J. Math. Biol. (1998) 36: 448–468

Journal of Mathematical Biology

Discrete chiasma formation models and their associated high order interference

H. Manos, U. Liberman*

School of Mathematical Sciences, Raymond and Beverly Sackler, Faculty of Exact Sciences, Tel-Aviv University, Ramat Aviv, 69978 Tel Aviv, Israel

Received: 22 January 1996 / Revised version: 17 September 1997

Abstract. We introduce some special chiasma formation processes. First a family of discrete chiasma formation processes is introduced and we determine the nature of higher order interference associated with those processes. Secondly we consider a two-stage chiasma formation process, where the associated recombination frequency between two markers depends not only on their map distance but also on their location along the chromosomes. We characterise under this process, in some cases, the nature of interference between two segments.

Key words: Recombination – Interference – Chiasma formation process

1 Introduction

The calculation of recombination rates among multiple markers on chromosomes is usually based on independence assumptions regarding the occurrence of recombination for all segments involved. These assumptions are unrealistic since the occurrence of recombination between two adjacent markers generally implies a reduced chance of recombination for the next marker. This phenomenon called *interference* is said to take place whenever crossover events fail to occur at random.

The basic measure of interference is the pairwise *coincidence measure* due to Haldane (1919). If A and B are two disjoint genomic regions, then the coincidence measure, $D_{A,B}$ for A and B is defined as

$$D_{A,B} = \frac{r(A\&B)}{r(A)r(B)} \tag{1}$$

* Supported in part by NIH grant GM 28016

where $r(A)$, $r(B)$ are the recombination values associated with A and B, respectively, and $r(A\&B)$ denotes the probability of simultaneous coccurrence of recombination in both A and B.

When $D_{A,B} < 1$ we say that there is *positive interference* between A and B, when $D_{A,B} > 1$ we say that there is *negative interference*, and the *non-interference* case corresponds to $D_{A,B} = 1$. Most cytological data show the effects of positive interference, but there are some suggestions that negative interference can occur.

Karlin and Liberman (1994) have the following generalization of Haldane's measure. Let $\boldsymbol{A} = \{A_1, A_2, \ldots, A_k\}$ and $\boldsymbol{B} = \{B_1, B_2, \ldots, B_k\}$ be two collections of mutually disjoint genomic regions and define

$$D_{A,B} = \frac{r(A_1\&A_2\&\ldots\&A_k\&B_1\&B_2\&\ldots B_l)}{r(A_1\&A_2\&\ldots\&A_k)r(B_1\&B_2\&\ldots B_l)} \tag{2}$$

as the *higher order interference measure* between the two collections $\boldsymbol{A}$ and $\boldsymbol{B}$, where *e.g.* $r(A_1\&A_2\&\ldots\&A_k)$ is the probability of simultaneous occurrence of recombination in all the genomic regions in the collection $\boldsymbol{A} = \{A_1, A_2, \ldots, A_k\}$. The higher order interference between $\boldsymbol{A}$ and $\boldsymbol{B}$ is said to be positive if $D_{A,B} < 1$, and negative if $D_{A,B} > 1$. The case of no interference between $\boldsymbol{A}$ and $\boldsymbol{B}$ corresponds to $D_{A,B} = 1$.

The nature of interference between two collections of genomic regions is determined by the chiasma formation process operating along the chromosome. Some analytical models for chiasma formation with no chromatidal interference have been suggested. Among them we find the following two models:

1. **The Count-Location (C-L) Process was introduced independently by Karlin and Liberman (1979a) and Risch and Lange (1979).** Let the chromosome under consideration have length l. The count-location (C-L) process is determined by a discrete probability sequence $\boldsymbol{c} = \{c_0, c_1, \ldots,: c_i \geqq 0, \sum_i c_i = 1\}$ and a continuous distribution function $F(x)$ with support $[0, l]$, subscribing to the following interpretations.
   (a) The probability that i crossovers are formed in $[0, l]$ is c_i for $i = 0, 1, 2, \ldots$
   (b) Conditioned that i crossovers occur, their locations along the chromosome are distributed as i independent realizations from the distribution $F(x)$.

   We refer to the probability sequence $\boldsymbol{c}$ as the *count distribution* and to $F(x)$ as the *location distribution*. An important feature of such processes is the existence of a genetic *map function*. A map function $r = M(x)$ expresses the recombination frequency r between two markers as a function of their map distance x, where x is the expected number of chiasmata between the markers. Evans et al. (1993) show that, with no chromatidal interference, the C-L process is the only chiasma formation point process that possesses a map function.

2. **Discrete (nodule) Chiasma location** (Karlin and Liberman, 1994). Here it is assumed that the possible points of exchange along the chromosome, known as *nodules* are predetermined. Therefore it is natural to consider a discrete chiasma process, where the possible locations of the chiasmata are the given nodules. Thus the *discrete* (*nodule*) *chiasma formation process* is described by a sequence $\{X_n\colon n = 1, 2, \ldots\}$ of dichotomous random variables where $X_n = 1$ or 0 depending whether there is a point of exchange at the n-th nodule or not. A simple version of the discrete process is when the sequence $\{X_n\}$ describes a Markov chain of dependent random variables.

In this paper we consider some special discrete chiasma formation processes and determine the nature of their associated higher order interference. The paper is organized as follows. In Sect. 2 we review some representations of multilocus recombination distributions and a classification of the higher order measure of interference in terms of linkage values. In Sect. 3 we introduce a family of discrete chiasma formation processes and determine the nature of higher order interference associated with those processes. In Sect. 4 we consider a two-stage chiasma formation process, where the associated recombination frequency between two markers depends not only on their map distance but also on their location along the chromosome. We characterize under this process, in some cases, the nature of interference between two segments.

2 Classification of higher order interference

Karlin and Liberman (1994) introduce a classification of higher order interference in terms of *linkage values*. The classification is based on the notions of multilocus *crosover distribution* and their associated *linkage values*. For ready reference we describe the notions and the main results here.

Let A be any genomic region. We associate with A an indicator random variable X such that $X = 1$ or $X = 0$, depending on whether recombination occurred or did not occur on A, respectively. We define:

Definition 2.1. *Crossover distribution. Let* $\boldsymbol{A} = \{A_1, A_2, \ldots, A_k\}$ *be a collection of pairwise disjoint genomic regions. Let* $X_1, \ldots, X_k$ *be the k random variables indicating the occurrence of recombination on* $A_1, A_2, \ldots, A_k$, *respectively. The crossover distribution* $\boldsymbol{C}^{\boldsymbol{A}}$ *is the joint probability distribution of the random variables* $X_1, \ldots, X_k$. *It is determined by the* 2^k *parameters* $\{\boldsymbol{C}^{\boldsymbol{A}}(\boldsymbol{x})\}$ *where*

$$\boldsymbol{C}^{\boldsymbol{A}}(\boldsymbol{x}) = \boldsymbol{C}^{\boldsymbol{A}}(x_1, \ldots, x_k) \qquad x_i = 0, 1 \quad \text{for } i = 1, 2, \ldots, k \tag{3}$$

is the probability $P(X_1 = x_1, \ldots, X_k = x_k)$. *The parameters* $\{\boldsymbol{C}^{\boldsymbol{A}}(\boldsymbol{x})\}$ *satisfy*

$$0 \leqq \boldsymbol{C}^{\boldsymbol{A}}(\boldsymbol{x}) \leqq 1 \quad \text{for all } \boldsymbol{x} \tag{4}$$

$$\sum_{\boldsymbol{x}} \boldsymbol{C}^{\boldsymbol{A}}(\boldsymbol{x}) = 1 \tag{5}$$

Remark 2.1. Let $\boldsymbol{A} = \{A_1, \ldots, A_k\}$ and $\boldsymbol{B} = \{B_1, \ldots, B_l\}$ be two collections of mutually disjoint genomic regions then clearly

