf{p}^{h_1}_{\mathcal{A}_1}\Big(\sum_{\mathbf{k}} q_{\mathbf{k}}\mathbf{T}^{h_1,h_2}(k_1)\cdots\mathbf{T}^{h_l,h_{l+1}}(k_l)\Big)\mathbf{1}'.$$

If there is no chiasma interference, this probability can be factored as

$$\mathbf{p}^{h_1}_{\mathcal{A}_1}(\mathbf{T}^{h_1,h_2}_{\mathcal{A}_1\mathcal{A}_2}\cdots\mathbf{T}^{h_l,h_{l+1}}_{\mathcal{A}_l\mathcal{A}_{l+1}})\mathbf{1}'.$$

In the above discussion it is assumed that there is at least one chiasma before $\mathcal{A}_1$, so the state of $\mathcal{A}_1$ can be defined by the order type of the strands and the pair involved in the previous chiasma. The above results still hold if each pair is assigned the same chance of being involved in the previous chiasma when no chiasmata have occurred before $\mathcal{A}_1$.

2.3. *Unordered tetrad data.* To analyze unordered tetrads, the most common type of tetrad data obtained from genetic experiments, we also begin with two markers $\mathcal{A}$ and $\mathcal{B}$. Recall that there are three possible types of unordered tetrad data with two markers. If the unordered tetrads are thought to be generated from ordered tetrads but with the order lost, the parental ditype would result

from ordered tetrads with order types at two markers being one of $(1,1)$, $(2,2)$, $\ldots$, $(6,6)$; nonparental ditype would result from ordered tetrads with order types being one of $(1,2)$, $(2,1)$, $\ldots$, $(6,5)$; and tetratype tetrads would result from all other order pairs. Let $\mathbf{p}_{\mathcal{A}}^{h}$ be the initial distribution at $\mathcal{A}$, and write $p^{p}(k), p^{np}(k)$, and $p^{t}(k)$ as the probability of parental ditype, nonparental ditype, and tetratype with k chiasmata. Then

$$p^{p}(k) = (\mathbf{p}_{\mathcal{A}}^{1}\mathbf{T}^{1,1}(k) + \mathbf{p}_{\mathcal{A}}^{2}\mathbf{T}^{2,2}(k) + \cdots + \mathbf{p}_{\mathcal{A}}^{6}\mathbf{T}^{6,6}(k))\mathbf{1}',$$

$$p^{np}(k) = (\mathbf{p}_{\mathcal{A}}^{1}\mathbf{T}^{1,2}(k) + \mathbf{p}_{\mathcal{A}}^{2}\mathbf{T}^{2,1}(k) + \cdots + \mathbf{p}_{\mathcal{A}}^{6}\mathbf{T}^{6,5}(k))\mathbf{1}',$$

$$p^{t}(k) = (\mathbf{p}_{\mathcal{A}}^{1}(\mathbf{T}^{1,3}(k) + \cdots + \mathbf{T}^{1,6}(k)) + \cdots + \mathbf{p}_{\mathcal{A}}^{6}(\mathbf{T}^{6,1}(k) + \cdots + \mathbf{T}^{6,4}(k)))\mathbf{1}'.$$

Define $\mathbf{u} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, it can be shown that:

$$\begin{aligned} p^{p}(k) &= \mathbf{u}\mathbf{T}^{1,1}(k)\mathbf{1}', \\ p^{np}(k) &= \mathbf{u}\mathbf{T}^{1,2}(k)\mathbf{1}', \\ p^{t}(k) &= 4\mathbf{u}\mathbf{T}^{1,3}(k)\mathbf{1}'. \end{aligned}$$

Suppose now that we "order" unordered tetrads by always assigning order type 1 to $\mathcal{A}$ and order type 3 to $\mathcal{B}$ for tetratype data. There will only be three possible ordered tetrad types, $(1,1)$, $(1,2)$ and $(1,3)$ for parental ditype, nonparental ditype and tetratype. Then $p^{p}(k)$, $p^{np}(k)$, and $p^{t}(k)$ can be shown to be: $p^{p} = \mathbf{u}\mathbf{T}^{1,1}\mathbf{1}', p^{np} = \mathbf{u}\mathbf{T}^{1,2}\mathbf{1}'$, and $p^{t} = 4\mathbf{u}\mathbf{T}^{1,3}\mathbf{1}'$, where $\mathbf{T}^{h_{\mathcal{A}},h_{\mathcal{B}}} = \sum_{k} q_{k}\mathbf{T}^{h_{\mathcal{A}},h_{\mathcal{B}}}(k)$ is as defined in the previous subsection.

