Genet. Res., Camb. (1981), **37**, *pp.* 239–252
Printed in Great Britain

Inbreeding for isogeneity by backcrossing to a fixed parent in haploid and diploid eukaryotes

By JOHN F. LESLIE

Department of Biological Sciences, Stanford University, Stanford, California 94305, U.S.A.

(*Received* 8 *July* 1980 *and in revised form* 29 *October* 1980)

SUMMARY

The consequences of repeated backcrossing to a fixed parent are examined for haploid eukaryotes having a transitory diploid phase. The isogenicity attained in the absence of selected markers depends on the number of chromosomes and the total genetic map length, while the isogenicity of a chromosome carrying a selected marker increases more slowly and depends on the size of the chromosome. As inbreeding proceeds, the remaining non-isogenic material is not distributed evenly to all of the progeny. Instead, the majority of the progeny are completely isogenic with the fixed parent (with the exception of a region surrounding each selected marker), while the non-isogenic material is concentrated in a minority of the progeny. Even when the average isogenicity of the progeny and the fixed parent exceeds 99%, a significant proportion of the progeny will contain tracts of non-isogenic material which average several map units in length. Minor modifications enable these results to be applied to diploids. Examples show how to determine the degree of isogenicity produced by a given number of backcrosses in several specific situations.

1. INTRODUCTION

It is frequently necessary to transfer a gene from one genetic background to another. In haploid eukaryotic organisms, such as the fungi, this transfer is most readily accomplished by backcrossing to a fixed parent. Such backcrossing can be used (1) to transfer alleles of a single locus from the original genetic background to a standard genetic background; (2) to screen natural populations for multiple interdependent recessive genes that affect the diploid portion of the life-cycle; and (3) to construct a series of isogenic strains, differing only in particular regions, to be used for studies of heterokaryon incompatibility genes (Mylyk, 1975, 1976), genes that control recombination (Catcheside, 1977), and genes from a natural population (Perkins, Turner & Barry, 1976; Leslie & Leonard, 1980).

The mathematics of inbreeding for isogenicity is well known for brother–sister matings in diploids (cf. Detlefsen, 1914; Jennings, 1916; Haldane, 1936; Fisher, 1965; Crow & Kimura, 1970; Stam, 1980). For haploids the mathematical calculations are as simple, but, to my knowledge, the formulas have not been previously

derived and are not available, for example, in the two most recent fungal genetics texts (Burnett, 1975; Fincham, Day & Radford, 1979). The purpose of this communication is to present these calculations, which may be of use to mycologists, biochemists, plant and animal breeders, and geneticists using haploid and diploid eukaryotes.

In each section the derivation of the appropriate equations is presented first, followed by one or more examples. The examples may be followed without an understanding of the derivations.

2. DEFINITIONS OF TERMS AND SYMBOLS

Allogenic. Two alleles, or chromosome segments, of independent origin. (Cotterman, as cited by Crow & Kimura, 1970.)

Isogenic. Two alleles, or chromosome segments, that are known to be identical by descent.

Generation. Haploid parents fuse to make a diploid zygote. Following meiosis haploid progeny are produced. The haploid parents are the parental or nth generation; the haploid progeny are the filial or $n+1$th generation.

Junction. The discontinuity connecting portions of strands of unlike origin. This discontinuity was generated by a crossover in an allogenic region (after Fisher, 1965).

$A(n)$. The average number of centimorgans per genome remaining allogenic after n generations of backcrossing to a fixed parent.

$a(n)$. The average number of centimorgans per genome not in tracts containing a selected marker that remain allogenic after n generations of backcrossing to a fixed parent.

$B(n)$. The average length of an allogenic tract not containing a selected marker after n successive crosses to a fixed parent.

c. Haploid chromosome number.

$C(n)$. Average number of chromosome tips remaining allogenic after n backcrosses to a fixed parent.

$J(n)$. The average number of junctions per centimorgan present after n successive backcrosses of progeny with the fixed parent.

$j(i)$. The number of junctions formed per centimorgan in the ith backcross to a fixed parent.

k. The number of segregating selected loci.

$L_l(x)$. The total number of centimorgans on chromosome x from the segregating marker to the left tip of the chromosome.

$L_r(x)$. The total number of centimorgans on chromosome x from the segregating marker to the right tip of the chromosome.

M. The total number of centimorgans in the genetic map of an organism.

$m(x)$. The total number of centimorgans in the genetic map of chromosome x.

n. The number of generations of backcrosses of the progeny with the fixed parent.

$Q(x,n)$. The average allogenicity of chromosome x after n backcrosses to the fixed parent.