1. $r(A_1 \& A_2 \& \ldots \& A_k) = \boldsymbol{C}^{\boldsymbol{A}}(1)$
2. $r(B_1 \& B_2 \& \ldots \& B_l) = \boldsymbol{C}^{\boldsymbol{B}}(1)$
3. $r(A_1 \& \ldots A_k \& B_1 \& \ldots \& B_l) = \boldsymbol{C}^{\boldsymbol{A} \cup \boldsymbol{B}}(1)$

where $\mathbf{1} = (1, \ldots, 1)$ is a vector with the corresponding dimension whose all components are units. Therefore, the higher order measure of interference $\boldsymbol{D}_{\boldsymbol{A},\boldsymbol{B}}$ is given by

$$\boldsymbol{D}_{\boldsymbol{A},\boldsymbol{B}} = \frac{\boldsymbol{C}^{\boldsymbol{A} \cup \boldsymbol{B}}(\mathbf{1})}{\boldsymbol{C}^{\boldsymbol{A}}(\mathbf{1})\boldsymbol{C}^{\boldsymbol{A}}(\mathbf{1})} . \tag{6}$$

Given $A_1, \ldots, A_k$, let A be a genomic region that is a union of some or all of $A_1, \ldots, A_k$. Thus $A = \bigcup_{i \in \boldsymbol{I}} A_i$, where $\boldsymbol{I}$ is a non-empty subset of $\{1, 2, \ldots, k\}$. Let $\boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_k)$, where $\Delta_i = 0, 1$ for $i = 1, 2, \ldots, k$, be the incidence vector of $\boldsymbol{I}$, namely

$$\boldsymbol{I} = \boldsymbol{I}(\boldsymbol{\Delta}) = \{i : \Delta_i = 1\} . \tag{7}$$

We then define:

Definition 2.2. *Recombination and Linkage Values. Let $\boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_k)$ be an incidence vector.*

(i) *When $\boldsymbol{\Delta} \neq \mathbf{0}$ $r(\boldsymbol{\Delta})$ is the recombination value associated with the genomic region $\bigcup_{i \in \boldsymbol{I}(\boldsymbol{\Delta})} A_i$.*

(ii) *When $\boldsymbol{\Delta} = \mathbf{0}$ the associated genomic region is empty and we define*

$$r(\mathbf{0}) = r(0, \ldots, 0) = 0 . \tag{8}$$

The linkage value $\gamma(\boldsymbol{\Delta})$ is the one corresponding to the recombination value $r(\boldsymbol{\Delta})$ and is therefore defined as

$$\gamma(\boldsymbol{\Delta}) = 1 - 2r(\boldsymbol{\Delta})$$

where accordingly $\gamma(\mathbf{0}) = 1$.

Liberman and Karlin (1984) establish the following relations between the crossover distribution and its associated linkage values.

Proposition 2.1. *Let $\boldsymbol{C}$ be the crossover distribution associated with the genomic regions $\{A_1, \ldots, A_k\}$ and let $\{\gamma(\boldsymbol{\Delta})\}$ be their associated linkage values. Then for all $\boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_k)$*

$$\gamma(\boldsymbol{\Delta}) = \sum_{\boldsymbol{x}} (-1)^{(\boldsymbol{x}, \boldsymbol{\Delta})} \boldsymbol{C}(\boldsymbol{x})$$

and for all $\boldsymbol{x} = (x_1, \ldots, x_k)$

$$\boldsymbol{C}(\boldsymbol{\Delta}) = \frac{1}{2^k} \sum_{\boldsymbol{\Delta}} (-1)^{(\boldsymbol{x}, \boldsymbol{\Delta})} \gamma(\boldsymbol{\Delta})$$

where $(\boldsymbol{x}, \boldsymbol{\Delta}) = \sum_{i=1}^{k} x_i \Delta_i$. The sums extend over all possible values of $\boldsymbol{x}$ and $\boldsymbol{\Delta}$ respectively.

Following the above definitions, Remark 2.1 and Proposition 2.1, we have the following classification of higher order interference due to Karlin and Liberman (1994).

Proposition 2.2. *Let* $\boldsymbol{A} = \{A_1, A_2, \ldots, A_k\}$ *and* $\boldsymbol{B} = \{B_1, B_2, \ldots, B_l\}$ *be two collections of pairwise disjoint genomic regions. Then*

$$\begin{aligned} \boldsymbol{D}_{A,B} < 1 \quad & \text{iff } \boldsymbol{C}^{A\cup B}(\mathbf{1}) < \boldsymbol{C}^{A}(\mathbf{1})\boldsymbol{C}^{B}(\mathbf{1}) , \\ \boldsymbol{D}_{A,B} = 1 \quad & \text{iff } \boldsymbol{C}^{A\cup B}(\mathbf{1}) = \boldsymbol{C}^{A}(\mathbf{1})\boldsymbol{C}^{B}(\mathbf{1}) , \\ \boldsymbol{D}_{A,B} > 1 \quad & \text{iff } \boldsymbol{C}^{A\cup B}(\mathbf{1}) > \boldsymbol{C}^{A}(\mathbf{1})\boldsymbol{C}^{B}(\mathbf{1}) . \end{aligned} \tag{9}$$

Equivalently if $\{\gamma(\boldsymbol{\Delta})\}$ *are the linkage values associated with* $\boldsymbol{C}^{A\cup B}$, *such that* $\boldsymbol{\Delta} = (\boldsymbol{\Delta}^{A}, \boldsymbol{\Delta}^{B})$, *where* $\boldsymbol{\Delta}^{A}$ *is that part of* $\boldsymbol{\Delta}$ *corresponding to* $\boldsymbol{A}$ *and* $\boldsymbol{\Delta}^{B}$ *to* $\boldsymbol{B}$, *then* $D_{A,B}$ *is less than* 1, *equal to* 1 *or bigger than* 1 *iff*

$$\sum_{\boldsymbol{\Delta}^{A},\boldsymbol{\Delta}^{B}} (-1)^{|\boldsymbol{\Delta}^{A}| + |\boldsymbol{\Delta}^{B}|} [\gamma(\boldsymbol{\Delta}^{A}, \boldsymbol{\Delta}^{B}) - \gamma(\boldsymbol{\Delta}^{A}, \mathbf{0})\gamma(\mathbf{0}, \boldsymbol{\Delta}^{B})] \tag{10}$$

is negtive, zero or positive, respectively. Here $|\boldsymbol{a}| = a_1 + a_2 + \ldots + a_m$ *when* $\boldsymbol{a} = (a_1, a_2, \ldots, a_m)$ *and* $\mathbf{0}$ *is the zero vector with the corresponding dimension.*

3 Discrete model for chiasma formation

Suppose that there is an aray of discrete positions along the chromosome denoted by $0, 1, 2, \ldots$ at which crossover can occur. Let P_0 be the probability of no crossover at the 0 nodule and suppose that for any $n \geqq 1$, if no crossover occurs at the $n-1$ nodule then a crossover occurs at the n nodule with probability $1 - a_n$ and if a crossover occurs at the $n-1$ nodule then with probability $1 - b_n$ a crossover occurs at the n nodule. The transition probability matrix from the $n-1$ position to the n position is thus given by:

$$\begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{pmatrix} a_n & 1-a_n \\ b_n & 1-b_n \end{pmatrix} \end{array}$$

Where "0" denotes no crossover and "1" denotes a crossover at the given position.

We would like to determine the nature of interference associated with this process under the assumption of no chromatidal interference.

Let P_k be the probability that there is no crossover at the k nodule. Then for $k \geqq 1$

$$P_k = a_k P_{k-1} + b_k(1 - P_{k-1}) = b_k + (a_k - b_k)P_{k-1}. \tag{11}$$

In the special case where $a_n = a$ and $b_n = b$ we get that for any $k \geqq 1$

$$P_k = b + (a - b)P_{k-1}. \tag{12}$$

Or equivalently

$$P_k = b \sum_{i=0}^{k-1} (a-b)^i + (a-b)^k P_0 \,. \tag{13}$$

This gives

$$P_k = \frac{b(1-(a-b)^k)}{1+b-a} + P_0(a-b)^k. \tag{14}$$

Therefore for the region A that consists of the n nodules $s+1$, $s+2, \ldots, s+n$

$$\gamma(A) = P_{s+1} a^{n-1} = \left\{ \frac{b(1-(a-b)^{s+1})}{1+b-a} + P_0(a-b)^{s+1} \right\} a^{n-1}. \tag{15}$$

This is so as $\gamma(A)$, under the assumption of no chromatidal interference, is the probability that no crossover occurs in $\boldsymbol{A}$. Once $\gamma(A)$ is known we can determine the nature of interference using (10).