For unordered tetrads with $l+1$ markers, we may assign the order type to each marker by assigning order type 1 to $\mathcal{A}_1$. If there are no tetratypes among the l intervals, the order types of other markers are uniquely determined, either 1 or 2. If $\mathcal{A}_j$ and $\mathcal{A}_{j+1}$ is the first pair starting from $\mathcal{A}_1$ having tetratype, we assign order type 3 to $\mathcal{A}_{j+1}$. There will be no ambiguity of assigning order types afterwards since the order type of the four strands is fixed after this step. A one to one correspondence can be established between unordered tetrad types and ordered tetrad types from this procedure. Let $\mathbf{h} = (1, h_2, \ldots, h_{l+1})$ be the ordered tetrad type corresponding to the unordered tetrad type $\mathbf{g}$. Then the probability of type $\mathbf{g}$ is $p^{\mathbf{g}} = p^{\mathbf{h}}$ when there is no tetratype among all l intervals, and $p^{\mathbf{g}} = 4p^{\mathbf{h}}$ otherwise.

The matrix $\mathbf{U}$ and its powers play an important role in the above discussion, but no simple expression for $\mathbf{U}^{k}$ has been obtained. In the case of unordered tetrads with two markers, explicit expressions for $p^{p}(k)$, $p^{np}(k)$, and $p^{t}(k)$ do exist.

Theorem 4. *Let* $a=\alpha+\gamma$, $c_1 = 3+a$, $c_2 = -1+a$, *and* $c_3 = \sqrt{-3+6a+a^2}$ *(c_3 could be a complex number). Then*

$$p^t(k) = \frac{2}{3} - (\frac{1}{2})^k \frac{1}{12c_3}((c_1 - c_3)(c_2 + c_3)^{k+1} - (c_1 + c_3)(c_2 - c_3)^{k+1}).$$

When $k > 0$ is even,

$$p^p(k) = \frac{1}{2}(1 - p^t(k) + (\alpha - \gamma)^{\frac{k}{2}}),$$

$$p^{np}(k) = \frac{1}{2}(1 - p^t(k) - (\alpha - \gamma)^{\frac{k}{2}}).$$

When k is odd,

$$p^p(k) = p^{np}(k) = \frac{1}{2}(1 - p^t(k)).$$

The proof is given in section 6. The number c_3 could be complex, but $p^t(k)$ is always a real number. Note that when $(\alpha, 2\beta, \gamma) = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$, i.e., there is no chromatid interference, $a=\frac{1}{2}$ and $p^t(k) = \frac{2}{3}(1 - (-\frac{1}{2})^k)$. This well-known result was first proved by Mather (1935). When $\alpha+\gamma = 1$, i.e., consecutive chiasmata always involve the same two strands or four strands, $p^t(k) = 1$ when k is odd and 0 when k is even. The probability that two markers are recombined on a single strand can also be derived from Theorem 4. Because one half of the strands on tetratype tetrads are recombined, and all strands on nonparental ditype tetrads are recombined, the recombination fraction r is $\frac{1}{2}p^t(k)+p^{np}(k) = \frac{1}{2}\{1 - (p^p(k) - p^{np}(k))\}$. Note that r is $\frac{1}{2}$ when k is odd, and $\frac{1}{2}(1 - (\alpha - \gamma)^{\frac{k}{2}})$ when k is even. This result was proved earlier for single spore data by another approach.

3. Data Analysis. We fit the Markov chromatid interference model to unordered tetrad *S. pombe*. The data were kindly provided by Peter Munz. In this organism, chiasma interference is thought to be absent, the chiasma process can be assumed to follow the Poisson process. The parameters α, 2β, γ, and genetic distances were estimated using maximum likelihood. The estimates of α, 2β and γ are summarized in Table 1. Four markers on chromosome *I* (*leu1, his7, mat, his5*) were used in cross XB1050. A total of 277 tetrads were genotyped. In Table 1, XB1050-1 used markers *leu1, his7,* and *mat*, whereas XB1050-2 used markers *his7, mat,* and *his5*. Seven markers on chromosome *II* (*mat, ura5, his3, tps13, leu3, ade1, lys4*) were used in cross XC2 which had 458 offspring. XC2-3

TABLE 1

Estimates of α, 2β, and γ (and their standard errors) from experimental crosses using S. pombe. The data were provided by Peter Munz.

