Theoretical and Applied Genetics 44, 270–271 (1974)
© by Springer-Verlag 1974

A Procedure for Computer Simulation of Linkage with Interference

J. L. FYFE

Scottish Plant Breeding Station, Pentlandfield, Roslin, Midlothian (Scotland)

Summary. In computer simulation of plant breeding procedures it is important to limit multiple recombination by allowing for interference. This can be approximated by using Kosambi's rule for relating coincidence to recombination.

Introduction

The procedure to be described was needed for a program to simulate the execution of different plant breeding strategies, by interactive work at a terminal of a multi-access computer system. The genetical situation is represented as two pairs of chromosomes, having l_1 and l_2 loci (marked with 0 or 1) delimiting $l_1 - 1$ and $l_2 - 1$ contiguous intervals. A fairly realistic representation of linkage was needed, with the probabilities of multiple recombination limited in approximation to the observed action of interference. The use of "masks" as described by Fraser and Burnell (1970) allows the probabilities of all possible recombinants to be calculated once and once only for a given situation, but their method of calculating the probabilities assumes zero interference. I could not find any published procedure simulating interference.

A comprehensive account of mapping functions is given by Bailey (1961). For present purposes, a simple modifying of the formulae proposed by Kosambi (1944) was considered adequate.

Following the notation used by Bailey

$$C = y_{1.2}/y_1 y_2 , \qquad (1)$$

where C = coincidence, $y_{1.2}$ = double recombination, y_i = recombination in interval i. Kosambi takes $C = 2y$, leading to the addition formula

$$y_{1+2} = (y_1 + y_2)/(1 + 4y_1 y_2) . \qquad (2)$$

This implies that C varies continuously between 0 and 1; for example, in a three-point linkage, with loci a and b delimiting interval 1, b and c delimiting interval 2, there would be no interference ($C = 1$) if a and c recombined freely ($y_{1+2} = 0.5$) but almost complete interference ($C \to 0$) if a and c recombined rarely ($y_{1+2} \to 0$).

The procedure described below assumes (a) that recombination proceeds sequentially, with events in interval i affecting those in interval j only if $j > i$; (b) that interference by a crossover does not extend past the next crossover and (c) that coincidence never exceeds unity.

Procedure

Instead of using y_i or $1 - y_i$ in calculating the probability of a mask representing recombination or its absence in interval i, a temporary array t_i of probabilities is used, initially having the values y_i. If a mask represents no crossover in interval i (..00.. or ..11..), multiplication is by $(1 - t_i)$ and the next interval is examined without further action. If the mask represents a crossover in interval i (..01.. or ..10..), multiplication is by t_i. Then intervals i and $i + 1$ are summed by formula (2) and C is given the value $2y_{i+(i+1)}$ or 1, whichever is the less. If $C = 1$, there is no further action in respect of the crossover in interval i but otherwise t_{i+1} is altered to Cy_{i+1}.

Adjustment of t_{i+2} for the influence of a crossover in interval i involves adding interval $i + 2$ to the compound interval $i + (i + 1)$ and calculating C again ($2y_{(i+(i+1))+(i+2)}$ or 1). Again, the consequences of a crossover in i have ended if $C = 1$, otherwise t_{i+2} becomes Cy_{i+2}. Ultimately, if C never becomes 1, adjustments resulting from a crossover in i extend to the last interval of the chromosome.

The next interval, $i + 1$, is then tested for representation of a crossover. If there is none, there is no effect on the values t_i. If a crossover is represented, new t_i values are calculated for the succeeding intervals exactly as above, referring back to the original y_i values, regardless of whether there had or had not been a previous crossover.

The masks are generated by counting down from $2^l - 1$ to $2^{(l-1)}$, e.g. with $l = 4$, from 1111 to 1000 in binary. After the probabilities of all masks have been computed, the next step is to order them in descending probability and compute the cumulative probabilities, as described by Fraser and Burnell. All the steps so far described make up a routine called "masker".

Every interval after the first suffers interference from preceding intervals and so the total recombination in each interval 2 to $l - 1$ has to be calculated, say y'_i. The routine "masker" is then called again with y_i replaced by $2y_i - y'_i$. The total recombina-

tion is again calculated, say y_i''. The routine is called a third time with y_i replaced by $y_i + y_i' - y_i''$. The total recombination in each interval is again calculated and printed, for comparison with the "target" values. In extensive tests, agreement has always been better than five parts per thousand.

Finally the masks and their cumulative probabilities are filed for future use by the program simulating plant breeding. Here they are used to produce representations of gametes as described by Fraser and Burnell, except that the starting chromosome is chosen by generating a random real number between 0 and 1 with 0.5 as the critical value.

Discussion

The Kosambi rule, $C = 2y$, makes some allowance for interference without invoking additional parameters. It is an improvement over the assumption of zero interference; in eight cases where the procedure described was tested with published data of three- and four-point experiments it always gave a better fit than assuming zero interference. However, the rule is only an approximation, regarded by A.R.G. Owen (see Bailey *loc. cit.* for references) as representing the situation near the middle of a chromosome arm; in three of the eight cases, the fit was still unsatisfactory.

For the present purposes, the procedure is probably accurate enough. The essential requirement is to avoid making a desired recombinant equally likely to occur in one round of gamete formation as in several, given the same total of gametes produced by heterozygotes. Reverting to the example, suppose the crop to be normally self-fertilized, the desired recombinant to be ABC and the starting point the heterozygote AbC/aBc. With $y_1 = y_2 = 0.05$, assuming no interference gives an expectation $y_{1.2} = = 0.0025$ but the Kosambi formulae give 0.00050, which is less by a factor of 5 (with $y_1 = y_2 = 0.1$ the expectations would be 0.01 and 0.0038 respectively). A strategy which allowed recombination to occur through several generations would have an advantage to set against the delay in reaching homozygosity and the advantage would be underestimated if interference were assumed to be zero.

Note: The program which carries out the whole procedure described is written in IMP, the Edinburgh Regional Computing Centre's development of Atlas Autocode. A listing can be provided on request.

Literature

Bailey, N. T. J.: Introduction to the mathematical theory of genetic linkage, pp. 298. Oxford: Clarendon Press 1961.

Fraser, A., Burnell, D.: Computer models in genetics, pp. 206. New York: McGraw Hill 1970.

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

Received April 1973

Communicated by R. W. Allard

Mr. J. L. Fyfe
Scottish Plant Breeding Station
Pentlandfield
Roslin, Midlothian EH25 9RF (Scotland)