Genet. Res., Camb. (1985), **45**, *pp.* 279–286
With 2 text-figures
Printed in Great Britain

Tests of hypotheses on recombination frequencies

By N. E. MORTON, C. J. MacLEAN and R. LEW

Population Genetics Laboratory, University of Hawaii, Honolulu, HI 96822

(*Received 12 June 1984 and in revised form 21 January 1985*)

SUMMARY

Data from *Neurospora, Drosophila*, and the mouse support the mapping parameter conventionally used for man, exclude the Haldane, Kosambi, and Carter–Falconer functions, and suggest a refinement for centromere mapping. Different sexes, chromosome arms, and types of data are surprisingly consistent. Double recombination frequencies are accurately predicted, but triple recombination frequencies are overestimated. The centromere appears to act on interference as an obligatory chiasma. Recombination across the centromere conforms to a simple approximation, based on the interval Markov assumption with a common mapping parameter. These results imply that mapping of n loci requires estimation of at most n parameters, and the relation between map distances and all recombination frequencies is explicit.

1. INTRODUCTION

Recombination values so small that multiple exchanges are negligible correspond to map distances. Generalization to larger recombination values requires a mapping function, characterized in the simplest case by a single parameter. If the mapping parameter is so variable that it must be estimated for each small set of loci the additivity of the map is destroyed, precision of estimation is lost, the order of loci may be in doubt, and the error frequency when flanking markers are used in genetic counselling must be determined empirically. On the contrary, if the mapping parameter is a known constant, multiple recombination frequencies may be predicted and genetic maps constructed with ease and accuracy. Information relevant to man is provided by experimental data and derivative mathematics.

In the 1950s the Cambridge school developed a theory of multiple recombination frequencies based on crossing-over beginning at the centromere and leading to the possibility of recombination values in excess of 1/2 (Owen, 1950). After these cytogenetic postulates were discredited, neglect of chiasma interference was tried and found wanting (Risch & Lange, 1983). To fill this vacuum we proposed a simple theory based on the interval Markov assumption that there is no interference across an interval in which an exchange occurs (Morton & MacLean, 1984). Extensions were made to tetrads and recombination across the centromere. None of these predictions has been tested in experimental data – a necessary first step to use of multiple recombination frequencies for gene mapping and genetic counselling in man.

2. MATERIALS AND RESULTS

Here we analyse available evidence to answer four questions.

(1) Are different organisms, sexes, chromosome arms, and types of data consistent with the same mapping parameter?

(2) Are multipoint recombination frequencies within a chromosome arm consistent with the interval Markov assumption, conditional on a single mapping parameter?

(3) Does the centromere act on interference as an obligatory chiasma?

(4) Does recombination across the centromere conform to a simple approximation, based on the interval Markov assumption with a common mapping parameter?

Favourable answers to these questions in experimental organisms provide a basis for applications to man.

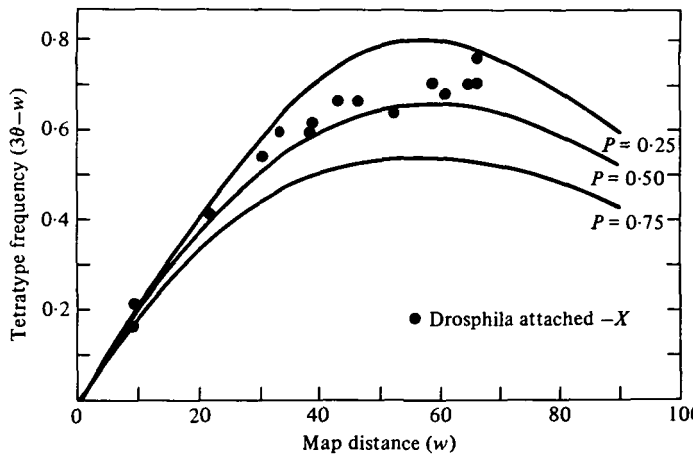

Fig. 1. Tetratype frequency by map distance.