$T(n)$. The average number of allogenic tracts remaining per genome after n generations of backcrossing to a fixed parent.

$t(n)$. The average number of allogenic tracts per genome not containing a selected marker remaining after n generations of backcrossing to a fixed parent.

$Z(0)$. The allogenicity of the two parental strains. One parent will be the fixed parent, while the other parent will carry the trait to be transferred to the genetic background of the fixed parent. If the two strains are unrelated, as will usually be the case, then $Z(0) = 1$.

$Z(n)$. The average allogenicity of the progeny of the nth generation with the fixed parent.

$z_l(x,n)$. The average allogenicity, after n backcrosses to a fixed parent, of the region on chromosome x from the left tip of the chromosome to $100/n$ centimorgans from the left side of the segregating marker.

$z_r(x,n)$. The average allogenicity after n backcrosses to a fixed parent of the region on chromosome x from the right tip of the chromosome to $100/n$ centimorgans from the right side of the segregating marker.

3. INBREEDING FOR CHROMOSOMES WITHOUT A SELECTED MARKER

In backcrossing to a fixed parent, if you consider only one given allele from the other parent, the probability of retaining it is $(\frac{1}{2})^n$, where n is the number of backcrosses. If, however, you want to be reasonably sure that you have eliminated *all* of the genes on that chromosome, then the problem is more complex, because crossing over prevents the whole chromosome from being inbred as a unit. Thus, after n generations of backcrossing, the loci that have not become isogenic will not be peppered at random throughout the genome, but will be linked together in tracts, and some progeny will carry these tracts while others will not. Consequently, even if inbreeding is continued until the *average* allogenicity is quite low, there will still be a significant proportion of the progeny that carry allogenic tracts.

This section considers the effects of inbreeding on chromosomes without selected markers. In later sections, when chromosomes carrying selected markers are considered, the calculations and formulas presented in this section will be modified. The discussion will now follow that of Fisher (1965).

Let $Z(0)$ be the initial allogenicity, let M be the total genetic map length, and let n be the number of generations that the progeny have been mated back to a fixed parent. Then, $Z(n)$, the allogenicity of the progeny of the nth generation with the fixed parent is

$$Z(n) = (\tfrac{1}{2})^n Z(0). \tag{1}$$

If the strains used to make the first cross are unrelated (i.e. the two parental strains are completely allogenic), then $Z(0)$, the initial allogenicity, is 1, and

$$Z(n) = (\tfrac{1}{2})^n. \tag{2}$$

Equation (2) may be used to compute the probability that a single gene will survive to the nth generation or to compute the average proportion of the genome which remains allogenic in the nth generation.

Each recombination event in an allogenic region will create a discontinuity connecting portions of strands of unlike origin. This discontinuity will be termed a junction. Backcrossing to a fixed parent generates only the type of junction characterized as 'external' by Fisher (1965). On one side of an external junction, the parents are isogenic with respect to the fixed parent, while on the other side of the junction they are allogenic. In the process of backcrossing to a fixed parent, all of the junctions, except for those close to selected markers, will eventually disappear.

The concepts of junctions and average allozygosity will now be related to each other and to the number of generations of backcrossing. Consider the number of junctions formed per centimorgan in the $i+1th$ cross to the fixed parent. By definition there are, on average, two recombination events per tetrad per 100 centimorgans (CM). New junctions can be formed only in that portion of the genome that is still allogenic. Thus, one new junction per strand will be generated for every 100 CM of allogenic region remaining. For the $i+1th$ backcross, the proportion of the genome remaining allogenic is $Z(i)$. Thus, the number of new junctions formed per centimorgan in the $i+1th$ backcross is $0{\cdot}01\,Z(i)$, which is

$$j(i+1) = 0{\cdot}01\,Z(i) = 0{\cdot}01(\tfrac{1}{2})^i. \tag{3}$$

The number of junctions generated per generation has been calculated. The average number of junctions current per centimorgan following n backcrosses, $J(n)$, will now be calculated. Only those junctions that survive from the generation in which they originated until generation n will contribute to the allogenicity of the progeny of the nth generation. The probability of a junction appearing in generation n surviving to generation $n+1$ is the same as the probability of a single locus surviving from generation n to generation $n+1$, or $\frac{1}{2}$. Thus the number of junctions current per centimorgan is one-half the number of junctions in the previous generation plus the number of junctions generated by the meiosis that produces the current generation, or