Consider now the general case where a_n and b_n depend on n. In this case it may be difficult to find $\gamma(A)$. However we can give two propositions which characterize the nature of interference between regions. Let $Y_n = 1$ denote the event that crossover occurs at the n module and $Y_n = 0$ denote the event that crossover does not occur at the n module. Let $P_m^n = P(Y_m = 0 \,|\, Y_n = 0)$.

Lemma 3.1. *Let A and B be any two regions along the chromosome. Assume that the last nodule in A is labeled by r and the first nodule in B is labeled by $m = r + s + 1$. Namely, between the regions A and B we have the nodules $r+1, \ldots r+s$. Then*

1. $\gamma(A \cup B) = \gamma(A)\gamma(B)\dfrac{P_m^r}{P_m}$.
2. *For any $j \geqq r$: $P_{j+1}^r - P_{j+1} = (P_j^r - P_j)(a_{j+1} - b_{j+1})$.*

Proof.

1. Let B' be the region B without its first nodule m. Then

$$\gamma(B) = P_m \gamma(B' \,|\, Y_m = 0) \tag{16}$$

where $\gamma(B' \,|\, Y_m = 0)$ is the conditional linkage value given $Y_m = 0$, namely the probability of no crossover in B' given that there is no chiasma formation at the m nodule.

Also

$$\gamma(B \,|\, A) = P_m^r \gamma(B' \,|\, Y_m = 0) \tag{17}$$

where $\gamma(B \,|\, A)$ denotes the probability of no crossover in the region B, given that there is no crossover in the region A. Since $\gamma(A \cup B) = \gamma(A)\gamma(B \,|\, A)$ (16) and (17) yield

$$\gamma(A \cup B) = \gamma(A)\gamma(B)\frac{P_m^r}{P_m} \,. \tag{18}$$

2. For any $j \geqq r$ we have by (11)

$$P_{j+1} = b_{j+1} + P_j(a_{j+1} - b_{j+1}) \tag{19}$$

and

$$P^r_{j+1} = b_{j+1} + P^r_j(a_{j+1} - b_{j+1}) \,. \tag{20}$$

Hence

$$P^r_{j+1} - P_{j+1} = (P^r_j - P_j)(a_{j+1} - b_{j+1}) \,. \tag{21}$$

Conclusion 1. Let A and B be two disjoint genomic regions. Denote the last nodule in A by r. Assume that for any j, $a_j > b_j$. Since $P^r_r = 1 > P_r$ we get from Lemma 3.1 by induction that for any $k \geqq r$, $P^r_k > P_k$. Hence the lemma implies that $\gamma(A \cup B) > \gamma(A)\gamma(B)$. Namely if $a_j > b_j$ for all j there is negative interference between any two genomic regions.

On the other hand if for all j, $b_j > a_j$ then the differences $P^r_k - P_k$ have alternating signs. In this case we cannot deduce in general the nature of interference. The following proposition gives the nature of pairwise interference for this case.

Proposition 3.2. *Let A and B be two disjoint genomic regions. Let r denote the last nodule in A and $m = r + s + 1$ denote the first nodule in the region B, namely between the regions A and B there are s nodules.*

Suppose that for all n, $b_n \geqq a_n$. Then, if s is even there is positive or zero interference between the regions A and B and if s is odd there is negative or zero interference between the regions A and B.

Proof. Let R be defined by

$$R = \gamma(A \cup B) - \gamma(A)\gamma(B) \,. \tag{22}$$

In order to prove the proposition we have to show that if s is even then $R \leqq 0$ and if s is odd then $R \geqq 0$.

From (18) it is sufficient to show that for any $s \geqq 0$, and $m = r + s + 1$, if s is even then $P^r_m \leqq P_m$ and if s is odd then $P^r_m \geqq P_m$.

It is clear that $P^r_r = 1 \geqq P_r$. Hence it is sufficient to prove that the sequence $\{P^r_j - P_j\}_j$ has alternating signs. This follows directly from (21), since for all $n, b_n \geqq a_n$.

Definition 3.1. "*Ordered collections*": *Let $\boldsymbol{S} = \{S_1, \ldots, S_{k_1}\}$ and $\boldsymbol{T} = \{T_1, \ldots, T_{k_2}\}$ be two collections of disjoint genomic segments, where $S_i = (x_i, x_i + s_i)$ for $1 \leqq i \leqq k_1$ and $T_j = (y_j, y_j + t_j)$ for $1 \leqq j \leqq k_2$. We say that the two collections $\boldsymbol{S}$ and $\boldsymbol{T}$ are "ordered collections" if*

$$x_1 < x_2 < \ldots < x_{k_1} < y_1 < y_2 < \ldots < y_{k_2} \,. \tag{23}$$

Namely there is no segment of the collection $\boldsymbol{T}$ between any two segments of the collection $\boldsymbol{S}$. Since the genomic regions are pairwise disjoint it is clear that

1. $x_i + s_i < x_{i+1}$ *for* $1 \leqq i < k_1$.
2. $y_i + t_i < y_{i+1}$ *for* $1 \leqq i < k_2$.
3. $x_{k_1} + s_{k_1} < y_1$.

We can prove that under the conditions $a_n \geqq b_n$ the result of Conclusion 1 for positive interference can be generalized for any pair of "ordered collections".

Proposition 3.3. *Suppose that for all n* $a_n \geqq b_n$. *Then the process involves higher order negative or zero interference for any pair of "ordered collections".*

Proof. Let $\boldsymbol{S} = \{S_1, \ldots, S_{k_1}\}$ *and* $\boldsymbol{T} = \{T_1, \ldots, T_{k_2}\}$ be two "ordered collections" of disjoint genomic regions. In view of (10) we have to prove that $R \geqq 0$ where R is defined by:

$$R = \sum_{\Delta} (-1)^{|\Delta|} [\gamma(\Delta^S, \Delta^T)] - \gamma(\Delta^S, 0)\gamma(0, \Delta^T)] \; . \tag{24}$$

For any $1 \leqq l \leqq k_1$ and $1 \leqq r \leqq k_2$ we define the set $E_{l,r}$ by

$$E_{l,r} = \{(\Delta^S, \Delta^T) \colon \textit{For } i > 1 \text{ and } j < r \; \Delta_i^S = \Delta_j^T = 0 \text{ and } \Delta_i^T = \Delta_l^S = 1\} \tag{25}$$

Then we can rewrite R in the form

$$R = \sum_{r,\, l} \sum_{\Delta \in E_{l,r}} (-1)^{|\Delta|} [\gamma(\Delta^S, \Delta^T) - \gamma(\Delta^S, \mathbf{0})\gamma(\mathbf{0}, \Delta^T)] \tag{26}$$

where $\Delta = 0$ was omitted from the sum as $\gamma(\mathbf{0}, \mathbf{0}) = 1$. In order to prove that $R \geqq 0$ it is sufficient to show that for any $1 \leqq l \leqq k_1$ and $1 \leqq r \leqq k_2$

$$\sum_{\Delta \in E_{l,r}} (-1)^{|\Delta|} [\gamma(\Delta^S, \Delta^T) - \gamma(\Delta^S, \mathbf{0})\gamma(\mathbf{0}, \Delta^T)] \geqq 0 \; . \tag{27}$$

For any Δ let $I(\Delta) = \{i \colon \Delta_i = 1\}$. Consider any $(\Delta^S, \Delta^T) \in E_{l,r}$. Then the last nodule in the region $A = \bigcup_{i \in I(\Delta^S)} S_i$ is the last module of S_l and the first nodule in $B = \bigcup_{i \in I(\Delta^T)} T_i$ is the first nodule of T_r. Let n denote the last nodule in the region S_l and $m = n + k + 1$ denote the first nodule in the region T_r. Then from Lemma 1 we get that

$$\gamma(\Delta^S, \Delta^T) = \frac{P_m^n}{P_m} \gamma(\Delta^S, \mathbf{0})\gamma(\mathbf{0}, \Delta^T) \; . \tag{28}$$

