Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 1423–1427, March 1986
Genetics

Statistical considerations for linkage analysis using recombinant inbred strains and backcrosses

JONATHAN SILVER AND CHARLES E. BUCKLER

Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892

Communicated by Horace Davenport, October 28, 1985

ABSTRACT **Recombinant inbred (RI) mouse strains are extremely useful for gene mapping, especially for establishing preliminary map positions for new loci. However, the usual statistical analysis applied to such experiments may lead to erroneous conclusions about linkage unless unusually stringent criteria are adopted for rejecting the null hypothesis. We describe a Bayesian statistical approach for determining the probability of linkage when no prior information is available about the location of the gene to be mapped (the test locus). We present a table that gives the probability of linkage, the most likely position of the test locus with respect to a marker locus, and the interval around the marker locus that has a 95% chance of containing the test locus, for all possible experimental results suggesting linkage in sets of up to 40 RI strains. These results show that for the probability of linkage to be greater than 95%, the number of RI strains inheriting chromosomes recombinant for the test and marker loci must be smaller than previously assumed. The formulas derived for RI strains can be applied, with only minor modifications, to the analysis of Mendelian backcrosses. Differences between the Bayesian approach advocated here and the more traditional analysis of linkage are discussed in detail.**

Recombinant inbred (RI) mouse strains are families of inbred mouse strains derived by inbreeding the progeny of a cross between two parental strains, designated A and B for convenience. RI strains provide a powerful tool for gene mapping (refs. 1, 2, and 3, pp. 131–141). If the parental strains carry different alleles at a locus one wants to map, one may be able to map the locus simply by determining, for each RI strain, which parental allele, A or B, it inherits at that locus. A table of these results is called a strain distribution pattern (SDP). By comparing the SDP of the new locus to a table of SDPs for other loci, one determines whether alleles at any other loci are inherited in a similar fashion. The more closely two loci are linked, the more likely it is that their SDPs will be identical or nearly identical.

In using RI strains for linkage analysis, it is common to find that the SDP for a new locus differs from the SDP for a previously described locus in a small proportion of the RI strains. One then wants to know the probability that the test and marker loci are linked and, more generally, the probability that the loci lie within *m* centimorgans of one another. These probabilities are not determined in the usual statistical approach (see ref. 3, pp. 131–141). The usual approach involves calculating the probability of the experimental results under various hypotheses, such as the null hypothesis that the loci are not linked. The null hypothesis is commonly rejected when results are obtained that are less likely than 0.05. However, this approach does not take into account the fact that randomly chosen loci are much more likely not to be linked (i.e., to be on different chromosomes) than to be linked (on the same chromosome). We show that, because of the low prior probability of linkage, the null hypothesis should only be rejected when results are obtained that are less likely than 0.004, or else loci that are not linked will frequently be deemed to be linked.

Bayesian statistical analysis (see ref. 4) provides a way of taking into account prior information about the probability of linkage in determining the probability of linkage after a set of experimental results is obtained. This communication describes a Bayesian approach for analyzing linkage experiments using RI strains. We derive formulas for the probability of linkage between test and marker loci, the most likely distance between them in centimorgans, and the probability that they are separated by up to *m* centimorgans, given *I* SDP differences in a set of *N* RI strains. These formulas were evaluated by computer for all values of *I* indicating a high probability of linkage with up to 40 RI strains. The results are easily generalizable to the analysis of Mendelian backcrosses. The advantages and disadvantages of the Bayesian approach are discussed.

METHODS

The numerical evaluation of various integrals and expressions was performed using the computer modeling software MLAB on a DECsystem 10 computer (Division of Computer Research and Technology, National Institutes of Health). To be sure that rounding errors were not leading to significant inaccuracies, many of the calculations were checked with an IBM personal computer using software that performs infinite precision rational arithmetic (Microsoft muMATH symbolic mathematics package, The Soft Warehouse, Honolulu, HI).

