CLUSTER MODEL OF CROSSING OVER

H. P. PAPAZIAN

Department of Zoology, University of Connecticut, Storrs, Connecticut

Received March 21, 1960

Many models of crossing over can be constructed that will fit more or less well with observed data of recombination frequencies and quadruplet frequencies. The model examined here is worthy of attention because it is a simple and attractive explanation of recent data of undoubted significance. When intervals which have a very low frequency of crossing over are studied, the frequency of multiple crossovers is too high. There is negative interference (Pritchard 1955).

To account for this, Pritchard proposed that crossovers occur in tight clusters. We can envisage the number of crossovers in a cluster to be constant for a species or for a chromosome. Map distance would be a function of the frequency of clusters.

The model: The frequency distribution of crossovers within a tight cluster will be taken to be a Poisson distribution with a mean of \( m_2 \), and the clusters themselves will be taken, in the first instance, to be in a Poisson distribution with a mean of \( m_1 \). This type of distribution (where a primary event occurs in a Poisson distribution and these primary events are themselves distributed in a Poisson) has been examined in several contexts where the primary and secondary events are respectively the number of passengers injured per automobile accident and the number of automobile accidents, the number of neutrons produced by an atomic collision and the number of collisions and, notably, the number of eggs laid by an insect and the number of layings. The latter case was examined by Geoffrey Beall (1940) to whom I am indebted for the argument leading to the present application of this so-called contagious distribution.

Distribution of tetratype and other frequencies: It has been shown (Papazian 1951) that if crossovers occur in a Poisson distribution then the frequency of tetratype quads equals \( 2/3 \left( 1 - e^{-3/2m} \right) \) where \( m \) is the mean frequency of crossovers. In extending this to the contagious distribution let the mean number of crossovers within a cluster be \( m_2 \), and let the mean number of clusters be \( m_1 \). In case there are \( S \) clusters the expected number of crossovers from all is Poisson with a mean \( Sm_2 \). In this case we should, from equation (1), get a proportion of \( 2/3 \left( 1 - e^{-3/2S_2} \right) \) tetratypes. The chance of there being \( S \) tight groups is, now, \( e^{-m_2} m_1^S \). This makes a total contribution of \( e^{-m_1} m_1^S 2/3 \left( 1 - e^{-3/2S_2} \right) \). We can

\[ \frac{S!}{S!} \]

\[ 1 \] Present address: 169 Cold Spring Street, New Haven, Connecticut.

Second Printing 1966 / University of Texas Printing Division, Austin
accordingly say that the over-all proportion of tetratypes, \( T \), is:

\[
\sum_{S=0}^{\infty} e^{-m} \frac{m^S}{S!} \left(1 - e^{-\frac{3}{2} m_2}\right)
\]

or

\[
T = \frac{e}{3} \left(1 - e^{-m_1} \left(1 - e^{-\frac{3}{2} m_2}\right)\right) - \frac{e}{3} \left(1 - e^{-m_2}\right)
\]

The frequency of nonparental ditype quads, \( N \), in the Poisson distribution of crossover model, has been shown to be

\[
N = 1/2 - 1/3 \left(1 - e^{-3m/2}\right) - (e^{-m})/2
\]

(Papazian 1952) where \( m \) is the mean frequency of crossovers.

By rearrangement: \( N = 1/6 + 1/3 \ e^{-3m/2} - 1/2 \ e^{-m} \) and according to the cluster model, by an argument similar to that used for tetratype frequencies,

\[
N = \sum_{S=0}^{\infty} e^{-m_1} \frac{m_2^S}{S!} \left(\frac{e}{3} + \frac{1}{3} e^{-\frac{3}{2} m_2} - \frac{e}{3} e^{-m_2}\right)
\]

and by a similar argument, from \( P = 1/2 - 1/3 \left(1 - e^{-3/2m}\right) + 1/2 e^{-m} \) (Papazian 1952) we get

\[
P = 1/6 + 1/3 \left(e^{-m_1} (1 - e^{-3/2m_2})\right) 1/2 \left(e^{-m_1} (1 - e^{-m_2})\right)
\]

where \( P \) is the frequency of parental ditype quads.

And from \( R = 1/2 (1 - e^{-m}) \) (Haldane 1919) we get

\[
R = 1/2 (1 - e^{-m_1} (1 - e^{-m_2}))
\]

Estimation of parameters: The cluster model requires two parameters, \( m_1 \) and \( m_2 \). If \( m_1 \) is a function of map distance and \( m_2 \) is a constant, we must examine how available or procurable data can be used to estimate \( m_2 \) and also to discriminate between a cluster model and a simple Poisson model.

