Theoretical and Applied Genetics 42, 196–207 (1972)

Coefficients of Inbreeding and Homozygosity in Recurrent Selection: The Case of *m* Linked Loci

R. L. CAIN and K. HINKELMANN

Virginia Polytechnic Institute and State University, Blacksburg, Virginia (USA)

Summary. Methods of calculating the coefficients of inbreeding and homozygosity in a finite population undergoing recurrent selection (self-select-intercross in succeeding generations) are investigated for the case of *m* linked loci and effective directional selection. These coefficients are derived in terms of vectors whose components reflect the various possible patterns of genes being identical at a given stage of the recurrent selection breeding program.

For the case of two linked loci the progress of the panmictic index and/or the index of total heterozygosity through twenty-five cycles of recurrent selection is traced by means of computer-simulated populations ranging in sizes from ten through one hundred, assuming varying recombination probabilities, and assuming both minimum and maximum inbreeding selection patterns.

Results indicate that the coefficient of relationship in the source population is extremely important in tracing the progress of the degree of inbreeding and/or total homozygosity, that linkage plays a major role in promoting heterozygosity in a recurrent selection system, and that careful intercrossing rather than random mating in alternate generations of the recurrent selection cycle is important in promoting maximum heterozygosity in the selected population. In the simulated populations the effect of small population sizes is observed and, in general, indications are that unless more than five complete recurrent cycles are contemplated, increasing the population size results in only relatively minor increases in panmixia, especially when linked loci are involved in the selected trait and when care is taken to avoid a maximum inbreeding selection pattern.

1. Introduction

The problem of tracing through several generations the progress of a population when linkage is involved has been investigated by Geiringer (1944), and Kimura (1963) has developed a method of handling inbreeding systems with linked genes in terms of probability rather than in terms of the path analysis of Wright (1921). Schnell (1961) generalized Malécot's (1948) inbreeding coefficient to embrace the effects of linkage, and Narain (1966) used Schnell's generalization to obtain expressions of the mean and variance of the number of loci homozygous by descent in selfed populations and to trace the progress of homozygosity by descent for an arbitrary number of linked loci under certain conditions. Cockerham and Weir (1968), and Weir and Cockerham (1969) introduced trigametic and quadrigametic measures of inbreeding in their discussion of sib and pedigree mating with two linked loci.

Shikata (1962, 1965a) showed that a generalized inbreeding coefficient for an arbitrary number *m* of linked loci could be represented as a vector with *m* components which was transformed through the progress of *n* generations by *n* successive multiplications by $m \times m$ matrices. The structure of the transformation matrices depends upon the mating system, the viability constants and the recombination probabilities. Shikata's method is restricted to pedigrees which may be expressed as single closed loops. If, in a pedigree, more than one closed loop ends in the same individual then there may be violation of the assumption that any parts of chromosomes contributed by individuals outside the closed loop do not contain genes that are identical by descent (Shikata 1965a).

This paper presents, for the case of *m* linked loci, the derivation of recurrence formulae for an average generalized inbreeding coefficient and an average generalized coefficient of parentage in the *n*th cycle of a recurrent selection (*RS*) program, and develops an index with which to measure the total degree of heterozygosity in a population under recurrent selection. Derivation of recurrence formulae for the one-locus case has been presented in an earlier paper (Cain and Hinkelmann 1970).

The notation used in this paper is similar to that used by Shikata (1962, 1963, 1966), who has shown (Shikata 1965a) the relationships between his notation and that of Schnell (1961), though Weir (1971) has pointed out some discrepancies in Shikata's work. Shikata (1965b) (and personal communication, 1968) has outlined the conditions which limit the applicability of his simplified method of determining the generalized inbreeding coefficient: that all genes introduced to the closed loop pedigree are identical by descent with probability zero to genes descended from the common parent of the closed loop, and/or to other genes introduced to the closed loop. In Section 4 it is shown that when these conditions are imposed, the generalized inbreeding coefficient derived for an individual of the population in Section 3 reduces to that given by Shikata's formulae; furthermore, it is shown that in the one-locus case, if one ignores the first generation of selfing, the generalized average inbreeding coefficient reduces to the coefficient given by Sprague, *et al.* (1952).

2. Assumptions

For this paper it is assumed that the individuals involved are diploid with only two alleles per locus, are capable of self-fertilization and breed in nonoverlapping generations. Where linkage is assumed, no interference among crossovers, recombination probabilities that are constant over generations, no differences in recombination probabilities between the male and female sides of the plant are assumed and, further, no maternal or paternal effect upon progeny is expected. It is assumed that mutation is negligible. Any possibility of differential viability is ignored and a constant number of offspring per mating is assumed.

The assumptions outlined above limit the applicability of the results to plants, and in particular to many of the more common crop plants for which hybridization is utilized in developing varieties. Specifically the methods of recurrent selection have been applied to experiments involving corn, tobacco, sweet clover, birdsfoot trefoil, oats, wheat, and bromegrass.

The assumptions of diploid individuals, self-fertilization, low mutation rate, and no differences in recombination probabilities between male and female sides of the plant are not restrictive as regards a large percentage of crop plants; and most annual plants conform to the assumption of nonoverlapping generations. As regards differential viability, in plant breeding the number of progeny per mating usually far exceed those needed for continuance of the program unless there exists inbreeding depression, in which case yield would be decreased resulting in the deletion of the individuals involved either by natural or by artificial selection. The assumption of no maternal or paternal effects is more restrictive, as maternal or paternal effects have been shown in several varieties of corn, particularly with respect to heterosis in yield. And the assumption of recombination probabilities that are constant over generations is unrealistic since it is well-known that selection progress affects the values of recombination probabilities (Bodmer and Parsons 1962 and Shikata 1967a); unfortunately the otherwise increased complexity of the mathematical formulation requires that this assumption be made.

The assumption of two alleles per locus would, under many circumstances, be restrictive; however, in the case of a hybrid resulting from the cross of two highly inbred (and thus near-homozygous) lines, for example, only two alleles per locus are introduced into the population under consideration. In the presence of multiple alleles instead of two alleles per locus the approach to homozygosity would be considerably decelerated.

The assumption of no interference among crossovers is known to be violated almost universally when more than two linked loci are involved, resulting in one crossover promoting the occurrence of another (negative interference) or inhibiting its occurrence (positive interference); however, with more than three loci, the mathematical complications are prohibitive and in practice the interference probabilities are usually unknown anyhow, and so this unrealistic but practical assumption has been made (see also Shikata 1964, 1967b, and 1968).

Selection is assumed to be directional with no intermediate equilibrium possible; it is assumed that all loci involved contribute equally to the desired characteristic, that there exists partial to complete dominance, and that selection is effective (i.e., if epistasis or genotype-environment interactions exist, their effects do not preclude effective selection). Since the effects of linkage have been incorporated in the analysis, the assumption of equal contribution is not restrictive. For example, suppose it is known that a trait is governed by two major genes and two equal-effect modifiers and that each of the major genes has twice the effect of the modifiers. Then a six-gene model can be postulated with complete linkage within two sets and the remaining two genes corresponding to that known to exist among the two major and two minor genes. The assumptions of partial to complete dominance and effective selection would be valid for many of the common desired characteristics in crop plants, including yield.

3. The Generalized Inbreeding Coefficient in a Recurrent Selection Program

In deriving the formulae of this section reference will be made to the pedigree in Fig. 1. This can be done without loss of generality since this is a typical pedigree arising in an *RS* program.

It is evident from Fig. 1 that any genes identical by descent at T must have been carried through the pedigree simultaneously from I and L and that only

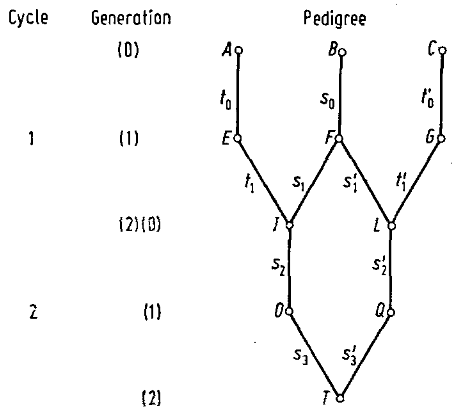

Fig. 1. Pedigree of an individual T in the fifth generation (second generation of the second cycle) of a recurrent series

at the first stage following selfing could any genes be introduced from outside the closed loop that might then be carried through the loop to yield genes identical by descent at T; hence, once the appropriate transformation is obtained for pedigree stage 1, the remainder of the stages could be handled by Shikata's method. Furthermore, if the initial source population is an open-pollinated variety, then it is with probability zero that genes identical by descent at T were introduced at stage 1 from outside the closed loop and Shikata's method would hold for the entire population by generalizing from a single closed loop pedigree.