RESULTS

Calculation of the Probability that a Test Locus Is Located Within *m* Centimorgans of a Marker Locus from Observations on a Set of RI Strains. Suppose that in a set of RI strains the SDPs for a marker locus and a test locus differ for *I* strains out of *N*. If I/N is small, the loci are likely to be linked, but this result could also arise by chance if the loci were not linked. The probability that the loci are within *m* centimorgans of one another can be calculated by imagining that one picks at random a very large number of loci and then determines, for all those loci whose SDPs have *I* or fewer differences from that of the marker locus, what proportion of these loci are within *m* centimorgans of the marker locus.

We use the following notation: L, linked; $\bar{L}$, not linked; m, test locus is within m centimorgans of the marker locus; I, I or fewer SDP differences in N RI strains; $P(a)$, probability of "a"; $P(a|b)$, probability of "a," given "b." Since loci are either linked or not linked,

$$P(I) = P(I|L)\,P(L) + P(I|\bar{L})\,P(\bar{L}). \qquad [1]$$

According to Bayes' theorem (4), the probability that a test

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: RI strains, recombinant inbred strains; SDP, strain distribution pattern.

locus is within m centimorgans of a marker locus, given I or fewer SDP differences, is

$$P(m|I) = P(I|m)\,P(m)/P(I). \qquad [2]$$

Similarly, the probability of linkage, given I or fewer SDP differences, is

$$P(L|I) = P(I|L)\,P(L)/P(I)$$
$$= \frac{1}{1 + P(\bar{L})\,P(I|\bar{L})/P(L)\,P(I|L)}. \qquad [3]$$

To evaluate the terms in formulas **1–3**, we assume that before the experiment is done, nothing is known about the position of the test locus. In this case it is reasonable to assume that the probability that the test locus lies in any region of length d centimorgans equals d/T, where T is the total length of the genome in centimorgans. This gives

$$P(m) = 2m/T. \qquad [4]$$

(The factor 2 arises because the test locus can be to either side of the marker locus; for simplicity, we limit ourselves to situations in which the marker locus does not lie within m centimorgans of the end of a chromosome.) Similarly, if ℓ is the length of the chromosome containing the marker locus,

$$P(L) = \ell/T. \qquad [5]$$

Since loci are either linked or not linked,

$$P(\bar{L}) = 1 - P(L) = 1 - (\ell/T). \qquad [6]$$

If the loci are not linked, the probability that the SDPs mismatch for J strains and match for $N - J$ strains is $\binom{N}{J}\left(\frac{1}{2}\right)^J\left(\frac{1}{2}\right)^{N-J}$, where $\binom{N}{J}$ is the binomial coefficient $N!/J!(N - J)!$. Therefore,

$$P(I|\bar{L}) = \sum_{J=0}^{I}\binom{N}{J}\left(\frac{1}{2}\right)^N. \qquad [7]$$

To calculate $P(I|m)$, we determine the probability of I or fewer SDP differences, given that the test locus is x centimorgans from the marker locus, and then integrate over x from 0 to m centimorgans on both sides of the marker locus. Let the distance x centimorgans correspond to a frequency of recombination, c, between the two loci in a single meiosis. c can be related to x by an empirical mapping function that takes into account the possibility of multiple recombination events (ref. 3, p. 107). One such mapping function is the Kosambi mapping function (ref. 3, p. 107):

$$c(x) = (e^{2x} - e^{-2x})/(2e^{2x} + 2e^{-2x}). \qquad [8]$$

The recombination frequency, $c(x)$, determines the expected proportion, R, of SDP differences in a set of RI strains:

$$R(x) = 4c(x)/[1 + 6c(x)]. \qquad [9]$$

Formula **9** was originally derived by Haldane and Waddington (5) by directly calculating the probability of recombination after multiple rounds of brother–sister mating. For loci separated by x centimorgans, the probability of J SDP differences in N RI strains is $\binom{N}{J}[R(x)]^J[1 - R(x)]^{N-J}$; the probability of I or fewer SDP differences is, therefore, $\Sigma_{J=0}^{I}\binom{N}{J}R(x)^J[1 - R(x)]^{N-J}$. Integrating over x from $x = 0$ to $x = m$ on both sides of the marker locus gives the probability of I or fewer SDP differences, given that the test locus is within m centimorgans of the marker locus. We need only provide a normalization factor of $1/2m$ so that the probability of N or fewer SDP differences equals 1. This gives

$$P(I|m) = \sum_{J=0}^{I}\binom{N}{J}(1/m)\int_0^m [R(x)]^J[1 - R(x)]^{N-J}dx. \qquad [10]$$

