

Maximal Power Tests for Detecting Defects in Meiotic Recombination

Thomas I. Milac,^{*,†} Frederick R. Adler^{‡,§} and Gerald R. Smith^{**,1}

Department of Applied Mathematics, University of Washington, Seattle, Washington 98195-2420, [†]The Genome Center, Department of Medicine, University of Washington, Seattle, Washington 98195-2145, [‡]Department of Mathematics, University of Utah, Salt Lake City, Utah 84112, [§]Department of Biology, University of Utah, Salt Lake City, Utah 84112 and ^{}Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109*

Manuscript received January 9, 2002
Accepted for publication April 12, 2002

ABSTRACT

We have determined the marker separations (genetic distances) that maximize the probability, or power, of detecting meiotic recombination deficiency when only a limited number of meiotic progeny can be assayed. We find that the optimal marker separation is as large as 30–100 cM in many cases. Provided the appropriate marker separation is used, small reductions in recombination potential (as little as 50%) can be detected by assaying a single interval in as few as 100 progeny. If recombination is uniformly altered across the genomic region of interest, the same sensitivity can be obtained by assaying multiple independent intervals in correspondingly fewer progeny. A reduction or abolition of crossover interference, with or without a reduction of recombination proficiency, can be detected with similar sensitivity. We present a set of graphs that display the optimal marker separation and the number of meiotic progeny that must be assayed to detect a given recombination deficiency in the presence of various levels of crossover interference. These results will aid the optimal design of experiments to detect meiotic recombination deficiency in any organism.

IN many studies of genetic recombination it is important to determine whether an individual organism or strain is recombination deficient. Identifying gene products involved in recombination, for example, requires testing appropriate mutants for their recombination proficiency. Similarly, testing whether a strain altered in some other aspect of cellular growth or metabolism is also altered in meiotic recombination requires assaying for a defect in meiotic recombination.

The frequency of recombination over some interval defined by a pair of genetic markers is measured by assaying a limited number of meiotic products (gametes) or their resultant progeny. It is relatively simple to determine if a mutant is strongly recombination deficient: Assay of a small sample of gametes using markers loosely linked in the wild type will reveal many recombinants from the wild type but few or none from the mutant. A statistical test will indicate if enough gametes have been tested to judge the difference in number of recombinants to be significant. It is more difficult to ascertain if a mutant is only modestly recombination deficient. In this case, both the mutant and wild type will produce roughly equal numbers of recombinants. If the markers used are closely linked, recombinants will be rare, and large numbers of gametes must be tested to accumulate enough recombinants for a difference in

number to be judged statistically significant. If the markers chosen are loosely linked, the mutant and wild type will both yield nearly 50% recombinants; again, large numbers of gametes must be tested to establish any difference. These considerations suggest that some intermediate marker separation (genetic distance) is most powerful for detecting a difference between the wild type and the putative mutant.

The problem of choosing markers to measure recombination deficiency has existed since the discovery of recombination along linear linkage maps by STURTEVANT (1913). Calculating the optimal marker separation could have been addressed any time after the formulation of a mapping function for recombination by HALLDANE (1919). To our knowledge, however, this problem has not been addressed or solved. Published studies measuring recombination deficiency state no basis for choosing markers beyond their availability and convenience in scoring. For example, GERSTEN and KEMP (1997), searching for an effect of a *p53* null mutation on meiotic recombination in mice, used markers “10–20 cM apart, to maximize the chance of observing either an increase or a decrease in recombinant frequency,” but stated no basis for this choice of interval size.

In this article, we calculate the optimal marker separation for detecting a reduction in recombination proficiency or in crossover interference. We find that the optimal separation depends only modestly on the degree of recombination deficiency and on the number of gametes to be tested.

¹Corresponding author: Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., Seattle, WA 98109.
E-mail: gsmith@fhcrc.org

METHODS

We take n to be the number of progeny or gametes tested in both the mutant and wild type if only a single interval is assayed in each. If it is assumed that recombination is uniformly decreased across a region of interest, and multiple independent intervals of the same genetic size (centimorgans) are assayed in each progeny, n is the number of intervals assayed times the number of progeny. For our results to apply in the presence of interference, the intervals assayed in each progeny must be sufficiently distant from one another so that recombination in one interval is independent of recombination in the others. In the absence of interference, the intervals chosen need be merely non-overlapping. Because recombination events among spores of a meiotic tetrad are not independent, the effective n for tetrad data is twice the number of tetrads assayed or, equivalently, one-half the total number of spores. For convenience, we describe n subsequently simply as the number of progeny or gametes.

