A MATHEMATICALLY TRACTABLE FAMILY OF GENETIC MAPPING FUNCTIONS WITH DIFFERENT AMOUNTS OF INTERFERENCE

JOSEPH FELSENSTEIN

Department of Genetics, University of Washington, Seattle, Washington 98195

Manuscript received September 22, 1978

ABSTRACT

By extension of the argument of KOSAMBI (1944), a family of mapping functions can be derived, which has a parameter regulating the intensity of interference. Different values of this parameter yield the HALDANE (1919) and KOSAMBI mapping functions as special cases. The parameter is the coincidence coefficient for nearby small intervals. The family includes mapping functions for negative interference. A simple rule for combining recombination fractions in adjacent intervals is also obtained.

MAPPING functions are formulas expressing the recombination fraction between two markers as a function of the map distance between them, or vice versa. They were pioneered by HALDANE (1919). Many mapping functions have been proposed since then, most of which are discussed by BAILEY (1961) in some detail. The functions proposed by CARTER and FALCONER (1951), STURT (1976), and by RAO, *et al.* (1977) are of particular interest for their realism. KOSAMBI's (1944) mapping function is less realistic than any of these, but has the advantage of mathematical tractability, which it shares with HALDANE's function. In particular, both of these latter have simple rules for computing the recombination fraction in a longer interval from the recombination fractions in two shorter intervals into which it can be subdivided. The difficulty seems to be that mapping functions which realistically express the relation between recombination fraction and map distance lack mathematical tractability, in that they do not predict simple rules for combining recombination fractions in adjacent intervals, nor do they allow the coincidence coefficient to be expressed simply as a function of map distance or recombination fraction.

This note presents a mapping function with one parameter determining the pattern of interference. It has a simple formula for combining recombination fractions. In fact, both the HALDANE and KOSAMBI functions are special cases of the present function. While it is undoubtedly less realistic than some of the other functions, the presence of an interference parameter may make this function a useful compromise between tractability and realism.

DERIVATION

The mapping function is derived by a direct extension of KOSAMBI's (1944) argument. This considers the recombination fraction $r(x + dx)$ in an interval

formed by adding a very short interval of length dx to an interval whose length (in Morgans) is x. Since recombination fraction should be equal to map distance for short intervals, we approximate $r(dx)$ by dx (this amounts to ignoring terms of order $(dx)^2$ in comparison to those of order dx, which is asymptotically valid as $dx \to 0$). Given that we know the recombination fraction $r(x)$ and the coefficient of coincidence $C(x,dx)$ between the short interval dx and the longer one x, the recombination fraction in the combined interval must be

$$\begin{aligned} r(x+dx) &= r(x) + r(dx) - 2C(x,dx)\, r(x)\, r(dx) \\ &= r(x) + dx - 2C(x,dx)\, r(x)\, dx \end{aligned} \tag{1}$$

so that

$$\frac{r(x+dx) - r(x)}{dx} = 1 - 2C(x,dx)\, r(x) \tag{2}$$

or as $dx \to 0$,

$$\frac{dr(x)}{dx} = 1 - 2C(x,dx)\, r(x) \ . \tag{3}$$

KOSAMBI's innovation was to assume that $C(x,dx)$ was a direct function of the recombination fraction $r(x)$. KOSAMBI assumed that coincidence rose from 0, when x was small, to 1 when x was large, and made the assumption that this rise was linear in $r(x)$, so that $C(x,dx) = 2\,r(x)$. The present mapping function makes the same assumption of linearity in $r(x)$, but assumes instead that when x is small, $C(x,dx) = K$, an arbitrary constant. When x is large, $C(x,dx)$ must be 1. We then have

$$C(x,dx) = K - 2\,(K-1)\, r(x) \ . \tag{4}$$

Substituting this into (3),

$$\frac{dr}{dx} = 1 - 2\,K\,r + 4\,(K-1)\, r^2 \ . \tag{5}$$

This is a differential equation that can be solved simply by separation of variables and integration by partial fractions:

$$\int \frac{dr}{[1 - 2\,K\,r + 4\,(K-1)\, r^2]} = \int dx \tag{6}$$

$$\int \frac{dr}{(1-2\,r)\,(1-2\,(K-1)\, r)} = x + C_0 \tag{7}$$

$$\int \left[-\frac{1}{(K-2)\,(1-2\,r)} + \frac{K-1}{(K-2)\,(1-2\,(K-1)\, r)} \right] dr = x + C_0 \tag{8}$$

$$\ln\,(1-2\,r) - \ln\,(1-2\,(K-1)\, r) = 2\,(K-2)\, x + C_1 \ . \tag{9}$$

Since we must have $r = 0$ when $x = 0$, it must be true that $C_1 = 0$, so that

$$x = \frac{1}{2\,(K-2)} \ln \left(\frac{1-2\,r}{1-2\,(K-1)\, r} \right) , \tag{10}$$

which gives map distance as a function of recombination fraction. This is easily solved to give

$$r = (1/2)\left(\frac{1-(K-1)e^{2(K-2)x}}{1-e^{2(K-2)x}}\right), \tag{11}$$

which is the mapping function we seek. As one might expect, when $K = 1$ this is precisely HALDANE's mapping function. When $K = 0$, it is KOSAMBI's function. When $K = 2$ the integration proceeds somewhat differently after (7), giving

$$r = \frac{x}{1+2x}. \tag{12}$$

Figure 1 shows the shape of the mapping function for four values of K. The dotted curve, for $K = -2$, shows the phenomenon of "map expansion" (HOLLIDAY 1964), which is the case of the recombination fraction for a long interval being greater than the sum of the recombination fractions in its subintervals. This is biologically impossible in an ordinary three-point cross, and its occur-

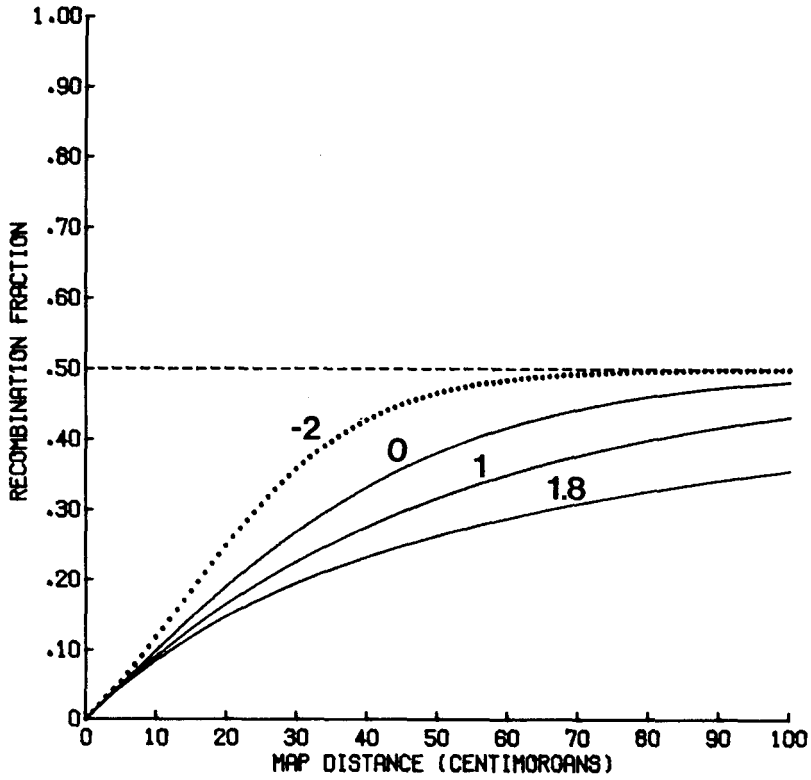

FIGURE 1.—The recombination fraction as a function of the map distance (in centiMorgans), plotted for various values of K. The values of K are given next to each curve. The dotted curve ($K = -2$) shows the phenomenon of map expansion.

rence in pairwise crosses when gene conversion is present seems to stem from marker effects on recombination (HOLLIDAY 1964). It will occur whenever $K < 0$. When $K > 2$ the mapping function approaches a plateau at $1/[2(K-1)]$, below $r = 1/2$, as map distance increases. Thus, the limits on K are $0 \leq K \leq 2$. Note from (4) that when $K > 1$, we have negative interference.