$$J(n+1) = \tfrac{1}{2}J(n)+j(n+1) = \tfrac{1}{2}J(n)+0{\cdot}01Z(n). \tag{4}$$

If completely allogenic strains are used in the initial cross, then $J(0) = 0$, $Z(0) = 1$, and

$$J(n) = 0{\cdot}01n(\tfrac{1}{2})^{n-1}. \tag{5}$$

After n generations of backcrossing to a fixed parent, let the remaining allogenic material be divided into $T(n)$ allogenic tracts. Each of these $T(n)$ tracts must end either at a chromosome tip or at a junction. If the haploid number of chromosomes is c, then $C(n)$, the average number of chromosome tips allogenic after n backcrosses, is

$$C(n) = 2c(\tfrac{1}{2})^n = c(\tfrac{1}{2})^{n-1}. \tag{6}$$

The average number of junctions present in generation n can be calculated using equation (5) to be $MJ(n)$. The average number of allogenic tracts remaining is one half the number of endpoints for the tracts, so the number of allogenic tracts remaining is

$$T(n) = \frac{[MJ(n)+C(n)]}{2} = \frac{0{\cdot}01nM(\frac{1}{2})^{n-1}+c(\frac{1}{2})^{n-1}}{2} = (0{\cdot}01nM+c)(\tfrac{1}{2})^{n}. \qquad (7)$$

The proportion of progeny with 0, 1, 2, ... allogenic tracts remaining can be estimated using a Poisson distribution with mean $T(n)$.

The average total length of the allogenic tracts can also be determined. After n backcrosses to a fixed parent

$$A(n) = MZ(n) = M(\tfrac{1}{2})^{n} \qquad (8)$$

centimorgans will remain allogenic. The average length, $B(n)$, of one of the remaining allogenic tracts is

$$B(n) = \frac{A(n)}{T(n)} = \frac{M(\frac{1}{2})^{n}}{(0{\cdot}01nM+c)(\frac{1}{2})^{n}} = \frac{M}{0{\cdot}01nM+c}. \qquad (9)$$

Example. Consider an organism that has 7 chromosomes and a total genetic map length of 1000 cM after 10 generations of backcrossing to a fixed parent, starting with unrelated parents. Then, by definition, $n = 10$ generations, $Z(0) = 1$, $M = 1000$ cM, and $c = 7$ chromosomes.

Using equation (1), $Z(10)$, the average proportion of the genome remaining allogenic can be calculated as

$$Z(10) = (\tfrac{1}{2})^{10} = 9{\cdot}77 \times 10^{-4}.$$

Using equation (7), $T(10)$, the average number of allogenic tracts remaining per genome can be calculated as

$$T(10) = [(0{\cdot}01 \times 10 \times 1000)+7](\tfrac{1}{2})^{10} = 0{\cdot}104.$$

Using equation (8), $A(10)$, the average number of centimorgans per genome can be calculated as

$$A(10) = 1000(\tfrac{1}{2})^{10} = 0{\cdot}977.$$

Using equation (9), $B(10)$, the average length of each of the remaining allogenic tracts can be calculated as

$$B(10) = \frac{1000}{(0{\cdot}01 \times 10 \times 1000)+7} = 9{\cdot}35 \text{ cM}.$$

The portion of the progeny carrying 0, 1, and more than 1 allogenic tracts is given by a Poisson distribution with mean $T(10) = 0{\cdot}104$, or 0 tracts allogenic – 90·1 %, 1 tract allogenic – 9·4 %, and more than 1 tract allogenic – 0·5 %.

Thus, after 10 generations of backcrossing to a fixed parent, the average allogenicity of the progeny is < 0·1 % and 90·1 % of the progeny will be completely isogenic with the fixed parent. The remaining 9·9 % of the progeny will have an average of at least 9·35 cM of allogenic material remaining. To be 99 % certain

that a sample from the progeny contains at least one isolate that is completely isogenic with the fixed parent, then at least 2 isolates must be analyzed, since $(9{\cdot}9\,\%)^2 < 1\,\%$.

EXAMPLE. The effects of genome size and *rec* genes.

For some haploid organisms, *rec* genes that affect the amount of recombination are known (Catcheside, 1977). In all known cases, low crossing-over is dominant to high crossing-over. Consequently, any high crossing-over *rec* genes in the fixed parent will cause increased crossing over only after they have become homozygous. The maximum effect of these genes can be estimated by assuming that the high crossing over *rec* genes are homozygous during the entire inbreeding process.

Consider an organism identical to that of the previous example, i.e. with an identical amount of DNA differing only in *rec* genes, such that the total genetic map length is now 4000 cM. After 10 generations of backcrossing to a fixed parent starting with unrelated parents, $n = 10$, $Z(0) = 1$, $M = 4000$ cM, and $c = 7$.