The genetic distance between markers, x , the relative (mutant to wild-type) recombination proficiency, λ , and the interference parameter, m , determine the recombination frequencies R_m^{mt} and R_m^{wt} of the mutant and wild type, respectively, via the mapping function in Equation 2. Because we chose x to have units of the wild-type genetic distance, we replaced x by λx in the mapping function when calculating the recombination frequency in the mutant.

For each combination of the parameters x , m , and n , we first calculated the joint probability distribution $P(k_1, k_2)$ of observing k_1 recombinants, $0 \leq k_1 \leq n$, in n draws from the wild-type population and k_2 recombinants, $0 \leq k_2 \leq n$, in n draws from the mutant population in the case that mutant and wild type are indistinguishable in their recombination proficiency; that is, $\lambda = 1$ and, consequently, $R_m^{\text{mt}} = R_m^{\text{wt}} = R$. We took the number of recombinants in the mutant or wild-type population to be governed by a binomial distribution with parameters R_m^{mt} and R_m^{wt} , respectively. We then identified the outcomes (k_1, k_2) least likely to occur in this null experiment and whose sum of probabilities most closely approached, but did not exceed, 0.05. We took this group of outcomes to comprise the rejection region. That is, we declared that if an experiment performed with an uncharacterized mutant yielded one of the outcomes in the rejection region, that outcome would constitute sufficient reason to reject the null hypothesis that mutant and wild type are indistinguishable with respect to their recombination proficiency.

Holding x , m , and n fixed, we then recalculated the distribution $P(k_1, k_2)$ with the recombination proficiency, λ , set to some value < 1 . To calculate the power, p , the probability of rejecting the null hypothesis, we summed the outcomes in this probability distribution belonging to the rejection region defined above.

To produce Figure 2, we repeated these calculations for values of x between 1 and 200 cM at intervals of 0.5 cM (1.0 cM for $n = 2000$ only) for $m = 0, 2, 5$ and for $n = 100, 250, 500, 1000, 2000$. For each combination of x , m , and n , we determined the maximum value of λ for which $p \geq 0.80$, $p \geq 0.90$, and $p \geq 0.95$ (that is, we found the upper boundary of each colored region) by an iterative method accurate to ± 0.0005 in λ . To produce Figure 3, we performed similar calculations with the wild type and mutant having different values of interference, m .

RESULTS AND DISCUSSION

We consider the wild type and the mutant to have a recombination proficiency that produces a certain mean number of crossovers in a given interval per meiotic product (chromatid or gamete). We use "crossover" to mean that point along a recombinant chromatid at which genetic information changes from that of one parent to that of the other. By definition, the genetic distance, x , between markers is equal to this mean number of crossovers. Two markers are separated by 1 M (or 100 cM) when the mean number of crossovers between them is 1. We assume that the mutant is recombination deficient relative to the wild type uniformly across the interval under consideration. That is, in an interval corresponding to a genetic distance x in the wild type, the mutant produces only λx mean crossovers per gamete, with $0 < \lambda \leq 1$. We refer to λ as the recombination proficiency (of the mutant relative to the wild type); if the mutant is hyperrecombination proficient, the roles of the mutant and wild type can be reversed. In special cases, a mutant may be recombination deficient in some regions of the genome but not in others. For example, a recombinational hotspot may be active in the wild type but not in the mutant. In these special cases, markers must be chosen according to the nature of the mutant, and the considerations in this article do not apply.

