# MAXIMUM LIKELIHOOD ESTIMATION OF LINKAGE AND INTERFERENCE FROM TETRAD DATA

RICHARD SNOW

*Department of Genetics, University of California, Davis, California 95616*

Manuscript received May 23, 1978
Revised copy received November 20, 1978

ABSTRACT

Maximum likelihood equations have been derived for estimation of map distance and interference from two-point and ranked tetrad data. The estimators have been applied to data from *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *S. cerevisiae* consistently shows quite strong interference over the mapped genome. In striking contrast, *S. pombe* consistently shows much weaker interference and many crosses exhibit negative interference. In neither species was there a conspicuous tendency for intervals spanning a centromere to show less interference than those that did not. Since the amount of recombination per microgram of DNA in the two species is similar, the difference in interference characteristics seems to be a reflection of some fundamental difference in the recombination process of the two species.

THE maximum likelihood method is widely used to estimate genetic parameters, especially linkage values. It has the very desirable property of extracting from the data the greatest amount of information available concerning the parameter in question, and so yields an estimate with the smallest variance (MATHER 1957). MATHER and BEALE (1942) applied this method to tetrad data, but their analysis was limited to cases where no more than two exchanges occurred between the linked loci and interference was not treated. A general treatment for any number of exchanges and for interference is presented here.

For organisms that produce tetrads, there are two types of data obtainable: those from two-point crosses and those from crosses with several segregating loci positioned such that the total number of exchanges between the ends of the interval can be more-or-less precisely determined. In the first case, the data are in the form of numbers of parental ditype (PD), nonparental ditype (NPD), and tetratype (T) tetrads. In the second case, the tetrads are classified as to "rank," that is, as to whether they have resulted from 0, 1, 2, . . . $r$ exchanges in the interval.

## DISCUSSION

### *Two-point crosses, interference not estimated*

In a genetic interval marked only at each end, three types of tetrads are possible, and we cannot determine directly the number of exchanges that have

occurred in any given tetrad, because the three types do not uniquely represent a given number of exchanges. Tetrads with no exchanges will of course be scored as PDs, but so also will those with two or more exchanges, the net effect of which is to produce the equivalent of two-strand double crossovers. Tetratypes arise from bivalents that had a single exchange, or from those with two or more exchanges, the net effect of which is to produce the equivalent of three-strand doubles. Nonparental ditypes arise from four-strand doubles as well as from higher orders, the net effect of which is to produce the equivalent of four-strand doubles.

In the absence of chromatid interference, for each four-strand double, there should be on average two three-strand doubles that will give tetratypes, and another two-strand double that will give a parental ditype. With each further exchange, those tetrads that would have been PD or NPD will be converted to T, while half of those that would have been T are converted with equal frequency into either PD or NPD. These relationships have been expressed by SHULT and LINDEGREN (1956) in the form

$$\begin{aligned} X(PD) &= T \\ X(NPD) &= T \\ X(T) &= T/2 + PD/4 + NPD/4 \end{aligned}$$

where X represents an exchange in a tetrad that would otherwise have been PD, NPD, or T.

If the interval is so short that the probability of exchanges greater than two is zero or very small, then the map distance can be calculated quite reliably from an expression derived by PERKINS (1949):

$$\text{map distance (cM)} = 1/2 \left[\frac{T + 6NPD}{PD+NPD+T}\right] 100.$$

The frequency of tetratypes for any number of exchanges, $r$, was given by MATHER (1935) as

$$p(T) = \frac{2}{3}\left[1 - \left(-\frac{1}{2}\right)^r\right]. \qquad (1)$$

Since with a large number of exchanges the proportions of PD and NPD should be equal (in the absence of chromatid interference), their proportions can be expressed as

$$p(PD) = p(NPD) = \frac{1}{2}\left\{1 - \frac{2}{3}\left[1 - \left(-\frac{1}{2}\right)^r\right]\right\} = \frac{1}{6} + \frac{1}{3}\left(-\frac{1}{2}\right)^r. \qquad (2)$$

When the interval is long enough so that there is a significant probability of exchanges greater than two, we assume that the probability of tetrads with 0, 1, 2, . . . $r$ exchanges between the two loci is given by that Poisson distribution which

has a mean equal to the mean exchange frequency in the interval (BARRATT *et al.* 1954):

$$p(r) = \frac{(2x)^r}{r!} e^{-2x} , \tag{3}$$

where $2x$ is the mean exchange frequency per tetrad in the marked interval; $x$ is the map length of the interval in map units $\times 10^2$ (*i.e.*, the mean exchange frequency per chromatid); and $r$ is the number of exchanges in the interval (tetrad rank).

The frequency of the three tetrad types is obtained by multiplying (1) and (2) by the sum of the appropriate Poisson terms:

$$\mathrm{P(PD)} = p(0) + \sum_{r=2}^{\infty} \frac{(2x)^r}{r!} e^{-2x} [p(\mathrm{PD})]$$

$$\mathrm{P(NPD)} = \sum_{r=2}^{\infty} \frac{(2x)^r}{r!} e^{-2x} [p(\mathrm{NPD})]$$

$$\mathrm{P(T)} = \sum_{r=1}^{\infty} \frac{(2x)^r}{r!} e^{-2x} [p(\mathrm{T})] .$$

Since

$$\sum_{r=1}^{\infty} \frac{(2x)^r}{r!} = e^{-2x} - 1, \text{ and } \sum_{r=1}^{\infty} \frac{(2x)^r}{r!} \left(-\frac{1}{2}\right)^r = e^{-x} - 1 ,$$

these three expressions reduce to:

$$\mathrm{P(PD)} = m_1 = \frac{1}{6} + \frac{1}{2} e^{-2x} + \frac{1}{3} e^{-3x} \tag{4}$$

$$\mathrm{P(NPD)} = m_2 = \frac{1}{6} - \frac{1}{2} e^{-2x} + \frac{1}{3} e^{-3x} \tag{5}$$

$$\mathrm{P(T)} = m_3 = \frac{2}{3} - \frac{2}{3} e^{-3x} . \tag{6}$$

These equations are the same as those obtained by HALDANE (1931) following a different line of reasoning, except that his exponents were $-x$ and $-3x/2$ because he used $x$ as the mean exchange frequency per bivalent (not per chromatid).