Let the inbreeding coefficients for m linked loci be denoted by either one of the three expressions

$$F_\alpha\,(X_1X_2\ldots X_m) = F_\alpha(\psi_j) = F\,(\{Y\}_\alpha)\,,$$

where X_k ($k = 1, \ldots, m$) is 1 or 0 according to whether or not the kth locus in the individual α has genes identical by descent, and where ψ_j ($j = 1, 2, \cdots, 2^m$) represents the jth ordered set of 1's and 0's corresponding to the 2^m possible sets $(X_1X_2\cdots X_m)$ and $\{Y\}_\alpha$ represents the set of genes that are identical by descent at α. There are thus 2^m possible inbreeding coefficients for m loci; for example, the inbreeding coefficients for an individual α with two linked loci would be

$$\begin{aligned} F_\alpha(11) &= F_\alpha(\psi_1) = F(\{Y_1, Y_2\}_\alpha)\\ F_\alpha(10) &= F_\alpha(\psi_2) = F(\{Y_1\}_\alpha)\\ F_\alpha(01) &= F_\alpha(\psi_3) = F(\{Y_2\}_\alpha)\\ F_\alpha(00) &= F_\alpha(\psi_4) = F(\Phi_\alpha)\,. \end{aligned}$$

Then the generalized inbreeding coefficient for an individual α may be defined as a vector expression with jth component ($j = 1, \ldots, 2^m$) equal to $F_\alpha(\psi_j)$. In the two-locus case,

$$\boldsymbol{F}_\alpha = \begin{bmatrix} F_\alpha(11)\\ F_\alpha(10)\\ F_\alpha(01)\\ F_\alpha(00) \end{bmatrix}$$

The panmictic index would be, in this case, $P_\alpha = F_\alpha(00)$. Note, however that this panmictic index, by definition, measures the probability that the genes at both loci are not homozygous by descent; i.e. either or both loci could have genes homozygous in state that were still not homozygous by descent.

In a similar manner, let the generalized coefficient of parentage for m linked loci, $\boldsymbol{R}_{\alpha\beta}$, have components

$$R_{\alpha\beta}\,(X_1\,X_2\cdots X_m) = R_{\alpha\beta}(\psi_j) = R\,(\{Y\}_{\alpha\beta})\,,$$

where X_k ($k = 1, \cdots, m$) is 1 or 0 according to whether or not the kth loci in individuals α and β have genes identical by descent, and where, as defined before, ψ_j ($j = 1, 2, \ldots, 2^m$) represents the jth ordered set of 1's and 0's indicating the pattern of genes identical by descent at α and β, and $\{Y\}_{\alpha\beta}$ denotes the set of genes identical by descent at α and β.

For later purposes it is convenient to introduce the reduced vectors $\boldsymbol{F}^*_\alpha$ and $\boldsymbol{R}^*_{\alpha\beta}$ which are obtained from $\boldsymbol{F}_\alpha$ and $\boldsymbol{R}_{\alpha\beta}$, resp., by omitting the last component; i.e. $F_\alpha\,(00\cdots0)$ and $R_{\alpha\beta}\,(00\cdots0)$, resp. Note also that the components of $\boldsymbol{F}_\alpha$ and $\boldsymbol{R}_{\alpha\beta}$ add to one.

Let $\{Y\}_{1,k}$ be the set of genes carried simultaneously on both sides of the closed loop through reproductive stage k ($k = 1, 2, 3$) in the loop pedigree from the loop parent (see Fig. 1). Let $\{Y\}_{2,l}$ be the set of genes introduced to the closed loop at I and L and carried in the closed loop pedigree through stage l ($l = 2, 3$) from some common ancestor outside the closed loop; and let $\{Y\}_{3,r}$ be the set of genes for which exactly one of each pair came from the common ancestor in the loop and exactly one came from an ancestor outside the loop ($r = 2, 3$). Denote by $\varphi\{Y\}_{1,k}$, $\varphi\{Y\}_{2,l}$, and $\varphi\{Y\}_{3,r}$ the set of 1's and 0's according to whether or not the genes at a locus were carried simultaneously in the closed loop pedigree (as described before); for example, if for the two loci case at the kth reproductive stage $\{Y\}_{1,k} = \{Y_1\}$; i. e., the first locus of individuals on both sides of the loop has genes identical by descent from F, then $\varphi\{Y\}_{1,k} = \{10\}$; or, if $\{Y\}_{3,r} = \{Y_1, Y_2\}$; i. e., the first and second locus both have genes identical by descent and one of each pair was transmitted from F and one of each pair came from outside the loop from either E or G, then $\varphi\{Y\}_{3,r} = \{11\}$.

Consider now the set of identical genes at T, $\{Y\}_T$. $\{Y\}_T$ is the union of four mutually exclusive sets:

A, the set of identical genes for which one of each pair at each locus came from E and one from G;
B, the set of identical genes for which one of each pair at each locus came from E and one from F;
C, the set of identical genes for which one of each pair at each locus came from G and one from F;
D, the set of identical genes for which both descended solely from F.
(3.1)

The reproductive process yielding A is such that

$$A \subset \bigcap_{l=2}^{3} \{Y\}_{2,l} \tag{3.2}$$

and that yielding B or C is included in

$$B \subset \bigcap_{r=2}^{3} \{Y\}_{3,r} \quad \text{or} \quad C \subset \bigcap_{r=2}^{3} \{Y\}_{3,r}\,. \tag{3.3}$$

The reproductive process yielding D is such that

$$D \subseteq \Big[\{Y\}_F \bigcap_{k=2}^{3} \{Y\}_{1,k}\Big] \cap \big[\sim\{Y\}_F \cap \{Y\}_{1,1}\big]\,, \tag{3.4}$$

where $\sim\{Y\}_F$ denotes the set of genes not identical by descent at F. The statement made by (3.2) is that in order for the set of genes carried through stages 2 and 3 to contain any that are identical by descent at T by reason of having descended to T from the individuals

E and G, these genes must have been passed to I and L in reproductive stages t_1 and t_1' by E and G, respectively, and thence have passed (and not been discarded) simultaneously through stages 2 (s_2 and s_2') and 3 (s_3 and s_3') to T. Statement (3.3) is identical to statement (3.2), except that the ancestors E and G are replaced by E and F and by F and G, respectively, for the two portions of the statement. Statement (3.4) takes into consideration the two paths available for a set of genes which are identical by descent at T having descended through the closed loop pedigree from the common parent F: these genes may or may not have been identical by descent at F. The first bracket includes the case in which the genes were identical by descent at F (and thus were passed to I and L with probability one); if so, then in order to be identical by descent at T, they must have passed simultaneously through reproductive stages 2 and 3. The second bracket includes the case that the genes were not identical by descent at F but, due to the fact that stage 1 (i.e. s_1 and s_1') is a selfing type of reproductive operation, identical genes were passed to I and L. Let the conditions corresponding to (3.2) be denoted by f_j; those corresponding to (3.3), g_j and h_j respectively; and those corresponding to (3.4), d_j and c_i for the first and second bracketed quantities respectively. Collecting all of these different conditions for m linked loci, there are possible 2^m each of the f's, g's, h's, and d's, and 2^{2m} c's. Thus for some i, j the set of identical genes at T can be represented as follows:

$$\{Y\}_T = (f_j \cup g_j \cup h_j) \cup (d_j \cap c_i) \qquad (3.5)$$

and, equivalently, for $i = 1, 2, \ldots, 2^{2m}$ and $j = 1, 2, \ldots, 2^m$, the sequence of three reproductive stages which yields $\{Y\}_T$ starting with $\{Y\}_{JK}$ (where $J = E$ or F and $K = F$ or G and where $\{Y\}_{JK}$ is the set of pairs of genes identical by descent with exactly one gene of each pair being in J and one being in K) is given by

$$\mathrm{I}\,(\{Y\}_T; 3; \{Y\}_{JK}) = [(\bigcup_j f_j) \cup (\bigcup_j g_j) \cup (\bigcup_j h_j)] \cup \\ \cup [(\bigcup_j d_j) \cap (\bigcup_i c_i)]\,, \qquad (3.6)$$

where the probability of the set expression in the first bracket yields the $2^m \times 1$ transition vector for the generalized coefficient of parentage $\boldsymbol{R}$ and the expression in the second bracket, the $2^m \times 2^m$ transition matrix for $\boldsymbol{F}$. Thus, with $P(\cdot)$ denoting the probability of the event in brackets, the elements of $\boldsymbol{F}_T^*$ are given by

$$F_T(\psi_k) = [P(f_k) + P(g_k) + P(h_k)] \times [\boldsymbol{R}(\{Y\}_{EG})]_k + \\ + \Big[\sum_S P(c)\,P(d)\Big]_k \boldsymbol{F}_F(\{Y\})\,, \\ (k = 1, 2, \ldots, 2^m - 1)\,, \qquad (3.7)$$

where S is the set of reproductive stages for which it is possible to develop $\{Y\}_{1,1}$ from the given $\{Y\}_F$ by means of condition "c" and then develop the specified $\{Y\}_T$ from $\{Y\}_{1,1}$ by means of condition "d", × indicates the product of the two vector elements, and the subscript k on the brackets indicates the kth row of the vector or matrix is involved, where $k = 1, \ldots, 2^m - 1$, and the $k = 2^m$th element of $\boldsymbol{F}_T$ is obtained by subtracting the sum of the $2^m - 1$ elements of $\boldsymbol{F}_T^*$ from one. Combining the three sets of conditions f_k, g_k, and h_k is possible since $R_{EG}(\psi_j) = R_{EF}(\psi_j) = R_{FG}(\psi_j)$ for all $j = 1, \ldots, 2^m$.

