Genet. Res., Camb. (1984), **44**, *pp.* 99–108
Printed in Great Britain

Multilocus recombination frequencies

By N. E. MORTON and C. J. MacLEAN

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

(*Received 26 October 1983 and in revised form 6 February 1984*)

SUMMARY

Multilocus recombination frequencies are expressed in terms of pairwise recombination frequencies for paracentric and pericentric intervals. Applications of this theory to mapping and genetic counselling are discussed.

1. PAIRS OF LOCI

Nearly all the present genetic information on the human linkage map has been derived from pairs of loci. If A, B are such a pair, we suppose a mapping function

$$w = f(\theta, p) \tag{1}$$

where θ is the recombination frequency and w is the mean number of exchanges in the segment AB, and p is a mapping parameter. The scalar w is called the map distance. In practice, θ differs between the sexes. In principle, different values of p might be required for different intervals or more than one mapping parameter might be required for a given interval. However, the existence of such variation has not been demonstrated.

The first mapping function, valid only for small distances, was $w = \theta$ (Sturtevant, 1914). This was refined by Haldane (1919) who showed that in the absence of interference,

$$w = (-1/2) \ln (1-2\theta). \tag{2}$$

He considered a compromise of the type

$$w = (1-p)\,\theta + p(-1/2) \ln (1-2\theta) \tag{3}$$

and suggested $p = 0{\cdot}3$ by inspection of Drosophila data. Other mapping functions have been proposed but seldom applied (reviewed by Karlin, 1984). Rao *et al.* (1977) generalized to

$$\begin{aligned} w = \{&p(2p-1)(1-4p) \ln (1-2\theta) + 16p(p-1)(2p-1) \tan^{-1}(2\theta) \\ &+ 2p(1-p)(8p+2) \tanh^{-1}(2\theta) + 6(1-p)(1-2p)(1-4p)\theta\}/6 \end{aligned} \tag{4}$$

which gives the most familiar special cases:

complete coincidence, $p = 1$ (Haldane, 1919)
moderate coincidence, $p = 1/2$ (Kosambi, 1944)
low coincidence, $p = 1/4$ (Carter & Falconer, 1951)
zero coincidence (complete interference), $p = 0$.

PGL Paper No. 298. This work was supported by grant GM 17173 from the U.S. National Institutes of Health.

Without allowing for chiasma movement or errors of localization, chiasma data of Hulten (1974) gave $p = 0{\cdot}351$ for the average chromosome in spermatogenesis (Morton *et al.* 1977). This value has been used to map human chromosomes (Rao *et al.* 1977; Keats *et al.* 1979; Sherman *et al.* 1984).

Mapping functions are not based on detailed understanding of recombination, and so they do not permit inference of 4-strand exchanges without other

Table 1. *Expectations in a 3-point cross*

Segments showing recombination	Event	Representative gamete	Probability
None	E_0	ABC	$1-\theta_1-\theta_2+\theta_{12}$
1	E_1	Abc	$\theta_1-\theta_{12}$
2	E_2	ABc	$\theta_2-\theta_{12}$
12	E_{12}	AbC	θ_{12}

assumptions. Their utility is to construct a genetic map for which loci are correctly ordered and distances are additive, and to predict recombination frequencies that may not have been observed. More theoretical approaches to mapping functions do not reflect known facts about interference and have not given an acceptable fit to genetic data (Sturt, 1976; Risch & Lange, 1983).

2. THREE PARACENTRIC LOCI

Three ordered loci A, B, C specify two segments, AB and BC, which we shall designate by 1 and 2, respectively. In a 3-point cross the genetic parameters are the recombination rates in segments 1 and 2 (say θ_1, θ_2) and the frequency of double recombination (θ_{12}). This situation can be reparametrized in various ways. Coincidence is defined as

$$C = \theta_{12}/\theta_1\theta_2 \tag{5}$$

where conventional (positive) interference implies $0 \leqslant C < 1$. There is experimental evidence that coincidence is greater across the centromere than within an arm. Here we shall assume that A, B, C are paracentric (i.e., lie in the same arm).

