CONTRIBUTORS TO VOLUME VI

RAYMOND W. BARRATT, Department of Biological Sciences, Stanford University, Stanford, California

LAURA GARNJOBST, Department of Biological Sciences, Stanford University, Stanford, California

BENTLEY GLASS, Mergenthaler Laboratories of Biology, The Johns Hopkins University, Baltimore, Maryland

JOHN HANCOCK, Ruakura Animal Research Station, New Zealand Department of Agriculture, Hamilton, New Zealand

PHILIP LEVINE, Ortho Research Foundation, Raritan, New Jersey

D. LEWIS, John Innes Horticultural Institution, Bayfordbury, Hertford, Herts, England

P. MICHAELIS, Max-Planck-Institut für Züchtungsforschung, Voldagsen über Elze, Germany

DOROTHY NEWMeyer, Department of Biological Sciences, Stanford University, Stanford, California

DAVID D. PERKINS, Department of Biological Sciences, Stanford University, Stanford, California

CHARLES L. REMINGTON, Osborn Zoological Laboratory, Yale University, New Haven, Connecticut

ALAN ROBERTSON, Institute of Animal Genetics, Edinburgh, Scotland
Map Construction in Neurospora Crassa

RAYMOND W. BARRATT, DOROTHY NEWMEYER, DAVID D. PERKINS, AND LAURA GARNJOBST

Department of Biological Sciences, Stanford University, Stanford, California

I. Introduction ................................................. 1
II. Problems in Estimating Linkage ......................... 2
   1. Uniform Treatment of Intervals .................... 3
   2. Correction for Undetected Multiple Crossovers by Means of Mapping Functions ................. 5
III. Linkage Groups and Chromosomes ...................... 20
    1. Status of the Linkage Groups .................... 20
    2. Cytogenetics ........................................... 24
IV. Data and Mapping ......................................... 26
    1. General Methods and Conventions ................ 26
    2. Presentation of Data ................................ 29
    3. Maps ................................................... 56
V. Discussion .................................................. 67
    1. Mapping Functions ................................... 67
    2. A Criterion for Independence Not Requiring Ordered Tetrads ..................... 78
    3. Data .................................................... 80
    4. Maps .................................................... 82
VI. Perspectives .............................................. 83
VII. Summary ................................................. 84
VIII. Acknowledgments ....................................... 85
IX. References ............................................... 86

I. INTRODUCTION

Neurospora was originally selected for studying biochemical genetics because it is an organism in which both formal genetic analysis and the investigation of nutritional mutants appeared possible (Beadle and Tatum, 1941). Its usefulness in both respects has been amply confirmed, but the genetic analysis has lagged considerably behind the biochemical studies. Published segregation data have been inadequate for constructing detailed linkage maps. This has handicapped the investigation of problems that require an accurate knowledge of linkage relations.

The situation has been changed to some extent by two recent publications which present extensive linkage data (Barratt and Garnjobst, 1949, and especially Houahan et al., 1949). In addition, the present
authors have had access to numerous unpublished data, both from Stanford University dissertations and from individual workers. Therefore an attempt has been made to compile here all available tetrads segregation data relevant to the estimation of linkage. Maps have been constructed on this basis, their reliability has been ascertained, and a variety of problems related to mapping have been considered. Several aspects of the work may be of interest not only for Neurospora but also for other organisms, and especially those in which tetrads can be analyzed.

Three major problems were encountered which are concerned with estimating linkage from tetrads segregations. First, the Neurospora tetrads data are from two different types of intervals (depending on whether the centromere is used as a marker), and it was necessary to combine both types into uniform maps. Second, many map intervals in Neurospora are so long that it appeared important to find a method of correcting for undetected multiple crossovers. Third, most of the data consist of small samples (partly because technical difficulties have prevented the analysis of large numbers of ascii), and criteria of reliability were therefore needed for application to linkage estimations. The small sample sizes have also made it necessary to evaluate carefully criteria for establishing linkage and allelism, and to make use of all available methods for determining gene order.