Let the transition vector be denoted by $\boldsymbol{V}$ and the transition matrix by $\boldsymbol{M}$. Then in terms of the $\boldsymbol{M}_0$ and $\boldsymbol{M}_1$ "mictic matrices" as defined by Shikata (1965a), $\boldsymbol{M}$ can be defined as follows, for the closed loop portion of the pedigree:

$$\boldsymbol{M} = (4\,\boldsymbol{M}_1^2)\,(4\,\boldsymbol{M}_1^2)\left(\frac{1}{4}\,\boldsymbol{M}_0\right).$$

Post-multiplying $\boldsymbol{M}$ by $\boldsymbol{M}_0$ takes the initial generation of selfing into account, resulting in

$$F_T(\psi_k) = [\boldsymbol{V} \times \boldsymbol{R}(\{Y\}_{EG})]_k + [\boldsymbol{M}\boldsymbol{M}_0\boldsymbol{F}(\{Y\}_F)]_k \qquad (3.8)$$

for $k = 1, 2, \ldots, 2^m - 1$.

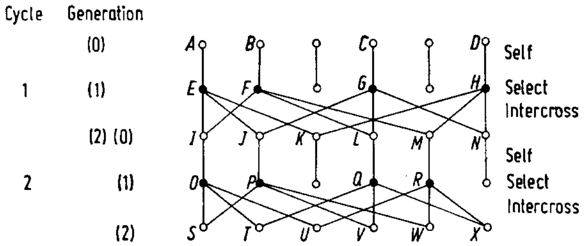

Fig. 2. A possible selection pattern illustrating derivation of the average inbreeding coefficient in the second cycle of recurrent selection

Fig. 2 is a possible RS pattern for which the path shown in Fig. 1 is a special case. Let $a_i^{(n+1)}$, $i = 1, \ldots, N$ be the number of times each line saved in cycle $n + 1$ is represented in the parents of progeny saved in cycle $n + 2$, where N is the total number of lines saved. Let $b_i^{(n+1)} = a_i^{(n+1)}$ if $a_i^{(n+1)} \geqq 2$ and let $\sum' a_i^{(n+1)}$ denote the sum of all $a_i^{(n+1)} = 1$. Then $\sum' a_i^{(n+1)} = \binom{N}{2} - \sum \binom{b_i^{(n+1)}}{2}$. Let $\boldsymbol{F}^{(n)}$ and $\boldsymbol{R}^{(n)}$ denote, respectively, the generalized inbreeding coefficient and generalized coefficient of parentage in the terminal (i.e. second) generation of the n th recurrent cycle, and let $\boldsymbol{F}^{*(n)}$ and $\boldsymbol{R}^{*(n)}$ denote the reduced coefficients; i.e., omitting the last element in $\boldsymbol{F}^{(n)}$ and $\boldsymbol{R}^{(n)}$, respectively.

Referring again to Fig. 2, it is evident that for a single-line descendent such as W with ancestors E, F, G, and H (R and P have no common grandparent), the contribution to the inbreeding coefficient will involve four sets of conditions analogous to the three expressions f_j, g_j, h_j. In other words, comparing the ancestry of W with that which generated the sets (3.1), since there is no common ancestor to produce a set such as D and thus no transformation for $\boldsymbol{F}^{(n)}$, there result four sets of conditions, which can be expressed in probability terms as a vector product with $\boldsymbol{R}^{(n)}$ since $R_{EH}(\psi_j) = R_{EG}(\psi_j) = R_{FG}(\psi_j) = R_{FH}(\psi_j)$ for all $j = 1, 2, \ldots, 2^m$. Call this vector $\boldsymbol{E}$ with $3\,\boldsymbol{E} = 4\,\boldsymbol{V}$. Thus

$$\boldsymbol{F}^{*(n+2)} = \frac{1}{\binom{N}{2}} \left\{ \left[\frac{N}{2} - \Sigma \binom{b_i^{(n+1)}}{2} \right] [\boldsymbol{E}^* \times \boldsymbol{R}^{*(n)}] + \right.$$

$$\left. + \left[\Sigma \binom{b_i^{(n+1)}}{2} \right] [\boldsymbol{V}^* \times \boldsymbol{R}^{*(n)} + [\boldsymbol{M}\boldsymbol{M}_0]^* \boldsymbol{F}^{(n)}] \right\}, \qquad (3.9)$$

and

$$\boldsymbol{R}^{*(n+1)} = \frac{N-3}{N+1} [\boldsymbol{E}^* \times \boldsymbol{R}^{*(n)}] +$$

$$+ \frac{4}{N+1} [\boldsymbol{V}^* \times \boldsymbol{R}^{*(n)} + [\boldsymbol{M}\boldsymbol{M}_0]^* \boldsymbol{F}^{(n)}], \qquad (3.10)$$

where $\boldsymbol{A}^*$ denotes the vector (matrix) $\boldsymbol{A}$ without its last element (row).

To obtain the probabilities involved, the probabilities for the formation of gametes (call them gametic probabilities) are denoted by p_j and p_j', resp., with $p_j = p_j'$ $(j = 1, 2, \ldots, 2^{m-1})$, where p_j and p_j' denote the probabilities for "complementary" gametes; e.g., the two parental gametes (p_1, p_1'), or the two gametes where the only cross-over occurs between the first and second locus (p_2, p_2'), etc. Note that since $p_j = p_j'$ it follows that $\sum p_j = 1/2$. For further examples, see Table 1. Given the gametic probabilities (which, of course, are functions of the various cross-over probabilities), the mating system of recurrent selection, and the conditions outlined by (3.2), (3.3), and (3.4), the transition expressions for the successive recurrent cycles may be obtained explicitly. Given $\boldsymbol{F}^{(0)}$ and $\boldsymbol{R}^{(0)}$, $\boldsymbol{F}^{(1)} = \boldsymbol{R}^{(0)}$ and these values enable one to trace the progress of the population.

Table 1. *Examples of subscripting of the gametic probabilities for $m = 1, 2, 3$ with 0 and 1 indicating the origin of the genes with regard to the parental chromosomes*

locus	$m = 1$		locus	$m = 2$			
	p_1	p_1'		p_1	p_1'	p_2	p_2'
1	1	0	1	1	0	1	0
			2	1	0	0	1

locus	$m = 3$							
	p_1	p_1'	p_2	p_2'	p_3	p_3'	p_4	p_4'
1	1	0	1	0	1	0	1	0
2	1	0	0	1	1	0	0	1
3	1	0	0	1	0	1	1	0

4. Special Cases

Consider the sets of identical genes given by (3.1). In the special case for which $R_{EF}(\Phi) = R_{FG}(\Phi) = 1$, $\boldsymbol{F}_T$ reduces to the generalized inbreeding coefficient given by Shikata (1965a, 1966). In this case (3.2) and (3.3) both yield the null set and (3.4) is equivalent to Formula 4 in Shikata's paper (1962) or, in probability terms, to his Formula 33 in his later paper (1965a).