Using equation (1), $Z(10)$, the average proportion of the genome remaining allogenic can be calculated as

$$Z(10) = (\tfrac{1}{2})^{10} = 9{\cdot}77 \times 10^{-4}.$$

Thus, the *rec* genes have no influence on the average proportion of the genome which remains allogenic.

Using equation (7), $T(10)$, the average number of allogenic tracts remaining per genome can be calculated as

$$T(10) = [(0{\cdot}01 \times 10 \times 4000) + 7](\tfrac{1}{2})^{10} = 0{\cdot}397.$$

Thus the *rec* genes result in an increase in the average number of allogenic tracts that is approximately proportional to the increase in the total genetic map length.

Using equation (8), $A(10)$, the average number of centimorgans remaining allogenic can be calculated as

$$A(10) = 4000(\tfrac{1}{2})^{10} = 3{\cdot}91.$$

Thus the *rec* genes result in an increase in the average number of centimorgans remaining allogenic, which is exactly proportional to the increase in the total genetic map length.

Using equation (9), $B(10)$, the average length of each of the remaining allogenic tracts can be calculated as

$$B(10) = \frac{4000}{(0{\cdot}01 \times 10 \times 4000) + 7} = 9{\cdot}83 \text{ cM}.$$

Thus, the *rec* genes result in only a minor increase in the average length of the remaining allogenic tracts.

The portion of the progeny carrying 0, 1 and more than 1 allogenic tracts is given by a Poisson distribution with mean $T(10) = 0{\cdot}397$, or 0 tracts allogenic –

67·2 %, 1 tract allogenic – 26·7 %, and more than 1 tract allogenic – 6·1 %. Thus the *rec* genes result in an increase in the portion of progeny carrying allogenic tracts that is not proportionate to the increase in total genetic map length.

With the altered *rec* gene constitution, the average allogenicity of the progeny is still $< 0{\cdot}1\,\%$ after 10 generations of backcrossing to a fixed parent, but only 67·2 % of the progeny now, as compared with 90·1 % of the progeny in the previous example, will be completely isogenic with the fixed parent. Thus, the proportionate increase in the total genetic map length caused by the *rec* genes does not change the physical amount of DNA that remains allogenic. However, instead of being limited to one or two allogenic tracts in less than 10 % of the progeny, almost one-third of the progeny, 32·8 %, will now have at least one allogenic tract remaining. Thus, to be 99 % certain that a sample from the progeny contains at least one isolate that is completely isogenic with the fixed parent, then at least 5 isolates must be analysed, since $(32{\cdot}8\,\%)^5 < 1\,\%$.

4. INBREEDING FOR A CHROMOSOME CARRYING A SELECTED MARKER

In this section the effects of backcrossing to a fixed parent are considered when a single chromosome carries a single selected marker. As expected, the allogenicity of the region surrounding the selected marker decreases slowly, while the allogenicity of the remainder of the genome decreases much more rapidly.

Consider a chromosome with a marker G/g that is L_l map units from the left end and L_r map units from the right end. The regions to the left and right of G/g are considered separately, and terms relating to left and right are distinguished by the subscripts l and r.

The breeding scheme differs from that in the previous section only with respect to G/g. Suppose the fixed parent is g and that the progeny selected for backcrossing are always G. The allogenic tracts will then be those regions descended from the original G parent. It is well known (cf. Crow & Kimura, 1970) that the average length of an allogenic region to the left of G after n generations of inbreeding is $100/n$ cM for $n \geqslant 10$.

The allogenicity of the remainder of the chromosome will be assumed independent of the interval around G. This assumption, although not formally correct, simplifies the mathematics and yields an approximate result that somewhat underestimates the true allogenicity. Therefore, the results obtained from the equations in this section and the next are the *lowest* amounts of allogenicity expected.

$z_l(x,n)$, the average allogenicity of the remainder of the chromosome to the left of G is

$$z_l(x,n) = \left(L_l(x) - \frac{100}{n}\right)(\tfrac{1}{2})^n. \tag{10}$$

A similar equation exists for the right end of the chromosome. Adding the allogenicity to the left ($z_l(x,n)$), the allogenicity to the right ($z_r(x,n)$), and the

allogenicity around G yields $Q(x,n)$, the total allogenicity of the chromosome, as

$$Q(x,n) = \frac{\left(L_l(x)-\frac{100}{n}\right)(\tfrac{1}{2})^n+\frac{100}{n}+\left(L_r(x)-\frac{100}{n}\right)(\tfrac{1}{2})^n+\frac{100}{n}}{L_l+L_r}$$

$$= \frac{\left(m(x)-\frac{200}{n}\right)(\tfrac{1}{2})^n+\frac{200}{n}}{m(x)}. \tag{11}$$

The segments of length $L_l(x)-(100/n)$ and $L_r(x)-(100/n)$ may be treated as independent chromosomes without segregating markers to determine the average length and number of the other remaining allogenic tracts.