The sum and integral in Eq. **10** occur frequently in the following analysis, so it is convenient to define

$$K_I(m) \equiv \sum_{J=0}^{I}\binom{N}{J}\int_0^m [R(x)]^J[1 - R(x)]^{N-J}dx. \qquad [11]$$

For computational purposes it is helpful to convert the integral over x to an integral over R, using $dx = dR/4(1 - R)(1 - 2R)$ from Eqs. **8** and **9**. This gives

$$K_I(m) = \frac{1}{4}\sum_{J=0}^{I}\binom{N}{J}\int_0^{R(m)} y^J(1 - y)^{N-1-J}/(1 - 2y)\,dy. \qquad [12]$$

The probability of I or fewer SDP differences given that the loci are linked, $P(I|L)$, is derived in the same way as $P(I|m)$ in Eq. **10**, except that the integral is over the entire chromosome bearing the marker locus. If the marker locus is λ centimorgans from the end of the chromosome, then

$$P(I|L) = (1/\ell)\,[K_I(\lambda) + K_I(\ell - \lambda)]. \qquad [13]$$

Computer evaluation of $P(I|L)$ for different values of N, I, ℓ, and λ shows that its value is not sensitive to changes in λ, provided that λ is not close to 0 or ℓ. Thus, for marker loci that are not close to the ends of chromosomes, some simplification is obtained without significant loss in accuracy by choosing $\lambda = \ell/2$.

Substituting these expressions for $P(m)$, $P(L)$, $P(\bar{L})$, $P(I|m)$, $P(I|L)$, and $P(I|\bar{L})$ in Eqs. **2** and **3** leads to

$$P(m|I) = \frac{K_I(m)}{K_I(\ell/2) + [(T - \ell)/2^{N+1}]\sum_{J=0}^{I}\binom{N}{J}} \qquad [14]$$

and

$$P(L|I) = \left[1 + \frac{(T - \ell)\sum_{J=0}^{I}\binom{N}{J}}{2^{N+1}K_I(\ell/2)}\right]^{-1} \qquad [15]$$

The Most Likely Position of the Test Locus with Respect to the Marker Locus Given I SDP Differences. $P(m|I)$ is the probability that the test locus lies anywhere from 0 to m centimorgans from the marker locus, given I. This can be thought of as an integral, over x from $x = 0$ to $x = m$, of the infinitesimal probability that the test locus lies between x and $x + dx$ centimorgans from the marker locus. Therefore, $dP(m|I)/dm$ is proportional to the probability that the test locus lies between m and $m + dm$ centimorgans from the marker locus, given I. By setting $d^2P(m|I)/dm^2 = 0$, one can show that $dP(m|I)/dm$ is maximal at $R(m) = I/N$; therefore, the value of m, designated $\hat{m}$, for which $R(m) = I/N$, is the most likely distance between the test and marker loci. This result is the same as that obtained by another maximal-likeli-