(i) *Attached*-X chromosomes in Drosophila

Assuming no chromatid interference and at most two exchanges between the centromere and a heterozygous locus, the probability of homozygosity for a particular allele is $(3\theta-w)/4$, where θ is the recombination fraction and w the centromere distance in morgans (Morton & MacLean, 1984, eq. 12). We use the Rao mapping function $w(\theta, p)$ where p is the mapping parameter (Rao *et al.* 1977):

$$w(\theta, p) = \{p(2p-1)\,(1-4p)\ln(1-2\theta)+16p(p-1)\,(2p-1)\tan^{-1}(2\theta) + 2p(1-p)\,(8p+2)\tanh^{-1}(2\theta)+6(1-p)\,(1-2p)\,(1-4p)\theta\}/6.$$

Table 1 gives published data on homozygosis for attached-X chromosomes in Drosophila. To estimate p, map distance w is taken as the standard for the X chromosome (Lindsley & Grell, 1968). Conditional on w, the likelihood ratio χ^2 tests goodness of fit for p. If a quadratic is fitted to n samples, $\chi^2 = a+bp+cp^2$, the maximum-likelihood estimator if the likelihood is $\exp(-\chi^2/2\sigma^2)$ is

$$p = -\frac{b}{2c} \pm \sqrt{\frac{\chi^2_{\min}}{(n-1)\,c}} = 0{\cdot}396 \pm 0{\cdot}018,$$

https://doi.org/10.1017/S0016672300022266 Published online by Cambridge University Press

in reasonable agreement with 0·351 from the chiasma distribution in human spermatogenesis (Rao *et al.* 1977) and intermediate between the values of 0·5 suggested for *Drosophila* (Kosambi, 1944) and 0·25 for the mouse (Carter & Falconer, 1951). The estimate of p is subject not only to errors of sampling, but also to differential viability of the markers and deviations from the standard map. These sources of variation are reflected in minimum $\chi^2_{14} = 199{\cdot}74$. Nevertheless, under our assumptions it seems that p for attached-X chromosomes in *Drosophila* is closer to the Kosambi than to the Carter–Falconer value (Fig. 1).

Table 1. *Reduction of attached-X chromosomes in* Drosophila

Gene	Homozygotes	Heterozygotes	Homozygosis (%)	Map distance w (cM)
y	8855	37766	19·0	66·0
sc	12165	56774	17·6	66·0
w	6883	32463	17·5	64·5
ec	11148	54324	17·0	60·5
rb	6799	31710	17·6	58·5
cv	4883	25132	16·3	52·3
ct	5255	26388	16·6	46·0
ptg	816	4098	16·6	42·8
t	698	3846	15·4	38·5
lz	124	709	14·9	38·3
v	4266	24271	14·9	33·0
m	137	876	13·5	29·9
g	3428	29453	10·4	21·6
f	1937	34609	5·3	9·3
B	21	493	4·1	9·0

These data came from Sturtevant (1931), Bonnier & Nordenskiold (1937), Beadle & Emerson (1935) and Welshons (1955). We omitted Emerson & Beadle (1933) and proximal markers of Beadle & Emerson (1935) because of reduced crossing-over.

(ii) *Second division segregation in* Neurospora

Ordered tetrads may be classified as first or second division segregation for each heterozygous locus. On the assumption of no chromatid interference and at most two exchanges between a locus and the centromere, the frequency of second division segregation is $3\theta - w$ (Morton & MacLean, 1984). Barratt *et al.* (1954) in their tables 4–10 summarized data on 73 noncentromeric loci in *N. crassa*. Centromere map distance was taken as the value corrected according to their mapping function. For the ascending part of the curve their parameter $k = 0{\cdot}2$ corresponds closely to Rao's parameter $p = 0{\cdot}4$. Agreement with attached-X chromosomes in *Drosophila* is remarkably good (Table 2).