We need to express θ_{12} in terms of w_1, w_2, and p. To do this, note in Table 1 that the probability of recombination between the flanking markers is the union of events E_1 and E_2, or

$$\theta_{1+2} = \theta_1+\theta_2-2\theta_{12}.$$

Rearrangement gives the familiar result

$$\theta_{12} = (\theta_1+\theta_2-\theta_{1+2})/2 \tag{6}$$

where $\theta_{1+2} = \theta(w_1+w_2, p)$. It is possible to estimate θ_{12} from 2-point data. However, 3-point data are obviously more efficient since the event E_{12} is observed directly. Combination of 2-point and 3-point data is advantageous.

3. FOUR PARACENTRIC LOCI

Robinson (1971) remarks 'At the present stage of theory, there is little information to be obtained from the four-point and higher crosses which cannot be gained from the three-point'. This is true *a fortiori* in man, where the calculations for multipoint mapping are heavy. However, the 4-point cross introduces some complications that baffled earlier workers, have now been solved, and are required to specify multipoint probabilities.

Ordered loci A, B, C, D define 3 segments which we have assumed to lie in the same chromosome arm. The 8 classes in a 4-point cross are functions of parameters w_1, w_2, w_3, and p, where the mapping parameter is assumed constant over the interval. All the probabilities in Table 2 except θ_{13} and θ_{123} can be expressed as in the last section. However, θ_{13} is the probability of recombination in segments 1 and 3, whether or not there is recombination in segment 2. A mapping function does not specify such probabilities without an ancillary assumption.

We suppose that coincidence is determined by the nearest recombinant and treat the segment in which it occurs as a geometric point. Then the conditional probability of a recombinant in segment 3, given a recombinant in segments 1 and 2 is

$$p(3 \mid 12) = p(3 \mid 2).$$

By definition,

$$p(3 \mid 12) = \theta_{123}/\theta_{12},$$
$$p(3 \mid 2) = \theta_{23}/\theta_{2}.$$

Therefore,

$$\theta_{123} = \theta_{12}\theta_{23}/\theta_{2} \tag{7}$$

which permits us to evaluate the union of events E_{12} and E_{13} as

$$\theta_{1(2+3)} = \theta_{12} + \theta_{13} - 2\theta_{123}$$

and so

$$\theta_{13} = \theta_{1(2+3)} - \theta_{12} + 2\theta_{123} \tag{8}$$

where by equation 6

$$\theta_{1(2+3)} = (\theta_{1} + \theta_{2+3} - \theta_{1+2+3})/2.$$

The Markov property invoked for θ_{123} is exact if w_2 is a geometric point and also at the opposite extreme, where w_2 is so great that $\theta_{12} = \theta_1\theta_2$, $\theta_{23} = \theta_2\theta_3$, and so $\theta_{123} = \theta_1\theta_2\theta_3$. For intermediate values of w_2 there is some approximation.

If higher accuracy is desired, the approach of Haldane (1919) may be useful. He defined marginal coincidence as

$$C_m = \lim_{w_2 \to 0} C = (1 + \partial\theta/\partial w)/2.$$

By integration over w_2 in the 3-point case, extension of the Markov property from a point to an interval may be avoided. With realistic mapping functions, integration must be numerical. For simplicity and for the following reasons, the

Markov assumption seems preferable. First, triple and higher order recombinants are rare: Morgan, Bridges & Schultz (1935) encountered only 61 triples and no higher order recombinants among 16,136 Drosophila *X* chromosomes. Therefore a solution that is accurate for double recombinants but approximate for higher order exchanges is satisfactory in practice. Secondly, most information in multipoint