(28) implies that

$$\sum_{\Delta \in E_{l,r}} (-1)^{|\Delta|} [\gamma(\Delta^S, \Delta^T) - \gamma(\Delta^S, \mathbf{0})\gamma(\mathbf{0}, \Delta^T)] = \left(\frac{P_m^n}{P_m} - 1\right) \sum_{\Delta \in E_{i,r}} (-1)^{|\Delta|} \gamma(\Delta^S, \mathbf{0})\gamma(\mathbf{0}, \Delta^T) \tag{29}$$

Using the fact that for any i, $a_i - b_i > 0$ and that $P_n^n = 1 \geqq P_n$ we get by induction from (21) that for any $j \geqq n$, $P_j^n \geqq P_j$. In particular $P_m^n \geqq P_m$. Therefore in order to prove (27) we have to show that for any $1 \leqq l \leqq k_1$ and $1 \leqq r \leqq k_2$, $M_{l,r} \geqq 0$ where $M_{l,r}$ is defined by

$$M_{l,r} = \sum_{\Delta \in E_{l,r}} (-1)^{|\Delta|} \gamma(\Delta^S, \mathbf{0})\gamma(\mathbf{0}, \Delta^T) = \sum_{\Delta^S \in F_l} (-1)^{|\Delta^S|} \gamma(\Delta^S, \mathbf{0}) \sum_{\Delta^T \in G_r} (-1)^{\Delta^T} \gamma(\mathbf{0}, \Delta^T) \tag{30}$$

where

$$F_l = \{\Delta^S \colon \; \Delta_i^S = 0 \; i > l \text{ and } \Delta_l^S = 1\} \tag{31}$$

and

$$G_r = \{\Delta^T \colon \; \Delta_i^T = 0 \; i < r \text{ and } \Delta_r^T = 1\} \; . \tag{32}$$

To show that $M_{l,r} \geqq 0$ it is sufficient to show that for any m regions $A_1, \ldots, A_m$ and $1 \leqq m_0 \leqq m$

$$\sum_{\Delta \in H_{m_0}} (-1)^{|\Delta|} \gamma(\Delta) \leqq 0 \tag{33}$$

where H_{m_0} is the set of all the Δ of order m with $\Delta_{m_0} = 1$. We may assume without loss of generality that $m_0 = 1$ and show that

$$\sum_{\Delta \in H_1} (-1)^{|\Delta|} \gamma(\Delta) \leqq 0 \tag{34}$$

where H_1 is the set of all Δ with $\Delta_1 = 1$.

Indeed let H be the set of all Δ of order m and H_0 be the set of all Δ of order m with $\Delta_1 = 0$. Then

$$\sum_{\Delta \in H_1} (-1)^{|\Delta|} \gamma(\Delta) = \sum_{\Delta \in H} (-1)^{|\Delta|} \gamma(\Delta) - \sum_{\Delta \in H_0} (-1)^{|\Delta|} \gamma(\Delta) \,. \tag{35}$$

Let $\mathbf{1}_j$ denote a vector of order j of the form $(-1, 1, \ldots, 1)$. Proposition 2.1 implies that

$$C(\mathbf{1}_m) = \frac{1}{2^m} \sum_{\Delta \in H} (-1)^{|\Delta|} \gamma(\Delta) \tag{36}$$

is the probability of simultaneously recombination in all the regions A_i where $1 \leqq i \leqq m$ and

$$C(0, \mathbf{1}_{m-1}) + C(1, \mathbf{1}_{m-1}) = \frac{1}{2^{m-1}} \sum_{\Delta \in H_0} (-1)^{|\Delta|} \gamma(\Delta) \tag{37}$$

is the probability of simultaneously recombination in all the regions A_i where $2 \leqq i \leqq m$. Equations (36) and (37) imply that

$$\begin{aligned} \sum_{\Delta \in H_1} (-1)^{|\Delta|} \gamma(\Delta) &= \sum_{\Delta \in H} (-1)^{|\Delta|} \gamma(\Delta) - \sum_{\Delta \in H_0} (-1)^{|\Delta|} \gamma(\Delta) \\ &= 2^{m-1} [2C(\mathbf{1}_m) - C(0, \mathbf{1}_{m-1}) - C(1, \mathbf{1}_{m-1})] \\ &= 2^{m-1} [C(\mathbf{1}_m) - C(0, \mathbf{1}_{m-1})] \,. \end{aligned} \tag{38}$$

Recall from Karlin and Liberman (1994) that for any $\boldsymbol{x} \geqq \boldsymbol{y}$, $C(\boldsymbol{x}) \leqq C(\boldsymbol{y})$, and in particular $C(\mathbf{1}_m) - C(0, \mathbf{1}_{m-1}) \leqq 0$. Hence (34) follows from (38).

This concludes the proof.

4 Two-stage model for chiasma formation

One of the problems with the C-L process is that the amount of interference between any pair of regions does not depend neither on the distance between them nor on their locations along the chromosome but just on their map lengths. On the other hand the deficiency of the discrete chiasma location process is the assumption that the nodules are in specified locations and also the fact that the model assumes a constant number of nodules. We describe

now a model that combines the C-L process and the discrete chiasma location process, which allows variable number of nodules at unspecified locations.

The model is composed of two stages. The first stage is a C-L process of nodules formation and the second one is some process that determines in which nodules crossover occur. More specifically,

Definition 4.1. *The two-stage model. In the first stage we have discrete probability sequence* $\boldsymbol{c} = \{c_0, c_1, \ldots, : c_i \geqq 0, \sum_i c_i = 1\}$ *and a continuous distribution function* $F(u)$ *with support* $[0, l]$, *determining the number and location of the nodules where*

1. *The probability that n nodules are formed in* $[0, l]$ *is* c_n *for* $n = 0, 1, 2, \ldots$
2. *Conditioned that n nodules are formed, their locations along the chromosome are distributed as n independent realization from the distribution* $F(u)$.

We refer to $\boldsymbol{c}$ *as the count distribution of the nodules formation process and to* $F(u)$ *as the location distribution of nodules.*

Given that there are n nodes, the second stage is a Markov discrete process for chiasma formation, that determines in which nodules crossover occurs. The second stage given by the pair $[M, P_1]$, *where* P_1 *is the probability of no crossover in the first nodule and the transition probabiity matrix M given by*

$$\begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{pmatrix} a & 1-a \\ b & 1-b \end{pmatrix} \end{array}.$$

If $P_1 = 0$ and $b = 0$ we get that in any nodule crossover must occur. This gives the regular C-L process determined by $[\boldsymbol{c}, F(x)]$.

Let $A = [c, d]$ be any segment along the chromosome with map length $t = \int_c^d F(u)du$. Let $x = \int_0^c F(u)du$ be the map length of the segment from the beginning of the chromosome to the beginning of the segment A. Namely A is a segment of the form $[x, x + t]$ in map units. Thus the linkage value of A is a function of x and t. Namely $\gamma(A) = \gamma(x, t)$.

Proposition 4.1. *Consider a two-stage process, where the first stage is a C-L process of nodules, determined by* $[\boldsymbol{c}, F(x)]$, *and the second one is a discrete chiasma location process determined by* $[M, P_1]$. *Let* $f(t)$ *be the generating function of the count distribution* $\boldsymbol{c}$ *and let* $g(t) = f(1 - t)$. *Then the linkage value of the segment* $A = [x, x + t]$ *is given by*

$$\begin{aligned} \gamma(x, t) = g(t) &+ \frac{b}{a(1 + b - a)}\{g(t - at) - g(t)\} \\ &+ \frac{1}{a}\left\{P_1 - \frac{b}{1 + b - a}\right\}\{g(t + x[1 + b - a] - at) \\ &- g(t + x[1 + b - a])\} \end{aligned} \tag{39}$$

Proof. Conditioned that n nodules are formed let $B_n(k_1, k_2)$ be the probability that k_1 of them appear in the region $[0, x]$, k_2 of them appear if $A = [x, x+t]$ and $n - k_1 - k_2$ of them appear in the region $[x+t, 1]$. Then

$$B_n(k_1, k_2) = \binom{n}{k_1}\binom{n-k_1}{k_2} x^{k_1} t^{k_2} (1-x-t)^{n-k_1-k_2} . \tag{40}$$

For $1 \leqq i \leqq n$ let P_i denote the probability of no crossover in the i nodule. Then

$$P_{i+1} = P_i a + (1-P_i) b = b + P_i(a-b) . \tag{41}$$

This implies that

$$P_{i+1} = \frac{b}{1+b \div a} + \left(P_1 - \frac{b}{1+b-a}\right)(a-b)^i . \tag{42}$$