(iii) *Unordered tetrads in* Neurospora

Unordered tetrads may be classified for two segregating loci as parental ditype (PD), tetratype (T), or non-parental ditype (NPD). Expected frequencies on the assumption of no chromatid interference and at most two exchanges (Morton & MacLean, 1984) satisfy the formula of Perkins (1949):

$$w = (\mathrm{T} + 6\mathrm{NPD})/2(\mathrm{PD} + \mathrm{T} + \mathrm{NPD}) \text{ morgans.}$$

However, this is less efficient than the maximum-likelihood estimate. Perkins (1962) and Bole-Gowda *et al.* (1962) reported large samples of paracentric 2-point data in *Neurospora*. With w given by maximum likelihood, the simultaneous estimate of p is 0·434. Minimum χ^2 is only 84 % as great as for the Perkins estimate.

Table 2. *Estimates of p for paracentric intervals*

Source	p	σ_p	n
(1) *Drosophila* attached-*X*	0·396	0·018	15
(2) *Neurospora* division II	0·397	0·006	73
(3) Unordered tetrads	0·434	0·031	18
(4) *Drosophila X* 3-point	0·338	0·010	84
(5) Male mouse	0·350	0·047	31
(6) Female mouse	0·346	0·034	29
(7) *Drosophila* III	0·293	0·036	24
Pooled 1–3	0·398	0·006	106
Pooled 4–7	0·341	0·008	168

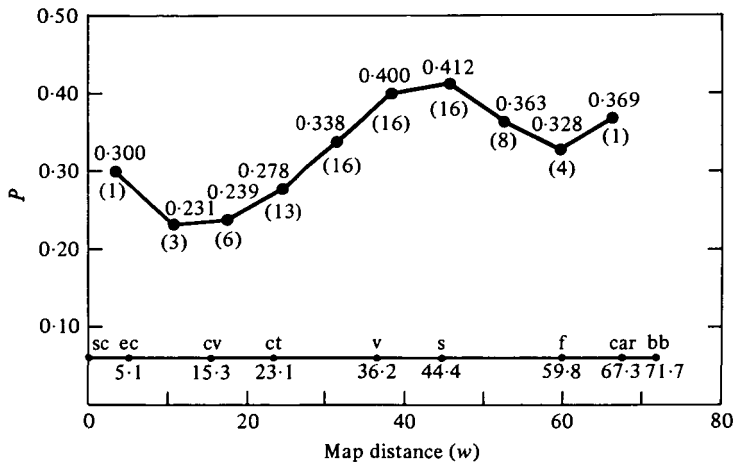

Fig. 2. Mapping parameter p along X chromosome of *D. melanogaster* (number of tests in parentheses).

(iv) Drosophila X *chromosome*

Morgan, Bridges & Schultz (1935) reported on crossing over among nine loci in 16136 gametes. Map length was 71·7, compared within the standard of 66·0. We generated all $\binom{9}{3}$ 3-point tests. Estimates of p were low terminally, rising to a maximum medially (Fig. 2). Allowing for this trend, the highest estimates were for small intervals. The multiple correlation coefficient was 0·84 for four highly significant independent variables (interval and a cubic of the mean map location). The standard deviation from regression was 0·033. Corrected for this attenuation, the standard deviation was 0·058.

In the $\binom{9}{4}$ 4-point tests, the mean value of the mapping parameter estimated simultaneously with map distances was 0·334. The multiple regression was similar to 3-point tests. The standard deviation from regression was 0·022. Corrected for

https://doi.org/10.1017/S0016672300022266 Published online by Cambridge University Press

this attenuation, the standard deviation was 0·032. The regression of χ^2 on map location was parabolic, with an estimated maximum at $w = 27{\cdot}1$. In this material only 61 triple exchanges were detected. The interval Markov assumption predicted 114·6 triple exchanges, which is clearly too high. However, the absolute error is only $(114{\cdot}6–61)/16136 = 0{\cdot}003$. Accurate prediction is more important for double recombination frequencies than for rare triple recombinants.