2 This formula is published by kind permission of Geoffrey Beall of the Gillette Safety Razor Blade Company, and of the Gillette Safety Razor Blade Co.
The making of genetic maps from recombination data will not be altered by a cluster theory inasmuch as the addition law still holds, for if \( m_2 \) is constant and \( c = 1 - e^{-m_2} \), then from (5) \( R = \frac{1}{2} (1 - e^{-m_1} c) \)

\[
m_1 c = \ln \frac{1}{1 - 2R} \quad \text{and if} \quad m_1(12) = m_1(1) + m_1(2)
\]

then

\[
\frac{1}{c} \ln \frac{1}{1 - 2R_{12}} = \frac{1}{c} \ln \frac{1}{1 - 2R_1} + \frac{1}{c} \ln \frac{1}{1 - 2R_2}
\]

or \( R_{12} = R_1 + R_2 - 2R_1 R_2 \)

Similarly, for the addition of tetratype, or second division segregation, frequencies, from (2):

\[
\frac{1}{c} \ln \frac{1}{1 - 3T_{12}} = \frac{1}{c} \ln \frac{1}{1 - 3T_1} + \frac{1}{c} \ln \frac{1}{1 - 3T_2}
\]

or \( T_{12} = T_1 + T_2 - \frac{3}{2} T_1 T_2 \)

which is the same as for the simple Poisson model (Papazian 1952). There remain three different kinds of data which can discriminate between the two models and also provide an estimate for \( m_2 \). The distribution of multiple crossovers when a very short interval is involved, the relation between tetratype and nonparental ditype frequencies and the frequency of chiasmata between two points where the frequency of crossing over is known.

**Multiple crossovers in short regions:** The data which suggested to Pritchard that crossovers occurred in clusters consisted of progeny which were selected for crossovers between two adenineless mutants in *Aspergillus nidulans*. The frequency of crossing over between the two \( ad \) mutants was very small, of the order of \( 10^{-5} \). Among progeny from crosses between \( ad8 \) and \( ad10 \) (Pritchard's Table 6) and between \( ad8 \) and \( ad12 \) (Pritchard's Table 5) which had a crossover in this narrow region, 116 plants did not have crossovers on either side of the \( ad \) loci, 47 plants had a crossover on one or the other side and four plants had a crossover on both sides.

If we consider that in this experiment a single crossover of a cluster has been isolated, and other members of the same cluster are represented by crossovers on either side of \( ad \), then these frequencies can give an estimate of \( m_2 \) for clusters involving \( ad \).

Estimation is complicated by the fact that we do not know the zero term for the Poisson series. If crossovers within a cluster occur in a Poisson distribution, then sometimes there will be effective pairing; there will be a cluster so to speak, but it will contain no crossovers. The maximum likelihood estimation of the mean of a truncated Poisson has been examined by David and Johnson (1952) in
connection with the mean number of factory workers injured where the number of uninjured was unknown.

In the present case, the numbers of single, double, and triple crossovers is 116, 47 and four with a total, excluding zeros, of 167. The mean of the truncated Poisson is then $\frac{116 \times 2 + 47 \times 3 + 4 \times 4}{167} = 1.329$. The probability of each term of a truncated Poisson is $P_x = \frac{e^{-m} m X}{1-e^{-m} X!}$ where $m$ is the mean. The maximum likelihood estimate of $m$ is given by $m = 0.6$, which means this data gives an estimate of 0.6 for $m_2$ or the number of crossovers within a cluster. From this estimate of the mean, the expected numbers of zeros, singles, doubles, etc., crossovers can be calculated. The expected and observed numbers are:

<table>
<thead>
<tr>
<th></th>
<th>Single</th>
<th>Double</th>
<th>Triple and more</th>
</tr>
</thead>
<tbody>
<tr>
<td>expected</td>
<td>121.98</td>
<td>36.59</td>
<td>8.43</td>
</tr>
<tr>
<td>observed</td>
<td>116</td>
<td>47</td>
<td>4</td>
</tr>
</tbody>
</table>

A $\chi^2$ goodness of fit test on these figures with one degree of freedom gives $P \sim 2.1$ percent.

Relation of $N$ and $T$ frequencies: On a simple Poisson model the relation between the frequencies of tetratypes and nonparental ditypes is given by $N = \frac{1}{2} (1 - T (1 - 3/2T)^{n/2})$ (PAPAZIAN 1952); this relationship is shown by the bottom curve in Figure 1.

On a cluster model, from (2) $1/3 e^{-m_2} (1 - e^{-3/2m_2}) = 1/3 - 1/2T$. If we let $c = 1/3 - 1/2T$ then, from (3), $N = 1/6 c - 1/2 e^{-m_1} (1 - e^{-m_2})$, and $c = N - 1/6 1/2e^{-m_1} (1 - e^{-m_2})$ = $1/3 - 1/2T$ from which $N = 1/2 (1 - T e^{-m_1} (1 - e^{-m_2}))$..............................................................................(6)

Curves showing this relationship when $m_2 = 0.5, 1, 3, \infty$ are given in Figure 1. These curves were constructed for each value of $m_2$ by calculating $T$ for appropriate values of $m_1$ in (2) then substituting these corresponding values of $m_1$ and $T$ in (6) to get the corresponding value of $N$.