To calculate $a(n)$, the average number of centimorgans per genome not containing a selected marker, equation (7) of the previous section must be slightly modified. First, M in equation (7) is replaced by $M-k(200/n)$, the total number of centimorgans in the genetic map minus the average number of centimorgans in allogenic regions containing selected markers, where k is the number of segregating selected loci. Second, c in equation (7) is replaced by $c+k$, the number of chromosomes not carrying a selected marker plus twice the number of chromosomes carrying a selected marker. With these changes, $T(n) = t(n)+k$ and

$$t(n) = \left[0{\cdot}01n\left(M-\frac{200k}{n}\right)+c+k\right](\tfrac{1}{2})^n. \tag{12}$$

The proportion of progeny with 0, 1, 2...allogenic tracts remaining that do not contain a selected marker can be estimated using a Poisson distribution with mean $t(n)$.

With a selected marker present, the average allogenicity, $Z(n)$, is no longer simply $(\frac{1}{2})^n$. Instead, $Z(n)$ will consist of two components; one component is the allogenicity of the chromosome carrying the selected marker weighted by the genetic map length of the chromosome, while the second component is the allogenicity of all of the remaining chromosomes weighted by their combined genetic map length. Thus $Z(n)$, the average allogenicity of the progeny after n generations of backcrossing is

$$Z(n) = \frac{Q(x,n)m(x)+(M-m(x))(\tfrac{1}{2})^n}{M}. \tag{13}$$

To calculate $a(n)$, the average number of centimorgans per genome not in tracts containing a selected marker remaining allogenic after n generations of backcrossing to a fixed parent, equation (8) must be modified by replacing M with $M-k\,(200/n)$. With this change $A(n)=a(n)+k(200/n)$ and

$$a(n) = \left(M-\frac{200k}{n}\right)(\tfrac{1}{2})^n. \tag{14}$$

To calculate $B(n)$, the average length of an allogenic tract that does not contain

https://doi.org/10.1017/S0016672300020243 Published online by Cambridge University Press

a selected marker after n generations of backcrossing to a fixed parent, $a(n)$ replaces $A(n)$, and $t(n)$ replaces $T(n)$ in equation (9) to give

$$B(n) = \frac{a(n)}{t(n)} = \frac{\left(M - \frac{200k}{n}\right)(\frac{1}{2})^n}{\left[0{\cdot}01n\left(M - \frac{200k}{n}\right) + c + k\right](\frac{1}{2})^n}$$

$$= \frac{M - (200k/n)}{0{\cdot}01n\left(M - \frac{200k}{n}\right) + c + k.} \tag{15}$$

Example. *Neurospora crassa* is a haploid ascomycete fungus that has 7 chromosomes and a map length of 1000 cM (Perkins & Barry, 1977). The gene for mating-type (A/a) in *N. crassa* is on linkage group I, approximately one quarter of the way from the left end. The total length of this linkage group is approximately 250 cM (Perkins & Barry, 1977). Suppose that a is the fixed parent and that the two initial parents are unrelated; then, after 10 generations of backcrossing $n = 10, Z(0) = 1, c = 7$ chromosomes, $M = 1000$ cM, and $m(\mathrm{I}) = 250$ cM.

Using equation (11), $Q(\mathrm{I},10)$, the average allogenicity of linkage group I A progeny with respect to the a parent, can be calculated as

$$Q(\mathrm{I},10) = \frac{\left[\left(250 - \frac{200}{10}\right)(\frac{1}{2})^{10}\right] + \frac{200}{10}}{250} = 0{\cdot}081 = 8{\cdot}1\,\%.$$

There will be an allogenic tract centred around mating type that averages 20 cM in length (100/10 cM on the left, plus 100/10 cM on the right). This allogenic tract alone is an average of 8% of the total genetic map length of linkage group I. Thus, on average, > 98% of the remaining allogenicity of linkage group I in A progeny with respect to the a fixed parent will be associated with the mating-type region.