Table 1. Computed probabilities of linkage

N	I	$\hat{m}$	m_{95}	$P(L\|I)$
7	0	0		0.484
8	0	0		0.607
9	0	0		0.724
10	0	0		0.819
11	0	0		0.889
12	0	0		0.934
13	0	0	20.3	0.963
14	0	0	11.7	0.979
15	0	0	9.0	0.989
16	0	0	7.6	0.994
	1	1.7	33.2	0.955
17	0	0	6.7	0.997
	1	1.6	15.3	0.974
18	0	0	6.1	0.998
	1	1.5	11.6	0.985
19	0	0	5.6	0.999
	1	1.4	9.8	0.991
	2	3.1	40.0	0.954
20	0	0	5.1	0.999
	1	1.3	8.6	0.995
	2	2.9	17.7	0.972
21	0	0	4.8	1.000
	1	1.2	7.8	0.997
	2	2.8	13.4	0.983
22	0	0	4.5	1.000
	1	1.2	7.1	0.999
	2	2.6	11.3	0.990
	3	4.2	37.2	0.955
23	0	0	4.2	1.000
	1	1.2	6.6	0.999
	2	2.5	10.0	0.994
	3	4.0	18.9	0.972
24	0	0	4.0	1.000
	1	1.1	6.2	1.000
	2	2.4	9.1	0.997
	3	3.9	14.6	0.983
25	0	0	3.8	1.000
	1	1.1	5.8	1.000
	2	2.3	8.4	0.998
	3	3.7	12.5	0.990
	4	5.3	32.0	0.958
26	0	0	3.6	1.000
	1	1.0	5.5	1.000
	2	2.2	7.8	0.999
	3	3.5	11.1	0.993
	4	5.0	19.3	0.973
27	0	0	3.4	1.000
	1	1.0	5.2	1.000
	2	2.1	7.3	0.999
	3	3.3	10.2	0.996
	4	4.8	15.4	0.983
28	0	0	3.3	1.000
	1	0.9	5.0	1.000
	2	2.0	6.9	1.000
	3	3.2	9.4	0.998
	4	4.6	13.4	0.990
	5	6.1	28.0	0.961
29	0	0	3.2	1.000
	1	0.9	4.7	1.000
	2	1.9	6.5	1.000
	3	3.1	8.8	0.999
	4	4.4	12.0	0.994
	5	5.8	19.2	0.975
30	0	0	3.0	1.000
	1	0.9	4.5	1.000
	2	1.9	6.2	1.000
	3	2.9	8.3	0.999
	4	4.2	11.0	0.996
	5	5.6	15.9	0.984
31	0	0	2.9	1.000
	1	0.8	4.3	1.000
	2	1.8	5.9	1.000
	3	2.8	7.8	1.000
	4	4.0	10.3	0.998
	5	5.3	14.0	0.990
	6	6.9	25.3	0.966
32	0	0	2.8	1.000
	1	0.8	4.1	1.000
	2	1.7	5.7	1.000
	3	2.7	7.4	1.000
	4	3.9	9.6	0.999
	5	5.1	12.7	0.994
	6	6.6	19.1	0.978
33	0	0	2.7	1.000
	1	0.8	4.0	1.000
	2	1.7	5.4	1.000
	3	2.6	7.1	1.000
	4	3.7	9.1	0.999
	5	4.9	11.7	0.996
	6	6.3	16.2	0.986
	7	7.8	40.8	0.954
34	0	0	2.6	1.000
	1	0.8	3.8	1.000
	2	1.6	5.2	1.000
	3	2.5	6.8	1.000
	4	3.6	8.6	1.000
	5	4.7	10.9	0.998
	6	6.0	14.4	0.991
	7	7.5	23.5	0.970
35	0	0	2.5	1.000
	1	0.7	3.7	1.000
	2	1.6	5.0	1.000
	3	2.5	6.5	1.000
	4	3.5	8.2	1.000
	5	4.6	10.3	0.999
	6	5.8	13.2	0.994
	7	7.2	18.9	0.980
36	0	0	2.4	1.000
	1	0.7	3.6	1.000
	2	1.5	4.8	1.000
	3	2.4	6.2	1.000
	4	3.3	7.8	1.000
	5	4.4	9.8	0.999
	6	5.6	12.3	0.996
	7	6.9	16.4	0.987
	8	8.4	31.4	0.960
37	0	0	2.4	1.000
	1	0.7	3.4	1.000
	2	1.5	4.6	1.000
	3	2.3	6.0	1.000
	4	3.2	7.5	1.000
	5	4.2	9.3	1.000
	6	5.4	11.5	0.998
	7	6.6	14.8	0.992
	8	8.1	22.3	0.973
38	0	0	2.3	1.000
	1	0.7	3.3	1.000
	2	1.4	4.5	1.000
	3	2.2	5.7	1.000
	4	3.1	7.2	1.000
	5	4.1	8.8	1.000
	6	5.2	10.9	0.999
	7	6.4	13.7	0.995
	8	7.8	18.6	0.982
	9	9.3	56.6	0.950
39	0	0	2.2	1.000
	1	0.7	3.2	1.000
	2	1.4	4.3	1.000
	3	2.2	5.5	1.000
	4	3.0	6.9	1.000
	5	4.0	8.5	1.000
	6	5.0	10.3	0.999
	7	6.2	12.8	0.997
	8	7.5	16.5	0.988
	9	8.9	27.2	0.966
40	0	0	2.2	1.000
	1	0.6	3.1	1.000
	2	1.4	4.2	1.000
	3	2.1	5.3	1.000
	4	2.9	6.6	1.000
	5	3.9	8.1	1.000
	6	4.9	10.0	0.999
	7	6.0	12.0	0.998
	8	7.2	15.1	0.992
	9	8.6	21.3	0.977