Expressions (4), (5), and (6) are the expectations used for maximum likelihood estimation of $x$. The log likelihood expression is

$$\log L = a_1 \log m_1 + a_2 \log m_2 + a_3 \log m_3 ,$$

where $a_1$, $a_2$, and $a_3$ are the observed numbers of PD, NPD, and T tetrads, respec-

tively. Differentiating and equating to zero gives the likelihood equation of estimation:

$$\frac{dL}{dx} = -a_1 \frac{e^{-2x}}{m_1}(1 + e^{-x}) + a_2 \frac{e^{-2x}}{m_2}(1 - e^{-x}) + a_3 \frac{2e^{-3x}}{m_3} = 0 \ . \qquad (7)$$

This equation can be solved by iteration and the standard error of $x$ obtained from the calculations used in the solution (MATHER 1957).

*Ranked data, interference not estimated*

When several closely spaced markers are available along the length of a chromosome, so that double crossing over between them is absent, the number of tetrads with various numbers of exchanges between the end markers can be determined. The expected distribution of ranks can be calculated from (3).

The value of $x$ could be estimated from the $p(0)$ term (the proportion with no exchanges), but that procedure will not utilize the information available from tetrads of higher rank. Since tetrads of rank 4 or higher will be rare or absent in most cases, only the terms $p(0)$, $p(1)$, $p(2)$, and $p(3)$ and the sum of the terms from $p(4)$ to $p(r)$ need be considered. Thus the expectations can be written as:

$$\begin{aligned} p(0) &= m_0 = e^{-2x} \\ p(1) &= m_1 = 2xe^{-2x} \\ p(2) &= m_2 = 2x^2e^{-2x} \\ p(3) &= m_3 = \frac{4}{3}x^3e^{-2x} \\ p(4\ldots r) &= m_4 = 1 - e^{-2x}[1 + 2x + 2x^2 + \frac{4}{3}x^3] \ . \end{aligned}$$

The log likelihood expression is:

$$\log L = a_0 \log m_0 + a_1 \log m_1 + a_2 \log m_2 + a_3 \log m_3 + a_4 \log m_4 \ .$$

Differentiating this expression and equating to zero gives the likelihood equation of estimation:

$$\frac{dL}{dx} = -2a_0 + a_1 \frac{1-2x}{x} + a_2 \frac{2-2x}{x} + a_3 \frac{3-2x}{x} + a_4 \frac{8}{3} \frac{x^3e^{-2x}}{m_4} = 0 \ , \qquad (8)$$

where $a_0$, $a_1$, $a_2$, $a_3$, and $a_4$ are the observed numbers of tetrads with 0, 1, 2, 3, and 4 or more exchanges, respectively.

The usual way of estimating $x$ from ranked data is to calculate the mean exchange frequency for a sample of tetrads. An example from Saccharomyces is given in Table 1. The average exchange frequency is 2194/2123 = 1.03344, which represents $2x$ in the Poisson expression (3). Hence $x = 0.51672$. The estimate obtained from (8) is 0.51688, the slight difference being due to rounding errors and to the truncation of the rank classification at 4. Hence, the usual procedure leads to the maximum likelihood estimate.

TABLE 1

*Analysis of ranked tetrads from* Saccharomyces cerevisiae *(data from* MORTIMER *and* FOGEL *1974)*

| Tetrad rank ($R_i$) | Number of tetrads ($a_i$) | Number of exchanges ($R_i a_i$) | Calculated number of tetrads (Poisson), no interference | Calculated number of tetrads, with interference ($k = 0.22785$) |
|---|---|---|---|---|
| 0 | 333 | 0 | 755.32 | 333.00 |
| 1 | 1406 | 1406 | 780.25 | 1438.73 |
| 2 | 367 | 734 | 402.99 | 303.62 |
| 3 | 14 | 42 | 138.76 | 42.71 |
| 4+ | 3 | 12 | 45.67 | 4.94 |
| | 2123 | 2194 | 2122.99 | 2123.00 |

This result means that an explicit solution should exist for the nontruncated form of the derivative of the log likelihood expression, which can indeed be derived:

$$\frac{dL}{dx} = -2a_0 + \frac{1}{x}\left[a_1(1-2x) + a_2(2-2x) + a_3(3-2x) + \ldots\right.$$

$$\left. + \quad a_n(n-2x)\right] = -2a_0 + \frac{1}{x}\left[a_1 + 2a_2 + \ldots na_n\right] - \left[2a_1 + 2a_2 + \ldots 2a_n\right],$$

$$= -2(a_0 + a_1 + \ldots a_n) + \frac{1}{x}\left[a_1 + 2a_2 + \ldots na_n\right] = 0 \ .$$

Thus,

$$x = \frac{1}{2}\,\frac{a_1 + 2a_2 + \ldots na_n}{a_0 + a_1 + \ldots a_n} = \frac{1}{2}\,\frac{\Sigma R_i a_i}{\Sigma a_i} \ .$$

Using $2x = 1.03344$, the expected distribution of tetrad ranks can be calculated from (3). The fourth column of Table 1 shows that there are far too many observed tetrads of rank 1, and too few of ranks 0, 2, 3, 4 and higher. Since the probability of the observed distribution being due to chance is less than 0.001, the data indicate strong chiasma interference.