Table 2. *Expectations in a 4-point cross*

Segments showing recombination	Event	Representative gamete	Probability
None	E_0	ABCD	$1-\theta_1-\theta_2-\theta_3+\theta_{12}+\theta_{13}+\theta_{23}-\theta_{123}$
1	E_1	Abcd	$\theta_1-\theta_{12}-\theta_{13}+\theta_{123}$
2	E_2	ABcd	$\theta_2-\theta_{12}-\theta_{23}+\theta_{123}$
3	E_3	ABCd	$\theta_3-\theta_{13}-\theta_{23}+\theta_{123}$
12	E_{12}	AbCD	$\theta_{12}-\theta_{123}$
13	E_{13}	AbcD	$\theta_{13}-\theta_{123}$
23	E_{23}	ABcD	$\theta_{23}-\theta_{123}$
123	E_{123}	AbCd	θ_{123}

data, and all applications to genetic counselling, come from trios of loci, where triple recombinants cannot be observed. Thirdly, no more accurate expression for multiple recombinants has so far been demonstrated. Finally, paternity errors, incomplete penetrance, viability effects, and other disturbances are likely to have greater effects on the apparent frequency of multiple recombinants than any second order correction of the expected frequency.

4. FIVE PARACENTRIC LOCI

These principles readily extend to more complex events. For example, in the 5-point cross.

$$\begin{aligned}\theta_{1234} &= \theta_{123}\theta_{234}/\theta_{23} \\ &= \theta_{12}\theta_{23}\theta_{34}/\theta_2\theta_3.\end{aligned} \tag{9}$$

In general the probability of every multiple recombinant can be expressed as a function of 2-point probabilities, and therefore of p and the w_i, without nuisance parameters. Small probabilities should not be assumed zero, else even a single recombinant or classification error could appear to rule out the true order of loci.

5. PERICENTRIC LOCI

Genetic data indicate that interference is reduced across the centromere, but cytological observations suggest that coincidence varies among organisms and is not complete (Sybenga, 1975). For simplicity we assume that the centromere acts as an obligatory chiasma, conferring the Markov property that crossing over is independent in different arms of the same chromosome. Thus if segments 1 and 3 are pericentric (i.e., on different chromosome arms), then $\theta_{13} = \theta_1\theta_3$. If the

centromere is included in one segment, say 1, dividing it so that a proportion k is pericentric from segment 2, then we tentatively assume from equation 6 that

$$\theta_{12} = k\theta_1\theta_2 + (1-k)(\theta_1+\theta_2-\theta_{1+2})/2$$

and that the Markov property holds for $\theta_1 = \theta(w_1)$ and higher order exchanges as in equation 7, with the same mapping parameter as for paracentric intervals. These rough expectations could be refined, but there is little point is doing so until agreement with data has been sought. Pericentric recombination frequencies merit more consideration than they have been given, although maps should ideally be constructed from paracentric intervals.

A more exact treatment on the assumption that the centromere acts as an obligatory exchange replaces $\theta(w_1)$ for pericentric loci by

$$\theta_1 = \theta(kw) + \theta(k-kw) - 2\theta(kw)\,\theta(k-kw)$$

with corresponding elaboration of multiple recombination probabilities. We hesitate to follow this approach, which alters mapping from pairs of loci. So far there is no compelling evidence against simplicity.

6. CENTROMERE MAPPING

Centromeric heteromorphisms occur on human chromosomes 1, 3, 4, 9, 16, and the acrocentrics, but for each chromosome the frequency of easily recognized variants is low. Genetic markers near the centromere are known for some chromosomes, and DNA polymorphisms will provide reliable markers for all centromeres. Here we explore the theory and application of centromere mapping from unordered and partial tetrads.

It is not feasible to determine centromere distance in this way for loci far from the centromere. Therefore we consider a region short enough so that the number of chromatid exchanges in meiosis does not exceed 2. Let c_i be the frequency of i exchanges ($i = 0, 1, 2$), and let p_j be the conditional probability that a double exchange involve j strands ($j = 2, 3, 4$). A 4-strand double exchange corresponds to 1 recombinant per strand (1 morgan), whereas other double exchanges correspond to 0·5 recombinant per strand (0·5 morgan). In the absence of chromatid interference,

$$(p_2, p_3, p_4) = (1/4, 1/2, 1/4).$$

Since genetic data from eukaryotes do not support chromatid interference, we assume its absence (Carter, 1954).