N, number of RI strains analyzed; I, number of SDP differences detected; $\hat{m}$, most likely distance between the loci; m_{95}, distance in centimorgans such that the interval extending m_{95} centimorgans to both sides of the marker locus has a 95% chance of including the test locus; $P(L|I)$, probability that the loci are linked, given I or fewer SDP differences. m_{95} exists only if $P(L|I) > 0.95$.

hood procedure (ref. 3, p. 137). Formulas **8** and **9** can be used to calculate $\hat{m}$, given I and N (Table 1).

Maximal Values of I for Which the Probability of Linkage Is >95%. Formula **15** can be used to determine the largest value of I for which the probability of linkage, given I or fewer SDP differences, is >95%. Substituting T = 1600, an estimate for the size of the mouse genome in centimorgans (ref. 3, p. 99), and ℓ = 115, the approximate size of the largest mouse chromosome (ref. 3, p. 99), we evaluated formula **15** by computer for various values of I and N (Table 1). As expected, the probability of linkage decreases as the proportion of SDP differences increases. The largest number of SDP differences for which the probability of linkage is >95% can be read from Table 1 for different values of N. To save space, for values of $N > 12$, we included in Table 1 only values of I for which the probability of linkage is >95%. The largest values of I from Table 1 that correspond to a probability of linkage >95% are smaller than the maximal numbers of SDP differences calculated by other methods (ref. 3, pp. 131–141). For $N < 13$, the probability of linkage is <95% even for $I = 0$. Thus, the use of fewer than 13 RI strains will never indicate linkage with >95% certainty. Further, Table 1 shows that for all combinations of I and N for which the probability of linkage is >95%, the most likely distance between test and marker loci, $\hat{m}$, is <9.3 centimorgans. Thus, only close linkage can be detected with confidence by use of RI strains.

Calculation of the Interval Around the Marker Locus That Has a 95% Chance of Containing the Test Locus. We used a computer to solve Eq. **14** for the value of m, designated m_{95}, for which $P(m|I) = 0.95$, using $T = 1600$ and $\ell = 115$ (Table 1). These values of m_{95} form a 95% confidence interval such that, given I or fewer SDP differences, there is a 95% probability that the interval extending m_{95} centimorgans to both sides of the marker locus contains the test locus.

Application to Analysis of Backcross Data. Formulas **1–3** and **5–7** apply equally well to a backcross, where I is the number of recombinants in a set of N backcross individuals. The conditional probability of I or fewer recombinants, given that the test locus is within m centimorgans of the marker locus, becomes (analogous to Eq. **10**)

$$P(I|m) = \sum_{J=0}^{I} \binom{N}{J}(1/m)\int_0^m [c(x)]^J[1 - c(x)]^{N-J}dx. \qquad [16]$$

$c(x)$ appears in Eq. **16** instead of $R(x)$ because $c(x)$ is the expected proportion of backcross individuals carrying chromosomes recombinant for the two loci and is analogous to $R(x)$, the expected proportion of RI strains carrying recombinant chromosomes. Substitution of these expressions in Eq. **2** gives the desired formula for $P(L|I)$.