*Two-point crosses, interference estimated*

The model for interference adopted here was presented by BARRATT *et al.* (1954). It is biologically reasonable, and can be handled rather easily mathematically. A model with several similar features for use with ordered tetrads has been developed by KUENEN (1962). With the BARRATT *et al.* (1954) model, the probability of nonexchange tetrads remains the same as with no interference, but interference progressively decreases the probability of tetrads of higher ranks in favor of those of lower ranks. This is achieved by multiplying each Poisson term except the $p(0)$ by $k^{r-1}$, where $k$ is an interference factor with a minimum of value 0 (complete interference). When $k = 1$ there is no interference, and values greater than one are indicative of negative interference.

If the probability of nonexchange tetrads is not changed by interference, then the sum of probabilities of ranks greater than zero must still equal $1-p(0)$ even with interference. This will be the case if each term greater than zero is multiplied by $k^{r-1}$ and also by a factor $S$, which is the sum of terms of rank greater than zero without interference divided by the sum of terms of rank greater than zero with interference. This ratio is:

$$S = \sum_{r=1}^{\infty} \frac{(2x)^r}{r!} e^{-2x} \Big/ \sum_{r=1}^{\infty} \frac{(2x)^r}{r!} e^{-2x} k^{r-1} = \frac{k(e^{2x}-1)}{e^{2kx}-1} .$$

In a manner analogous to the case with no interference, the expectations for the three tetrad types can be expressed as:

$$\mathrm{P(PD)} = p(0) + \sum_{r=2}^{\infty} \frac{(2x)^r}{r!} e^{-2x} S k^{r-1} [p(\mathrm{PD})]$$

$$\mathrm{P(NPD)} = \sum_{r=2}^{\infty} \frac{(2x)^r}{r!} e^{-2x} S k^{r-1} [p(\mathrm{NPD})]$$

$$\mathrm{P(T)} = \sum_{r=1}^{\infty} \frac{(2x)^r}{r!} e^{-2x} S k^{r-1} [p(\mathrm{T})] .$$

These expressions reduce to:

$$\mathrm{P(PD)} = m_1 = e^{-2x} + \frac{1}{6} \frac{(1-e^{-2x})(e^{2kx}+2e^{-kx}-3)}{e^{2kx}-1}$$

$$\mathrm{P(NPD)} = m_2 = \frac{1}{6} \frac{(1-e^{-2x})(e^{2kx}+2e^{-kx}-3)}{e^{2kx}-1}$$

$$\mathrm{P(T)} = m_3 = \frac{2}{3} \frac{(1-e^{-2x})(e^{-kx})(e^{3kx}-1)}{e^{2kx}-1} ,$$

which are the expectations used for estimation.

There are now two ways to proceed. One way is to make use of the fact that when the number of parameters to be estimated is the same as the number of degrees of freedom, then the equations for estimation can be obtained by setting the expectations equal to the observations (BAILEY 1951). In this case we have:

$$m_1 = a_1/N$$
$$m_2 = a_2/N$$
$$m_3 = a_3/N ,$$

where $N$ is the total number of tetrads. Simultaneous solution of any two of these three equations for $x$ and $k$ will give the maximum likelihood estimates.

The second method is the usual one of taking the partial derivatives of the log likelihood expression with respect to $x$ and $k$, and finding values of these param-

eters which make the partial derivatives simultaneously equal to zero. Thus we have:

$$\frac{\partial L}{\partial x} = \frac{a_1}{m_1}\frac{1}{3}\left\{k\alpha\beta + e^{-2x}\left[\frac{2e^{-kx} - 5e^{2kx} + 3}{e^{2kx} - 1}\right]\right\}$$
$$+ \frac{a_2}{m_2}\frac{1}{3}\left\{k\alpha\beta + e^{-2x}\left[\frac{e^{2kx} + 2e^{-kx} - 3}{e^{2kx} - 1}\right]\right\}$$
$$+ \frac{a_3}{m_3}\frac{2}{3}\left\{-k\alpha\beta + 2e^{-2x}\left[\frac{e^{-kx}(e^{3kx} - 1)}{e^{2kx} - 1}\right]\right\} = 0 \ , \qquad (9)$$

$$\frac{\partial L}{\partial k} = \left[\frac{a_1}{m_1} + \frac{a_2}{m_2}\right]\frac{1}{3}x\alpha\beta - \frac{a_3}{m_3}\frac{2}{3}x\alpha\beta = \frac{x\alpha\beta}{3}\left[\frac{a_1}{m_1} + \frac{a_2}{m_2} - \frac{2a_3}{m_3}\right] = 0 \ , \qquad (10)$$

where

$$\alpha = \frac{e^{-kx}(1 - e^{-2x})}{e^{2kx} - 1}, \text{ and } \beta = \frac{2e^{3kx} - 3e^{2kx} + 1}{e^{2kx} - 1} \ .$$

Iterative solution of these simultaneous equations can be accomplished as described by MATHER (1957), using efficient corrections after each round. Standard errors can be obtained from the inverse of the information matrix, using values obtained in the last round of iteration.

Because of the nature of tetrad data and the conditions of the interference model, the estimate of $x$ obtained by (9) and (10) is equal to $-(1/2)\ln p(0)$ (*i.e.*, $p(0) = e^{-2x}$), where $p(0)$ is the proportion of tetrads with no exchanges, that is, PD−NPD/$N$. For example, a Saccharomyces cross involving the markers *mat1* and *his4* on chromosome 3 (see Table 2) produced 97 PD, 7 NPD, and 174 T. The maximum likelihood estimate of $x$ with interference is 0.56390, which is $-(1/2)\ln(90/278)$. ($k$ for this data is 0.29483, while $x$ estimated without interference is 0.58340. $x$ calculated from the PERKINS (1949) formula is 0.38843.) This simple relation can be used instead of (9) and (10) to calculate $x$.