(v) *Other paracentric data*

Robinson (1972) gives references to 3-point and 4-point data on the mouse. Our preliminary analysis identified two matings with an unusual number of apparent double crossovers. Both involved pintail (Pt). Lane (1963) wrote: 'It is possible that there may be negative interference in this region of the linkage map, but a more likely explanation is misclassification for Pt. At the time the animals were classified the order of the loci was not known. The double recombinants were not recognized as unusual and were discarded without progeny testing.' Omitting these data, we obtained the results in Table 2. Although standard errors are large for the mouse, the estimates are consistent with *Drosophila*.

As a check on the large *Drosophila* X material, we abstracted paracentric 3-point and 4-point crosses from the magnificent presentation of Bridges & Morgan (1923), of which Wallace (1975) remarked: 'A scrutiny of the published *Drosophila* three-points reveals a degree of heterogeneity which renders impossible an accurate interpretation of the operation of interference. In some cases there is heterogeneity of output between heterozygous types, in others between matings of the same type; and the source of heterogeneity is sometimes found to be due to (possibly real) fluctuation in recombination values, and sometimes cannot be traced at all, due to inadequate presentation of the data. Perhaps the most surprising feature is that the balanced design has not been used extensively.' Despite such inevitable reservations about the shoulders on which we stand, there is good agreement with other evidence. The mapping parameter reveals a pattern that eludes casual inspection or estimates of unstable parameters like coincidence.

(vi) *Pericentric intervals*

Many of the intervals on chromosome III of *Drosophila melanogaster* are pericentric (Bridges & Morgan, 1923). To these we applied the simple approximation of Morton & MacLean (1984). Allowance for the centromere halves χ^2 and reduces p to values in fair agreement with paracentric data (Table 3). When the centromere is neglected, the estimate of p for 3-point data approaches unity (no interference). The estimate of p for 4-point data, neglecting the centromere, corresponds roughly to the weighted mean for two paracentric and one pericentric segment,

$$(2/3)\ (0{\cdot}341) + (1/3) = 0{\cdot}561.$$

All pericentric data were reduced to 2-point tests and assigned the standard map distance (Lindsley & Grell, 1968). For 36 pairs of loci, the estimate of p was $0{\cdot}642 \pm 0{\cdot}043$, but this is an average over different centromere locations. The data suggest a function like

$$p = 0{\cdot}341 + 1{\cdot}318\ K,$$

where $0 \leqslant K \leqslant 0{\cdot}5$ is the relative position of the centromere. For the centromere at the midpoint, $K = 0{\cdot}5$ and $p = 1$, corresponding to no interference. For the centromere close to one locus, $K = 0$ and $p = 0{\cdot}341$, corresponding to paracentric intervals.

Table 3. *Estimates of p for pericentric intervals*

Source	p	σ_p	n
3-pt as paracentric	1·000	0·060	17
3-pt as pericentric	0·513	0·207	17
4-pt as paracentric	0·465	0·032	76
4-pt as pericentric	0·292	0·022	76

Table 4. *Lod scores for centromere mapping*

	$g(\theta) = 3\theta - w$, $p = 0{\cdot}398$			$g(\theta) = (2/3)\,[\sin(3/2)\sin^{-1}(2\theta)]$		
θ	$g(\theta)$	Z_N	Z_R	$g(\theta)$	Z_N	Z_R
0·001	0·002	−2·5229	0·4763	0·002	−2·5229	0·4763
0·05	0·100	−0·8241	0·4314	0·100	−0·8248	0·4315
0·10	0·199	−0·5242	0·3805	0·198	−0·5266	0·3811
0·20	0·394	−0·2288	0·2599	0·386	−0·2375	0·2654
0·30	0·572	−0·0667	0·1088	0·548	−0·0850	0·1321
0·40	0·699	0·0207	−0·0446	0·656	−0·0071	0·0138

Each case of reduction is scored Z_R, and each case of non-reduction is scored Z_N.