Consider next the special case $m = 1$. Then $\boldsymbol{F}_T = \begin{bmatrix} F_T(1) \\ F_T(0) \end{bmatrix}$ or, equivalently, $\boldsymbol{F}_T^* = F_T(1)$ and $p_1 = p_1' = 1/2$. Since, e.g., f_1 represents the condition that two identical genes are passed simultaneously through reproductive stages 2 and 3 on both sides of the loop via any two opposing paths of the four possible paths in the loop,

$$P(f_1) = P(\{Y\}_{2,2} \cap \{Y\}_{2,3} = \{Y_1\}_T) = 4\ (1/2)^4$$

and, by analogy,

$$P(f_1) = P(g_1) = P(h_1) = P(d_1) = 4\ (1/2)^4$$

and, consequently,

$$P(f_2) = P(\{Y\}_{2,2} \cap \{Y\}_{2,3} = \Phi_T) = 1 - 4\ (1/2)^4$$

with

$$P(f_2) = P(g_2) = P(h_2) = P(d_2) = 1 - 4\ (1/2)^4 .$$

Further,

$$P(c_1) = P(\{Y_1\}_F \cap \{Y\}_{1,1} = \{Y_1\}_{IL}) = $$
$$= p_1 + p_1' = 1$$

$$P(c_2) = P(\{Y_1\}_F \cap \{Y\}_{1,1} = \Phi_{IL}) = 0$$

$$P(c_3) = P(\Phi_F \cap \{Y\}_{1,1} = \{Y_1\}_{IL}) = $$
$$= p_1^2 + p_1'^2 = 1/2$$

$$P(c_4) = P(\Phi_F \cap Y_{1,1} = \Phi_{IL}) = 1 - (p_1^2 + p_1'^2) = 1/2.$$

Hence $\boldsymbol{V}^* = 3/4$ and $\boldsymbol{M}^* = (1/4,\ 1/8)$, where the elements of $\boldsymbol{M}^*$ are obtained from (s. formula (3.7))

$$m_{11} = P(c_1)\ P(d_1)$$

$$m_{12} = P(c_3)\ P(d_1)\ .$$

It then follows from (3.8) (ignoring the initial generation of selfing that

$$\boldsymbol{F}_T^* = F_T(1) = \frac{3}{4} R_{EG}\ (1) + \frac{1}{8} [1 + F_F(1)]\ . \qquad (4.1)$$

With the assumption of no inbreeding and no relationships in the source population (4.1) yields the coefficient given by Sprague, *et al.* (1952) as an average generalized coefficient of inbreeding (s. also Cain and Hinkelmann 1970).

5. A Two-Loci Example

Listed in Table 2 are the values of $P(f_j)$, $P(g_j)$, $P(h_j)$ and $P(d_j)$; in Table 3 the values of $P(c_i)$; and in Table 4, the values of $\Sigma\, P(d_j)\, P(c_i)$ which are nonzero for a two-loci example. Summarized in Table 5 are the values of $P(\{Y\}_T;\, 3;\, \{Y\}_T)$.

For example, in Table 2, $P(d_2)$ is obtained by observing that the formation of gametes described by p_1 or p_2 will preserve the identical genes at locus one and thus with probability $4\,(p_1 + p_2)^4 = 4(1/2)^4$ the genes at locus one are identical by descent through reproductive stages 2 and 3, and hence $4\,(p_1 + p_2)^4 - 4p_1^4$ is the conditional probability that the genes at

Table 2. *Values of* $P(d_j)$

$(X_1X_2)_T$	j	Possible values of $\varphi(\{Y\}_{1,1})$	Formula for $P(d_j)$	Value of $P(d_j)$
(11)	1	(11)	$4\,p_1^4$	$4\,p_1^4$
(10)	2	(11) (10)	$4\,(p_1 + p_2)^4 - 4\,p_1^4$	$\frac{1}{4} - 4\,p_1^4$
(01)	3	(11) (01)	$4\,(p_1 + p_2')^4 - 4\,p_1^4$	$\frac{1}{4} - 4\,p_1^4$
(00)	4	(11) (10) (01) (00)	$(p_1+p_1'+p_2+p_2')^4 - 8\,(p_1+p_2')^4+4\,p_1^4$	$\frac{1}{2} + 4\,p_1^4$

Table 3. *Values of* $P(c_i)$

$\varphi(\{Y\}_{1,1})$	$(X_1X_2)_F$	i	Formula for $P(c_i)$	Value of $P(c_i)$
(11)	(11)	1	$(p_1+p_1'+p_2+p_2')^2$	1
(10)	(11)	2	0	0
(01)	(11)	3	0	0
(00)	(11)	4	0	0
(11)	(10)	5	$(p_1^2+p_2^2+p_1'^2+p_2'^2) + 2\,(p_1p_2'+p_1'p_2)$	1/2
(10)	(10)	6	$2\,(p_1p_1'+p_2p_2' + p_1p_2 + p_1'p_2')$	1/2
(01)	(10)	7	0	0
(00)	(10)	8	0	0
(11)	(01)	9	same as c_5	1/2
(10)	(01)	10	0	0
(01)	(01)	11	same as c_6	1/2
(00)	(01)	12	0	0
(11)	(00)	13	$p_1^2 + p_1'^2 + p_2^2 + p_2'^2$	$2\,(p_1^2+p_2^2)$
(10)	(00)	14	$2\,(p_1p_2 + p_1'p_2')$	$4\,p_1p_2$
(01)	(00)	15	$2\,(p_1p_2' + p_1'p_2)$	$4\,p_1p_2$
(00)	(00)	16	$2\,(p_1p_1' + p_2p_2')$	$2\,(p_1^2+p_2^2)$

Table 4. *Non-zero values of* $\sum P(d_j)\,P(c_i)$

$(X_1X_2)_T$	$\varphi(\{Y\}_{1,1})$	$(X_1X_2)_F$	$\sum P(d_j)\,P(c_i)$
(11)	(11)	(11)	$P(d_1)\,P(c_1)$
(10)	(11)	(11)	$P(d_2)\,P(c_1)$
(01)	(11)	(11)	$P(d_3)\,P(c_1)$
(00)	(11)	(11)	$P(d_4)\,P(c_1)$
(11)	(11)	(10)	$P(d_1)\,P(c_5)$
(10)	(11)	(10)	$P(d_2)\,P(c_5)$
(01)	(11)	(10)	$P(d_3)\,P(c_5)$
(00)	(11)	(10)	$P(d_4)\,P(c_5)$
(10)	(10)	(10)	$[P(d_1) + P(d_2)]\,P(c_6)$
(00)	(10)	(10)	$[P(d_3) + P(d_4)]\,P(c_6)$
(11)	(11)	(01)	$P(d_1)\,P(c_9)$
(10)	(11)	(01)	$P(d_2)\,P(c_9)$
(01)	(11)	(01)	$P(d_3)\,P(c_9)$
(00)	(11)	(01)	$P(d_4)\,P(c_9)$
(01)	(01)	(01)	$[P(d_1) + P(d_3)]\,P(c_{11})$
(00)	(01)	(01)	$[P(d_2) + P(d_4)]\,P(c_{11})$
(11)	(11)	(00)	$P(d_1)\,P(c_{13})$
(10)	(11)	(00)	$P(d_2)\,P(c_{13})$
(01)	(11)	(00)	$P(d_3)\,P(c_{13})$
(00)	(11)	(00)	$P(d_4)\,P(c_{13})$
(10)	(10)	(00)	$[P(d_1) + P(d_2)]\,P(c_{14})$
(00)	(10)	(00)	$[P(d_3) + P(d_4)]\,P(c_{14})$
(01)	(01)	(00)	$[P(d_1) + P(d_3)\,]P(c_{15})$
(00)	(01)	(00)	$[P(d_2) + P(d_4)]\,P(c_{15})$
(00)	(00)	(00)	$[P(d_1) + P(d_2) + P(d_3) + P(d_4)]\,P(c_{16})$

locus one but not at locus two of both individuals are identical by descent (the multiplicative factor of "4" arises because of the paths occasioned by selfing at stages s_2 and s_2'). In Table 3, $P\,(c_5)$ is the probability of the case which yields both genes identical by descent after stage 1 while only the first locus has genes identical by descent at F.

The transition matrix given in Table 5 yields the matrix multiplier of $\boldsymbol{F}_F$ and the values given in Table 2 are the elements of the vector multiplier for R_{EG}; however, for the complete recurrent selection scheme,

Table 5. *Values of* $P(\{Y\}_T;\,3;\,\{Y\}_F)$*, and their row and column positions in the transition matrix* $\boldsymbol{M}$