Upper Limit for the Probability of Linkage Given I or Fewer Recombinants in a Backcross or I or Fewer SDP Differences in a Set of RI Strains. It is useful to have an upper limit for the probability of linkage that does not require a computer for evaluation. Formula **3** can be simplified by substituting $P(L) = \ell/T$ and $P(\bar{L}) = 1 - \ell/T$ and noting that $P(I|L) \leq 1$. This gives

$$P(L|I) \leq 1/[1 + (T - \ell)(1/\ell)P(I|\bar{L})]. \qquad [17]$$

The probability on the right hand side of Eq. **17**, $P(I|\bar{L})$, plays a critical role in the usual statistical analysis of backcrosses and RI strains. $P(I|\bar{L})$ is the probability of the experimental results given the null hypothesis. This probability may be calculated approximately from a χ^2 analysis (see *Appendix* and ref. 6). Designating the result P_χ, we have the result

$$P(L|I) \leq 1/[1 + (T - \ell)(1/\ell)P_\chi]. \qquad [18]$$

Two examples illustrating the use of formula **18** are given in the *Appendix*.

DISCUSSION

The major difference between the Bayesian approach used here and the more traditional analysis of linkage (ref. 3, pp. 33–34) is that the Bayesian approach takes into account that, when nothing is known to begin with about the position of the test locus, the prior probability of linkage to any particular marker locus is small. In the mouse this probability, $P(L)$, varies from about 0.03 for marker loci on the smallest mouse chromosome to about 0.07 for marker loci on the largest. The low prior probability of linkage means that there is a greater chance of misclassifying nonlinked loci as linked unless stricter criteria than usual are adopted for rejecting the null hypothesis. To see this, note that Eq. **17** may be inverted to give

$$P(I|\bar{L}) \leq \{[1/P(L|I)] - 1\}\ell/(T - \ell). \qquad [19]$$

If we require that $P(L|I)$, the probability of linkage given the experimental results, be >95%, then $P(I|\bar{L})$, the probability of the observed results assuming the null hypothesis, must be less than $\ell/[19(T - \ell)]$, or about 0.001 for marker loci on the smallest mouse chromosome to about 0.004 for those on the largest. This contrasts with the cut-off value $P(I|\bar{L}) < 0.05$ commonly used to reject the null hypothesis.

When RI strains are used to screen for linkage, it may make sense to consider as potentially linked those loci for which there is only a moderate probability of linkage. This is equivalent to choosing a fairly large cut-off for $P(I|\bar{L})$, such as 0.05, as the criterion for rejecting the null hypothesis. In this case, additional experiments would need to be done to determine which of the possibly linked loci were, in fact, linked. If no further experiments were planned, then it would be appropriate to refrain from concluding that linkage exists, unless the number of SDP differences were sufficiently small that the probability of linkage were greater than, say, 95% (Table 1).

The Bayesian approach requires an assumption about the prior likelihood of linkage. In the absence of any information, it is reasonable to assume that the test locus has a uniform prior probability distribution over the genome. Although uniform prior probability distributions are controversial, their use has been staunchly defended by Bayesian theorists (7). If the test locus were known to be on a certain chromosome from analysis of somatic-cell hybrids, then the formulas described above could be modified by setting $P(L) = 1$ and $P(\bar{L}) = 0$ for marker loci on this chromosome. On the other hand, if more detailed information about position on a particular chromosome were available (e.g., from *in situ* hybridization or analysis of a backcross), then a Bayesian approach would be more difficult to apply because of increased computational complexity.

From a conceptual point of view, it is useful to contrast the questions posed by the Bayesian approach and the more traditional statistical analysis of linkage data (ref. 3, pp. 33–34). The Bayesian approach answers the question "Given the observed results, what is the probability that the two loci are separated by up to m centimorgans?" The more traditional analysis answers the question "If the loci were separated by m centimorgans, how unlikely would the observed results be?" We believe that the answer to the former question corresponds more closely to what the experimenter wants to know.

Both the traditional analysis and the Bayesian approach lead to 95% confidence intervals for predictions of map position (Table 1 and ref. 8). The Bayesian confidence interval answers the question "Given the observed results, what interval around the marker locus has a 95% probability of containing the test locus?" The more traditional confidence interval answers the question "What is the greatest distance between the test and marker loci for which the probability of the observed results is still greater than 5%?" It is remarkable that, for most values of I and N, these confidence intervals are reasonably close, despite their very different meanings.