DISCUSSION

From this analysis it appears that the mapping parameter p is remarkably stable for different organisms, sexes, chromosome arms, and types of data. The only significant difference for paracentric intervals is between centromere distances ($p = 0{\cdot}398$) and other loci ($p = 0{\cdot}341$), for which we do not have a convincing explanation. A possible reason for this discrepancy is that absence of chromatid interference is assumed for centromere distances. An excess of two-strand doubles reduces interference and has been reported for small map distances (Emerson, 1963; cf. Perkins, 1962). The assumption of at most two crossovers in an interval is not critical because of the rarity of higher order recombinants. Interference may be reduced near the centromere, compared to the distal region in which loci cluster (Fig. 2). If centromere mapping becomes important in man, the practical solution is to use lod scores for $p = 0{\cdot}398$ (Table 4). These agree closely with the trigonometric formula of Ott *et al.* (1976) for recombination values up to 0·25.

There is extraordinarily close agreement between the p value of 0·351 from chiasma distributions in human spermatogenesis (Rao *et al.* 1977) and the present estimate of 0·341. Conventional use of 0·351 for both sexes in man appears well justified (Keats *et al.* 1979; Sherman *et al.* 1984). The observed variation within a chromosome arm is clearly significant, but the magnitude is small and its pattern is not understood, although in this sample there is a tendency for p to be small (and therefore interference greater) in regions of high chiasma frequency (Fig. 2).

Undoubtedly the theory for multiple recombination frequencies could be refined, but a critical test of alternative hypotheses requires fastidious control over

viability, penetrance, misclassification, age, environment, and residual genotype. Until such experiments satisfying tests of homogeneity and goodness of fit are performed, the present theory may prove adequate.

The interval Markov assumption clearly overestimates triple crossovers, but the absolute error for this rare class is small. Moreover, it is unlikely that analysis of 4-point and higher order recombination will be of much consequence for gene mapping and genetic counselling in man. The interval Markov assumption is not invoked for paracentric 2-point and 3-point frequencies.

The centromere appears to act on interference as an obligatory chiasma. A simple approximation based on the interval Markov assumption serves rather well for pericentric intervals. Ideally centromeric markers will be identified for every chromosome in man, so that pericentric mapping will be unnecessary. Meanwhile, the assumption of reduced interference ($p = 0{\cdot}642$) is appropriate for pericentric intervals, although the discrepancy from $p = 0{\cdot}351$ is not appreciable unless the interval is large.

On this evidence we propose to treat the mapping parameter in man as a known constant, approximately 0·398 for centromere mapping and 0·351 otherwise. Information to refine these values can be presented as lod scores for p, which may be tested for heterogeneity and optionally pooled over different chromosome regions and both sexes. Coincidences and other nuisance parameters need not (and for efficiency should not) be estimated except as functions of map distances and the mapping parameter. The practical consequence of this result is that n-point linkage is described rather well by only n parameters, instead of the $2^{n-1}-1$ that would be required if coincidences and other nuisance parameters were estimated. Any attempt to estimate many parameters in a reasonable sample size risks zero estimates of nonzero probabilities and has no defined relation to map distances.

PGL No. 301. This work was supported by grant GM 17173 from the U.S. National Institutes of Health.

REFERENCES

BARRATT, R. W., NEWMEYER, D., PERKINS, D. D. & GARNJOBST, L. (1954). Map construction in *Neurospora crassa*. *Advances in Genetics* **6**, 1–93.

BEADLE, G. W. & EMERSON, S. (1935). Further studies of crossing over in attached-X chromosomes of *Drosophila melanogaster*. *Genetics* **20**, 192–206.

BOLE-GOWDA, B. N., PERKINS, D. D. & STRICKLAND, W. N. (1962). Crossingover and interference in the centromere region of linkage group I of *Neurospora*. *Genetics* **47**, 1243–1252.