Because only odd numbers of crossovers between markers produce an observed recombinant, a mapping function is needed to convert the observed recombination frequency, R , between markers to genetic distance, x . Under the assumption that crossovers along chromatids and among gametes are randomly (Poisson) distributed and therefore do not interfere with one another, HALDANE (1919) determined the relation between recombination frequency, R , and genetic distance, x , to be

$$R = \frac{1}{2} [1 - e^{-2x}]. \quad (1)$$

In most organisms, however, meiotic crossovers do interfere with one another such that fewer multiple crossovers are observed than predicted from the Haldane function, Equation 1. To account for interference, numerous mapping functions have been derived. The

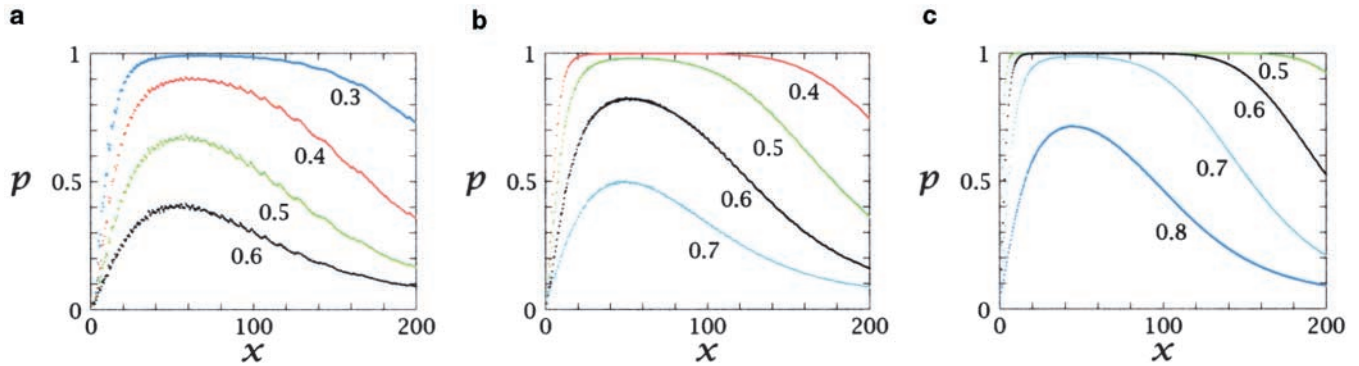


FIGURE 1.—The probability or power (p) of detecting a difference in the recombination proficiency (λ) of a mutant relative to wild type is greatest over a limited range of marker separations (x) measured in centimorgans. Recombination proficiencies from 0.8 to 0.3 are indicated next to the curves. The number of progeny or independent intervals (n) analyzed is (a) 100, (b) 250, and (c) 1000. Shown are results for a wild type and mutant having no crossover interference ($m = 0$).

function of Foss *et al.* (1993) incorporates a single parameter, m , that can be varied to fit the recombination data observed for a given organism. Their mapping function is

$$R_m = \frac{1}{2} \left[1 - e^{-y} \sum_{i=0}^m \frac{y^i}{i!} \left(1 - \frac{i}{m+1} \right) \right], \quad (2)$$

where $y = 2(m+1)x$. For organisms with no interference, $m = 0$ and the mapping function in (2) reduces to that of Haldane (1). Interference is an increasing function of the parameter m , $m = 1, 2, 3, \dots$. The mapping function in (2) fits recombination data from *Neurospora crassa* reasonably well with $m = 2$ and fits data from *Drosophila melanogaster* with $m = 4$ or 5 (Foss *et al.* 1993).

As detailed in METHODS, we calculated the statistical power (a probability), p , of rejecting the hypothesis that mutant and wild type have the same recombination proficiency in an experiment in which n gametes (or independent intervals) from both the mutant and the wild type are assayed. We calculated the power for experiments conducted with marker separations, x , ranging from 1 to 200 cM; for recombination proficiencies, λ , between 0.001 and 1.0; for degrees of interference, m , equal to 0, 2, and 5; and for the number of gametes tested, n , ranging from 100 to 2000. We considered markers separated by ≤ 200 cM because the genetic length of few eukaryotic chromosomes exceeds this value.