Maps constructed from the compiled data indicate that many uncertainties exist regarding linkage relations. It is hoped that presentation of maps at this time will prove valuable for visualizing experiments necessary to improve their reliability, but even more for planning basic investigations in gene action and cytogenetics.

No attempt has been made to deal comprehensively with aspects of the formal genetics of Neurospora which are not directly concerned with map construction or which have been adequately treated elsewhere. For example, special aspects of interference have been discussed as fully as present data warrant by Whitehouse (1942), Lindegren and Lindegren (1942), and Papazian (1952). For background material on genetic analysis in Neurospora and similar forms, the reader is referred to the publications of Whitehouse (1942), Lindegren (1942), Beadle (1945), Ryan (1950), Ludwig (1937), Knapp (1937), Mather (1938), Mather and Beale (1942), Catchetside (1951), and Papazian (1952).

II. Problems in Estimating Linkage

A number of problems will be discussed below that were encountered in the process of estimating map distances. Estimation of linkage values is, of course, meaningful only for genes known to be linked; the criteria used for establishing linkage are presented in Section IV.
1. Uniform Treatment of Intervals

a. Types of Segregation Data. Segregants from Neurospora crosses may be collected either as random single spores or as tetrads, which in turn may be obtained either in order or unordered. The great majority of existing segregation data are recorded from ordered tetrads, which consist of spores that have been dissected according to their linear order in the ascus. It would be possible to weigh and combine the few data based on random isolates with the tetrad linkage data (Papazian, 1952; see also Mather and Beale, 1942), but this would make it difficult to solve a much more pressing problem, namely the integration of centromere and linkage data (see Section II, 1, c below). Map distances computed in this paper have therefore been based solely on tetrads, although random isolates have been used for such other purposes as determining allelism and establishing linkage.

b. Types of Intervals. Analysis of ordered tetrads makes it possible to map centromeres (Lindegren, 1933), since the occurrence of second-division segregation is revealed by the arrangement of ascospores. The frequency of second-division segregations (postreductions) provides an approximate measure of the recombination frequency between a gene and a centromere (the so-called centromere distance), because homologous centromeres are known to segregate at the first meiotic division (prereduction).

Thus two types of intervals may be encountered when an ordered tetrad is analyzed. One type is bounded by a gene and a centromere, and is measured by one-half the percentage of second-division segregation (Lindegren, 1933); the second is defined by recombination between two genes. Map distances as ordinarily computed for these two types are not equivalent, because 4-strand double crossovers are detectable as recombinants in gene-gene intervals, but not in gene-centromere intervals. If data from both types of interval are to be plotted on the same map, it is necessary to put them on an equal basis.

Figure 1 illustrates the problem for a zygote from a cross of AB × ab. Three segregation types are possible: parental ditype (AB + AB + ab + ab), tetratype (AB + Ab + aB + ab), and nonparental ditype (Ab + Ab + aB + aB). These will be abbreviated PD, T, and NPD throughout the text. Four-strand double exchanges are detectable as nonparental ditype segregations when both ends of the interval are marked by genes. But in the absence of a marker at the centromere, like gene A in the figure, it would be impossible to distinguish a tetrad with a 4-strand double exchange from a tetrad having no exchange between the centromere and locus B, because both the PD and NPD tetrads give
first-division segregations for the B alleles. Thus, first-division segregations for a gene-centromere interval are equivalent to the sum of PD plus NPD segregations for a gene-gene interval, whereas second-division segregations are equivalent to tetratypes. This equivalence also holds for the results of higher multiple exchanges.

c. Integration of Data from Gene-Centromere and Gene-Gene Intervals. Uniform treatment of the two types of data could be achieved by ignoring the difference between the two ditype classes in the case of gene-gene intervals, and using as a measure of linkage either the frequency of tetraparental ditype or of second-division segregations, depending on the type of interval. If this be done, those few multiple crossovers that can be detected in gene-gene intervals, but which appear as noncrossovers in gene-centromere intervals, are scored consistently as noncrossovers in both cases.