$(X_1X_2)_T$	$(X_1X_2)_F$	Position in $\boldsymbol{M}$	$\sum P(d_j)\,P(c_i)$	Value of $P(\{Y\}_T;\,3;\,\{Y\}_F)$
(11)	(11)	(1,1)	$P(d_1)\,P(c_1)$	$4\,p_1^4$
(10)	(11)	(2,1)	$P(d_2)\,P(c_1)$	$1/4 - 4\,p_1^4$
(01)	(11)	(3,1)	$P(d_3)\,P(c_1)$	$1/4 - 4\,p_1^4$
(00)	(11)	(4,1)	$P(d_4)\,P(c_1)$	$1/2 + 4\,p_1^4$
(11)	(10)	(1,2)	$P(d_1)\,P(c_5)$	$2p_1^4$
(10)	(10)	(2,2)	$P(d_2)\,P(c_5) + [P(d_1) + P(d_2)]\,P(c_6)$	$1/4 - 2\,p_1^4$
(01)	(10)	(3,2)	$P(d_3)\,P(c_5)$	$1/8 - 2\,p_1^4$
(00)	(10)	(4,2)	$P(d_4)\,P(c_5) + [P(d_3) + P(d_4)]\,P(c_6)$	$5/8 + 2\,p_1^4$
(11)	(01)	(1,3)	$P(d_1)\,P(c_9)$	$2\,p_1^4$
(10)	(01)	(2,3)	$P(d_2)\,P(c_9)$	$1/8 - 2\,p_1^4$
(01)	(01)	(3,3)	$P(d_3)\,P(c_9) + [P(d_1) + P(d_3)]\,P(c_{11})$	$1/4 - 2\,p_1^4$
(00)	(01)	(4,3)	$P(d_4)\,P(c_9) + [P(d_2) P(d_4)]\,P(c_{11})$	$5/8 + 2\,p_1^4$
(11)	(00)	(1,4)	$P(d_1)\,P(c_{13})$	$8\,p_1^4\,(p_1^2 + p_2^2)$
(10)	(00)	(2,4)	$P(d_2)\,P(c_{13}) + [P(d_1) + P(d_2)]\,P(c_{14})$	$p_1p_2 + (1/2 - 8\,p_1^4) \times (p_1^2 + p_2^2)$
(01)	(00)	(3,4)	$P(d_3)\,P(c_{13}) + [P(d_1) + P(d_3)]\,P(c_{15})$	$p_1p_2 + (1/2 - 8\,p_1^4 \times (p_1^2 + p_2^2)$
(00)	(00)	(4,4)	$P(d_4)\,P(c_{13}) + [P(d_3) + P(d_4)]\,P(c_{14}) + [P(d_2) + P(d_4)]\,P(c_{15}) + [P(d_1) + P(d_2) + P(d_3) + P(d_4)]\,P(c_{16})$	$2\,(p_1^2 + p_2^2)\,(3/2 + 4\,p_1^4) + 6\,p_1p_2$

these must be adjusted to account for the initial generation of selfing. This initial generation of selfing has no effect on the coefficient of parentage (i.e. $R_{AC} = R_{EG}$); however the multiplier of $\boldsymbol{F}_F$ must be post-multiplied by

$$\boldsymbol{M}_0 = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & 2\,(p_1^2 + p_2^2) \\ 0 & \frac{1}{2} & 0 & 4p_1\,p_2 \\ 0 & 0 & \frac{1}{2} & 4p_1\,p_2 \\ 0 & 0 & 0 & 2\,(p_1^2 + p_2^2) \end{bmatrix}, \qquad (5.1)$$

as given on page 135 of Shikata (1962). Also the possible contributions of all lines represented singly must be included. Thus the average generalized inbreeding coefficient, weighted according to the number of times a given line is represented in the pedigree, for the two-locus case is given by (3.9) and the generalized coefficient of parentage by (3.10), where

$$\boldsymbol{V}^* = \begin{bmatrix} 12p^4 \\ \frac{3}{4} - 12p_1^4 \\ \frac{4}{3} - 12p_1^4 \end{bmatrix} \text{ and } \boldsymbol{E}^* = \frac{4}{3}\,\boldsymbol{V}^*; \qquad (5.2)$$

and $[\boldsymbol{MM}_0]^*$ has elements

$$\left.\begin{aligned} &m_{11} = 4\,p_1^4\,;\;\; m_{12} = 3\,p_1^4\,;\;\; m_{13} = 3\,p_1^4\,; \\ &m_{14} = 3\,p_1^4 - 16\,p_1^5 p_2 + 64\,p_1^6 p_2^2 \\ &m_{21} = \frac{1}{4} - 4\,p_1^4\,;\;\; m_{22} = \frac{1}{4} - 3\,p_1^4\,; \\ &m_{23} = \frac{3}{16} - 3\,p_1^4\,; \\ &m_{24} = \frac{3}{16} - 3\,p_1^4 + 16\,p_1^5 p_2 - 64\,p_1^6 p_2^2 \\ &m_{31} = \frac{1}{4} - 4\,p_1^4\,;\;\; m_{32} = \frac{3}{16} - 3\,p_1^4\,; \\ &m_{33} = \frac{1}{4} - 3\,p_1^4\,; \\ &m_{34} = \frac{3}{16} - 3\,p_1^4 + 16\,p_1^5 p_2 - 64\,p_1^6 p_2^2\,. \end{aligned}\right\} \qquad (5.3)$$

Using (5.2) and (5.3) and assuming equal representation of all lines ($b_i^{(n+1)} = 2$ for all i and n), the progress of a population through fifty recurrent selection generations (25 cycles) was simulated by computer for various population sizes, various combinations of $\boldsymbol{F}^{(0)}$ and $\boldsymbol{R}^{(0)}$, and varying values of the recombination probabilities. Representative ones of

Table 6. *Progress of populations of varying sizes, varying initial coefficients of parentage, and varying gametic probabilities through twenty-five cycles of recurrent selection assuming a minimum inbreeding selection pattern with $P^{(0)} = 1$*

Initial Conditions			Values of the Panmictic Index($P^{(n)} = [\boldsymbol{F}^{(n)}]_4$) at the End of the Indicated Recurrent Cycle											
$\boldsymbol{R}^{(0)}$	p_1	N	1	2	3	4	5	6	7	8	10	15	20	25
$\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$	.30	10	1.00	.92	.82	.74	.67	.62	.57	.54	.49	.42	.40	.39
		15	1.00	.95	.88	.82	.77	.73	.70	.67	.63	.57	.55	.54
		25	1.00	.97	.93	.89	.86	.83	.81	.79	.76	.72	.70	.69
		100	1.00	.99	.98	.97	.96	.95	.95	.94	.93	.92	.91	.91
$\begin{bmatrix}.25\\.25\\.25\\.25\end{bmatrix}$	.30	10	.25	.48	.48	.46	.44	.43	.42	.42	.40	.39	.39	.38
		15	.25	.50	.52	.52	.53	.53	.53	.53	.53	.53	.53	.53
		25	.25	.51	.56	.58	.60	.61	.62	.63	.65	.67	.68	.68
		100	.25	.53	.60	.65	.68	.71	.74	.76	.80	.86	.88	.90
$\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$	.35	10	1.00	.92	.84	.79	.75	.72	.70	.68	.66	.65	.65	.65
		15	1.00	.95	.89	.85	.82	.80	.79	.78	.76	.75	.75	.75
		25	1.00	.97	.94	.91	.89	.88	.87	.86	.85	.84	.84	.84
		100	1.00	.99	.98	.98	.97	.97	.96	.96	.96	.96	.96	.96
$\begin{bmatrix}.25\\.25\\.25\\.25\end{bmatrix}$	.35	10	.25	.51	.57	.59	.61	.62	.63	.63	.64	.64	.65	.65
		15	.25	.53	.61	.65	.68	.70	.71	.72	.73	.74	.74	.75
		25	.25	.54	.65	.70	.74	.76	.78	.80	.81	.83	.84	.84
		100	.25	.56	.68	.76	.81	.84	.87	.89	.92	.95	.95	.96
$\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$	.45	10	1.00	.94	.90	.88	.86	.86	.85	.85	.85	.85	.85	.85
		15	1.00	.96	.93	.92	.91	.90	.90	.90	.90	.90	.90	.90
		25	1.00	.99	.95	.94	.94	.94	.94	.94	.94	.94	.94	.94
		100	1.00	.99	.99	.99	.99	.98	.98	.98	.98	.98	.98	.98
$\begin{bmatrix}.25\\.25\\.25\\.25\end{bmatrix}$	.45	10	.25	.62	.75	.79	.82	.83	.84	.84	.84	.85	.85	.85
		15	.25	.64	.78	.83	.86	.87	.88	.89	.89	.89	.89	.89
		25	.25	.65	.80	.86	.89	.91	.92	.93	.93	.93	.94	.94
		100	.25	.66	.82	.90	.93	.95	.96	.97	.98	.98	.98	.98

the results in terms of the panmictic index component of $\boldsymbol{F}^{(n)}$, are given in Table 6 and illustrated by Fig. 3. In Table 7 are listed results in the case that all conditions are the same as those for Table 6 except a maximum instead of a minimum inbreeding selection pattern is assumed. Table 8 lists representative results for the case of independent segregation of loci ($p_1 = p_2 = 1/4$).

Although various values for $\boldsymbol{F}^{(0)}$ were considered initially, the general result observed was that the values for $\boldsymbol{R}^{(0)}$ have a profound effect upon the progress of $\boldsymbol{F}^{(n)}$ (since $\boldsymbol{F}^{(1)} = \boldsymbol{R}^{(0)}$), though $\boldsymbol{F}^{(0)}$ does not. Consequently, the results presented assume $[\boldsymbol{F}^{(0)}]_4 = P^{(0)} = 1$ throughout. Although various values for initial sample size were simulated, only $N = 10$, $N = 15$, $N = 25$ (since these are most often used in practice) and $N = 100$ (to simulate an "infinite" population) are presented.