Using equation (13), $Z(10)$, the average remaining allogenicity of the entire genome of the A progeny with respect to the a fixed parent can be calculated as

$$Z(10) = \frac{[(0{\cdot}081 \times 250) + (750(\frac{1}{2})^{10})]}{1000} = \frac{21{\cdot}0}{1000} = 0{\cdot}021 = 2{\cdot}1\,\%.$$

Thus, after 10 generations of backcrossing to a fixed parent, the allogenic tract centred at mating type (2% of the total map length) accounts for > 95% of the remaining allogenicity.

Using equation (12), $t(10)$, the average remaining number of allogenic tracts per genome not containing the mating type locus, can be calculated as

$$t(10) = \left[0{\cdot}01 \times 10\left(1000 - \frac{200}{10}\right) + 7 + 1\right](\frac{1}{2})^{10} = 0{\cdot}104.$$

$T(10)$, the average number of allogenic tracts remaining per genome is

$$T(10) = 0{\cdot}104 + 1 = 1{\cdot}104.$$

Using equation (14), $a(10)$, the average number of centimorgans per genome that remain in tracts not containing the mating type locus can be calculated as

$$a(10) = \left(1000 - \frac{200}{10}\right)(\tfrac{1}{2})^{10} = 0{\cdot}957 \text{ cM}.$$

$A(10)$, the average number of centimorgans remaining allogenic is

$$A(10) = 0{\cdot}957 + 20 = 20{\cdot}957 \text{ cM}.$$

Using equation (15), $B(10)$, the average length of each of the remaining allogenic tracts not containing the mating type locus, can be calculated as

$$B(10) = \frac{a(10)}{t(10)} = \frac{0{\cdot}957}{0{\cdot}104} = 9{\cdot}24 \text{ cM}.$$

The proportion of progeny carrying allogenic tracts other than the 20 centimorgan tract centred at mating type can be estimated using a Poisson distribution with mean $t(10) = 0{\cdot}104$. Thus the portion of the progeny with 0 additional allogenic tracts is 90·1 %, with 1 additional allogenic tract is 9·4 %, and with more than 1 additional allogenic tract is 0·5 %.

Thus, after 10 generations of backcrossing to a fixed a parent, the average allogenicity of the A progeny with respect to the a parent is 2·1 %, and > 95 % of this allogenicity is accounted for by the 20 cM allogenic tract containing A. Of the A progeny, 90·1 % are completely isogenic with the a parent, except for the 20 cM around mating type, and the remaining 9·9 % of the A progeny will have an average of at least 9·24 cM of allogenic material in addition to the 20 cM tract containing mating type. To be 99 % certain that a sample from the A progeny contains at least one isolate that is completely isogenic to the a parent, except for the mating-type region, then at least 2 isolates must be analysed since $(9{\cdot}9\,\%)^2 < 1\,\%$.

5. INBREEDING WITH MULTIPLE SEGREGATING LOCI

The breeding scheme considered here differs from that of the previous section in that multiple unlinked loci are considered instead of a single locus. Linked loci will not be considered. Introgression of traits controlled by multiple loci may be necessary for the study of complex developmental and regulatory traits. One such trait, initiation of haploid fruiting in *Schizophyllum commune* (Leslie & Leonard, 1979*a*, *b*; Esser, Saleh & Meinhardt, 1979) is known to be controlled by at least nine different loci. Additional loci may also be present in natural populations (Leslie & Leonard, 1980).

Suppose there are k segregating loci, X_1/x_1, X_2/x_2, . . . , X_k/x_k, each in a different linkage group. After n generations of crossing to a fixed parent each locus will

be at the centre of an allogenic region of average length $200/n$ for $n \geqslant 10$. The average allogenicity of each chromosome may be calculated from equation (11) of the previous section. The total allogenicity remaining after n backcrosses to a fixed parent is

$$Z(n) = \left[\sum_{i=1}^{k} Q(i,n)m(i) + \sum_{j=k+1}^{c} m(j)(\tfrac{1}{2})^n\right] \Big/ M. \tag{16}$$

Again, segments to the left and right of the marker on each chromosome are to be treated as independent chromosomes without segregating markers to determine the approximate average length and number of the other remaining allogenic tracts. $T(n)$, $t(n)$, $A(n)$, $a(n)$ and $B(n)$ are calculated as in the previous section.