Let J be any genomic region. Conditioned that n nodules are formed along the chromosome, let $M_n(J)$ be the probability of no chiasma formed in J, and $T_n(J)$ be the probability that at least one of the nodules is in the region J, but there are no chiasma in this region. Then $M_n(A)$ is given by

$$M_n(A) = (1-t)^n + T_n(A) \tag{43}$$

where

$$T_n(A) = \sum_{k_1=0}^{n-1} \sum_{k_2=0}^{n-k_1} P_{k_1+1} a^{k_2-1} B_n(k_1, k_2) . \tag{44}$$

Substituting (40) in (44) yields

$$T_n(A) = \sum_{k_1=0}^{n-1} \binom{n}{k_1} x^{k_1} \sum_{k_2=1}^{n-k_1} \binom{n-k_1}{k_2} t^{k_2} (1-t-x)^{n-k_1-k_2} a^{k_2} \frac{P_{k_1+1}}{a} \tag{45}$$

Recall that

$$\begin{aligned} \sum_{k_2=1}^{n-k_1} \binom{n-k_1}{k_2}\left(\frac{at}{a-t-x}\right)^{k_2} &= -1 + \sum_{k_2=0}^{n-k_1} \binom{n-k_1}{k_2}\left(\frac{at}{1-t-x}\right)^{k_2} \\ &= -1 + \left(1 + \frac{at}{1-t-x}\right)^{n-k_1} \end{aligned} \tag{46}$$

Or equivalently

$$\sum_{k_2=1}^{n-k_1} \binom{n-k_1}{k_2}\left(\frac{at}{a-t-x}\right)^{k_2} = -1 + \left[\frac{1-x-(1-a)t}{1-t-x}\right]^{n-k_1} . \tag{47}$$

Substituting (47) in (45) and using the fact that for $k_1 = n$ the right-hand side of (47) vanishes yields that

$$\begin{aligned} T_n(A) = & -\frac{(1-t-x)^n}{a} \sum_{k_1=0}^{n} \binom{n}{k_1} P_{k_1+1} \left(\frac{x}{1-t-x}\right)^{k_1} \\ & + \frac{[1-x-(1-a)t]^n}{a} \sum_{k_1=0}^{n} \binom{n}{k_1} P_{k_1+1} \left[\frac{x}{1-x-(1-a)t}\right]^{k_1} . \end{aligned} \tag{48}$$

From (42) we get that for any W

$$\sum_{k_1=0}^{n}\binom{n}{k_1}P_{k_1+1}\,W^{k_1}=\frac{b}{1+b-a}\sum_{k_1=0}^{n}\binom{n}{k_1}W^{k_1}+\left[P_1-\frac{b}{1+b-a}\right]\sum_{k_1=0}^{n}\binom{n}{k_1}[W(a-b)]^{k_1}. \quad (48)$$

Or equivalently

$$\sum_{k_1=0}^{n}\binom{n}{k_1}P_{k_1+1}\,W^{k_1}=\frac{b}{1+b-a}(1+W)^n+\left[P_1-\frac{b}{1+b-a}\right](1+(a-b)W)^n \quad (50)$$

Using (50) we get from (48) that

$$T_n(A)=\frac{b}{a(1+b-a)}[(1-t+at)^n-(1-t)^n]+\frac{1}{a}\left[P_1-\frac{b}{1+b-a}\right]\times[1-t-x(1+b-a)+at)^n-(1-t-x(1+b-a))^n] \quad (51)$$

Recall that $g(x)=\sum_{n=0}^{\infty}c_n(1-s)^n$ and for any segment $A=[x,x+t]$ $\gamma(A)=\gamma(x,t)=\sum_{n=0}^{\infty}c_nM_n(A)$, we conclude by (43) and by (51) that

$$\gamma(x,t)=g(t)+\frac{b}{a(1+b-a)}\{g(t-at)-g(t)\}+\frac{1}{a}\left\{P_1-\frac{b}{1+b-a}\right\}\times\{g(t+x[1+b-a]-at\}-g(t+x[1+b-a])\}\,. \quad (52)$$

This proves the proposition.

Nature of interference far from the centromere

It is difficult to characterize in general the natural of interference for two or more segments since it is arduous to calculate the linkage value of the region $A\cup B$ where A and B are any two disjoint genomic region. However this is possible in the case where the process is stationary, $P_1=\frac{b}{1+b-a}$ and consequently by (42) for any k, $P_k=P_1$. This is approximately the case when the segments under consideration are far from the centromere.

Consider two disjoint segments A and B with map lengths t_1 and t_2 respectively. Let $\Delta=(\Delta^S,\Delta^T)=(\varepsilon_1,\varepsilon_2)$ where $\varepsilon_i=0,1$. Let $l_\Delta=\varepsilon_1t_1+\varepsilon_2t_2$ and let $l_{\Delta^s}=\varepsilon_1t_1$ and $l_{\Delta^T}=\varepsilon_2t_2$. We prove

Proposition 4.2. *Consider a two-stage process such that the first stage is determined by* $[\boldsymbol{c},F(x)]$ *and the second stage by* $[M,P_1]$ *where* M *is the transition probability matrix given by*

$$\begin{array}{c} \\ 0 \\ 1 \end{array}\begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{pmatrix} a & 1-a \\ b & 1-b \end{pmatrix} \end{array} \quad (53)$$

and $P_1=\frac{b}{(1+b-a)}$.

Assume that $a \geqq b$ and that the C-L process $[\boldsymbol{c}, F(x)]$ involves negative interference for any two disjoint genomic regions. Then the two-stage process involves negative interference for any two disjoint genomic regions.

Proof. Let $A = [x_1, x_1 + t_1]$ and $B = [x_1 + t_1 + x_2, x_1 + t_1 + x_2 + t_2]$ be two disjoint genomic regions represented in map units. We have to prove that

$$\gamma(A)\gamma(B) - \gamma(A \cup B) < 0 \; . \tag{54}$$

We determine first the linkage value $\gamma(A \cup B)$ of the region $A \cup B$. Conditioned that n nodules are formed the probability of no chiasma formation in the region $A \cup B$ is given by

$$M_n(A \cup B) = P_n(0,0) + P_n(0,1) + P_n(1,0) + P_n(1,1) \tag{55}$$

where

1. $P_n(0,0)$ is the conditional probability that given that n nodules are formed, no one of them is located in the region $A \cup B$. $P_n(0,0)$ is given by

$$P_n(0,0) = (1 - t_1 - t_2)^n \; . \tag{56}$$

2. $P_n(1,0)$ is the probability that conditioned that n nodules are formed at least one of them appears in the region A, no one of them appears in the region B, but there is no crossover in this region. $P_n(1,0)$ is given by

$$P_n(1,0) = (1 - t_2)^n T_n(\tilde{A}) \tag{57}$$

where $\tilde{A} = [\frac{x_1}{1-t_2}, \frac{x_1+t_1}{1-t_2}]$, and $T_n(\tilde{A})$ is the probability that at least one nodule is located at $\tilde{A}$, but there are no chiasma in this region. Since $P_1 = \frac{b}{1+b-a}$, (51) implies that

$$P_n(1,0) = -\frac{b}{a(1+b-a)} \sum_{\Delta^S} (-1)^{|\Delta^S|} (1 - t_1 - t_2 + a l_{\Delta^S})^n \tag{58}$$

where $\Delta^S = 0$ or 1.

3. $P_n(0,1)$ is the probability that conditioned that n nodules are formed, no nodule appears in the region A, at least one of them appears in the region B, but there is no crossover in this region. $P_n(0,1)$ is given by

$$P_n(0,1) = (1 - t_1)^n T_n(\tilde{B}) \tag{59}$$

where $\tilde{B} = [\frac{x_1+x_2}{1-t_1}, \frac{x_1+x_2+t_2}{1-t_1}]$. (51) implies that

$$P_n(0,1) = -\frac{b}{a(1+b-a)} \sum_{\Delta^T} (-1)^{|\Delta^S|} (1 - t_1 - t_2 + a l_{\Delta^T})^n \tag{60}$$

where $\Delta^T = 0$ or 1.