Table 7. *Progress of populations of varying sizes, varying initial coefficients of parentage, and varying gametic frequencies through fifteen cycles of recurrent selection assuming a maximum inbreeding selection pattern with* $P^{(0)} = 1$

Initial Conditions			Values of the Panmictic Index ($P^{(n)} = [\boldsymbol{F}^{(n)}]_4$) at the End of the Indicated Recurrent Cycle									
$\boldsymbol{R}^{(0)}$	p_1	N	1	2	3	4	5	6	7	8	10	15
$\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$	.30	10	1.00	.71	.63	.55	.49	.44	.39	.36	.31	.25
		15	1.00	.69	.63	.57	.53	.49	.46	.43	.39	.34
		25	1.00	.67	.64	.59	.56	.54	.52	.50	.47	.43
		100	1.00	.65	.64	.61	.61	.60	.59	.59	.58	.57
$\begin{bmatrix}.25\\.25\\.25\\.25\end{bmatrix}$	.30	10	.25	.34	.31	.30	.28	.27	.26	.25	.24	.23
		15	.25	.32	.30	.31	.31	.31	.31	.31	.31	.31
		25	.25	.31	.30	.33	.34	.35	.36	.37	.38	.39
		100	.25	.30	.30	.35	.37	.40	.42	.44	.47	.51
$\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$	.35	10	1.00	.73	.66	.60	.56	.53	.51	.50	.48	.47
		15	1.00	.71	.66	.61	.58	.56	.55	.54	.52	.51
		25	1.00	.69	.66	.62	.60	.59	.58	.57	.56	.56
		100	1.00	.67	.66	.64	.63	.63	.62	.62	.62	.62
$\begin{bmatrix}.25\\.25\\.25\\.25\end{bmatrix}$	.35	10	.25	.38	.40	.43	.44	.45	.45	.46	.46	.46
		15	.25	.36	.39	.43	.46	.47	.48	.49	.50	.51
		25	.25	.35	.39	.45	.47	.49	.51	.52	.54	.55
		100	.25	.33	.38	.45	.49	.52	.55	.56	.59	.61
$\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$	.45	10	1.00	.79	.75	.72	.71	.70	.70	.70	.69	.69
		15	1.00	.77	.75	.72	.71	.71	.70	.70	.70	.70
		25	1.00	.76	.74	.72	.71	.71	.71	.71	.71	.71
		100	1.00	.74	.74	.72	.72	.72	.72	.72	.71	.71
$\begin{bmatrix}.25\\.25\\.25\\.25\end{bmatrix}$	.45	10	.25	.52	.60	.64	.67	.68	.69	.69	.69	.69
		15	.25	.51	.59	.64	.67	.68	.69	.69	.70	.70
		25	.25	.50	.58	.64	.67	.68	.69	.70	.70	.70
		100	.25	.49	.57	.64	.67	.69	.70	.70	.71	.71

In Table 6, one may observe the effect of linkage in maintaining a high degree of panmixia in the population. If $p_1 = .5$ in the two-locus case, the result is the same as for the one-locus case and the speed of the progress of the panmictic index towards an eventual value of zero depends upon the initial population size; if $p_1 = .25$, implying independent segregation, the value of $P^{(n)}$ tends to zero as n becomes large, as can be seen in Table 8; however, for intermediate values of p_1 the recurrent selection system results in larger terminal values of $P^{(n)}$, regardless of initial values of $\boldsymbol{R}$ and/or $\boldsymbol{F}$, regardless of population size, and regardless of whether the breeding pattern results in minimum or maximum inbreeding. When $\boldsymbol{R}^{(0)}$ is large initially, this selection system results in an increase in successive values of P; in some instances an increase is observed for a few cycles followed by a decrease, as illustrated by Fig. 3. Several researchers have reported increases in variability following one or two cycles of recurrent selection (Johnson and El Banna (1957), Lonnquist (1951, 1961), and Penny, *et al.* (1963)); if the source population were composed of inbred lines or consisted of related lines, this could well be the reason for the observed phenomenon.

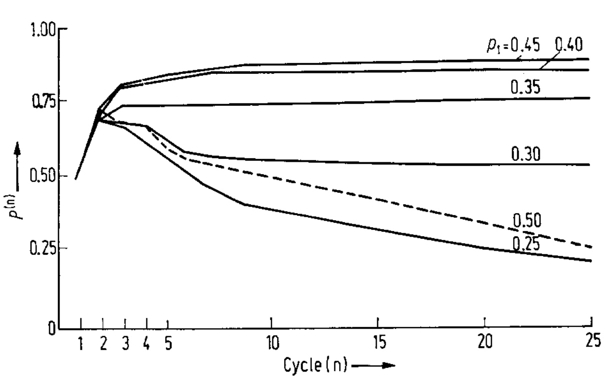

Fig. 3. Progress of the panmictic index over 25 cycles of recurrent selection assuming minimum inbreeding, $N = 15$, and

$$\boldsymbol{F}^{(0)} = \boldsymbol{R}^{(0)} = \begin{bmatrix}.250\\.125\\.125\\.500\end{bmatrix}$$

6. Mean and Variance of the Number of Homozygous Loci

In the case of m loci, the mean number of loci homozygous by descent in an individual α at the end of

Table 8. *Progress of populations of varying initial inbreeding coefficients and varying coefficients of parentage and various population sizes for both minimum and maximum inbreeding selection patterns for the case of independent segregation of two loci* ($p_1 = p_2 = 1/4$)

Initial Conditions			Progress of $P^{(n)}$ Assuming Minimum Inbreeding							Progress of $P^{(n)}$ Assuming Maximum Inbreeding						
$\boldsymbol{R}^{(0)}$	$\boldsymbol{F}^{(0)}$	N	1	2	3	4	5	10	15	1	2	3	4	5	10	15
$\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$	10	1.00	.92	.80	.70	.62	.32	.17	1.00	.71	.61	.51	.43	.16	.02
		15	1.00	.95	.87	.80	.73	.50	.37	1.00	.68	.62	.54	.48	.28	.16
		25	1.00	.97	.92	.87	.83	.67	.57	1.00	.66	.62	.57	.53	.39	.30
		100	1.00	.99	.98	.97	.96	.91	.88	1.00	.64	.63	.60	.59	.55	.52
$\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$	$\begin{bmatrix}.25\\.25\\.25\\.25\end{bmatrix}$	10	1.00	.91	.79	.69	.60	.31	.16	1.00	.66	.59	.50	.42	.15	.02
		15	1.00	.94	.85	.78	.72	.49	.36	1.00	.63	.61	.53	.48	.27	.15
		25	1.00	.96	.91	.87	.82	.67	.57	1.00	.61	.62	.56	.53	.39	.30
		100	1.00	.99	.98	.97	.97	.91	.88	1.00	.58	.63	.60	.59	.55	.52
$\begin{bmatrix}.25\\.25\\.25\\.25\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$	10	.25	.46	.41	.36	.31	.16	.09	.25	.32	.24	.21	.16	.02	.00
		15	.25	.48	.46	.43	.40	.31	.26	.25	.30	.24	.23	.20	.12	.07
		25	.25	.50	.50	.49	.48	.45	.44	.25	.29	.24	.25	.24	.21	.19
		100	.25	.51	.54	.56	.58	.59	.61	.25	.28	.24	.27	.28	.34	.37
$\begin{bmatrix}.25\\.25\\.25\\.25\end{bmatrix}$	$\begin{bmatrix}.25\\.25\\.25\\.25\end{bmatrix}$	10	.25	.45	.39	.34	.30	.16	.08	.25	.27	.23	.19	.15	.01	.00
		15	.25	.47	.45	.42	.39	.31	.26	.25	.25	.23	.22	.19	.11	.06
		25	.25	.49	.49	.48	.48	.45	.43	.25	.24	.24	.24	.23	.21	.19
		100	.25	.51	.54	.56	.58	.59	.69	.25	.22	.24	.27	.28	.34	.37

the kth cycle of a recurrent selection program is given by

$$E(\nu, \alpha) = \sum_{i=1}^{2^m-1} \nu(i, \alpha)\, [\boldsymbol{F}_\alpha^{(k)}]_i\,, \tag{6.1}$$

and the variance of the number of loci homozygous by descent is given by

$$\mathrm{Var}\,(\nu, \alpha) = \sum_{i=1}^{2^m-1} [\nu\,(i, \alpha)]^2\, [\boldsymbol{F}_\alpha^{(k)}]_i - [E\,(\nu, \alpha)]^2\,, \tag{6.2}$$

where $\nu\,(i, \alpha)$ is the number of non-zero elements in $(X_1\, X_2 \ldots X_m)_\alpha$ (see Section 3). For example, in the two-locus case given in Section 5,