Figure 1 illustrates how the power of an experiment to distinguish a recombination-deficient mutant from its wild type varies with marker separation for selected values of recombination proficiency. Note that for fixed λ , n , and m there is an optimal marker separation, x , for detecting recombination deficiency. For example, when $n = 100$, $m = 0$, and $\lambda = 0.3$, a test is most powerful ($p \geq 0.95$) when $x = 30$ –130 cM (Figure 1a). In other words, if the mutant is reduced in recombination proficiency by a factor of 3.3, and 100 gametes each from

the wild type and mutant are tested, in $\geq 95\%$ of experiments the difference in recombination proficiency will be detected if the markers used are separated by 30–130 cM. Choosing a smaller interval reduces the probability of detecting the recombination deficiency: For example, with $x = 30$ cM, $p = 0.95$, but with $x = 15$ cM, $p = 0.75$. The range of optimal marker separation is narrower if the recombination proficiency is greater ($\lambda > 0.3$). In some cases, marker separations over a broad range are, to good approximation, equally powerful in detecting recombination deficiency.

By increasing n to 250, a twofold reduction in recombination proficiency ($\lambda = 0.5$) can be detected readily ($p \geq 0.95$) if $x = 30$ –90 cM (Figure 1b). Thus, a modest reduction in recombination proficiency can be detected with high probability provided the markers used are separated by the appropriate distance.

Figure 2 summarizes in a compact form the power of numerous experimental designs to detect recombination deficiency. Color is used to distinguish experimental designs in which different numbers of gametes, n , are assayed. Each colored region indicates combinations of λ and x for which the power of the experimental design exceeds the threshold value shown on the left in the figure; the value of the interference parameter used in calculating each part is shown along the top.

The results in Figure 2 can be used to optimize the design of experiments measuring recombination deficiency. From the graphs one can read the optimal marker separation x and the number of gametes n that must be tested in both the mutant and the wild type to detect, with a desired power, p , a suspected level of recombination proficiency λ in the mutant when interference is at a certain level m . No calculations need be performed. For example, if one is working with an organism manifesting no interference ($m = 0$) and suspects that a mutant has approximately twofold reduction in recombination proficiency ($\lambda = 0.5$), Figure 2a shows that ≥ 250 gametes of both the mutant and wild