The relation $x = -(1/2)\ln p(0)$ is essentially the formula derived by SHULT and LINDEGREN (1956) for calculating map distances in units they called "stranes." In their notation, the right side of the expression was multiplied by 100 to convert "stranes" into units comparable to centimorgans, *e.g.*, stranes $(D) = -50 \ln (a_1 - a_2/N)$.

The expected proportion of NPD's as a function of T's in two-point crosses without interference was derived by PAPAZIAN (1952):

$$\text{NPD} = (1/2)\left[1 - \text{T} - \left(1 - \frac{3\text{T}}{2}\right)^{2/3}\right] \ .$$

If interference is operative, the proportion of NPD's in the sample will decrease, so that with complete interference there will be no NPD's whatsoever. An estimate of interference might be derived as the ratio of the observed proportion of NPD's to the proportion expected on the basis of the PAPAZIAN (1952) formula (hereafter called the NPD ratio). With no interference, this ratio will be 1,

while with complete interference it will be 0. It turns out that the interference values estimated by equations (9) and (10) are almost the same as the NPD ratios. For the *mat1-his4* data, $k$ is 0.29482, while the NPD ratio is 0.23007. The regression of the NPD ratios against $k$ for 28 Saccharomyces crosses and 30 Schizosaccharomyces crosses gave a coefficient of 1.0657, with 95% confidence limits of 1.0539 to 1.0784. Thus, the $k$ values for two-point crosses obtained by the maximum likelihood equations are essentially measures of the discrepancy of the observed proportion of NPDs compared to the proportion expected on the PAPAZIAN (1952) formula. It is probable that the regression is not exactly one because of the finite sample sizes.

For two-point crosses, interference will increase the proportion of T tetrads at the expense of NPD's. Since nonexchange tetrads contain no information about interference, the information about $k$ comes from the T and NPD tetrads. When there are no NPD's, $k$ cannot be estimated by (9) and (10) because of the division by zero that occurs.

*Ranked data, interference estimated*

As in the case of ranked data without interference, we consider only five terms of the Poisson series. The probability of nonexchange tetrads is again based on the premise that the $p(0)$ term is not changed by interference. Terms above $p(0)$ are multiplied by $Sk^{r-1}$. Thus, in a manner analogous to no interference, the expectations can be written as:

$$p(0) = m_0 = e^{-2x}$$

$$p(1) = m_1 = \frac{2kx(1-e^{-2x})}{e^{2kx}-1}$$

$$p(2) = m_2 = \frac{2k^2x^2(1-e^{-2x})}{e^{2kx}-1}$$

$$p(3) = m_3 = \frac{4}{3}\,\frac{k^3x^3(1-e^{-2x})}{e^{2kx}-1}$$

$$p(4\ldots r) = m_4 = 1-e^{-2x}-\frac{2kx(1-e^{-2x})}{e^{2kx}-1}\left[1+kx+\frac{2}{3}k^2x^2\right].$$

Differentiating the log likelihood expression with respect to $x$ and $k$, and equating the partial derivatives to zero gives the likelihood equations of estimation:

$$\frac{\partial L}{\partial x} = -2a_0 + \frac{a_1}{m_1}2k\left[\frac{2xe^{-2x}-e^{-2x}+1}{e^{2kx}-1} - \varepsilon\right]$$

$$+\frac{a_2}{m_2}2k^2x\left[\frac{2xe^{-2x}-2e^{-2x}+2}{e^{2kx}-1} - \varepsilon\right]$$

$$+\frac{a_3}{m_3}\,\frac{4}{3}k^3x^2\left[\frac{2xe^{-2x}-3e^{-2x}+3}{e^{2kx}-1} - \varepsilon\right]$$

$$+\frac{a_4}{m_4}\left\{2e^{-2x}-\left[\frac{\partial m_1}{\partial x}+\frac{\partial m_2}{\partial x}+\frac{\partial m_3}{\partial x}\right]\right\}=0\ , \tag{11}$$

$$\frac{\partial L}{\partial k}=-\frac{a_1}{m_1}2x(1-e^{-2x})\left[\frac{2kxe^{2kx}-e^{2kx}+1}{(e^{2kx}-1)^2}\right]-\frac{a_2}{m_2}2kx^2(1-e^{-2x})$$
$$\left[\frac{2kxe^{2kx}-2e^{2kx}+2}{(e^{2kx}-1)^2}\right]-\frac{a_3}{m_3}\frac{4}{3}k^2x^3\,(1-e^{-2x})\left[\frac{2kxe^{2kx}-3e^{2kx}+3}{(e^{2kx}-1)}\right]$$
$$-\frac{a_4}{m_4}\left[\frac{\partial m_1}{\partial k}+\frac{\partial m_2}{\partial k}+\frac{\partial m_3}{\partial k}\right]=0\ , \tag{12}$$

where $\varepsilon$ equals

$$\frac{2kxe^{2kx}\,(1-e^{-2x})}{(e^{2kx}-1)^2}\ .$$

As before, these equations are solved simultaneously for $x$ and $k$, and the standard errors are obtained from the calculations used in solution.

With interference, the probability of tetrads of ranks greater than zero is:

$$p(r)=\frac{(2x)^r}{r!}e^{-2x}Sk^{r-1}\ .$$

This relation can be used to calculate the rank distribution after an estimate has been obtained for $x$ and $k$. For the data in Table 1, solution of the above two partial derivatives yielded $x = 0.92622 \pm 0.02515$ and $k = 0.22785 \pm 0.02160$. The expected distribution of ranks using these estimates is given in column 5 of the table, the number of rank 0 tetrads being set at 333.00. Although the calculated distribution fits the observations better, there is a deficiency in rank 3 and an excess in rank 2, so that $\chi^2$, with 3 degrees of freedom, is highly significant ($p < 0.001$). This suggests that an improved interference model might be derived by multiplying terms above $p(0)$ by some factor other than $k^{r-1}$, for instance, by $k^{r^2-1}$.