The proposed procedure unfortunately prevents incorporation of data from random isolates into the same maps, except for short intervals. However, not only do gene-centromere intervals constitute the majority of all information available for mapping, but also there are so few intercrosses that at present it is impossible to map most loci without using gene-centromere intervals. There is therefore little choice except to exclude the few data on random isolates, as well as the information on the few detectable multiple crossovers, and integrate the two types of tetrad data in the manner proposed.
Maps constructed directly from data so integrated are completely uncorrected for multiple crossovers, i.e., they assume complete interference. A correction for multiple crossovers that is applicable to both gene-centromere and gene-gene intervals will now be described.

2. Correction for Undetected Multiple Crossovers by Means of Mapping Functions

a. Introduction. Few Neurospora crosses involve intervals short enough to preclude a significant amount of multiple crossing over, so that it has not been possible to build up accurate maps from data based on a series of short intervals. It has therefore seemed essential, as a temporary expedient, to seek a correction for the errors due to undetected multiple exchanges within the long intervals which contribute the bulk of existing data. This might be accomplished by a suitable mapping function. (Other possible methods of estimating the frequency of undetected crossovers (Whitehouse, 1942; Perkins, 1949) either provide incomplete corrections or are impractical for present data because of large sampling errors.) Although a mapping function might be only of theoretical interest in Drosophila or maize, where data suitable for mapping are abundant, it becomes a practical necessity for Neurospora, where data are fragmentary, and have mostly been collected incidentally to studying gene action.

Mapping function signifies a mathematical expression relating the observable results of crossing over to the actual map distance. In the case of single strands the observable results are recombination frequencies. In the case of tetrads, second-division or tetratype segregation frequencies are the only observable results common to all intervals; expressions relating these to map distance will be referred to as tetrads mapping functions.

Existing single-strand functions are inadequate for the Neurospora data. A simple model has therefore been proposed from which corresponding functions for both single strands and tetrads can be derived for any specified interference intensity. From among these possibilities a specific function has been chosen which agrees satisfactorily with various single-strand and tetrad data. Neurospora maps corrected by this function will later be presented for comparison with the uncorrected maps.

b. Inadequacy of Existing Mapping Functions. The true extent of interference in any Neurospora linkage group cannot be known until it is so well mapped that a correction is unnecessary. The best that can be done at present is to use as a correction a function that is based on the amount of interference found in some better known organism, and that agrees with the available Neurospora data.
A number of single-strand mapping functions have been proposed at various times for relating recombination frequencies to map distance. Several of these are shown in Fig. 2, where curve A describes the theoretical relation that would obtain with complete interference, and curve B, with zero interference (Haldane, 1919, 1931). Curve C is stated to agree with data for the mouse (Carter and Falconer, 1951). Curve D, proposed by Ludwig (1934) on the basis of his "gene curves," fits extensive data for the X-chromosome and individual autosomal arms in Drosophila. Curve E was proposed on the basis of mathematical simplicity (Kosambi, 1944) and is consistent with rather inadequate data in several plant species. For discussions of these and other mapping functions see Owen (1950), Glass (1951), Carter and Falconer (1951), and Carter and Robertson (1952).

In order to be generally useful for ordered tetrads, a function should be capable of relating map distance to second-division (or tetratype)
segregation frequency rather than merely to recombination frequency, which cannot be determined accurately from centromere-distance data. Consequently, a single-strand function, even though it described recombination correctly, would be useful for present needs only if a corresponding tetrad function could be derived from it. This is impossible for the empirical curves shown in Fig. 2, since they do not predict the distribution of exchanges among bivalents. A few single-strand functions do exist that are based on theoretical models (e.g., Owen, 1950; Carter and Robertson, 1952; Rademacher, 1932). These involve special assumptions of uncertain general validity, are mathematically complicated, and have been fully developed only for application to single strands.