4. $P_n(1,1)$ is the probability that given that n nodules are formed at least one of them appears in B, and one of them in A, but there is no crossover in the region $A \cup B$. Let $B_n(k_1, k_2, k_3, k_4)$ be the probability of k_1 nodules in the

segment $[0, x_1]$, k_2 nodules in the segment $A = [x_1, x_1 + t_1]$, k_3 nodules in the segment $[x_1 + t_1, x_1 + t_1 + x_2$, k_4 nodules in the segment $B = [x_1 + t_1 + x_2, S]$ and, $n - k_1 - k_2 - k_3 - k_4$ nodules in the segment $[S, 1]$, where $S = x_1 + x_2 + t_1 + t_2$. Then

$$B_n(k_1, k_2, k_3, k_4) = \binom{n}{k_1}\binom{n-k_1}{k_2}\binom{n-k_1-k_2}{k_3}\binom{n-k_1-k_2-k_3}{k_4} \times x_1^{k_1} t_1^{k_2} x_2^{k_3} t_2^{k_4} (1-S)^{n-k_1-k_2-k_3-k_4} \tag{61}$$

and $P_n(1,1)$ is given by

$$P_n(1,1) = \sum_{\boldsymbol{K}} B_n(k_1, k_2, k_3, k_4) P_{k_1+1} a^{k_2-1} P^k_{k_3+k+1} a^{k_4-1} . \tag{62}$$

Where $k = k_1 + k_2$ and $\boldsymbol{K}$ is the set

$$\boldsymbol{K} = \{(k_1, k_2, k_3, k_4); \quad k_1 + k_2 + k_3 + k_4 \leqq n \text{ and } k_2, k_4 \geqq 1\} \tag{63}$$

We recall that $P^k_{k_3+k+1}$ is the probability of no crossover in the $k + k_3 + 1$ nodule conditioned that there is no crossover in the k nodule. Equation (41) implies that

$$P^k_{k_3+k+1} = \frac{b}{a+b-a} + \frac{(1-a)(a-b)^{k_3+1}}{1+b-a} . \tag{64}$$

Substituting (61) in (62) yields,

$$\begin{aligned} P_n(1,1) = \frac{1}{a^2} \sum_{k_1=0}^{n} \binom{n}{k_1} P_{k_1+1} x_1^{k_1} \sum_{k_2=1}^{n-k_1} \binom{n-k_1}{k_2} (at_1)^{k_2} \sum_{k_3=0}^{n-k_1-k_2} P^k_{k_3+k+1} \\ \times \binom{n-k_1-k_2}{k_3} x_2^{k_3} \sum_{k_4=1}^{n-k_1-k_2-k_3} \binom{n-k_1-k_2-k_3}{k_4} \\ \times (at_2)^{k_4} (1-S)^{n-k_1-k_2-k_3-k_4} \end{aligned} \tag{65}$$

Recall that

$$\sum_{k_4=1}^{n-k_1-k_2-k_3} \binom{n-k_1-k_2-k_3}{k_4} (at_2)^{k_4} (1-S)^{n-k_1-k_2-k_3-k_4} = \sum_{\Delta^T} (-1)^{|\Delta^T|+1} (1 - S + al_{\Delta^T})^{n-k_1-k_2-k_3} . \tag{66}$$

We conclude that

$$\begin{aligned} P_n(1,1) = \frac{1}{a^2} \sum_{k_1=0}^{n} \binom{n}{k_1} P_{k_1+1} x_1^{k_1} \sum_{k_2=1}^{n-k_1} \binom{n-k_1}{k_2} \\ \times (at_1)^{k_2} \sum_{k_3=0}^{n-k_1-k_2} P^k_{k_3+k+1} \binom{n-k_1-k_2}{k_3} x_2^{k_3} \sum_{\Delta^T} (-1)^{|\Delta^T|+1} \\ \times (1 - S + al_{\Delta^T})^{n-k_1-k_2-k_3} \end{aligned} \tag{67}$$

(64) implies that for any W,

$$\sum_{k_3=0}^{m}\binom{m}{k_3}P^{k}_{k_3+k+1}W^{k_3}=\frac{b}{a+b-a}\sum_{k_3=0}^{m}\binom{m}{k_3}W^{k_3}$$

$$+\frac{(a-b)(1-a)}{1+b-a}\sum_{k_3=0}^{m}\binom{m}{k_3}[W(a-b)]^{k_3}$$

$$=\frac{b}{1+b-a}(1+W)^m+\frac{(1-a)(a-b)}{1+b-a}$$

$$\times[1+(a-b)W]^m. \tag{68}$$

Recalling that $S=x_1+x_2+t_1+t_2$ we conclude that

$$\sum_{k_3=0}^{n-k_1-k_2}P^{k}_{k_3+k+1}\binom{n-k_1-k_2}{k_3}x_2^{k_3}\sum_{\Delta^T}(-1)^{|\Delta^T|+1}(1-S+al_{\Delta^T})^{n-k_1-k_2-k_3}$$

$$=-\frac{b}{1+b-a}\sum_{\Delta^T}(-1)^{|\Delta^T|}(1-x_1-t_1-t_2+al_{\Delta^T})^{n-k_1-k_2}$$

$$-\frac{(a-b)(1-a)}{1+b-a}\sum_{\Delta^T}(-1)^{|\Delta^T|}(1-S+(a-b)x_2+al_{\Delta^T})^{n-k_1-k_2}. \tag{69}$$

Substituting (69) in (67) and summing over k_2 imply that

$$P_n(1,1)=\frac{1}{a^2}\sum_{k_1=1}^{n}\binom{n}{k_1}P_{k_1+1}x_1^{k_1}\left\{\frac{b}{1+b-a}\sum_{\Delta}(-1)^{|\Delta^T|}\right.$$

$$\times(1-t_1-t_2-x_1+al_{\Delta^T})^{n-k_1}+\frac{(a-b)(1-a)}{1+b-a}$$

$$\left.\times\sum_{\Delta}(-1)^{|\Delta|}(1-t_1-t_2-(1+b-a)x_2+al_{\Delta^T})^{n-k_1}\right\}. \tag{70}$$

Using (49) and the fact that $P_1=\frac{b}{1+b-a}$ yields that

$$P_n(1,1)=\frac{1}{a^2}\left\{\left(\frac{b}{1+b-a}\right)^2\sum_{\Delta}(-1)^{|\Delta|}(1-t_1-t_2+al_{\Delta^T})^n+\right.$$

$$\left.\times\frac{(a-b)(1-a)}{(1+b-a)^2}\sum_{\Delta}(-1)^{|\Delta|}(1-t_1-t_2-(1+b-a)x_2+al_{\Delta^T})^n\right\}.$$

(71)

Recall that $\gamma(A\cup B)=\sum_{n=0}^{\infty}c_nM_n(A\cup B)$ and that $g(s)=\sum_{n=0}^{\infty}c_n(1-s)^n$ we conclude by (55), (56), (58), (60), and (71) that for $P_1=\frac{b}{1+b-a}$ the linkage value of $A\cup B$ is given by

$$\begin{aligned}\gamma(A\cup B)&=g(t_1+t_2)-\frac{b}{a(1+b-a)}\\&\quad\times\left[\sum_{\Delta^s}(-1)^{|\Delta^s|}g(t_1+t_2-al_\Delta)+\sum_{\Delta^\tau}(-1)^{|\Delta^\tau|}g(t_1+t_2-al_\Delta)\right]\\&\quad+\frac{b}{a^2(1+b-a)^2}\left\{\left(b\sum_{\Delta}(-1)^{|\Delta|}g(t_1+t_2-al_\Delta)+(a-b)(1-a)\right.\right.\\&\quad\left.\left.\times\sum_{\Delta}(-1)^{|\Delta|}g(t_1+t_2+(1+b-a)x_2-al_\Delta)\right)\right\}\end{aligned}\tag{72}$$

By Proposition 4.1 if $P_1=\frac{b}{1+b-a}$ then $\gamma(A)\gamma(B)$ is given by

$$\begin{aligned}\gamma(A)\gamma(B)&=\left\{g(t_1)-\frac{b}{a(1+b-a)}\sum_{\Delta^s}(-1)^{|\Delta^s|}g(t_1-al_{\Delta^s})\right\}\\&\quad\times\left\{g(t_2)-\frac{b}{a(1+b-a)}\sum_{\Delta^\tau}(-1)^{|\Delta^\tau|}g(t_2-al_{\Delta^\tau})\right\}.\end{aligned}\tag{73}$$