As in the case of two-point crosses with an estimate of interference, the values of $x$ obtained from the equations for ranked data with interference are equal to $-(1/2)\ln p(0)$ because of the assumptions of the model.

*Estimation of linkage parameters for tetrad data for Saccharomyces and Schizosaccharomyces*

Table 2 contains estimates of map distance with and without interference for some of the extensive two-point data of Saccharomyces for which at least 100 tetrads are available. A Wang model 720C Programmable Calculator was used. Iterations were continued until values of $x$ and $k$ were found that caused the estimators to differ from zero by less than $10^{-5}$. The solutions were confirmed to be those that made the logarithm likelihood expressions a maximum, because slight changes of $x$ or $k$ on either side of the estimated values made the expressions smaller (in this case more negative).

TABLE 2

*Estimates of map distance and interference for selected two-point crosses of* Saccharomyces

| Chromosome | Gene pair | $x(P)$ | $x$ + s.e. | | $x(i)$ + s.e. | | $k$ + s.e. | | Reference |
|---|---|---|---|---|---|---|---|---|---|
| *2* | *cyh1–gal1** | 0.198 | 0.232 | 0.034 | 0.230 | 0.033 | 0.337 | 0.341 | b |
| | *gal1–lys2* | 0.518 | 0.828 | 0.079 | 0.795 | 0.042 | 0.488 | 0.108 | b |
| | *lys2–tyr1* | 0.352 | 0.516 | 0.045 | 0.500 | 0.034 | 0.245 | 0.101 | b |
| | *tyr1–his7* | 0.425 | 0.751 | 0.119 | 0.705 | 0.063 | 0.194 | 0.111 | b |
| | *SUP45–lys2* | 0.259 | 0.331 | 0.053 | 0.327 | 0.048 | 0.258 | 0.260 | c |
| | *SUP45–tyr1* | 0.185 | 0.174 | 0.033 | 0.174 | 0.035 | 1.817 | 1.406 | c |
| *3* | *his4–mat1** | 0.388 | 0.583 | 0.057 | 0.563 | 0.039 | 0.294 | 0.112 | a |
| | *his4–leu2* | 0.175 | 0.199 | 0.016 | 0.198 | 0.016 | 0.370 | 0.216 | b |
| | *leu2–mat1** | 0.345 | 0.453 | 0.033 | 0.446 | 0.027 | 0.429 | 0.127 | b |
| | *mat1–thr4* | 0.214 | 0.249 | 0.020 | 0.248 | 0.020 | 0.404 | 0.205 | b |
| | *thr4–MAL2* | 0.295 | 0.389 | 0.037 | 0.382 | 0.032 | 0.294 | 0.149 | b |
| *4* | *SUP35–aro1* | 0.208 | 0.239 | 0.041 | 0.237 | 0.041 | 0.468 | 0.477 | c |
| | *trp1–cdc2** | 0.460 | 0.878 | 0.171 | 0.809 | 0.075 | 0.204 | 0.116 | d |
| | *asp1–trp4* | 0.180 | 0.209 | 0.026 | 0.208 | 0.026 | 0.284 | 0.287 | d |
| *5* | *his1–trp2* | 0.248 | 0.311 | 0.035 | 0.308 | 0.032 | 0.277 | 0.198 | a |
| | *ura3–hom3** | 0.347 | 0.555 | 0.063 | 0.532 | 0.043 | 0.115 | 0.082 | b |
| *6* | *SUP11–his2* | 0.219 | 0.258 | 0.043 | 0.256 | 0.042 | 0.386 | 0.392 | b |
| *7* | *trp5–ade6** | 0.632 | 0.820 | 0.049 | 0.833 | 0.095 | 1.333 | 0.698 | a |
| | *ade5,7–tyr3* | 0.638 | 1.031 | 0.178 | 1.010 | 0.084 | 0.764 | 0.241 | b |
| | *tyr3–lys5* | 0.086 | 0.080 | 0.016 | 0.080 | 0.017 | 2.355 | 2.464 | b |
| | *cyh2–trp5* | 0.440 | 0.745 | 0.104 | 0.705 | 0.058 | 0.265 | 0.119 | b |
| | *leu1–ade6** | 0.339 | 0.479 | 0.034 | 0.467 | 0.026 | 0.271 | 0.091 | b |
| | *MAL1–ade3* | 0.474 | 0.707 | 0.104 | 0.685 | 0.065 | 0.495 | 0.197 | b |
| *8* | *pet1–CUP1* | 0.456 | 0.794 | 0.095 | 0.747 | 0.049 | 0.271 | 0.096 | a |
| | *thr1–CUP1* | 0.243 | 0.323 | 0.024 | 0.318 | 0.021 | 0.112 | 0.079 | b |
| | *CUP1–pet3* | 0.358 | 0.543 | 0.057 | 0.524 | 0.040 | 0.210 | 0.105 | b |
| *9* | *his6–lys1** | 0.435 | 0.628 | 0.051 | 0.611 | 0.035 | 0.461 | 0.116 | b |
| *10* | *SUP4–SUP7** | 0.522 | 0.834 | 0.039 | 0.801 | 0.063 | 0.498 | 0.159 | c |
| *11* | *met14–met1* | 0.440 | 0.610 | 0.096 | 0.597 | 0.068 | 0.545 | 0.258 | d |
| | *met1–MAL4* | 0.289 | 0.337 | 0.047 | 0.335 | 0.045 | 0.633 | 0.380 | d |
| *15* | *ser1–ade2* | 0.273 | 0.384 | 0.042 | 0.375 | 0.036 | 0.098 | 0.098 | b |
| | *ade2–cyh4* | 0.326 | 0.426 | 0.044 | 0.419 | 0.037 | 0.397 | 0.181 | b |
| | *pet17–ade2* | 0.469 | 0.764 | 0.090 | 0.728 | 0.050 | 0.367 | 0.119 | d |
| *17* | *met2–pha2* | 0.373 | 0.518 | 0.069 | 0.506 | 0.053 | 0.386 | 0.198 | d |
| | *pet2–pha2* | 0.473 | 0.681 | 0.094 | 0.664 | 0.062 | 0.551 | 0.208 | d |

$x(P)$ = map distance calculated from formula of PERKINS (1949); $x$ + s.e. = map distance and standard error estimated from maximum likelihood equation (7); $x(i)$, $k$ = map distance and interference estimated from equations (9) and (10). Map distances are given in Morgans. Intervals spanning a centromere are indicated with an asterisk. In certain cases changes have been made in gene nomenclature to conform to more recent usage.