Example. *Schizophyllum commune* is a basidiomycete fungus with two unlinked mating factors with multiple alleles. The total map length of the 11 chromosomes in this organism has been estimated as 800 cM (Carmi *et al.* 1978). Assume that the two parents are unrelated; that no recombination occurs within the A and B factors, and that one parent is A_1B_1 and that the fixed parent is A_2B_2. Suppose that the A factor is on the longest chromosome ($m(1) = 105$ centimorgans) and that the B factor is on the next longest chromosome ($m(2) = 90$ cM). After 10 generations of backcrossing to a fixed parent $n = 10$ generations, $Z(0) = 1, c = 11$ chromosomes, $M = 800$ cM, $m(1) = 105$ cM, $m(2) = 90$ cM, and $k = 2$. Substituting into equation (11), after 10 backcrosses to the fixed A_2B_2 parent the allogenicity of the A_1 chromosome with respect to the A_2 chromosome is

$$Q(1,10) = \frac{[105 - (200/10)](\tfrac{1}{2})^{10} + (200/10)}{105} = 0{\cdot}191 = 19{\cdot}1\,\%,$$

while the allogenicity of the B_1 chromosome with respect to the B_2 chromosome is

$$Q(2,10) = \frac{[90 - (200/10)](\tfrac{1}{2})^{10} + (200/10)}{90} = 0{\cdot}223 = 22{\cdot}3\,\%.$$

Using equation (16), then the total remaining allogenicity for the entire genome is

$$Z(10) = \frac{(0{\cdot}191 \times 105) + (0{\cdot}223 \times 90) + (605(\tfrac{1}{2})^{10})}{800} = 0{\cdot}051 = 5{\cdot}1\,\%.$$

Using equation (12), $t(10)$, the average number of allogenic tracts per genome not containing either of the mating type loci can be calculated as

$$t(10) = \left[0{\cdot}01 \times 10\left(800 - \frac{400}{10}\right) + 11 + 2\right](\tfrac{1}{2})^{10} = 0{\cdot}087.$$

$T(10)$, the average number of allogenic tracts remaining per genome, is

$$T(10) = 0{\cdot}087 + 2 = 2{\cdot}087.$$

Using equation (14), $a(10)$, the average number of centimorgans per genome

that remain in tracts not containing either of the mating type loci, can be calculated as

$$a(10) = \left(800 - \frac{400}{10}\right)(\tfrac{1}{2})^{10} = 0{\cdot}742 \text{ cM}.$$

$A(10)$, the average number of centimorgans remaining allogenic, is

$$A(10) = 0{\cdot}742 + 40 = 40{\cdot}742 \text{ cM}.$$

Thus, after 10 generations of backcrossing to the fixed A_2B_2 parent, 98% of the remaining allogenicity is accounted for by the tracts containing the A_1 and the B_1 loci.

Using equation (15), $B(10)$, the average length of each of the remaining allogenic tracts not containing either of the mating loci, can be calculated as

$$B(10) = \frac{a(10)}{t(10)} = \frac{0{\cdot}742}{0{\cdot}087} = 8{\cdot}53 \text{ cM}.$$

The proportion of progeny carrying allogenic tracts other than those centred at the mating type loci can be estimated using a Poisson distribution with mean $t(10) = 0{\cdot}087$. Thus the portion of the progeny with 0 additional allogenic tracts is 91·7%, with 1 additional allogenic tract is 8·0%, and with more than 1 additional allogenic tracts is 0·3%.

Thus, after 10 generations of backcrossing to the fixed A_2B_2 parent, the average allogenicity of the A_1B_1 progeny with respect to the A_2B_2 parent is 5·1% and > 98% of this allogenicity is accounted for by the allogenic tracts containing the A_1 and the B_1 factors. Of the A_1B_1 progeny, 91·7% are completely isogenic with the A_2B_2 parent, except for the tracts containing the A and B factors, and the remaining 8·3% of the A_1B_1 progeny have an average of at least 8·53 cM of allogenic material in addition to the tracts containing the A and the B factors. To be 99% certain that a sample from the A_1B_1 progeny contains at least one isolate that is completely isogenic to the A_2B_2 parent, except for the tracts containing the A and B factors, then at least 2 isolates must be analysed, since $(8{\cdot}3\%)^2 < 1\%$.

6. DISCUSSION

The breeding scheme described here was designed for haploid inbreeding problems. The results, however, can be easily modified to fit certain diploid inbreeding situations. For example, consider a diploid organism where an isogenic strain, I, is available. Cross I with any other strain, say H, and collect the F_1 progeny. Half of the chromosomes of each of the F_1 progeny will be isogenic with I, while the other half, originally from parent H, will be completely allogenic with I. Each of these F_1 progeny, then, is equivalent to the first cross between the fixed haploid parent and the completely unrelated parent. When the diploid F_1 progeny are crossed to the isogenic parent, then the resulting F_2 progeny are analogous to the progeny of the first cross between the fixed haploid parent and the completely unrelated parent. If backcrossing to the isogenic diploid parent

is continued, then the isogenicity of the progeny of the $n+1$th diploid generation will be the same as the relative isogenicity of the progeny of the nth haploid generation compared to the fixed haploid parent. Thus, for this example, the calculations made here for haploids can be readily modified and applied to diploids. Fisher's theory of junctions has recently been extended by Stam (1980) to finite random-mating diploid populations and by Bos (1976, 1980) to autotetraploids.