By definition, the map distance of the region under consideration (in morgans) is $w = c_1/2 + c_2$, whereas the recombination fraction is

$$\begin{aligned}\theta &= c_1/2 + (p_3/2 + p_4)\,c_2 \\ &= (1-c_0)/2.\end{aligned}$$

From these two equations and the constraint that

$$\sum_{i=0}^{2} c_i = 1$$

https://doi.org/10.1017/S0016672300026276 Published online by Cambridge University Press

we obtain

$$\left.\begin{aligned} c_0 &= 1-2\theta, \\ c_1 &= 4\theta-2w, \\ c_2 &= 2(w-\theta). \end{aligned}\right\} \tag{10}$$

A tetrad is a parental ditype (PD) if there is no exchange or a 2-strand double exchange, with probability

$$\begin{aligned} p(PD) &= c_0 + p_2 c_2 \\ &= 1-2{\cdot}5\theta+0{\cdot}5w. \end{aligned}$$

A tetrad is a nonparental ditype (NPD) if there is a 4-strand double exchange,

$$\begin{aligned} p(NPD) &= p_4 c_2 \\ &= (w-\theta)/2. \end{aligned}$$

Finally, a tetrad is a tetratype (T) if there is a single exchange or a 3-strand double exchange,

$$\begin{aligned} p(T) &= c_1 + p_3 c_2 \\ &= 3\theta-w. \end{aligned} \tag{11}$$

Now consider the case where the genetic region extends to the centromere. Let q_{II} be the conditional probability if there is no nondisjunction in meiosis II that a heterozygous marker at the other end of the region gives rise to a heterozygous gamete: retained heterozygosity is called *nonreduction*. This will happen if there is a single exchange or a 3-strand double exchange. Therefore

$$\begin{aligned} q_{II} &= c_1 + p_3 c_2 \\ &= 3\theta-w \end{aligned}$$

which is the same as the tetratype frequency $p(T)$. Let q_I be the conditional probability of nonreduction if there is nondisjunction in meiosis I. Then

$$q_I = 1-q_{II}$$

if nonreduction in meiosis I is unrelated to crossingover. If, as seems probable, nondisjunction in meiosis I is due to failure of crossingover, θ and w are reduced and $q_I > 1-q_{II}$.

With attached-X chromosomes a heterozygous recessive gene may become homozygous in a gamete. This is reduction in meiosis I with recovery of a particular allele. The freqency is

$$h = q_{II}/4 = (3\theta-w)/4 \tag{12}$$

assuming normal crossingover in attached-X chromosomes.

Trisomies also recover 2 members of a tetrad. Consider a centromeric heteromorphism or DNA polymorphism that establishes the mode of origin of a trisomic zygote as maternal or paternal nondisjunction in meiosis I or II, assuming no crossingover between heteromorphism and centromere (Jacobs & Morton, 1977). Then the probabilities q_I, q_{II} apply to other loci that may be linked to the centromere of the trisomic chromosome. Maternal triploids are even more useful, since every chromosome is potentially informative. A hypothetical third type of

twin, resulting from fertilization of an egg and a polar body, would give the same information as maternal trisomy, with allowance for contributions of 2 sperm.