References: (a) HAWTHORNE and MORTIMER 1960; (b) MORTIMER and HAWTHORNE 1966; (c) HAWTHORNE and MORTIMER 1968; (d) MORTIMER and HAWTHORNE 1973.

Several points emerge from a review of the data in Table 2. First, the PERKINS (1949) formula considerably underestimates the map distance for large genetic intervals because a considerable amount of triple and higher order crossing over is not taken into account. Second, the estimates for $x(i)$ are somewhat smaller than those for $x$, in keeping with expectations of the interference model. Third, except for three cases where $k$ exceeds one, the other estimates of this factor are usually less than 0.5. The average $k$ for all 35 crosses is 0.482; if the three exceptional crosses are omitted, the average is 0.355. This is similar to values of $k$ that have been obtained by plotting (R. K. MORTIMER, personal communication). Fourth, the regression of $k$ on $x(i)$ for this data is $-0.3757$; however, this is not significantly different from zero. Therefore there is no tendency for $k$ to vary in a consistent manner with length of the genetic interval under consideration. Fifth, the standard errors of the estimates of $k$ are quite large. This is due to the fact that the NPD tetrads contribute the most information to his estimate, but there are usually relatively few of them.

The only region of the genome where it seems possible that $k$ values consistently different from the average may occur is in the left arm and the proximal part of the right arm of chromosome 7. The crosses *trp5-ade6*, *ade5,7-tyr3*, and *tyr3-lys5* have $k$ values considerably higher than average, and higher values also occurred in several other crosses for which fewer than 100 tetrads were analyzed. However, not all crosses in this region have high $k$ values, for example, *lys5-cyh2*, *cyh2-trp5*, *and leu1-ade6*. The reduced interference in this region may be real and a more detailed investigation may well prove rewarding.

There is no conspicuous tendency for intervals spanning a centromere to show higher $k$ values than those which do not. The average $k$ for the nine crosses of this type in Table 2 is 0.438. If the *trp5-ade6* cross with the exceptionally high $k$ of 1.333 is omitted, the average $k$ is 0.326.

Ranked tetrad data are much rarer than two-point data. Through the kindness of R. K. MORTIMER and SEYMOUR FOGEL, I was fortunate to obtain access to the immense amount of ranked data from their gene conversion studies (Table 3). The crosses usually involve well over 100 tetrads, often several thousand. The pooled data from chromosome *6*, for instance, are based on 1,076 tetrads, those for chromosome *8* on 14,907.

The map distance for the centromere-*CUP1* interval on the right arm of chromosome *8* is very uniform, except for diploid 5313 (where it is much higher), and diploids 5420, 5420–1, 5420–1–1, and 5475 (where it is lower). The $k$ values for these crosses also vary from the norm in the same direction. The higher $x$ value for 5313 results because rank 2 and 3 tetrads are relatively more frequent than in other crosses, while the lower $x$ values for the other diploids are due to a smaller than usual proportion of rank 2 tetrads. Though different from the rest, the data for the 5420 diploids are quite homogeneous. 5420 is heterozygous *arg4–3*/+, 5420–1 was derived from this strain by a conversion event that produced a homozygous *ARG* strain, and 5420–1–1 was derived from the latter as a single cell isolate.

TABLE 3

*Estimates of map distance and interfersence from ranked data of* Saccharomyces *(data from* MORTIMER *and* FOGEL, *personal communication)*

| Chromosome | Diploid | $x$ + s.e. | | $x(i)$ + s.e. | | $k$ + s.e. | |
|---|---|---|---|---|---|---|---|
| *6* | 4335 | 0.306 | 0.029 | 0.427 | 0.043 | 0.154 | 0.059 |
| | 4338 | 0.302 | 0.028 | 0.390 | 0.039 | 0.285 | 0.086 |
| | 4339 | 0.331 | 0.030 | 0.467 | 0.046 | 0.188 | 0.061 |
| | 4350 | 0.363 | 0.032 | 0.531 | 0.053 | 0.198 | 0.059 |
| | 4351 | 0.326 | 0.031 | 0.479 | 0.049 | 0.119 | 0.049 |
| | 4352 | 0.286 | 0.027 | 0.394 | 0.040 | 0.122 | 0.055 |
| | Pooled | 0.318 | 0.021 | 0.443 | 0.018 | 0.180 | 0.025 |
| *8* | M141 | 0.546 | 0.014 | 1.078 | 0.038 | 0.204 | 0.013 |
| | M150 | 0.543 | 0.013 | 0.993 | 0.031 | 0.242 | 0.014 |
| | 3653 | 0.509 | 0.013 | 0.904 | 0.031 | 0.226 | 0.016 |
| | 5246 | 0.504 | 0.012 | 0.916 | 0.028 | 0.205 | 0.013 |
| | 5275 | 0.533 | 0.015 | 0.979 | 0.037 | 0.229 | 0.017 |
| | 5276 | 0.511 | 0.013 | 0.929 | 0.032 | 0.214 | 0.015 |
| | 5276+ | 0.515 | 0.015 | 0.928 | 0.034 | 0.223 | 0.017 |
| | 5276-3 | 0.520 | 0.015 | 0.942 | 0.035 | 0.224 | 0.017 |
| | 5308 | 0.561 | 0.015 | 1.002 | 0.036 | 0.273 | 0.018 |
| | 5310 | 0.517 | 0.018 | 0.941 | 0.042 | 0.219 | 0.020 |
| | 5313 | 0.847 | 0.023 | 1.354 | 0.066 | 0.494 | 0.030 |
| | 5420 | 0.450 | 0.016 | 0.790 | 0.034 | 0.164 | 0.018 |
| | 5420-1 | 0.426 | 0.014 | 0.699 | 0.027 | 0.181 | 0.019 |
| | 5420-1-1 | 0.453 | 0.021 | 0.761 | 0.043 | 0.198 | 0.027 |
| | 5475 | 0.282 | 0.011 | 0.389 | 0.016 | 0.108 | 0.021 |
| | 5497 | 0.508 | 0.009 | 0.862 | 0.021 | 0.256 | 0.012 |
| | Pooled* | 0.523 | 0.004 | 0.941 | 0.009 | 0.232 | 0.004 |