The present mapping function with negative K may be useful in map construction when map expansion is present in a set of pairwise crosses. Alternatively, one could make use of $K < 0$ in ordinary multipoint cross data if a mapping function more extreme than KOSAMBI's, with more interference over short distances, were desired, and if one were prepared to ignore small amounts of map expansion.

ADDITION LAW FOR RECOMBINATION FRACTIONS

One desirable property for both HALDANE's and KOSAMBI's functions is that they yield simple rules for predicting the recombination fraction in a longer interval from the recombination fractions in two subintervals into which it is divided. The present mapping function gives rise to an addition law for recombination fractions which generalizes HALDANE's and KOSAMBI's rules. It is easily obtained. Note that (10) is of the form $x = f(r)$. Now suppose that we have three markers, 1, 2, and 3, which are arranged on a linkage map in the order 1-2-3. Let us put subscripts on x and r to indicate which pair of markers they refer to. Now we must have that $x_{12} + x_{23} = x_{13}$. So $f(r_{12}) + f(r_{23}) = f(r_{13})$. Substituting for (10), we have

$$\frac{1}{2(K-2)} \ln\left(\frac{1-2r_{12}}{1-2(K-1)r_{12}}\right) + \frac{1}{2(K-2)} \ln\left(\frac{1-2r_{23}}{1-2(K-1)r_{23}}\right) = \frac{1}{2(K-2)} \ln\left(\frac{1-2r_{13}}{1-2(K-1)r_{13}}\right) \tag{13}$$

Taking exponentials of both sides of (13)

$$\left(\frac{1-2r_{12}}{1-2(K-1)r_{12}}\right)\left(\frac{1-2r_{23}}{1-2(K-1)r_{23}}\right) = \left(\frac{1-2r_{13}}{1-2(K-1)r_{13}}\right), \tag{14}$$

which is easily solved for r_{13} giving the desired rule for combining recombination fractions of adjacent intervals:

$$r_{13} = \frac{r_{12} + r_{23} - 2Kr_{12}r_{23}}{1 - 4(K-1)r_{12}r_{23}} . \tag{15}$$

This formula is valid even in the extreme case $K = 2$, although a different derivation must be used in that case, starting with (12).

THE COINCIDENCE FUNCTION

We obtained the mapping function by assuming a particular functional form for the "marginal coincidence" function $C(x, dx)$ but we did not specify the coincidence between adjacent intervals of arbitrary length. We can use (15) to

give the coincidence between adjacent intervals in terms of the recombination fractions r_{12} and r_{23} in those intervals. By definition of coincidence

$$C_{12,23} = \frac{r_{12} + r_{23} - r_{13}}{2\, r_{12}\, r_{23}} \,. \tag{16}$$

Substituting r_{13} from (15), we can express (16) entirely in terms of r_{12} and r_{23}, simplifying (16) to

$$C_{12,23} = \frac{K - 2\,(K-1)\,(r_{12} + r_{23})}{1 - 4\,(K-1)\,r_{12}\,r_{23}} \,. \tag{17}$$

As one would expect of a coincidence function, this becomes 1 when either r_{12} or r_{23} is 1/2. When both r_{12} and r_{23} are small the coincidence approaches K, as it must. One could alternatively express the coincidence (17) as a function of map distances. However, the coincidence is more easily expressed in terms of r_{13} alone. From (15) and (17) one can prove straightforward that

$$C_{12,23} = 1 + (K-1)\,(1 - 2r_{13}) \tag{18}$$

which shows that the coincidence depends only on the recombination fraction in the whole interval. Substituting (11) into (18) for r_{13} we find that the dependence on total map distance is

$$C_{12,23} = \frac{1 - (K-1)^2\, e^{2(K-2)x_{13}}}{1 - (K-1)\; e^{2(K-2)x_{13}}} \,, \tag{19}$$

x_{13} being the total map distance $x_{12} + x_{23}$.