(c) Mapping Functions for Tetrad. The only tetrad mapping functions which have been proposed in the past describe the two theoretical limiting cases of complete and zero interference (curves A and B, Fig. 5). The complete interference relation, \( y_t = 2x \), has been used for mapping ever since 1933, when Neurospora centromere distances were first computed by Lindegren. (The term \( y_t \) signifies the frequency of second-division or of tetratype segregations; \( x \) signifies map units times \( 10^{-3} \).) The zero interference tetrad function \( y_t = \frac{3}{4}(1 - e^{-x}) \) was first given in explicit form by Riset and Engelman (1949) and by Papazian (1951).

Although there is evidence that interference across the centromere may be negative in Neurospora (Lindegren and Lindegren, 1942; Whitehouse, 1942), the few available data indicate that interference is positive but incomplete within chromosome arms. What is needed for use within arms is a tetrad function intermediate between the existing complete and zero interference curves, analogous to the single-strand functions devised by Ludwig or Carter and Falconer. Such functions have been derived by making simple assumptions regarding the effect of interference on the distribution of exchanges among bivalents in a model chromosome segment.

d. Development of New Functions. Assumptions underlying a family of functions will be stated, and the hypothetical model on which they are based will be described before equations are derived.

(i) Description of hypothetical model. We have taken as a working model a chromosome segment marked genetically at both ends. Assume that interference between exchanges (‘chiasma’ interference) within this segment can take on any intensity from zero to complete interference, and that chromatid interference is absent.

When there is no interference, exchanges will be distributed at random among bivalents, following a Poisson distribution whose mean depends on the map length of the segment. We have assumed that when positive interference is introduced into the model the probability of nonexchange
tetrad remains unchanged, but that interference progressively decreases the proportion of multiple-exchange tetrads, relative to the frequency of singles. It is proposed to approximate the new proportions of tetrads that will have \( r \) exchanges \((r > 0)\) by multiplying the original (zero interference) probability of each successive Poisson term by a factor \( k^{r-1} \). The term \( k \) is a coefficient proportional to coincidence, and can

![Diagram of tetrad distribution](image)

**Fig. 3.** Effect of changing interference on the frequency distribution of exchanges in a model chromosome segment described in the text. Ordinates: probability. Abscissa: number of exchanges. M.D.: map distance. \( k \): coefficient proportional to coincidence, described in the text. Solid histograms: probability of exchanges in the model. Open histograms (frame E): distribution of exchange frequencies in the \( sc - g \) interval, *Drosophila melanogaster*, computed from data compiled by Anderson and Rhoades (1931). Dashed lines indicate the frequencies that would obtain with zero interference in segments having the same map lengths.

assume any value between zero (complete interference) and one (no interference).

The effect of interference on the distribution of exchanges, according to these assumptions, is shown by the histograms in Fig. 3, where the results of introducing different levels of interference into a model chromosome segment are shown in successive frames. As interference increases in this model segment from zero to complete (frames A to F), the probability of exchanges of small rank increases at the expense of exchanges of larger rank, but nonexchange tetrads retain their original probability. Therefore for each level of interference a new map length is realized by the model. For comparison the distribution of exchanges that would
be found in a segment having the same map length, but zero interference, is indicated in each frame by dashed lines.

(ii) Derivation of equations based on the model. Zero interference. Consider first the situation when zero interference obtains. The probability of bivalents with 0, 1, 2, . . . r exchanges in the model chromosome segment is given by the Poisson distribution that has a mean equal to the mean exchange frequency per bivalent.

\[ p(r) = \frac{(2x)^r}{r!} e^{-2x} \]  

(1)

where \(2x\) = mean exchange frequency per bivalent in the segment;
\(x\) = map length of segment (in map units \(\times \) 10\(^{-3}\)) = mean crossover frequency per strand;
\(r\) = number of exchanges in individual tetrads (or tetrad “rank”);
Weinstein, 1936);
\(p(r)\) = probability of occurrence of a tetrad of rank \(r\);
\(e\) = base of natural logarithms.