Cross	α	se_α	2β	$se_{2\beta}$	γ	se_γ
XB1051-1	0.34	0.03	0.46	0.02	0.21	0.03
XB1051-2	0.14	0.03	0.77	0.02	0.09	0.02
XC2-3	0.25	0.02	0.45	0.02	0.30	0.02
XC2-5	0.26	0.02	0.48	0.02	0.26	0.02

used markers *tps13, mat* and *leu3*. XC2-5 used markers *leu3, ade1* and *lys4*. The standard errors were calculated from the numerical approximation of the Fisher information. Except for XB1051-2, estimates of α, 2β, γ are close to $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$, which correspond to no chromatid interference.

4. Incorporating chromatid and chiasma interference. The chiasma process has to be specified if we are to fit this chromatid interference model to single spore and tetrad data. Due to the difficulty of separating chromatid interference from chiasma interference and the fact that chiasma interference is observed in many organisms, a suitable model for the chiasma process is essential to the estimation of α, β, and γ. If the chiasma model is misspecified, the estimates of α, β, and γ might indicate the presence of chromatid interference even when it is, in fact, absent.

Among different chiasma process models, the chi-square model has been found to give good fit to data from a variety of organisms (Zhao, Speed and McPeek 1995). The chi-square model can be traced back to Fisher, Lyon and Owen (1947) and has generally been of interest due to its mathematical tractability. Recently the chi-square model was suggested as a plausible biological model by Foss et al. (1993). However there are now doubts concerning the appropriateness of this motivation (Foss and Stahl 1995). The model is represented in the form $Cx(Co)^m$, as follows: assume that chiasma intermediates (C events) are randomly distributed along the four-strand bundle, and every C event will either resolve in a chiasma (Cx) or not (Co). When a C resolves as a Cx, the next m C's must resolve as Co events, and after m Co's the next C must resolve as a Cx, i.e., the C's resolve in a sequence $\cdots Cx(Co)^m Cx(Co)^m \cdots$. To make the process stationary, given a set of C events, the leftmost C has an equal chance to be one of $Cx(Co)^m$. We say a C event is in state k if it is the kth C after a Cx, i.e., the state "0" means that this C event is a Cx and "k" ($k \neq 0$) means this C event is the kth Co after a Cx. If we start counting C events from the leftmost marker, then the state of the nth C, s_n, forms a homogeneous Markov Chain.

Under the chi-square model, the state of a marker along the chromosome can be defined to be the same as the state of the last C event before the marker. If we also consider strands involved in each chiasma, a joint state of a marker $\mathcal{A}$ can be defined as $(r,"y")$ if the last C event before $\mathcal{A}$ is in state r and that this strand was involved in the last Cx event. Marker $\mathcal{A}$ has joint state $(r,"n")$ if the last C event before $\mathcal{A}$ is in state r and this strand was not involved in the last Cx event. Given $\mathcal{A}$ being in state $s_{\mathcal{A}}$ and there being k chiasmata between $\mathcal{A}$ and $\mathcal{B}$, define $p^o_{s_{\mathcal{A}},s_{\mathcal{B}}}(k)$ as the conditional probability that $\mathcal{B}$ is in state $s_{\mathcal{B}}$ and that the strand is involved in an odd number of chiasmata. Define $p^e_{s_{\mathcal{A}},s_{\mathcal{B}}}(k)$ as the conditional probability that $\mathcal{B}$ is in state $s_{\mathcal{B}}$ and that the strand is involved in an even number of chiasmata. Recursive relationships among the $p^o_{s_{\mathcal{A}},s_{\mathcal{B}}}(k)$ and $p^e_{s_{\mathcal{A}},s_{\mathcal{B}}}(k)$ can be easily established. For example, under the $CxCo$ model, the following relationships hold:

$$\begin{pmatrix} p^o_{s_{\mathcal{A}},(0,y)}(k+1) \\ p^o_{s_{\mathcal{A}},(1,y)}(k+1) \\ p^o_{s_{\mathcal{A}},(0,n)}(k+1) \\ p^o_{s_{\mathcal{A}},(1,n)}(k+1) \\ p^e_{s_{\mathcal{A}},(0,y)}(k+1) \\ p^e_{s_{\mathcal{A}},(1,y)}(k+1) \\ p^e_{s_{\mathcal{A}},(0,n)}(k+1) \\ p^e_{s_{\mathcal{A}},(1,n)}(k+1) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \eta & 0 & 1-\eta \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-\eta & 0 & \eta & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \eta & 0 & 1-\eta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1-\eta & 0 & \eta \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} p^o_{s_{\mathcal{A}},(0,y)}(k) \\ p^o_{s_{\mathcal{A}},(1,y)}(k) \\ p^o_{s_{\mathcal{A}},(0,n)}(k) \\ p^o_{s_{\mathcal{A}},(1,n)}(k) \\ p^e_{s_{\mathcal{A}},(0,y)}(k) \\ p^e_{s_{\mathcal{A}},(1,y)}(k) \\ p^e_{s_{\mathcal{A}},(0,n)}(k) \\ p^e_{s_{\mathcal{A}},(1,n)}(k) \end{pmatrix}.$$

These relationships can be used to derive closed form expressions for multilocus recombination probabilities. Because the techniques are essentially the same as before, we omit the details here. Similarly, closed form multilocus tetrad probabilities can be derived under this Markov chromatid interference model and the chi-square chiasma interference model.

5. Discussion. In this paper, we studied a Markov model for chromatid interference. Although both chromatid and chiasma interference are commonly assumed to be absent in analyzing genetic data, crossover interference has been observed in almost all organisms studied, including humans. Thus a reasonable mathematical model, which can capture the main features of the genetic data, may help the understanding of the underlying biological mechanisms and the construction of genetic maps.

Genetic map functions, $r = M(d)$, are often used to relate the unobservable map distance (d) to the observable recombination fraction, r, from single spore data. Map functions are also used to infer the map distance between the centromere and the marker to the proportion of second division segregants (SDS)

at the marker using ordered tetrad data. Most of the map functions proposed in the literature were constructed to deal with crossover interference, and some do fit data well. But when there are more than three markers present in the data, multilocus recombination/tetrad probabilities are not, in general, uniquely determined by the map function. Map functions under different chromatid and chiasma interference models were compared in Zhao and Speed (1998). For single spore data, it was found that recombination fraction is no longer a monotone function of map distance when both types of interference are present. In general, there is no one to one correspondence between the map distance and the recombination fraction, unless the NCI assumption holds. The presence of chiasma interference diminishes the effect of chromatid interference. For ordered tetrad data, for different chromatid interference models, the SDS proportion never exceeds $\frac{2}{3}$ under the Poisson model. When consecutive chiasmata always involve the same pair or different pairs of strands, the SDS proportion never goes above $\frac{2}{3}$. When consecutive chiasmata always involve three strands, except for the Poisson model, the SDS proportions all rise above $\frac{2}{3}$. As with single spore data, in general, there is no one to one correspondence between the map distance and the SDS proportion, unless NCI holds. Therefore, in the presence of genetic interference, extra caution should be taken when gene centromere distance is estimated from the SDS proportion, especially when the observed SDS proportion is large.

Both the Markov chromatid interference model and the chi-square model discussed in this paper are based on discrete time Markov chains, which are equivalently definable as one-dimensional random fields. Therefore, there is no preferred directonality in theory for both models.

The chromatid interference model discussed in this paper relies on the simplified assumption that the interference parameters do not depend on the distance between two crossovers. When the distance increases, the degree of interference might decrease. Recall that for single spore data, chromatid interference is determined by $\alpha - \gamma$. Carter and Robertson (1952) and Stam (1979) proposed that $\alpha - \gamma$ is a function $g(t)$ of the distance between two consecutive chiasmata, and considered the special form $g(t) = g(0)e^{-ct}$. For tetrad data, α, β, and γ have to be specified as a function of the distance. A simple model is

$$\begin{aligned}
\alpha(t) &= \alpha(0)e^{-ct} + \tfrac{1}{4}(1 - e^{-ct}), \\
\beta(t) &= \beta(0)e^{-ct} + \tfrac{1}{4}(1 - e^{-ct}), \\
\gamma(t) &= \gamma(0)e^{-ct} + \tfrac{1}{4}(1 - e^{-ct}).
\end{aligned}$$

However, no closed form expressions for multilcous probabilities have been obtained for this model.