Some ovarian teratomas result from suppression of meiosis *II* or fusion of polar body *II* with the secondary oocyte (Linder *et al.* 1975), while others represent

Table 3. *The frequency of double recombination* θ_{12} *as a function of* θ_1, θ_2, *and* p

		p				
θ_1	θ_2	0·114	0·25	0·351	0·5	1·0
0·001	0·001	~ 0	~ 0	~ 0	~ 0	0·000001
0·001	0·010	~ 0	~ 0	~ 0	~ 0	0·000010
0·001	0·050	~ 0	~ 0	0·000001	0·000005	0·000050
0·001	0·100	~ 0	0·000001	0·000007	0·000020	0·000100
0·001	0·200	~ 0	0·000013	0·000037	0·000080	0·000200
0·001	0·300	0·000012	0·000065	0·000114	0·000180	0·000300
0·010	0·010	~ 0	~ 0	~ 0	0·000004	0·000100
0·010	0·050	~ 0	0·000001	0·000015	0·000060	0·000500
0·010	0·100	~ 0	0·000010	0·000073	0·000219	0·001000
0·010	0·200	~ 0	0·000141	0·000393	0·000833	0·002000
0·010	0·300	0·000137	0·000687	0·001178	0·001838	0·003000
0·050	0·050	~ 0	0·000017	0·000156	0·000495	0·002500
0·050	0·100	~ 0	0·000108	0·000553	0·001471	0·005000
0·050	0·200	~ 0	0·001032	0·002479	0·004808	0·010000
0·050	0·300	0·001192	0·004287	0·006763	0·009906	0·015000
0·100	0·100	~ 0	0·000478	0·001671	0·003846	0·010000
0·100	0·200	0·000100	0·003204	0·006411	0·011111	0·020000
0·100	0·300	0·004277	0·011040	0·015755	0·021429	0·030000
0·200	0·200	0·004637	0·013197	0·019566	0·027586	0·040000
0·200	0·300	0·020816	0·033111	0·040252	0·048387	0·060000
0·300	0·300	0·055118	0·065711	0·072127	0·079412	0·090000

The symbol ~ 0 signifies $< 0{\cdot}5 \times 10^{-6}$.

failure of meiosis *I* (Carritt *et al.* 1982). Ovarian contamination and teratomas due to endoreduplication of haploid products of meiosis *II* must be excluded. DNA polymorphisms promise to eliminate uncertainties caused by other modes of origin and ovarian contamination. Once recognized, teratomas due to meiotic failure can be used for centromere mapping, with due caution about possibly reduced crossingover in the ovary.

DISCUSSION

These generalities apply to either sex in any organism, with appropriate choice of a mapping function. If the mapping parameter p is found to be nearly constant for different data, it may be used to constrain the human map. The efficiency of inferring order and distance from 2, 3, and more points must be examined, and ways must be found to communicate this information as concisely as lod scores encode 2-point data.

Flanking markers around a disease locus provide a good basis for genetic

counselling, given information on phase from relatives. Prediction will be in error if there is a double recombinant, with probability $\theta_{12} = C\theta_1\theta_2$. Therefore the coincidence C, but not higher order recombinant frequencies, promises to play a major role in genetic counselling. In Table 3 we present the probablity of double recombination for various values of θ_1, θ_2, and p. Since $C = 1$ for $p = 1$, these probabilities may be converted to coincidence by dividing with the figure in the last column.

When transmission is inferred, the probability of error is θ_{12}. This could lead to selective termination of a normal fetus, or unwarranted fear in a normal adult who believes himself to be a carrier. When transmission is excluded, the probability of error is also θ_{12}. This could lead to unintentional propagation of the disease gene and to deception of an individual believed normal. Fortunately the conditions for flanking markers to provide a high level of security are rather broad. Even with recombination of 10% between the disease locus and each flanking marker, the error frequency is less than 1%. If we accept the rough approximation that 1 cMo corresponds to 10^6 base pairs, a flanking marker thousands of loci away should still provide adequate control on θ_{12} even in a region of high recombination.