Multiplying each term by its rank, and summing, will give back the mean exchange frequency per bivalent; dividing by two will give the map length. Thus, map length is given by the equation

\[ x = \frac{1}{2} \sum_{r=1}^{\infty} r \cdot \frac{(2x)^r}{r!} e^{-2x}. \]  

(2)

The proportion of recombinant single strands expected from tetrads of any rank whatever except zero is uniformly \(\frac{1}{2}\) (Emerson and Rhoades, 1933). Therefore single-strand recombination frequency, \(y_0\), can be obtained by multiplying all terms except the \(p(0)\) term of equation 1 by \(\frac{1}{2}\)

\[ y_0 = \sum_{r=0}^{\infty} \frac{(2x)^r}{r!} e^{-2x} \cdot \frac{1}{2}. \]  

(3)

The frequency of tetratype (or second-division segregation) tetrads, \(y_t\), expected at zero interference can be obtained by multiplying each successive term in equation 1 by the proportion of tetrads of each successive rank that are expected to be tetratypes. This proportion is \(\frac{1}{2}[1 - (-\frac{1}{2})^r]\) (Mather, 1935) and the equation is

\[ y_t = \sum_{r=1}^{\infty} \frac{(2x)^r}{r!} e^{-2x} \cdot \frac{2}{3} \left[ 1 - \left( -\frac{1}{2} \right)^r \right]. \]  

(4)
Equation 3 is equivalent to \( y_\nu = \frac{1}{2}[1 - p(0)] \) and thus to Haldane's expression \( y_\nu = \frac{1}{2}(1 - e^{-2x}) \), plotted as curve B in Fig. 2. Similarly equation 4 is equivalent to the simpler tetrad equation \( y_\nu = \frac{1}{2}(1 - e^{-2x}) \) plotted as curve B in Fig. 5. However, the uncontracted equations 3 and 4 will be more useful for predicting the effect of interference, which is expected to change the relative frequencies of different specific ranks.

When interference is absent \((k = 1)\) the sum of terms of rank \(> 0\) is given by the summation of Poisson terms

\[
S_{k=1} = \sum_{r=1}^{r=\infty} \frac{(2x)^r}{r!} e^{-2x}.
\]  

(5)

Positive interference. Consider now the effect of interference on the probabilities of exchanges of different rank within the same model segment. When interference is positive \((k < 1)\), the factor \(k^{-1}\) is introduced and the sum of terms of rank \(> 0\) is correspondingly decreased.

\[
S_k = \sum_{r=1}^{r=\infty} \frac{(2x)^r}{r!} e^{-2x} \cdot k^{-1}.
\]

(6)

On our hypothesis, the probability of nonexchange tetrads, \(p(0)\), is not changed by interference. Therefore, the sum of probabilities for ranks greater than zero must continue to equal \(1 - p(0)\), even when interference obtains. This will be the case if each term is multiplied not only by \(k^{-1}\) but also by the factor \(S_{k=1}/S_k\), which will reconvert the term to a probability. In the presence of interference, the probability of tetrads of any rank greater than zero can then be expressed

\[
p(r) = \frac{(2x)^r}{r!} e^{-2x} k^{-1} \cdot \frac{S_{k=1}}{S_k}.
\]

(7)

Interference decreases the probability of higher multiple exchanges and must therefore decrease the map length from \(x\) (the value at zero interference) to a new value, \(x'\). Boost (1939) has used the term "a priori map length" to describe the map length, \(x\), that would obtain at zero interference in a model similar to ours. Since \(p(0) = e^{-2x}\), the a priori distance, \(x\), equals \(-\frac{1}{2} \log p(0)\). A priori map length depends only on the probability of nonexchange tetrads, \(p(0)\), and is independent of interference. Its value can be obtained from observed recombination frequencies by the equation \(x = -\frac{1}{2} \log (1 - 2y_\nu)\).