Like any mathematical model, the models discussed above have to be tested and validated using real data sets. A variety of goodness-of-fit tests can be performed to examine whether the model provides a reasonable fit to the data sets (Read and Cressie 1988).

Because the existence of crossover interference has been well established in many organisms, chromatid interference should be considered together with chiasma interference. Although two types of interference are not separable using single spore data (Zhao and Speed 1996), they can be distinguished from tetrad data as demonstrated in this paper. When chromatid interference is present, genetic mapping assuming the absence of chromatid interference can lead to incorrect genetic maps. In such cases, the model studied in this paper provides a useful approach to incorporating chromatid interference.

6. Proofs.

6.1. *Proof of Theorem 1.* Define $\mathbf{S}_k = \mathbf{T}_k^1 + \mathbf{T}_k^0$ and $\mathbf{D}_k = \mathbf{T}_k^1 - \mathbf{T}_k^0$. We have,

$$\mathbf{S}_{k+1} = \mathbf{S}_k \begin{pmatrix} \eta & 1-\eta \\ 1-\eta & \eta \end{pmatrix} = \mathbf{S}_k \mathbf{U},$$

and

$$\mathbf{D}_{k+1} = \mathbf{D}_k \begin{pmatrix} -\eta & 1-\eta \\ -(1-\eta) & \eta \end{pmatrix} = \mathbf{D}_k \mathbf{V}.$$

Thus,

$$\mathbf{S}_k = \mathbf{S}_0 \mathbf{U}^k \qquad \text{and} \qquad \mathbf{D}_k = \mathbf{D}_0 \mathbf{V}^k,$$

where

$$\mathbf{S}_0 = \mathbf{T}_0^1 + \mathbf{T}_0^0 = \begin{pmatrix} +1 & 0 \\ 0 & +1 \end{pmatrix},$$

and

$$\mathbf{D}_0 = \mathbf{T}_0^1 - \mathbf{T}_0^0 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

It is easy to show that

$$\mathbf{U}^k = \frac{1}{2} \begin{pmatrix} 1+(2\eta-1)^k & 1-(2\eta-1)^k \\ 1-(2\eta-1)^k & 1+(2\eta-1)^k \end{pmatrix}.$$

When k is even,

$$\mathbf{V}^k = \begin{pmatrix} \delta^{\frac{k}{2}} & 0 \\ 0 & \delta^{\frac{k}{2}} \end{pmatrix}.$$

When k is odd,

$$\mathbf{V}^k = \begin{pmatrix} -\eta\delta^{\frac{k-1}{2}} & (1-\eta)\delta^{\frac{k-1}{2}} \\ -(1-\eta)\delta^{\frac{k-1}{2}} & \eta\delta^{\frac{k-1}{2}} \end{pmatrix}.$$

It is then straightforward to arrive at the expressions for $\mathbf{T}_k^1$ and $\mathbf{T}_k^0$.

6.2 *Proof of Theorem 4.* Without loss of generality, assume $\mathcal{A}$ is in state $[1,(0,0)]$. Let $p^{s_{\mathcal{B}}}(k)$ be the conditional probability that $\mathcal{B}$ is in state $s_{\mathcal{B}}$ given k chiasmata between $\mathcal{A}$ and $\mathcal{B}$. Define

$$p(k) = p^{[1,(0,0)]}(k) + p^{[1,(0,1)]}(k) + p^{[1,(1,0)]}(k) + p^{[1,(1,1)]}(k),$$

$$np(k) = p^{[2,(0,0)]}(k) + p^{[2,(0,1)]}(k) + p^{[2,(1,0)]}(k) + p^{[2,(1,1)]}(k),$$

$$t_1(k) = p^{[3,(0,0)]}(k) + p^{[4,(0,0)]}(k) + p^{[5,(0,0)]}(k) + p^{[6,(0,0)]}(k),$$

$$t_2(k) = p^{[3,(0,1)]}(k) + p^{[4,(0,1)]}(k) + p^{[5,(0,1)]}(k) + p^{[6,(0,1)]}(k),$$

$$t_3(k) = p^{[3,(1,0)]}(k) + p^{[4,(1,0)]}(k) + p^{[5,(1,0)]}(k) + p^{[6,(1,0)]}(k),$$

$$t_4(k) = p^{[3,(1,1)]}(k) + p^{[4,(1,1)]}(k) + p^{[5,(1,1)]}(k) + p^{[6,(1,1)]}(k).$$