Most calculations for inbreeding in diploids are for sib matings. (I have not encountered the analogous calculations for haploids.) For diploids the calculation of isogenicity resulting from sib mating can be quite complex. Qualitatively, additional concepts and terms are required to understand the diploid procedure (cf. Fisher, 1965; Crow & Kimura, 1970). Quantitatively, the relative proportion of seven different types of matings must be considered, and a knowledge of matrix algebra is required to follow the mathematical arguments. These complications can all be ignored when the consequences of backcrossing haploids to a fixed parent are considered. Thus, the calculations presented in this paper, which require a minimum of mathematical training, should be useful for teaching purposes in addition to their practical applications.

I thank C. Denniston, T. J. Leonard, J. H. Meade, B. Turner, and W. Watt for fruitful discussions and for critically reading the manuscript. This work was supported by U.S. Public Health Service Research Grant AI-01462.

REFERENCES

Bos, I. (1976). Selection for a qualitative trait in an autotetraploid crop. *Euphytica* **25**, 161–166.

Bos, I. (1980). About the advisable number of backcrosses for autotetraploid crops. *Euphytica* **29**, 9–15.

Burnett, J. H. (1975). *Mycogenetics*. New York: Wiley.

Carmi, P., Holm, P. B., Koltin, Y., Rasmussen, S. W., Sage, J. & Zickler, D. (1978). The pachytene karyotype of *Schizophyllum commune* analyzed by three dimensional reconstruction of synaptonemal complexes. *Carlsberg Research Communications* **43**, 117–132.

Catcheside, D. G. (1977). *The Genetics of Recombination*. London: Arnold.

Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*, 1st ed. New York: Harper and Row.

Detlefsen, J. A. (1914). Genetic studies on a cavy species cross. *Carnegie Institution of Washington*, publication no. 205.

Esser, K., Saleh, F. & Meinhardt, F. (1979). Genetics of fruit body production in higher basidiomycetes. II. Monokaryotic and dikaryotic fruiting in *Schizophyllum commune*. *Current Genetics* **1**, 85–88.

Fincham, J. R. S., Day, P. R. & Radford, A. (1979). *Fungal Genetics*, 4th ed. Oxford: Blackwell Scientific Publications.

Fisher, R. A. (1965). *The Theory of Inbreeding*, 2nd ed. New York: Academic Press.

Haldane, J. B. S. (1936). The amount of heterozygosis to be expected in an approximately pure line. *Journal of Genetics* **32**, 375–391.

Jennings, H. S. (1916). The numerical results of diverse systems of breeding. *Genetics* **1**, 53–89.

Leslie, J. F. & Leonard, T. J. (1979*a*). Three independent genetic systems that control initiation of a fungal fruiting body. *Molecular and General Genetics* **171**, 257–260.

LESLIE, J. F. & LEONARD, T. J. (1979*b*). Monokaryotic fruiting in *Schizophyllum commune*: Genetic control of the response to mechanical injury. *Molecular and General Genetics* **175**, 5–12.

LESLIE, J. F. & LEONARD, T. J. (1980). Monokaryotic fruiting in *Schizophyllum commune*: Survey of a population from Wisconsin. *American Midland Naturalist* **103**, 367–374.

MYLYK, O. M. (1975). Heterokaryon incompatibility genes in *Neurospora crassa* detected using duplication producing chromosome rearrangements. *Genetics* **80**, 107–124.

MYLYK, O. M. (1976). Heteromorphism for heterokaryon incompatibility genes in natural populations of *Neurospora*. *Genetics* **83**, 275–284.

PERKINS, D. D. & BARRY, E. G. (1977). The cytogenetics of *Neurospora*. *Advances in Genetics* **19**, 133–285.

PERKINS, D. D., TURNER, B. C. & BARRY, E. G. (1976). Strains of Neurospora collected from nature. *Evolution* **30**, 281–313.

STAM, P. (1980). The distribution of the fraction of the genome identical by descent in finite random mating populations. *Genetical Research* **35**, 131–155.