The algebraic solution of the expression of $N$ as a function of $T$ is heavy. As regards the recombination frequency, an algebraic solution is feasible and gives

$$R = -1/2 (1 - 3/2T) \left( \frac{1 - e^{-m_2}}{1 - e^{-3/2m_2}} \right) + 1/2.$$

Chiasmata and crossing over: If a chiasma, visible in tetrads, corresponds to regions where a cluster of crossovers occurs then $m_1$ should correspond to the frequency of chiasmata. From (2)

$$m_1 = \frac{ln(1/2) T}{1-e^{-3/2m_2}}$$

and from (5) $m_1 = \frac{ln(1/2) R}{1-e^{-m_2}}$ which gives the relationship
CROSSING OVER IN CLUSTERS

Figure 1.—Theoretical relationships between tetratype and nonparental ditype frequencies. Top four long horizontal curves give $T$ and $N$ for four values of $m_2$ on a cluster model, bottom curve on a simple Poisson model. Five, more vertical, shorter curves are isograms giving positions of five $m_1$ values on the cluster model curves. The short straight curve represents the cluster model with complete interference between clusters, only single clusters occurring.

between chiasmata and tetratype frequencies and chiasmata and recombination frequencies, respectively.

In Figure 1 the short, more nearly vertical, curves, give $N$ and $T$ values for values of $m_1$ of 0.5, 1.0, 1.5, 2.0 and 3.0. Where one of these five curves crosses a long curve for a certain $m_2$ value, the $T$ and $N$ value along each axis give $T$ and $N$ frequencies for that value of $m_1$, or chiasma frequency and $m_2$. Since $R = N + \frac{1}{2}T$, $R$ values can be read from these curves.

DISCUSSION

This study of the cluster model has neglected the troublesome consideration of the extent of the cluster. In the derivation of (2), (3), (4) and (5) the crossovers are assumed to occur all at a point whereas in Pritchard’s data the cluster has a definite but unknown extent on either side of the $ad$ complex.

As regards the fitting of data to this model: the calculated numbers of multiple crossovers are not outrageously different from Pritchard’s observations. The deviations are in the direction of not enough singles and triples, too many doubles. No plausible explanation of this occurs to the author. Data from Strickland (1958) in which quads were analyzed give $N$ and $T$ values which fall below the $m_2 = 0.5$ curve in Figure 1.

From his Table 2, $T = \frac{156}{392} = 0.398$, and $N = \frac{5}{392} = 0.013$; and from his Table 4, $T = \frac{203}{574} = 0.352$, and $N = \frac{21}{574} = 0.037$. If $m_2$ is constant for an organism then for $A. nidulans$ it should be around 0.5 from Pritchard’s multiple crossover data.
The cluster model can accommodate discrepancies of data from more widely separated loci such as STRICKLAND's, by allowing positive interference between clusters. If such interference is complete, then $T = m_1 \frac{2}{3} \left( 1 - e^{-3/m_2} \right)$, and $N = m_1 \left( \frac{1}{6} + \frac{1}{3} e^{-3/m_2} - \frac{1}{2} e^{-m_2} \right)$; the items in brackets being constants, the relationship is linear. This curve for $m_2 = 0.5$ is drawn in Figure 1. It will be noticed that in this case, $T$ can never rise above 0.36 and $N$ never above 0.022, instead of $2/3$ and $1/6$ which is true for all other curves shown in Figure 1.

There are few studies on the relationship between chiasmata and crossing over. Those of DARLINGTON (1934) on corn deviate in the direction of the crossover value being too low. This deviation may, and has, been diminished by more complete genetic data but can be also explained by a cluster theory of crossing over because you would sometimes get a cluster with zero crossovers. Of course whether such zero clusters would show up as chiasma is a moot point.

**SUMMARY**

The consequences of a theory of crossing over in which crossovers occur in tight clusters is examined. Formulas relating recombination, quadruplet, and chiasmata frequencies are developed and parameters roughly estimated from published data.

**ADDENDUM**

I would like to draw attention to a paper by R. H. PRITCHARD (Localized negative interference and its bearing on models of gene recombination. Genetical Research 1: 1–24) which appeared after this went to press.

PRITCHARD calculates the mean number of crossovers per cluster or effective pairing region using new data and a different approach. His approach makes use of the concept and the value of the length of an effective pairing region but in general the assumptions in PRITCHARD's paper appear to be the same as those made here, notably a Poisson distribution of crossovers within a cluster. I believe PRITCHARD's model, as used to calculate the mean crossovers per effective pairing region, is essentially equivalent to that used here.

PRITCHARD arrives at the value of 0.6 crossovers per effective pairing region which is the same as that found above. The exact coincidence of values must be regarded as coincidental.

**ACKNOWLEDGMENT**

I wish to thank Dr. ALAN JAMES of the Department of Mathematics, Yale University for guidance in dealing with the truncated Poisson.

**LITERATURE CITED**


HALDANE, J. B. S., 1919  The combination of linkage values and the calculation of distances between loci of linked factors. J. Genet. 8: 299–309.


1952  The analysis of tetrad data. Genetics 37: 175–188.