The simplicity of multipoint recombination probabilities depends on the assumption that coincidence is determined by the nearest recombinant interval. This interval Markov assumption is not the same as the point Markov postulate expressed by Bailey (1961, p. 158) as 'a cross-over point once established effectively divides the chromosome into regions which do not interfere with each other, although there may still be interference within each region'. This point Markov postulate is appealing, but mathematically intractable. It may easily be verified that the results of Owen (1950, 1953), recapitulated by Bailey, do not satisfy either Markov assumption. For example, equation 10·39 of Bailey (1961, p. 159) is (in our notation)

$$\begin{aligned}\theta_{12} &= 2K\theta_1\theta_2(\theta_1+\theta_2)\\ \theta_{23} &= 2K\theta_2\theta_3(\theta_2+\theta_3)\\ \theta_{123} &= 4K^2\theta_1\theta_2\theta_3\{(\theta_1+\theta_2)(\theta_2+\theta_3)-\theta_2^2/3\} \neq \theta_{12}\theta_{23}/\theta_2\end{aligned}$$

where $K > 1$ implies that recombination between two loci on the same chromosome cannot approach 1/2, and $K < 1$ implies that recombination between systenic loci can exceed 1/2. For large w_2 the relative underestimate of θ_{123} reaches 1/3. The difference from the Markov assumption may in practice be trivial, but greatly complicates the theory. It is completely avoided by abandoning the Kosambi coefficient, with its ancillary hypotheses that chiasmata form in a regular time sequence, starting from the centromere (Mather, 1936), and that the frequency distribution of the interval between successive crossovers gives rise to recombination fractions greater than 50% (Fisher, Lyon & Owen, 1947). We have restricted our treatment to the simplest predictions required for multipoint mapping and genetic counselling, that may be verified empirically. Our results are not limited to any particular mapping function, which should be chosen to fit observations.

REFERENCES

BAILEY, N. T. J. (1961). *Introduction to the Mathematical Theory of Genetic Linkage.* London: Oxford University Press.

CARRITT, B., KING, J. & WELCH, H. M. (1982). Gene order and localization of enzyme loci on the short arm of chromosome 1. *Annals of Human Genetics* **46**, 329–335.

CARTER, T. C. (1954). A search for chromatid interference in the male house mouse. *Zeitschrift für indukt. Abstammungs. und Vererbungslehre* Bd. **86**, S. 210–223.

CARTER, T. C. & FALCONER, D. S. (1951). Stocks for detecting linkage in the mouse and the theory of their design. *Journal of Genetics* **50**, 307–323.

FISHER, R. A., LYON, M. F. & OWEN, A. R. G. (1947). The sex chromosome in the house mouse. *Heredity* **1**, 355–365.

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.

HULTEN, M. (1974). Chiasma distribution at diakinesis in the normal human male. *Hereditas* **76**, 55–78.

JACOBS, P. A. & MORTON, N. E. (1977). A chiasma map of man. *Human Heredity* **27**, 38–51.

KARLIN, S. (1984). Theoretical aspects of genetic map functions in recombination processes. *C. C. Li Symposium on Human Population Genetics*, ed. A. Chakravarti (in the Press.)

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

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

LINDER, L., MCCAW, B. K. & HECHT, F. (1975). Parthenogenic origin of benign ovarian teratomas. *New England Journal of Medicine* **292**, 63–66.

MATHER, K. (1936). The determination of position in crossing-over. *Journal of Genetics* **33**, 207–235.

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

MORTON, N. E. (1955). Sequential tests for the detection of linkage. *American Journal of Human Genetics* **7**, 277–318.

MORTON, N. E., RAO, D. C., LINDSTEN, J. & YEE, S. (1977). A chiasma map of man. *Human Heredity* **27**, 38–51.

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

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* (In the Press.)

ROBINSON, R. (1971). *Gene Mapping in Laboratory Mammals, Part A, The Biometrical Approach.* London–New York: Plenum Press.

SHERMAN, S. L., JACOBS, P. A. & TURNER, G. (1984). Marker X syndrome: a cytogenetic and genetic analysis. *Annals of Human Genetics* **48**, 21–37.

STURT, E. (1976). A mapping function for human chromosomes. *Annals of Human Genetics* **40**, 147–163.

STURTEVANT, A. H. (1914). The reduplication hypothesis applied to Drosophila. *American Naturalist* **XLVIII**, 535.