BONNIER, G. & NORDENSKIOLD, M. (1937). Studies in *Drosophila* with attached-*X*'s. I. Crossingover values, frequencies of reciprocal and nonreciprocal exchange, chromatid interference. *Hereditas* **23**, 257–278.

BRIDGES, C. B. & MORGAN, T. H. (1923). The third chromosome group of mutant characters of *Drosophila melanogaster*. *Carnegie Institute of Washington Yearbook* **327**, 251.

CARTER, T. C. & FALCONER, D. S. (1951). Stocks for detecting linkage in the mouse and the theory of their design. *Journal of Genetics* **50**, 307–373.

EMERSON, S. (1963). Meiotic recombination in fungi with special reference to tetrad analysis. *Methodology in Basic Genetics* (ed. W. J. Burdette), pp. 167–208. San Francisco: Holden-Day.

EMERSON, S. & BEADLE, G. W. (1933). Crossing over near the spindle fiber in attached-X chromosomes of *Drosophila melanogaster*. *Zeitschrift für induktive Abstammungs und Vererbungslehre* **65**, 129–140.

HALDANE, J. B. S. (1919). The combination of linkage values and the calculation of distances between the loci of linked factors. *Journal of Genetics* **8**, 299–309.

KEATS, B. J. B., MORTON, N. E., RAO, D. C. & WILLIAMS, W. R. (1979). *A Source Book for Linkage in Man*. Baltimore: Johns Hopkins University Press.

KOSAMBI, D. D. (1944). The estimation of map distances from recombination values. *Annals of Eugenics (London)* **12**, 172–175.

LANE, P. W. (1963). Whirler mice, a recessive behavior mutation in linkage group VII. *Journal of Heredity* **54**, 263–266.

LINDSLEY, D. L. & GRELL, E. H. (1968). Genetic variations of *Drosophila melanogaster*. *Carnegie Institute of Washington Yearbook* **627**, 471.

MORGAN, T. H. & BRIDGES, C. B. (1916). Sex-linked inheritance in *Drosophila*. *Carnegie Institute of Washington Yearbook* **237**, 87.

MORGAN, T. H., BRIDGES, C. B. & SCHULTZ, J. (1935). Constitution of the germinal material in relation to heredity. *Carnegie Institute of Washington Yearbook* **34**, 284–291.

MORTON, N. E. & MACLEAN, C. J. (1984). Multilocus recombination frequencies. *Genetical Research* **44**, 99–108.

OTT, J., LINDER, D., MCCAW, B. K., LOVRIEN, E. W. & HECHT, F. (1976). Estimating distances from the centromere by means of benign ovarian teratomas in man. *Annals of Human Genetics, London* **40**, 191–196.

OWEN, A. R. G. (1950). The theory of genetical recombination. *Advances in Genetics* **3**, 117–157.

PERKINS, D. D. (1962). Crossingover and interference in a multiply marked chromosome arm of *Neurospora*. *Genetics* **47**, 1253–1274.

RAO, D. C., MORTON, N. E., LINDSTEN, J., HULTEN, M. & YEE, S. (1977). A mapping function for man. *Human Heredity* **27**, 99–104.

RISCH, N. & LANGE, K. (1983). Statistical analysis of multilocus recombination. *Biometrics* **39**, 949–963.

ROBINSON, R. (1972). *Gene Mapping in Laboratory Mammals, part B*. London: Plenum Press.

SHERMAN, S. L., KING, J., ROBSON, E. B. & YEE, S. (1984). A revised map of chromosome 1. *Annals of Human Genetics* **48**, 243–251.

STURTEVANT, A. H. (1931). Two new attached-*X* lines of *Drosophila melanogaster* and further data on the behavior of heterozygous attached-*X*'s. *Carnegie Institute of Washington Yearbook* **421**, 61–81.