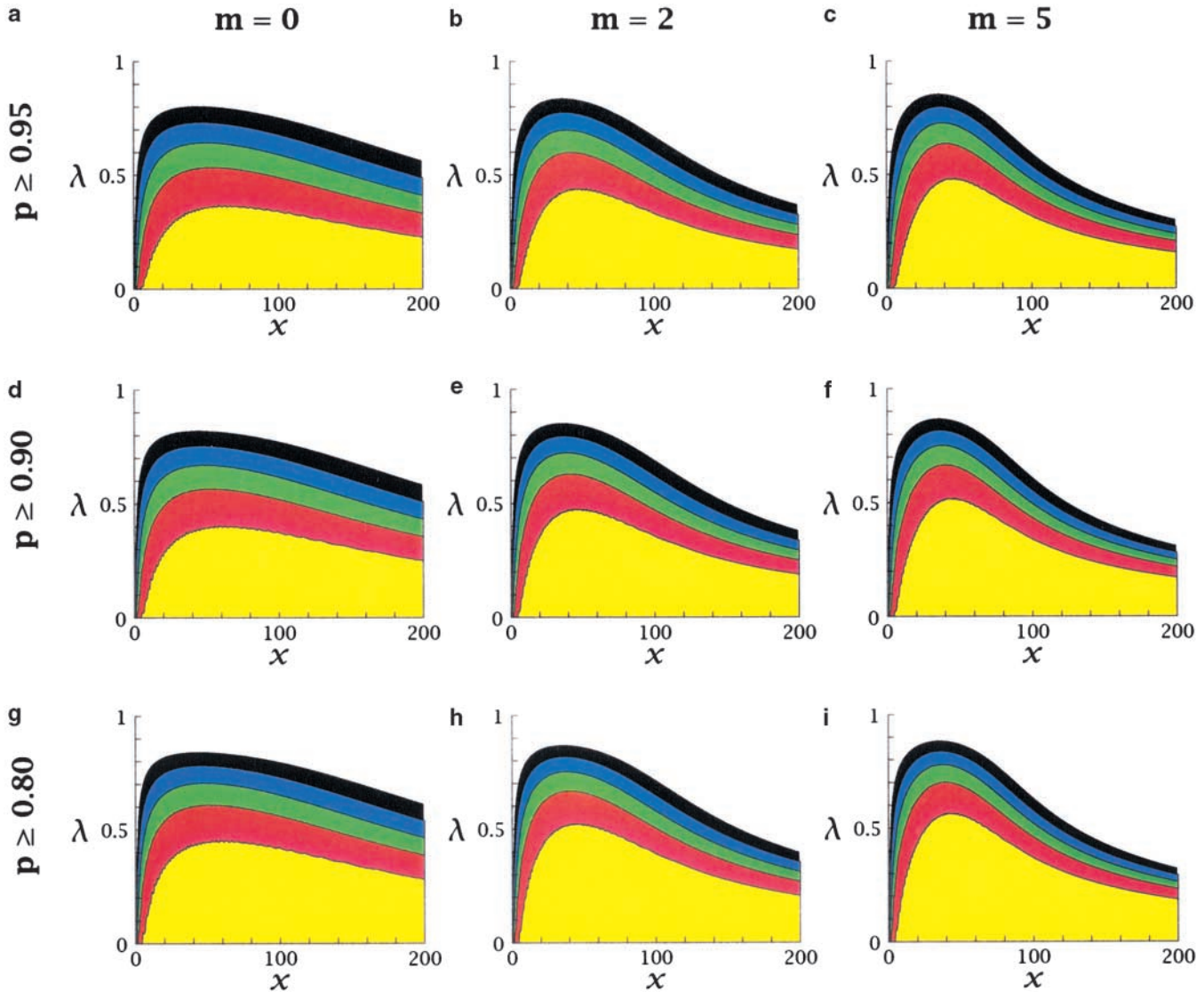


FIGURE 2.—The ability to detect recombination deficiency depends on the marker separation (x), the recombination proficiency of the mutant relative to its wild type (λ), and the number of progeny or independent intervals (n) analyzed. Results are shown for different combinations of the desired power (p) and crossover interference parameter (m). Color indicates the number of progeny analyzed: black, 2000; blue, 1000; green, 500; red, 250; yellow, 100. p is greater than or equal to the indicated value for all combinations of λ and x within and below the appropriate colored area. The power of an experiment conducted with $n = 1000$ progeny, for example, is greater than or equal to the power threshold shown on the left for all values of λ and x encompassed by the regions shaded blue, green, red, and yellow.

type must be tested to have a $\geq 95\%$ chance of revealing this difference. If 250 gametes are tested from each, the markers must be separated by 30–90 cM; if 500 gametes are tested, markers separated by 10–150 cM are suitable. As n increases, lesser reductions in recombination proficiency can be detected, and the optimal marker separation for detecting them shifts slightly to smaller values. For example, to detect a 30% reduction in proficiency ($\lambda = 0.7$) requires $n \geq 1000$ and $x = 25$ –80 cM.

Relaxing the required power of a test allows detecting slightly greater mutant recombination proficiency with the same number of gametes assayed [compare Figure 2, g–i ($p \geq 0.80$) with Figure 2, d–f ($p \geq 0.90$) and

Figure 2, a–c ($p \geq 0.95$)]. The optimal range of marker separation does not change greatly, however.

For a fixed value of suspected recombination proficiency, λ , the presence of interference often reduces the required number of gametes that needs to be assayed to achieve a chosen level of power. This is a consequence of interference decreasing the probability of double crossovers and, hence, effectively increasing the number of observed (single crossover) recombinants. If, as in the example considered above, one suspects a twofold reduction in recombination proficiency ($\lambda = 0.5$), with $m = 0$ one needs to assay 250 gametes of the mutant and wild type with markers separated by 20–115 cM to