In contrast, the new map length \(x'\), as well as the new tetratype frequency, \(y_\nu'\), does vary with interference. Their values can be obtained by
substituting the new \( p(r) \) values from equation 7 for the Poisson terms in the original \( x \) and \( y \) equations (equations 2 and 4).

\[
x' = \frac{1}{2} \sum_{r=1}^{\infty} r \cdot \frac{(2x)^r}{r!} e^{-2x} k^{r-1} \frac{S_{k-1}}{S_k};
\]

\[
y' = \frac{2}{3} \sum_{r=1}^{\infty} r \cdot \frac{(2x)^r}{r!} e^{-2x} \left[ 1 - \left( -\frac{1}{2} \right)^r \right] k^{r-1} \frac{S_{k-1}}{S_k}.
\]

The single-strand recombination frequency, \( y_s \), is not changed when new \( p(r) \) terms from equation 7 are substituted in equation 3; \( y \), depends only on \( p(0) \), and thus resembles a priori map length in being independent of interference.

By substituting various values of \( x \) in equations 3, 8, and 9, we have obtained the curves shown in Figs. 4 and 5, where corresponding single-strand and tetrad mapping functions based on our model are shown for a
series of interference levels. This not only gives a series of tetrad functions from which to choose, but enables the extensive single-strand data available from other organisms to be used in making the selection.

e. Choice of a Specific Function for Neurospora. The shapes of the single-strand mapping function curves agree qualitatively with expecta-

![Diagram](image_url)

**Fig. 5.** Tetrad mapping functions based on the model described in the text. Curves are shown for seven graded interference intensities between the theoretical limiting curves A and B; these are the same intensities used for intermediate curves in Fig. 4. The sources of the two limiting curves are given in the text. Three sets of points are plotted: (a) open circles: postreduction frequencies computed from single-strand data, *D. melanogaster* X-chromosome (Anderson and Rhoades, 1931); (b) postreduction frequencies observed (solid circles) and calculated (triangles) from Drosophila attached-X-chromosome data (Beadle and Emerson, 1935).

...tions based on experience, and suggest that the model may be adequate as a first approximation. Use of a particular curve to estimate Neurospora map intervals must be justified, however, by quantitative correspondence with actual data.

(i) Agreement with Drosophila data. Probably the best data from any organism that relate map distance to recombination frequency and record all exchanges within a long chromosome segment, are those compiled by
Anderson and Rhoades (1931) for the X-chromosome of *Drosophila melanogaster*. These data can be used in at least four different ways to check the soundness of the model and to determine what particular function shows best agreement.

1. Recombination frequencies can be plotted against map distance, as was done originally by Anderson and Rhoades (Fig. 4). Their data, comprising 26,908 chromosomes tested by Bridges and Obrycht (1926) and by Anderson (1925), give good agreement with the $k = 0.2$ curve.

2. Postreduction frequencies can be computed from the single-strand data (Sax, 1932; Beadle and Emerson, 1935) and plotted against map distance (Fig. 5). Agreement of computed values with the $k = 0.2$ tetrad function is fair except for the two rightmost points (for $sc - f$ and $sc - f$).

3. Computed postreduction frequencies can be plotted against recombination frequencies (Fig. 6). The data fall near the 0.2 and 0.3 curves.

4. Frequency distributions of exchanges among bivalents can be computed from single-strand data (Sax, 1932; Mather, 1936) and compared with probabilities derived from the model (Fig. 3). Probabilities are tabulated below for bivalents with different numbers of exchanges in the $sc - f$ interval (Mather, 1936) and in the $sc - g$ interval. These are compared with the probabilities expected in models having a priori map distances similar to the two intervals and having interference intensities corresponding to $k = 0.2$ and $k = 0.3$.