* Diploids 5313, 5420-1, 5420-1-1, and 5475 omitted.

The intervals marked for chromosome *6* were centromere-*his2-SUP6-met10*; for chromosome *8*, centromere-*arg4-thr1-CUP1*, except for crosses 5420, 5420-1, 5420-1-1, and 5497, which were centromere-*pet1-arg4-thr1-CUP1*. $x$ + s.e. = map distance and standard error estimated from equation (8); $x(i)$ + s.e. and $k$ + s.e. = map distance and interference with standard errors estimated from equations (11) and (12), respectively.

The data for the centromere-*met10* interval on the right arm of chromosome *6* are also quite homogeneous, and the $k$ values are about the same as those for chromosome *8*.

Table 4 contains a selection from the extensive two-point data published by KOHLI, *et al.* (1977) for Schizosaccharomyces. Intervals were chosen that covered nearly all the mapped regions of the three chromosomes and for which usually 100 or more tetrads were available. It is obvious that there is much less interference in this yeast than in Saccharomyces, the average $k$ for the 30 crosses being 1.332. If the two atypical crosses with $k$ values of 4.691 and 7.693 are eliminated, the average $k$ is 0.985. Since, however, 102 and 337 tetrads, respectively, were scored for the two crosses, the high $k$ values may be significant. Although occasional crosses show low $k$ values characteristic of Saccharomyces, for the most part all chromosomes have much less interference. This consistency leaves no doubt that the two yeasts differ fundamentally in this respect.

TABLE 4

*Estimates of map distance and interference for selected two-point crosses of* Schizosaccharomyces pombe *(data from* KOHLI et. al. *1977)*

| Chromosome | Gene pair | $x(P)$ | $x$ + s.e. | | $x(i)$ + s.e. | | $k$ + s.e. | |
|---|---|---|---|---|---|---|---|---|
| *1* | *cyh1–cdc1* | 0.521 | 0.803 | 0.124 | 0.777 | 0.070 | 0.545 | 0.196 |
| | *cdc1–leu2* | 0.500 | 0.688 | 0.105 | 0.677 | 0.071 | 0.698 | 0.277 |
| | *his1–leu2* | 0.142 | 0.114 | 0.025 | 0.115 | 0.028 | 4.691 | 4.101 |
| | *sup3–aro3* | 0.602 | 0.751 | 0.101 | 0.762 | 0.070 | 1.350 | 0.538 |
| | *ura2–ade2* | 0.321 | 0.358 | 0.030 | 0.357 | 0.028 | 0.888 | 0.277 |
| | *ade2–ade4* | 0.703 | 1.220 | 0.194 | 1.205 | 0.078 | 0.869 | 0.234 |
| | *lys3–ura1* | 0.210 | 0.199 | 0.004 | 0.199 | 0.014 | 1.771 | 0.491 |
| | *ura1–lys5* | 0.328 | 0.394 | 0.054 | 0.328 | 0.049 | 0.716 | 0.379 |
| | *pro1–ade3* | 0.435 | 0.495 | 0.078 | 0.497 | 0.068 | 1.087 | 0.530 |
| | *ade3–pro2* | 0.651 | 0.968 | 0.083 | 0.967 | 0.044 | 0.985 | 0.185 |
| *2* | *ade7–ura5* | 0.121 | 0.134 | 0.004 | 0.134 | 0.012 | 0.229 | 0.230 |
| | *ade7–his3* | 0.554 | 0.694 | 0.059 | 0.698 | 0.043 | 1.117 | 0.259 |
| | *glu1–his3* | 0.518 | 0.668 | 0.077 | 0.665 | 0.055 | 0.911 | 0.277 |
| | *his3–mat1** | 0.676 | 1.127 | 0.034 | 1.110 | 0.044 | 0.837 | 0.133 |
| | *tsl24–mat1* | 0.677 | 0.961 | 0.094 | 0.976 | 0.051 | 1.261 | 0.338 |
| | *leu1–his5* | 0.264 | 0.282 | 0.024 | 0.282 | 0.024 | 0.951 | 0.337 |
| | *his5–leu3* | 0.495 | 0.620 | 0.102 | 0.618 | 0.077 | 0.921 | 0.417 |
| | *ade1–his4* | 0.395 | 0.424 | 0.030 | 0.427 | 0.109 | 1.249 | 0.238 |
| | *his4–trp1* | 0.710 | 1.090 | 0.227 | 1.106 | 0.108 | 1.227 | 0.674 |
| | *ade8–arg4* | 0.286 | 0.313 | 0.038 | 0.312 | 0.037 | 0.894 | 0.426 |
| *3* | *ade10–fur1* | 0.161 | 0.175 | 0.009 | 0.175 | 0.029 | 0.581 | 0.591 |
| | *ade10–ade6** | 0.294 | 0.317 | 0.036 | 0.317 | 0.036 | 0.974 | 0.429 |
| | *fur1–sin2** | 0.237 | 0.254 | 0.030 | 0.253 | 0.030 | 0.892 | 0.469 |
| | *fur1–min5** | 0.221 | 0.224 | 0.028 | 0.224 | 0.029 | 1.177 | 0.628 |
| | *ade6–min5* | 0.054 | 0.044 | 0.008 | 0.044 | 0.009 | 7.693 | 6.074 |
| | *tsl5–arg1* | 0.319 | 0.431 | 0.063 | 0.423 | 0.052 | 0.302 | 0.216 |
| | *arg1–ade5* | 0.537 | 0.849 | 0.042 | 0.827 | 0.071 | 0.645 | 0.212 |
| | *arg1–aro4* | 0.552 | 0.563 | 0.110 | 0.580 | 0.092 | 2.523 | 2.655 |
| | *trp3–aro4* | 0.194 | 0.196 | 0.032 | 0.196 | 0.033 | 1.151 | 0.858 |
| | *ade5–wee1* | 0.522 | 0.690 | 0.143 | 0.685 | 0.100 | 0.850 | 0.454 |