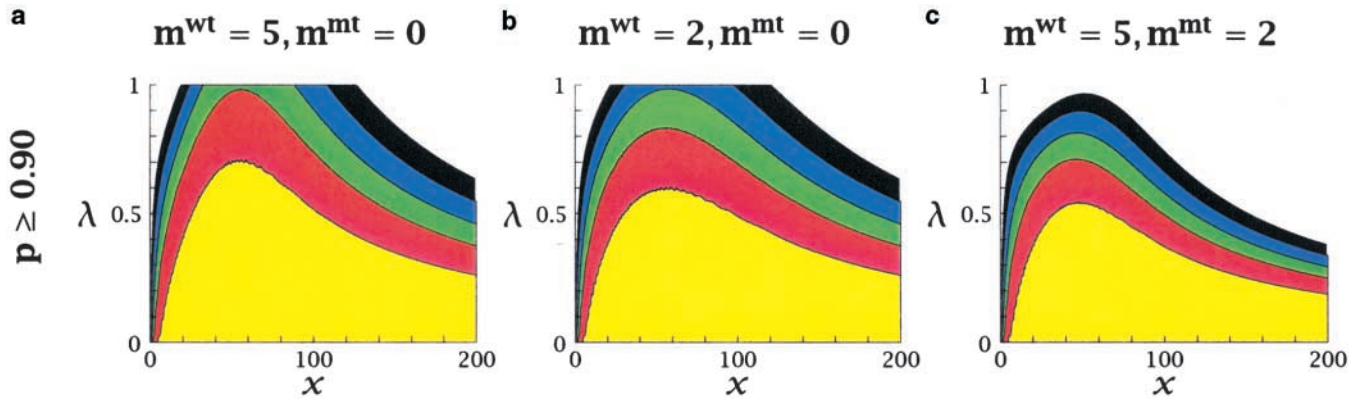


FIGURE 3.—Detection of a reduction in crossover interference with or without a reduction in recombination proficiency. The crossover interference in the wild type (m^{wt}) and mutant (m^{mt}) is as shown. Color indicates the number of progeny analyzed: black, 2000; blue, 1000; green, 500; red, 250; yellow, 100. The results are otherwise presented as in Figure 2.

have a $\geq 90\%$ chance of revealing this recombination deficiency (Figure 2d). With $m = 5$, however, one needs to assay only 100 gametes of the mutant and wild type with markers separated by 35–55 cM to achieve nearly the same level of power (Figure 2f).

Some mutations may reduce or abolish interference with or without a reduction in recombination proficiency. Figure 3 summarizes optimal experimental designs to detect such mutations with power ≥ 0.90 . In the examples shown, the wild type has interference, $m = 2$ or 5, whereas the mutant has none, $m = 0$, or a reduced level, $m = 2$. Figure 3a shows that if the wild type has $m = 5$, a mutant with $m = 0$ would be detected even if it had no reduction in recombination proficiency; *i.e.*, $\lambda = 1$, provided $x = 35$ – 90 cM and $n \geq 500$. With $n = 250$, a mutant with a loss of interference and a slight reduction in recombination proficiency, $\lambda = 0.97$, would be detected if an interval of 55 cM were chosen. LIN (2001) also noted, among those marker separations she tested, that 55 cM minimized the number of progeny required to detect a loss of interference ($m = 4$ in the wild type). Determining whether a mutant is reduced in recombination proficiency or interference or both requires additional tests with multiple markers.

If only a limited number of phenotypic markers were available, one might not be able to test for recombination deficiency using markers with the optimal separa-

tion. Genome projects, however, have greatly expanded the list of available DNA markers, including restriction fragment length polymorphisms, single-nucleotide polymorphisms, and repeat length differences, thus allowing one to choose marker separations more or less at will. These markers and the results presented here will facilitate the design of more powerful tests for detecting meiotic recombination deficiency.

We thank Li Hsu, Leonid Kruglyak, Phil Meneely, David Perkins, John Philpot, Bob Robbins, Suzanne Rutherford, and an anonymous reviewer for helpful comments. This work was supported by grant R01-GM32194 from the National Institutes of Health (NIH) to G.R.S. and K01-HG00032 from the NIH to T.I.M.

LITERATURE CITED

- FOSS, E., R. LANDE, F. W. STAHL and C. M. STEINBERG, 1993 Chiasma interference as a function of genetic distance. *Genetics* **133**: 681–691.
- GERSTEN, K. M., and C. J. KEMP, 1997 Normal meiotic recombination in *p53*-deficient mice. *Nat. Genet.* **17**: 378–379.
- HALDANE, J. B. S., 1919 The combination of linkage values, and the calculation of distances between the loci of linked factors. *J. Genet.* **8**: 299–309.
- LIN, S., 2001 Sample size requirements for detecting interference under the chi-square model. *Hum. Hered.* **52**: 201–209.
- STURTEVANT, A. H., 1913 The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J. Exp. Zool.* **14**: 43–59.

Communicating editor: M. LICHTEN