<table>
<thead>
<tr>
<th>Interval</th>
<th>Frequency of bivalents of rank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Interference 0 1 2 3 4 &gt;4 A priori Realized</td>
</tr>
<tr>
<td>$sc - g$</td>
<td>observed 0.143 0.701 0.152 0.002 0.001 — 97 50.9</td>
</tr>
<tr>
<td>model</td>
<td>$k = 0.2$ 0.135 0.703 0.140 0.018 0.003 — 100 52.5</td>
</tr>
<tr>
<td>model</td>
<td>$k = 0.3$ 0.135 0.663 0.189 0.037 0.005 — 100 57.1</td>
</tr>
<tr>
<td>$sc - f$</td>
<td>observed 0.072 0.629 0.287 0.011 0.001 — 132 52.0</td>
</tr>
<tr>
<td>model</td>
<td>$k = 0.2$ 0.074 0.706 0.184 0.032 0.004 — 130 59.0</td>
</tr>
<tr>
<td>model</td>
<td>$k = 0.3$ 0.074 0.610 0.237 0.063 0.013 0.003 — 130 71.5</td>
</tr>
</tbody>
</table>

Agreement with frequencies predicted from the model are good for the $sc - g$ data and $k = 0.2$. These values are plotted as open bars in frame 5 of Fig. 3. The $sc - f$ data agree rather poorly with frequencies expected from $k = 0.3$. (This discrepant interval is represented by the rightmost point in Fig. 5, where it also deviates in the same direction from $k = 0.2$.)

Drosophila recombination data from several sources other than Anderson and Rhoades (Morgan et al., 1935 (X-chromosome); Reck, 1937
(X-chromosome at 13.5°, 25°, 30°C); Bridges and Morgan, 1919, 1923 (autosome arms) have also been plotted against map distance as in Fig. 4 but have not been shown. The points mostly fall near the 0.2 curve, although agreement may be better with \( k = 0.3 \) in some cases (notably Morgan et al., 1935).

![Graph](image)

**Fig. 6.** Curves relating tetrad type segregation frequency to single-strand recombination frequency based on the theoretical model described in the text. Two sets of points are plotted: (a) open circles: observed values for *Neurospora crassa* (from Tables 11-16). All intervals of 10 map units or more, within chromosome arms, and for which at least 25 tetrads have been recorded, are used; (b) solid circles: observed recombination and computed postreduction values for *Drosophila melanogaster* X-chromosome (Anderson and Rhodes, 1931). Curves are shown for the limiting cases of zero interference \((k = 1)\) and complete interference \((k = 0)\) and for interference corresponding in intensity to \( k = 0.2 \) and to \( k = 0.3 \) in the model.

A somewhat different type of comparison can be made using homozygosis frequencies from attached-X stocks of *Drosophila*. The most detailed analysis of this type was made by Beadle and Emerson (1935). Their observed and calculated recessive homozygosis values, multiplied by four so as to correspond with postreduction frequencies, have been plotted in Fig. 6. These could be interpreted to support use of either the \( k = 0.2 \) or the \( k = 0.3 \) curves.
(ii) Agreement with Neurospora data. Available Neurospora data cannot be used to choose a mapping function by testing recombination or second-division segregation frequencies against map distances, because the map distance axis is unknown; determination of map intervals is in fact the purpose of the investigation. But the two observable values, recombination frequency and tetratype frequency, can be used as coordinates for plotting the data so as to provide information regarding interference and regarding correspondence with the model, without its being necessary to make any assumptions whatever about map distance. This is identical with method 3 applied to the Drosophila data above. All data from crosses comprising 25 or more tetrads have been taken from Tables 11 through 16 for Neurospora genes known to be linked within individual chromosome arms. These are plotted in Fig. 6. As might be expected, they are inadequate for choosing between the curves that come closest to describing the results of crossing over in Drosophila, but do not seem to be inconsistent with either the \( k = 0.2 \) or \( k = 0.3 \) functions.

As a result of all these considerations, the \( k = 0.2 \) mapping function was adopted for converting observed tetratype or second-division segregation frequencies in Neurospora into "corrected" map distances. An enlarged curve, suitable for carrying this out in practice, is reproduced as Fig. 7.

An attempt was made to check the fit of the \( k = 0.2 \) function with the Neurospora data by testing the agreement of the sum of corrected short intervals with the length of the corrected over-all interval in all possible cases. No conclusions could be reached because of sampling errors.