Symbols same as in Table 1.

As with *S. cerevisiae*, there is no strong tendency for intervals spanning a centromere to show higher $k$ values than others. A total of 12 such intervals are available in the published literature (four of them are indicated in Table 4). The average $k$ is 1.110.

The question arises as to whether the two species have about the same total amount of recombination per unit of DNA, but differ only in the distribution of

crossovers, or whether the level of recombination differs. This can be approached by calculating the map units per microgram of DNA. Taking the amount of DNA per haploid spore for Saccharomyces and Schizosaccharomyces as $2.1 \times 10^{-8}$ and $1.46 \times 10^{-8}$ $\mu$g, respectively (HARTWELL 1970; BOSTOCK 1970), and the total map units as 3700 and 1300 (FOGEL and MORTIMER 1971; KOHLI *et al.* 1977), one finds about $1.76 \times 10^{11}$ compared to $0.89 \times 10^{11}$ map units per $\mu$g DNA for the two organisms. In view of the fact that the *S. pombe* map will certainly be lengthened by future work, perhaps even doubled, it does not seem likely that the total amount of recombination is considerably different in the two.

MORTIMER and FOGEL (1974) found that in Saccharomyces gene convertants recombined for their flanking markers showed interference in adjacent intervals, but those convertants that retained the parental outside marker combinations did not. They also found that half the conversion events were associated with adjacent marker recombination. Since $k$ values greater than one are a sign of negative interference, parallel gene conversion studies with *S. pombe* should be very profitable.

It is a pleasure to acknowledge the most helpful discussions I have had with SUBODH JAIN, who derived the explicit solution for equation (8), and with MICHAEL TURELLI. Thanks are also due R. K. MORTIMER and SEYMOUR FOGEL for the use of their ranked tetrad data and to DAVID PERKINS and JULIAN ADAMS for their comments on the manuscript. This work was supported by funds assigned to California Agricultural Experiment Station Project CA-D*-GEN-3161-H.

## LITERATURE CITED

BAILEY, N. T. J., 1951 Testing the solubility of maximum likelihood equations in the routine application of scoring methods. Biometrics **7**: 268–274.

BARRATT, R. W., D. NEWMEYER, D. D. PERKINS and L. GARNJOBST, 1954 Map construction in *Neurospora crassa*. Advan. Genet. **6**: 1–93.

BOSTOCK, C. J., 1970 DNA synthesis in the fission yeast *Schizosaccharomyces pombe*. Exp. Cell Res. **60**: 12–26.

FOGEL, S. and R. K. MORTIMER, 1971 Recombination in yeast. Ann. Rev. Genet. **5**: 219–236.

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

HARTWELL, L. H., 1970 Biochemical genetics of yeast. Ann. Rev. Genet. **4**: 373–396.

HAWTHORNE, D. C. and R. K. MORTIMER, 1960 Chromosome mapping in Saccharomyces: Centromere-linked genes. Genetics **45**: 1085–1110. ——, 1968 Genetic mapping of nonsense suppressors in yeast. Genetics **60**: 735–742.

KOHLI, J., H. HOTTINGER, P. MUNZ, A. STRAUSS and P. THURIAUX, 1977 Genetic mapping in *Schizosaccharomyces pombe* by mitotic and meiotic analysis and induced haploidization. Genetics **87**: 471–489.

KUENEN, R., 1962 Ein modell zur analyse der crossover-interferenz. Zeit. Vererbungs. **93**: 35–65.

MATHER, K., 1935 Reduction and equational separation of the chromosomes in bivalents and multivalents. J. Genet. **30**: 53–78. ——, 1957 *The Measurement of Linkage in Heredity*, second edition. Methuen and Co., London.

MATHER, K. and G. H. BEALE, 1942 The calculation and precision of linkage values from tetrad analysis. J. Genet. **43**: 1–30.

MORTIMER, R. K. and S. FOGEL, 1974 Genetical interference and gene conversion. pp. 263–275. In: *Mechanisms in Recombination.* Edited by R. F. GRELL. Plenum Press, New York.

MORTIMER, R. K. and D. C. HAWTHORNE, 1966 Genetic mapping in Saccharomyces. Genetics **53:** 165–173. ——, 1973 Genetic mapping in Saccharomyces. IV. Mapping of temperature sensitive genes and use of disomic strains in localizing genes. Genetics **74:** 33–54.

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

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

SHULT, E. E. and C. C. LINDEGREN, 1956 A general theory of crossing-over. J. Genet. **54:** 343–357.

Corresponding editor: F. SHERMAN