$$E(\nu, \alpha) = 2\, F_\alpha(\psi_1) + F_\alpha(\psi_2) + F_\alpha(\psi_3)$$

and

$$\mathrm{Var}\,(\nu, \alpha) = 4\, F_\alpha(\psi_1) + F_\alpha(\psi_2) + F_\alpha(\psi_3) - [E(\nu,\alpha)]^2,$$

where, to simplify notation, $F = F^{(k)}$. Furthermore, the mean number of heterozygous loci in this case is $F_\alpha(\psi_2) + F_\alpha(\psi_3) + 2\, F_\alpha(\psi_4)$. Given independent segregation with $F =$ Prob {the genes at one locus are identical by descent}, then $F_\alpha(\psi_1) = F^2$, $F_\alpha(\psi_4) = 1 - F$, $F_\alpha(\psi_2) + F_\alpha(\psi_3) = 2\, F(1 - F)$, so that

$$2\, F_\alpha(\psi_1) + F_\alpha(\psi_2) + F_\alpha(\psi_3) = 2\, F^2 + 2\, F\,(1 - F) = 2\, F$$

and

$$F_\alpha(\psi_2) + F_\alpha(\psi_3) + 2\, F_\alpha(\psi_4) = 2\,(1 - F)\,,$$

as indicated by Kempthorne (1957); however with linkage and selection these relationships do not necessarily hold. Also, assuming independence, the variance of the number of heterozygous loci is $2\, F\,(1 - F)$; though with linkage and selection an increase or decreacse in the quantity would depend upon the selection pattern and the linkage parameters.

7. The Generalized Coefficient of Homozygosity and the Generalized Heterozygosity Index

Let the generalized homozygosity coefficients for m linked loci be denoted by

$$H_\alpha\,(X_1 X_2 \ldots X_m) = H\,(\{Y\}_\alpha)\,,$$

where X_k, $k = 1, \ldots, m$ is 1 or 0 according to whether or not the kth locus in the individual α has genes identical in state or by descent and where $\{Y\}_\alpha$ represents the set of genes identical by descent or in state at α; similarly, denote by

$$T_{\alpha\beta}\,(X_1 X_2 \ldots X_m) = T\,(\{Y\}_{\alpha\beta})$$

the coefficients measuring the degree of homozygosity between individuals of a population; H and T are analogous to F and R, repectively, measuring degrees of total homozygosity—i.e. homozygosity either in state or by descent. Then $\boldsymbol{H}_\alpha$ and $\boldsymbol{T}_{\alpha\beta}$ may be defined as vector quantities having $H\,(\{Y\}_\alpha)$ and $T\,(\{Y\}_{\alpha\beta})$, respectively, as components in the same manner as in Section 3, and the derivation of formulae analogous to (3.9) and (3.10) follows the same line of reasoning as in Section 3.

Thus, given $\boldsymbol{H}^{(0)}$ and $\boldsymbol{T}^{(0)}$, $\boldsymbol{H}^{(1)} = \boldsymbol{T}^{(0)}$, and the progress of total homozygosity in a recurrent selection program may be traced. Since, unter intense selection, the total homozygosity is more likely to affect the genetic variance than inbreeding alone, it may be this index of total homozygosity that the plant breeder would find more useful.

Indicative of results for the two-locus case would be those of the example given in Section 5 for the cases $[\boldsymbol{T}^{(0)}]_4 \neq 1$ and the results shown in Fig. 3. Since, by assumption of two alleles per locus, it would be impossible for $[\boldsymbol{T}^{(0)}]_4 = 1$, it follows that Shikata's method would not hold for this index and the generalized method of Section 3 would be required.

A specific example was simulated on the computer to compare the progress of total heterozygosity in a population

$$\text{for which } \boldsymbol{H}^{(0)} = \boldsymbol{T}^{(0)} = \begin{bmatrix} .25 \\ .25 \\ .25 \\ .25 \end{bmatrix} \text{ with that in a popula-}$$

$$\text{tion for which } \boldsymbol{H}^{(0)} = \boldsymbol{T}^{(0)} = \begin{bmatrix} .29 \\ .28 \\ .28 \\ .15 \end{bmatrix}, \text{ varying the po-}$$

pulation sizes and the recombination probabilities, and observing the progress of $[\boldsymbol{H}^{(n)}]_4$ in the case of both maximum and minimum inbreeding. The effect of differing population sizes and differing recombination probabilities may be observed in Table 9, and comparison of Table 9 with Table 6 shows how minor is the effect of varying $\boldsymbol{H}^{(0)}$ (or $\boldsymbol{F}^{(0)}$). In Fig. 4 one may observe a comparison of the results of minimum versus maximum inbreeding in the cases of complete independence, complete linkage, and an intermediate recombination probability; again, the effect of linkage (if it is not complete linkage) in promoting heterozygosity is noticeable.

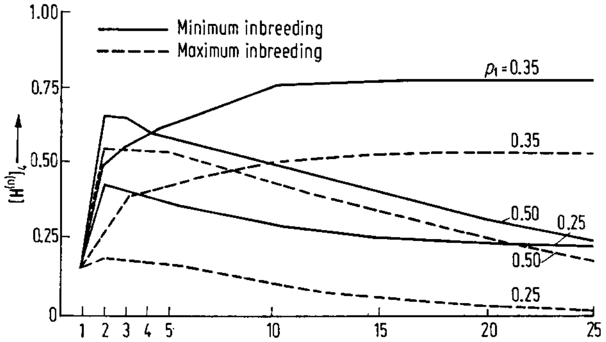

Fig. 4. Comparison of $[\boldsymbol{H}^{(n)}]_4$ for the cases of minimum and maximum inbreeding for $N = 15$ and varying recombination probabilities. $\boldsymbol{H}^{(0)} = \boldsymbol{T}^{(0)} = \begin{bmatrix} .29 \\ .28 \\ .28 \\ .15 \end{bmatrix}$ is assumed

Table 9. *Progress of populations of varying sizes, varying degrees of initial homozygosity (in state or by descent), and varying degrees of linkage, assuming a minimum inbreeding pattern*

Initial Conditions			Values of $[\boldsymbol{H}^{(n)}]_4$ at the End of the Indicated Recurrent Cycle											
$\boldsymbol{H}^{(0)} = \boldsymbol{T}^{(0)}$	p_1	N	1	2	3	4	5	6	7	8	10	15	20	25
$\begin{bmatrix} .25 \\ .25 \\ .25 \\ .25 \end{bmatrix}$	.30	10	.25	.47	.46	.45	.44	.43	.43	.41	.40	.39	.38	.38
		15	.25	.49	.51	.52	.52	.52	.52	.52	.53	.53	.53	.53
		25	.25	.51	.55	.57	.59	.61	.62	.63	.64	.67	.68	.68
		100	.25	.53	.60	.65	.68	.71	.74	.76	.80	.85	.88	.90
$\begin{bmatrix} .29 \\ .28 \\ .28 \\ .15 \end{bmatrix}$	.30	10	.15	.41	.42	.41	.41	.40	.40	.40	.39	.39	.38	.38
		15	.15	.43	.47	.48	.49	.50	.50	.51	.51	.52	.53	.53
		25	.15	.45	.51	.54	.56	.58	.59	.61	.63	.66	.67	.68
		100	.15	.47	.56	.61	.65	.68	.71	.74	.78	.85	.88	.89
$\begin{bmatrix} .25 \\ .25 \\ .25 \\ .25 \end{bmatrix}$	.35	10	.25	.50	.56	.59	.60	.62	.63	.63	.64	.64	.65	.65
		15	.25	.52	.60	.65	.68	.69	.71	.72	.73	.74	.74	.75
		25	.25	.54	.64	.70	.73	.76	.78	.80	.81	.83	.84	.84
		100	.25	.55	.68	.76	.81	.84	.87	.89	.92	.95	.95	.96
$\begin{bmatrix} .29 \\ .28 \\ .28 \\ .15 \end{bmatrix}$	.35	10	.15	.44	.52	.56	.59	.60	.62	.62	.64	.64	.65	.65
		15	.15	.47	.57	.62	.66	.68	.70	.71	.73	.74	.74	.75
		25	.15	.48	.60	.67	.72	.75	.77	.79	.81	.83	.84	.84
		100	.15	.50	.65	.73	.79	.83	.86	.88	.91	.95	.95	.95
$\begin{bmatrix} .25 \\ .25 \\ .25 \\ .25 \end{bmatrix}$	.45	10	.25	.61	.74	.79	.82	.83	.84	.84	.84	.84	.85	.85
		15	.25	.63	.77	.83	.86	.87	.88	.89	.89	.89	.90	.90
		25	.25	.65	.80	.86	.89	.91	.92	.93	.93	.93	.94	.94
		100	.25	.66	.82	.90	.93	.95	.96	.97	.98	.98	.98	.98
$\begin{bmatrix} .29 \\ .28 \\ .28 \\ .15 \end{bmatrix}$	.45	10	.15	.57	.72	.78	.81	.82	.83	.84	.84	.85	.85	.85
		15	.15	.59	.75	.82	.85	.87	.88	.88	.89	.89	.90	.90
		25	.15	.60	.77	.85	.88	.90	.92	.92	.93	.93	.94	.94
		100	.15	.61	.80	.88	.92	.95	.96	.97	.98	.98	.98	.98

As an index of total heterozygosity within or between individuals in a population, the 2^mth component of $\boldsymbol{H}$ or of $\boldsymbol{T}$ respectively should be used. The means and variances of the number of loci homozygous by descent or in state or heterozygous by descent or in state would be derived in the same manner as in Section 6.