The Bayesian approach is clearly not limited to the analysis of data from RI strains. As indicated above, it can be applied with very little modification to analysis of a backcross. The major limitations in comparison with more traditional analysis of linkage are greater computational difficulty when exact results are required and the problem of determining prior probabilities. However, powerful computational resources are now fairly widespread, and in some situations, such as when nothing is known about the location of the locus to be mapped, the prior probability distribution is straightforward. Since the Bayesian approach takes into account more information about the experimental situation, we suggest that it be used more frequently in the statistical analysis of genetic experiments.

APPENDIX

The following examples illustrate the use of formula **18** and the more conservative interpretation provided by the Bayesian analysis.

Example 1: Suppose that in a set of 26 RI strains derived from parental strains A and B, one observes the following pattern of inheritance of parental alleles at test and marker loci:

Marker locus	Test locus	No. of strains
A	A	10
B	B	11
A	B	2
B	A	3

If the loci were not linked, one would expect one-fourth of the strains in each group. Calculating χ^2 (with Yates correction; see ref. 6) from these data,

$$\chi^2 = \sum \left(|E - 0| - \frac{1}{2}\right)^2 / E = 7.7.$$

For 1 degree of freedom, this gives $P \approx 0.005$. Thus, these results would be very unlikely if the loci were not linked. However, this does not imply a correspondingly high probability of linkage. Suppose the marker locus lay on a mouse chromosome of length $\ell = 100$ centimorgans. Then from Eq. **18**, $P(L|I) \leq 1/[1 + (1600 - 100)(1/100)(0.005)] = 0.93$. Thus, even though the experimental results would be very unlikely if the loci were not linked, the probability of linkage is still <0.95. A more detailed calculation of $P(L|I)$ from Eq. **15** yields $P(L|I) = 0.91$, indicating that the upper limit from Eq. **18** is fairly close to the actual value.

Example 2: Suppose that the progeny of a backcross A × (A × B) are found to have inherited alleles from the heterozygous parent as follows:

Marker locus	Test locus	No. of individuals
A	A	32
B	B	30
A	B	18
B	A	20

If the loci were not linked, one would expect 25 animals (one-fourth of the total) in each group. The usual statistical analysis gives $\chi^2 = \Sigma(E - 0)^2/E = 5.92$, which corresponds to $P \approx 0.02$. These results would ordinarily be interpreted as providing reasonable evidence for linkage. However, suppose the marker locus lay on mouse chromosome 19, which is about 42 centimorgans long (ref. 3, p. 99). Then from Eq. **18**, $P(L|I) \leq 1/[1 + (1600 - 42)(1/42)(0.02)] = 0.57$. Thus, the Bayesian result indicates that the probability of linkage is, in fact, only about 50%. A more detailed calculation from Eqs. **2** and **16**, using the Kosambi mapping function yields, for $N = 100$ and $I = 38$, $P(L|I) = 0.56$.

These examples demonstrate the need for caution before concluding that two loci are linked, when the standard χ^2 analysis is used and $P(I|\bar{L})$ is not <0.004.

We thank Dr. David Alling and Dr. Earl Green for critical review of previous versions of the manuscript.

1. Taylor, B. A. (1978) in *Origins of Inbred Mice*, ed. Morse, H. C., III (Academic, New York), pp. 423–438.
2. Bailey, D. W. (1981) in *The Mouse in Biomedical Research*, eds. Foster, H. L., Small, J. D. & Fox, J. G. (Academic, New York), Vol. 1, pp. 223–239.
3. Green, E. L. (1981) *Genetics and Probability in Animal Breeding Experiments* (Oxford Univ. Press, New York).
4. Lindley, D. V. (1965) *Introduction to Probability and Statistics from a Bayesian Viewpoint* (Cambridge Univ. Press, London), p. 20.
5. Haldane, J. B. S. & Waddington, C. H. (1931) *Genetics* **16**, 357–374.
6. Snedecor, G. W. & Cochran, W. G. (1972) *Statistical Methods* (The Iowa State Univ. Press, Ames, Iowa), 6th Ed., pp. 211–213.
7. Jeffreys, H. (1961) *Theory of Probability* (Oxford University Press, Oxford), 3rd Ed., pp. 117–192, 401–424.
8. Silver, J. (1985) *J. Hered.* **76**, 436–440.