It is apparent both in Drosophila (see Ludwig, 1934, 1938; Mather, 1938) and in Neurospora (Whitehouse, 1942; Lindegren and Lindegren, 1942) that a correction of this type is justified only for intervals within chromosome arms, because of centromere effects on interference. Consequently, the curve has not been applied to data from intervals spanning the centromere.

3. Determination of Confidence Limits

Most of the crosses involve such small numbers of asci (often 20 or fewer) that it seemed essential to have some measure of the magnitude of sampling error. For each observed map interval we have therefore determined the limits within which the true value of the map distance may reasonably be expected to lie, using the 95\% level of significance. To facilitate this task, a series of confidence limits curves has been constructed, from which the limits for each datum may be read directly.

Points for constructing the curves were obtained by the rapid graphical method of Mosteller and Tukey (1949), using a photographic en-
largement of their binomial probability paper. Limits were thus obtained for tetratype (or second-division) segregation percentages of 2, 5, 10, . . . 75 at 5-unit intervals, and for values of $N$ (number of asci) of 10, 20, 30, 50, 75, 100, 200, 400, and 600. The resulting points were plotted

Fig. 7. Tetratype mapping function curve based on the model discussed in the text, with interference corresponding to $k = 0.2$.

to give the curves shown in Fig. 8, in which each sample size is represented by two curves, corresponding to the upper and lower confidence limits. $N = \infty$ is a single curve separating the two sets of limits. A semi-logarithmic plot was used to spread the curves for greater readability at the lower tetratype frequencies. (The curves for $N = 75$ and $N = 400$
have been omitted from Fig. 8 to permit clearer reproduction.) Similar, accurately computed, curves have been published elsewhere (Eisenhart et al., 1947) but were not useful for our purpose because they included too few sample sizes and had too small a scale.

From Fig. 8 the approximate upper and lower limits can be read directly for any observed per cent tetratypes (or second-division segregations) by using the curves for the appropriate value of \( N \) and interpolating where necessary. The limits for each interval in the tables of centromere and linkage data (Tables 4 to 16) have been determined in this manner. In order to convert into uncorrected map distance, both the observed per cent tetratypes and the corresponding confidence limits were divided by two. To convert into corrected map distance, both the observed per cent tetratypes and the confidence limits were corrected by the \( k = 0.2 \) mapping-function curve (Fig. 7).

Precise confidence limits from the binomial and Poisson distributions may be obtained from a table prepared by W. L. Stevens (Fisher and Yates, 1949, p. 48); the use of this table is described more fully by Stevens (1942). However, reading the limits from our curves is considerably more rapid than computing them from the numbers in Stevens' table and is sufficiently accurate for the present purpose. Representative limits obtained from binomial probability paper fall within one-tenth of a map unit of those from Stevens' table when the range (the distance between the upper and lower limits) is 7 units or less; they never differ from his limits by more than \( \frac{1}{2} \) unit except when the ranges are so great (20 to 30 map units) that the error is negligible.

Where the observed tetratype frequency is zero, binomial probability paper is not sufficiently accurate, and the upper limits have therefore been calculated directly from the first term of the binomial expansion by the formula \( 1 - 0.025^{1/2} \), which is equivalent to the formula of Stevens. (The lower limits in such cases are zero.) Binomial probability paper is also inadequate for certain points (5%, \( N = 10 \); and 2%, \( N = 10, 20, \) and 30). Approximate values for these points were obtained by graphical interpolation on an arithmetic plot. These points do not actually occur in the discontinuous binomial distribution and thus cannot be calculated, but were needed to give smooth curves for interpolation. The portions of the curves derived from these points are indicated by dashed lines in the figure.

For sample sizes of more than 600 ascii, the limits have been calculated from the standard error. The standard error has not been used in the other cases because, as pointed out by Stevens (1942), it becomes extremely inaccurate when the distribution is not normal—in this case, when the tetratype frequency deviates appreciably from 50% and the sample size is small.
Fig. 8. Confidence limits curves (95% significance).
(cont.) drawn as described in the text.