8. Summary and Conclusions

For the case of m linked loci with effective selection, the generalized inbreeding coefficient in the $(n+2)$ *n*d generation was given by formulae (3.9) and (3.10). Tables 6 and 7 illustrate the progress of simulated populations for the two-locus case when various initial conditions are assumed for minimum and maximum inbreeding recurrent selection patterns.

The progress of populations assuming m independently segregating loci may be traced as a special case of m linked loci by appropriate specification of the recombination probabilities; for example, in the two-locus case $p_1 = p_2 = 1/4$ implies independent segregation and Table 8 outlines the progress of populations under both minimum and maximum inbreeding in this case. Fig. 3 and 4 illustrate the effect of linkage on the degree of panmixia when large initial coefficients of parentage are assumed; an increase in panmixia occurs in the first few cycles even in the presence of selection.

In the two-linked-locus situation simulated by computer, it was found that in most instances, very little change occurred in the panmictic index after fifteen generations, but the size of the panmictic index assuming minimum inbreeding was from about twenty-five percent greater than that assuming maximum inbreeding with the same initial conditions. Thus for a recurrent selection program in this case linkage tended to promote heterozygosity, particularly with a minimum inbreeding selection pattern.

The plant breeder is concerned with measuring the degree of inbreeding in a population under recurrent selection. When pedigree information is known, the methods of Section 3 yield far more precise results than have been possible previously; with no available pedigree information, a minimum degree of inbreeding may still be obtained using these methods and assuming equal representation of all lines. If the experimenter desires a generalized heterozygosity index in order to obtain an estimate of the variability and expected progress of the selected population in succeeding generations, then the methods of Section 7 would apply. Specifically, Table 6, 7, and 8 can be used to estimate minimum and maximum inbreeding coefficients in experiments for which the outlined assumptions can be made for two-locus cases in a recurrent selection program. Likewise the portions of these tables which do not assume $[\boldsymbol{R}^{(0)}]_4 = 1$ can be used to estimate minimum and maximum total heterozygosity coefficients in a population under recurrent selection. In order to trace the progress of overall homozygosity, either by descent or in state in a population, one may utilize the same derivations as were developed to trace the homozygosity by descent.

From equation (3.9) one can observe the large relative contribution of the coefficients of relationship; assuming this coefficient to be zero, as is so often done in practice in calculating the degree of inbreeding in a recurrent selection program, may very well result in major errors. The constant multipliers representing the effect of selection are

$$\left[\binom{N}{2} - \Sigma\binom{b_i}{2}\right] \Big/ \binom{N}{2}, \qquad (8.1)$$

and

$$\Sigma\binom{b_i}{2} \Big/ \binom{N}{2} \qquad (8.2)$$

where (8.1) is the coefficient of $\boldsymbol{R}^{(n)}$ and (8.2) the coefficient of an expression involving both $\boldsymbol{R}^{(n)}$ and $\boldsymbol{F}^{(n)}$. Under minimum inbreeding conditions (8.1) equals $\frac{N-3}{N-1}$ and (8.2) equals $\frac{2}{N-1}$; thus as N becomes large the major contribution to the inbreeding coefficient in cycle n is made by the term containing only the coefficient of parentage in cycle $n-2$. Under maximum inbreeding conditions, (8.1) equals $\frac{2(N-2)}{N(N-1)}$ and (8.2) equals $\frac{N^2-3N+4}{N(N-1)}$, so that as N becomes large the major contribution to the inbreeding coefficient in cycle n comes from the term involving both the inbreeding coefficient and the coefficient of relationship in cycle $n-2$. Extending this to the generalized heterozygosity index, the recurrent selection pattern is such that the importance of the degree of heterozygosity present between the individuals of a breeding population far outweighs the importance of that within individuals in the previous cycles in determining the degree of heterozygosity for a given cycle.

References

1. Bodmer, W. F., Parsons, P. A.: Linkage and Recombination in Evolution. Adv. in Gen. **11**, 1–99 (1962). — 2. Cain, R. L., Hinkelmann, K.: Coefficients of Inbreeding and Homozygosity in Recurrent Selection: The One-Locus Case. Theor. appl. Gen. **40**, 327–335 (1970). — 3. Cockerham, C. C., Weir, B. S.: Sib Mating with two Linked Loci. Genetics **60**, 629–640 (1968). — 4. Geiringer, H.: On the Probability Theory of Linkage in Mendelian Heredity. Ann. math. Statistics **15**, 25–57 (1944). — 5. Johnson, I. J., El Banna, A. S.: Effectiveness of Successive Cycles of Phenotypic Recurrent Selection in Sweetclover. Agronomy J. **49**, 120–125 (1957). — 6. Kempthorne, O.: An Introduction to Genetic Statistics. New York: John Wiley & Sons 1957. — 7. Kimura, M.: A Probability Method for Treating Inbreeding Systems, Especially with Linked Genes. Biometrics **19**, 1–17 (1963). — 8. Lonnquist, J. H.: Recurrent Selection as a Means of Modifying Combining Ability in Corn. Agronomy J. **43**, 311–315 (1951). — 9. Lonnquist, J. H.: Progress from Recurrent Selection Procedures for the Improvement of Corn Populations. Univ. of Nebraska College of Agric. Res. Bull. No. 197 (1961). — 10. Malécot, G.:

Les Mathématiques de l'Hérédité. Paris: Masson et Cie. 1948. — 11. Narain, P.: Homozygosity in a Selfed Population with an Arbitrary Number of Linked Loci. J. of Gen. **46**, 254—266 (1966). — 12. Penny, L. H., Russell, W. A., Sprague, G. F., Hallauer, A. R.: Recurrent Selection. Statistical Genetics and Plant Breeding, National Academy of Science, NRC Publ. No. **982**, 352—367 (1963). — 13. Schnell, F. W.: Some General Formulations of Linkage Effects in Inbreeding. Genetics **46**, 947—957 (1961). — 14. Shikata, M.: The Generalized Inbreeding Coefficient in a Markov Process. Reports of Statistical Applications in Research, JUSE **9-A** (4), 127—136 (1962). — 15. Shikata, M.: Representation and Calculation of Selfed Population by Group Ring. J. of theor. Biol. **5**, 142—160 (1963). — 16. Shikata, M.: Inference Effect of Cross-overs in Selfed Populations. J. of theor. Biol. **7**, 181—223 (1964). — 17. Shikata, M.: A Generalization of the Inbreeding Coefficient. Biometrics **21**, 665—681 (1965a). — 18. Shikata, M.: Limit of Application of Generalized Inbreeding Coefficient and Generalized Coefficient of Parentage. Ann. Rep. Res. Group on Biophysics in Japan, **5**, 65—66 (1965b). — 19. Shikata, M.: Transformation of Generalized Inbreeding Coefficient in Components of Pedigree. J. of theor. Biol. **10**, 11—14 (1966). — 20. Shikata, M.: Effect of Selections upon Recombination Probabilities in Selfed Populations. J. of theor. Biol. **14**, 59—65 (1967a). — 21. Shikata, M.: An Effect of Interference of Cross-overs on Inbreeding Coefficient in Selfed Populations. Jap. J. of Breeding **17**, 245—250 (1967b). — 22. Shikata, M.: An Effect of Interference of Cross-overs on Inbreeding Coefficient in Closed Loop Pedigrees. Jap. J. of Breeding **18**, 1—6 (1968). — 23. Sprague, G. F., Miller, P. A., Brimhall, B.: Additional Studies of the Relative Effectiveness of Two Systems of Selection for Oil Content of the Corn Kernel. Agron. J. **44**, 329—331 (1952). — 24. Weir, B. S.: Note. J. of theor. Biol. **30** (1971) (In press). — 25. Weir, B. S., Cockerham, C. C.: Pedigree Mating with Two Linked Loci. Genetics **61**, 923—940 (1969). — 26. Wright, S.: Systems of Mating I—V. Genetics **6**, 111—178 (1921).

Received November 22, 1971

Communicated by W. Seyffert

Dr. R. L. Cain
Morehead State University
Morehead, Kentucky 40351 (USA)

Dr. K. Hinkelmann
Virginia Polytechnic Institute
and State University
Blacksburg, Virginia 24061 (USA)