Or equivalently

$$\begin{aligned}\gamma(A)\gamma(B)&=g(t_1)g(t_2)+\left(\frac{b}{a(1+b-a)}\right)^2\sum_{\Delta}(-1)^{|\Delta|}g(t_1-al_{\Delta^s})g(t_2-al_{\Delta^\tau})\\&\quad-\frac{bg(t_2)}{a(1+b-a)}\sum_{\Delta^s}(-1)^{|\Delta|}g(t_1-al_{\Delta^s})\\&\quad-\frac{bg(t_1)}{a(1+b-a)}\sum_{\Delta^\tau}(-1)^{|\Delta|}g(t_2-al_{\Delta^\tau})\end{aligned}\tag{74}$$

(72) and (74) imply that

$$\begin{aligned}\gamma(A)\gamma(B)-\gamma(A,B)&=g(t_1)g(t_2)-g(t_1+t_2)\\&\quad+\left(\frac{b}{a(1+b-a)}\right)^2\sum_{\Delta}(-1)^{|\Delta|}\left[g(t_1-al_{\Delta^s})g(t_2-al_{\Delta^\tau})-g(t_1+t_2-al_\Delta)\right]\\&\quad+\frac{b}{a(1+b-a)}\sum_{\Delta^s}(-1)^{|\Delta^s|}\left[g(t_1+t_2-al_{\Delta^s})-g(t_2)g(t_1-al_{\Delta^s})\right]\\&\quad+\frac{b}{a(1+b-a)}\sum_{\Delta^\tau}(-1)^{|\Delta^\tau|}\left[g(t_1+t_2-al_{\Delta^\tau})-g(t_1-al_{\Delta^\tau})\right]\\&\quad+\frac{b(b-a)(1-a)}{(1+b-a)^2}\sum_{\Delta}(-1)^{|\Delta|}\left[g(t_1+t_2+(1+b-a)x_2-al_\Delta)\right.\end{aligned}\tag{75}$$

Let $y = t_1 + t_2 + (1 + b - a)x_2 - a(t_1 + t_2)$.

$$\sum_{\Delta} (-1)^{|\Delta|} g(t_1 + t_2 + (1 + b - a)x_2 - al_{\Delta}) =$$

$$[g(y) + g(y + at_1 + at_2)] - [g(y + at_2) + g(y + at_1)] > 0 \tag{76}$$

and it is positive since $g'' > 0$. Hence if $a \geqq b$ the last term in the right-hand side of (75) is non-positive and therefore by direct computation

$$\begin{aligned}\gamma(A)\gamma(B) - \gamma(A, B) \leqq (1 - V)^2[g(t_1)g(t_2) - g(t_1 + t_2)] \\ + V(1 - V)\{[g(t_1)g(t_2 - at_2) - g(t_1 + t_2 - at_2)] \\ + [g(t_2)g(t_1 - at_1) - g(t_1 + t_2 - at_1)]\} \\ + V^2[g(t_1 - at_1)g(t_2 - at_2) - g(t_1 + t_2 - a(t_1 + t_2))]\end{aligned} \tag{77}$$

where $V = \frac{b}{a(1+b-a)}$.

If $a \geqq b$ then $V = \frac{b}{a(1+b-a)} \leqq 1$. If the C-L process $[c, F(x)]$ involves negative interference for any two genomic regions then from Karlin and Liberman (1994) we know that for any x, y

$$g(x)g(y) - g(x + y) < 0 \,. \tag{78}$$

Hence all the terms in the right-hand side of (77) are negative and this in turn implies that

$$\gamma(A)\gamma(B) - \gamma(A \cup B) < 0 \tag{79}$$

as desired.

Example 1: Two-stage process with Poisson count

Consider a two-stage process with Poisson count of nodules formation namely.

$$c_k = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, 2, \ldots \tag{80}$$

It is known that the Poisson C-L processes is the only process that involve non-interference for any configuration of loci (see Karlin and Liberman 1994).

We show that in this case the two-stage process involves positive or negative interference depending only on the transition probability matrix M and specifically depending whether $a < b$ or $a > b$.

The generating function of the Poisson count is given by

$$f(s) = e^{-\lambda(1-s)} \,. \tag{81}$$

Therefore $g(x) = f(1 - s) = e^{-\lambda s}$ and Proposition 4.1 yields that

$$\gamma(x, t) = e^{-\lambda t}(1 - P + Pe^{\lambda at}) + Q(e^{\lambda at} - 1)e^{-\lambda t}e^{-\lambda x(1+b-a)} \,. \tag{82}$$

where $P = \frac{b}{a(1+b-a)}$ and $Q = \frac{1}{a}(P_1 - \frac{b}{1+b-a})$. The associated recombination value is

$$r(x, t) = \tfrac{1}{2} - \tfrac{1}{2}e^{-\lambda t}(1 - P + Pe^{\lambda at}) - \tfrac{1}{2}Q(e^{\lambda at} - 1)e^{-\lambda t}e^{-\lambda x(1+b-a)} \,. \tag{83}$$

Let $P_1 = \frac{b}{1+b-a}$ then (82) implies that

$$\gamma(x,t) = (1-P)\mathrm{e}^{-\lambda t} + P\mathrm{e}^{-\lambda(1-a)t}\,. \tag{84}$$

Since for any $x, y, g(x)g(y) - g(x+y) = 0$, (75) implies that

$$\gamma(A)\gamma(B) - \gamma(A \cup B) = \frac{b(b-a)(1-a)}{(1+b-a)^2}\mathrm{e}^{-\lambda(t_1+t_2+(1+b-a)x_2)}(1-\mathrm{e}^{a\lambda t_2}) \tag{85}$$

Therefore if $b > a$ then

$$\gamma(A)\gamma(B) - \gamma(A \cup B) > 0 \tag{86}$$

and the two-stage process involves positive interference for any two disjoint genomic regions, and if $b < a$ then

$$\gamma(A)\gamma(B) - \gamma(A \cup B) < 0 \tag{87}$$

and the two-stage process involves negative interference for any two disjoint genomic regions.

Example 2: Two-stage process with Binomial count
In Proposition 4.2 we show that at the stationary case where $P_1 = \frac{b}{1+b-a}$, if $a \geqq b$ and the C-L process $[\boldsymbol{c}, F(x)]$ involves negative interference for any two genomic regions, then also the corresponding two-stage process involves negative interference for any two genomic regions.

Consider the case where $b \geqq a$ and the C-L process $[\boldsymbol{c}, F(x)]$ involves positive interference for any two genomic regions. Unlike Proposition 4.2 we cannot characterize the nature of interference at the stationary case. Although when $b \geqq a$ the last term in (75) is nonnegative and we get an opposite inequality in (77), in this case $V = \frac{b}{a(1+b-a)} \geqq 1$ and this implies that the second term in the right-hand side of (77) is negative. Hence we cannot use (77) directly to show that $\gamma(A)\gamma(B) - \gamma(A \cup B) > 0$.

However we show that in the special case where $\boldsymbol{c}$ is a bionomial count and $b \geqq a$ then the corresponding two-stage process involves (at the stationary state) positive interference for any two genomic regions. It is known from Liberman and Manos (1995) that C-L process with Binomial count involves higher order positive interference

The bionomial count distribution $\boldsymbol{c}$ is given by

$$c_k = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & 0 \leqq k \leqq n \\ 0 & k > n \end{cases} \tag{88}$$

The generating function of the Binomial count is given by $f(s) = (1 - p + ps)^n$. We prove

Proposition 4.3. *Consider a two-stage process such that the first stage is determined by* $[\boldsymbol{c}, F(x)]$ *where* $\boldsymbol{c}$ *is a Binomial count and the second stage is specified by* $[M, P_1]$ *where* M *is the transition probability matrix given by*

$$\begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{pmatrix} a & 1-a \\ b & 1-b \end{pmatrix} \end{array} \tag{89}$$

and $P = \frac{b}{(1+b-a)}$.