The dependence of the coefficient of coincidence only on the recombination fraction in the total interval (and thus only on the sum of the map distances) is a striking feature of our model, and fairly easy one to check in any biological case in which it is proposed to use this mapping function.

OTHER MAPPING FUNCTIONS

Barratt *et al.* (1954) have also presented a family of mapping functions with one parameter, which yields the Haldane function as a special case. Rao *et al.* (1977) have presented a one-parameter family of functions, which yields the Haldane, Kosambi and Carter-Falconer functions as special cases. The present family of mapping functions is considerably simpler than either of these. It has the additional advantage of mathematical tractability: neither of the other families of mapping functions provides a simple addition law for recombination fractions, neither yields simple expressions for the coincidence between two adjacent intervals, and neither allows both map distance to be expressed as a function of recombination fraction and recombination fraction to be expressed as a function of map distance.

LIMITATIONS

Our mapping function is purely phenomenological, derived without any underlying model of the recombination process. Its parameter K is of only limited

meaning biologically, giving us little more insight than do the raw data. A more serious limitation of this mapping function is that it does not permit prediction of the frequencies of recombinant classes at more than three loci. To be able to predict, say, the frequencies of all sixteen classes of gametes formed in a quadruple heterozygote, we need to know more than the pairwise coincidence coefficients among adjacent chromosome segments. There is an extra degree of freedom in the problem, which is not determined even if we know the recombination fractions of all possible pairs of markers. This limitation is a serious one. If we have data on (say) four markers in a quadruple backcross and we have observed all sixteen classes of progeny, the most straightfordward method of estimating map distances would be to obtain a maximum-likelihood or minimum χ^2 fit of observed to expected numbers of progeny. The values of x_{12}, x_{23}, x_{34}, and K would then be joint estimates based on all of the data. However, our mapping function does not allow prediction of progeny class frequencies for more than three markers at a time. If we had only three markers, the maximum likelihood estimates of x_{12}, x_{23}, and K would be obtained simply by finding those values that fit the frequencies of the four classes of recombinants exactly. If we analyze the data on markers 1-2-3 separately from the data on 2-3-4, we are then faced with the problem of how to combine our two different estimates of x_{23}, and also our two different estimates of K.

There are a number of papers in the genetic literature in which specific mapping functions are derived which are predicted by particular models of the recombination process. These include the models of STURT (1976) and COBBS (1978), and many previous efforts cited by them. While these models have the advantage of precision, they run the risk of being made irrelevant by advances in our understanding of the recombination process. In this respect the very lack of precision of the present phenomenological approach makes it practically invulnerable to disproof.

I thank BARBARA NORTH BECK for bringing this problem to my attention, and for testing the useability of the results in a case of high negative interference. I also wish to thank LARRY SANDLER and JÜRG OTT for valuable discussion and references. This work was supported by U. S. Department of Energy contract number EY-76-S-06-2225-5 with the University of Washington.

LITERATURE CITED

BAILEY, N. T. J., 1961 *Introduction to the Mathematical Theory of Genetic Linkage*. Clarendon Press, Oxford.

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

CARTER, T. C. and D. S. FALCONER, 1951 Stocks for detecting linkage in the mouse and the theory of their design. J. Genet. **50**: 307–323.

COBBS, G., 1978 Renewal process approach to the theory of genetic linkage: case of no chromatid interference. Genetics **89**: 563–581.

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

HOLLIDAY, R., 1964 A mechanism for gene conversion in fungi. Genet. Res. **5**: 282–304.

KOSAMBI, D. D., 1944 The estimation of map distance from recombination values. Ann. Eugen. **12**: 172–175.

RAO, D. C., N. E. MORTON, J. LINDSTEN, M. HULTEN and S. YEE, 1977 A mapping function for man. Human Hered. **27**: 99–104.

STURT, E., 1976 A mapping function for human chromosomes. Ann. Human Genet. **40**: 147–163.

Corresponding editor: D. L. HARTL