Theorem 4 can be proved using the following recursive relationships:

$$\begin{aligned}
p(k+1) &= \beta t_1(k) + \gamma t_2(k) + \alpha t_3(k) + \beta t_4(k), \\
np(k+1) &= \beta t_1(k) + \alpha t_2(k) + \gamma t_3(k) + \beta t_4(k), \\
t_1(k+1) &= \alpha t_1(k) + \beta t_2(k) + \beta t_3(k) + \gamma t_4(k), \\
t_2(k+1) &= np(k), \\
t_3(k+1) &= p(k), \\
t_4(k+1) &= \gamma t_1(k) + \beta t_2(k) + \beta t_3(k) + \alpha t_4(k).
\end{aligned}$$

Acknowledgments. We thank Dr. Peter Munz for providing the S. Pombe data for our analysis and two referees for their insightful comments.

REFERENCES

CARTER, T. C. and ROBERTSON, A. (1952). A mathematical treatment of genetical recombination using a four-strand model. *Proc. Roy. Soc., B* **139** 410–426.

FISHER, R. A., LYON, M. F. and OWEN, A. R. G. (1947). The sex chromosome in the house mouse. *Heredity* **1** 335–365.

FOSS, E. and STAHL, F. W. (1995). A test of a counting model for chiasma interference. *Genetics* **139** 1201–1209.

FOSS, E., LANDE, R., STAHL, F. W. and STEINBERG, C. M. (1993). Chiasma interference as a function of genetic distance. *Genetics* **133** 681–691.

GRIFFITHS, A. J. F., MILLER, J. H., SUZUKI, D. T., LEWONTIN, R. C. and GELBART, W. M. (1996). *An Introduction to Genetic Analysis*, 6 *th* edition. W. H. Freeman and Company, New York.

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

MATHER, K. (1938). Crossing-over. *Biological Reviews* **13** 252–292.

MCPEEK, M. S. and SPEED, T. P. (1995). Modeling interference in genetic recombination. *Genetics* **139** 1031–1044.

READ, T. R. C. and CRESSIE, N. A. C. (1988) *Goodness-of-Fit Statistics for Multivariate Discrete Data*. Springer-Verlag, New York.

SPEED, T. P. (1995). What is a genetic map function? in *Genetic Mapping and Sequencing*, edited by T. P. SPEED and M. S. WATERMAN. IMA volumes on Mathematics and its Applications. Springer-Verlag, New York.

SPEED, T. P., MCPEEK, M. S. and EVANS, S. N. (1992). Robustness of the no-interference model for ordering genetic markers. *Proc. Natl. Acad. Sci. USA* **89** 3103–3106.

STAM, P. (1979). Interference in genetic crossing over and chromosome mapping. *Genetics* **92** 573–594.

STURT, E. and SMITH, C. A. B. (1976). The relationship between chromatid interference and the mapping function. *Cytogen. Cel. Genet.* **17** 212–220.

WEINSTEIN, A. (1938). Mathematical study of multiple-strand crossing over and coincidence in the chromosomes of Drosophila. *Am. Phil. Soc. Yearbook* 1937. 227–228.

ZHAO, H. and SPEED, T. P. (1996). On Genetic Map Functions. *Genetics* **142** 1369–1377.

ZHAO, H. and SPEED, T. P. (1998). Stochastic modeling of the crossover process during meiosis. *Communications in Statistics, Theory and Methods*. In Press.

ZHAO, H., MCPEEK, M. S. and SPEED, T. P. (1995). Statistical analysis of chromatid interference. *Genetics* **139** 1057–1065.

ZHAO, H., SPEED, T. P. and MCPEEK, M. S. (1995). Statistical analysis of crossover interference using the chi-square model. *Genetics* **139** 1045–1056.

DIVISION OF BIOSTATISTICS
YALE UNIVERSITY SCHOOL OF MEDICINE
NEW HAVEN, CT 06520
HONGYU.ZHAO@YALE.EDU

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA AT BERKELEY
BERKELEY, CALIFORNIA 94720
TERRY@STAT.BERKELEY.EDU