Assume that $a \leqq b$. *Then the two-stage process involves positive interference for any two disjoint genomic regions.*

Proof. Let s_1 and s_2 be the map lengths of the two genomic regions A and B respectively. Since $b \geqq a$ the last term in the right-hand side of (75) is nonnegative and using $g(s) = f(1-s) = (1-ps)^n$, we conclude that

$$\begin{aligned} &\gamma(A)\gamma(B) - \gamma(A,B) \\ &\quad \geqq (V-1)^2[(1-t_1)^n(1-t_2)^n - (1-t_1-t_2)^n] \\ &\quad - V(V-1)[(1-t_1)^n(1-t_2+at_2)^n - (1-t_1-t_2+at_2)^n \\ &\quad + (1-t_2)^n(1-t_1+at_1)^n - (1-t_1-t_2+at_1)^n] \\ &\quad + V^2[(1-t_1+at_1)^n(1-t_2+at_2)^n - (1-t_1-t_2+at_1+at_2)^n] \end{aligned} \tag{90}$$

where $t_1 = ps_1$ and $t_2 = ps_2$ and $V = \frac{b}{a(1+b-a)} \geqq 1$. Let $X = 1 - t_1 - t_2$ and $Y = 1 - t_1 - t_2 + t_1 t_2$. Then (90) yields

$$\begin{aligned} \gamma(A)\gamma(B) - \gamma(A,B) &\geqq (V-1)^2 t_1 t_2 \, t \sum_{k=1}^{n} Y^{k-1} X^{n-k} \\ &\quad - V[(V-1)(1-a)t_1 t_2 \sum_{k=1}^{n} (Y + at_1 - at_1 t_2)^{k-1}(X + at_1)^{n-k} \\ &\quad + (Y + at_2 1 - at_1 t_2)^{k-1}(X + at_2)^{n-k} + V^2(1-a)^2 t_1 t_2 \\ &\quad \times \sum_{k=1}^{n} (Y + at_1 + at_2 - a(2-a)t_1 t_2)^{k-1}(X + at_1 + at_2)^{n-k}. \end{aligned} \tag{91}$$

Let $h_1 = at_1(1-t_2)$ and $h_2 = at_2(1-\mathrm{t}_1)$. In order to prove that $\gamma(A)\gamma(B) - \gamma(A,B) > 0$ it is sufficient to prove that for any m, r

$$\begin{aligned} L = (V-1)^2 Y^m X^r &+ V^2(1-a)^2(Y + h_1 + h_2 + a^2 t_1 t_2)^m \\ &\times (X + at_1 + at_2)^r - V(V-1)(1-a)[(Y+h_1)^m \\ &\times (X + at_1)^r + (Y + h_1)^m (X + at_2)^r] > 0 \,. \end{aligned} \tag{92}$$

Recall that

$$\begin{aligned} V^2(1-a)^2(Y + h_1 + h_2 + a^2 t_1 t_2)^m (X + at_1 + at_2)^r \\ \geqq V(V-1)(1-a)(Y + h_1 + h_2)^m (X + at_1 + at_2)^r \\ + [V^2(1-a)^2 - V(V-1)(1-a)] Y^m X^r \end{aligned} \tag{93}$$

We conclude using (92) that

$$L \geqq [(V-1)^2 + V^2(1-a)^2 - V(V-1)(1-a)]\, Y^m X^r + V(V-1)(1-a)[(Y+h_1+h_2)^m (X+at_1+at_2)^r - (Y+h_1)^m (X+at_2)^r - (Y+h_2)^m (X+at_2)^r] \quad (94)$$

By direct computation

$$(V-1)^2 + V^2(1-a)^2 - V(V-1)(1-a) > V(V-1)(1-a)\,. \quad (95)$$

Hence (94) implies that

$$L \geqq V(V-1)(1-a)\{Y^m X^r + (Y+h_1+h_2)^m (X+at_1+at_2)^r - (Y+h_1)^m (X+at_1)^r - (Y+h_2)^m (X+at_2)^r\}\,. \quad (96)$$

Suppost $t_1 \geqq t_2$. It is clear that,

$$(Y+h_1+h_2)^m (X+at_1+at_2)^r \geqq (Y+h_1+h_2)^m (X+at_1)^r + Y^m[(X+at_1+at_2)^r - (X+at_1)^r] \quad (97)$$

Recall that

$$(X+at_1+at_2)^r - (X+at_1)^r \geqq (X+at_2)^r - X^r \quad (98)$$

We conclude using (97) and (98) that

$$Y^m X^r + (Y+h_1+h_2)^m (X+at_1+at_2)^r \geqq (Y+h_1+h_2)^m (X+at_1)^r + Y^m (X+at_2)^r\,. \quad (99)$$

Hence (96) coupled with (99) implies that

$$L \geqq V(V-1)(1-a)(X+at_1)^r\{(Y+h_1+h_2)^m - (Y+h_1)^m\} - V(V-1)(1-a)(X+at_2)^r\{(Y+h_2)^m - Y^m\}\,. \quad (100)$$

Since $t_1 \geqq t_2$ we conclude that

$$L \geqq V(V-1)(1-a)(X+at_1)^r\{(Y+h_1+h_2)^m + Y^m - (Y+h_1)^m - (Y+h_2)^m\}\,. \quad (101)$$

It is clear that

$$Y^m + (Y+h_1+h_2)^m - (Y+h_1)^m - (Y+h_2)^m > 0 \quad (102)$$

and $V = \frac{b}{a(1+b-a)} \geqq 1$ as $b \geqq a$, and (101) implies that $L \geqq 0$. We conclude that $\gamma(A)\gamma(B) - \gamma(A \cup B) > 0$. Hence the process involves positive interference for any two genomic regions.

5 Discussion

Recombination does not occur uniformly across chromosomes in most species. In some chromosomal areas, chiasmata are rarely seen. It is important,

therefore, in modelling the process of recombination, and especially in interpreting data with such models, that this lack of uniformity be considered. The count-location process, while a first step in the modelling, entails that interference does not depend on the distance between chromosomal regions or their location. By superposition of the second stage of our model, namely a process that determines in which nodules crossovers occur, we generalize the possibilities for interference considerably.

Our second stage is a discrete Markov process for chiasma formation specified by the simple matrix M. We have seen in Proposition 4.2 and Examples 1 and 2 that if $a > b$ in (53) negative interference is generated. Examples 1 and 2 demonstrate that $b > a$ produces positive interference. In light of the rarity of observed negative interference, this suggests that biologically reasonable versions of M would have $b \geqq a$. It would, of couse, be interesting to estimate these parameters from observed pairs of chromosomal regions assuming Poisson or Binomial counts. This would allow eventual comparison of observed and expected levels of interference.

References

1. Evan, S. N., McPeek, M. S., Speed, T. P. (1993) A characterisation of crossover models that possess map functions. Theor. Popul. Biol. **43**, 80–90
2. Haldane, J. B. S. (1919) The combination of linkage values and the calculation of distance between the loci of linkage factors. J. Genet. **8**, 299–309
3. Karlin, S. and Liberman, U. (1979a) Classifications and comparisons of multilocus recombination distributions. Proc. Nat. Acad. Sci. USA **75**, 6332–6336
4. Karlin, S. and Liberman, U. (1979b) A natural class of multilocus recombination processes and related measures of crossover interference. Adv. Appl. Prob. **11**, 479–501
5. Karlin, S. and Liberman U. (1994) Theoretical recombination processes incorporating interference effects. Theor. Popul. Biol. **46**, 198–231
6. Liberman, U. and Karlin, S. (1984) Theoretical model of genetic map functions. Theor. Popul. Biol. **25**, 331–346
7. Liberman, U. and Manos, H. (1996) C-L processes involving higher order positive (negative) interference for any two collections of disjoint chromosomal regions. J. Math. Biol. **34**, 773–788
8. Mather, K. (1936) The determination of position in crossing over. J. Genet. **33**, 207–235
9. Owen, A. R. G. (1950) The theory of genetical recombination. Advan. Genet. **3**, 111–157
10. Risch, N. and Lange, K. (1979) An alternative model of recombination and interference. Ann. Human Genet. **43**, 61–70
11. Speed, T. P., McPeak, M. S. and Evans, S. N. (1992) Robustness of the no-interference model for ordering genetic marker. Proc. Natl. Acad. Sci. USA **89**, 3103–3106