Theoretical and Applied Genetics 40, 327–335 (1970)

# Coefficients of Inbreeding and Homozygosity in Recurrent Selection: The One-Locus Case

**R. L. Cain**[1] and **K. Hinkelmann**

Virginia Polytechnic Institute, Blacksburg, Virginia (USA)

**Summary.** For selection programs which can be represented by successive self-select-intercross cycles (such as recurrent selection or reciprocal recurrent selection) general recurrence formulae are developed for obtaining the coefficients of inbreeding and homozygosity in each cycle. The formula for the coefficient of inbreeding is a generalization of a result given by Sprague, *et al.* (1952). It is shown that the coefficient of parentage in the source population has a major effect on the coefficient of inbreeding in the following cycles as does the population size. The relationship of both types of coefficients and their importance in practical work are discussed.

## 1. Introduction

By recurrent selection (*RS*) is meant a basic selection pattern which involves self-select-intercross cycles. There are four basic types of *RS* as distinguished by the manner in which the desirable individuals are located:

1. Simple recurrent selection (*SRS*): individuals are divided into groups of "discards" or "selects" solely on the basis of their own phenotypes (Jenkins 1935, 1940).
2. Recurrent selection for general combining ability (*RSGCA*): plants or animals are classified on the basis of the phenotypes of their progeny when crossed with a heterozygous tester stock (Jenkins 1935).
3. Recurrent selection for specific combining ability (*RSSCA*): plants or animals are classified on the basis of the phenotypes of their progeny when crossed with a homozygous tester stock (Hull 1945).
4. Reciprocal recurrent selection (*RRS*): selection for general combining ability is made simultaneously in two populations by observation of the phenotypes of progeny resulting from crosses involving two heterozygous source populations (Comstock, *et al.* 1949 and Robinson, *et al.* 1955).

A program of *RS* enables the experimenter to maintain a level of inbreeding which is lower than that inherent in many other types of breeding programs; thus it is possible to continue effective selection progress over a longer period of time if the selected trait is controlled by many loci. If the trait is controlled by only a few loci and selection is artificial (thus being relatively intense) and if selection is effective, then *RS* will result in genetic advance for only a few cycles. However both inbreeding and homozygosity in state have an increasing effect upon the progress of a population under selection as the intensity of selection increases, since genetic advance is dependent upon the variability present in the population and with intense effective selection, genetic fixation is rapidly approached. Thus, homozygosity by descent and homozygosity in state both contribute to the same end result: genetic fixation and cessation of selection advance. *RS* was developed as a method whereby plant breeders especially could utilize the most intense form of inbreeding, selfing, to select effectively the superior individuals and retain their superior genes in the gene pool, while alternately utilizing the least intense form of inbreeding, crossing in all possible combinations, to maintain the genetic variability necessary to enable continued selection progress over several selection cycles.

*SRS* is effective only if the character being selected has high heritability and if the character is governed by a relatively large number of loci (if more than two or three cycles are planned). *RSSCA* requires sound judgment as to the value of the tester stock: a poor choice will invalidate the results of many years of experimentation, since a single homozygous tester stock must be chosen initially and perpetuated by selfing for use at every other year of the selection cycle. It is generally concluded that *SRS*, *RSSCA*, and *RSGCA* are best suited to their own specific purposes, while *RRS* enhances the genetic diversity of breeding stocks and offers the most advantages in an overall program of selection in crop plants, particularly when selection involves yield as one of the selected characteristics and when the crop involved is known to show heterosis upon crossing pure lines.

Illustrated in Figure 1 is the mating pattern for *SRS*, *RSSCA*, *RSGCA* and *RRS*. Following the notation of Shikata (1966) the four basic components of pedigree may be observed in Figure 1:

(i) the path $A-E$ illustrates selfing;

(ii) the path $EF-I$ illustrates that phase of intercrossing for which each offspring is a result of a cross of two parents;

(iii) the path $E-IK$ illustrates that phase of intercrossing for which a single parent contributes genetic material to more than one offspring or line, and

(iv) the path $E-IK-OQ-V$ illustrates the combining of the 3 components listed above to form

---

[1] Now at Morehead State University.

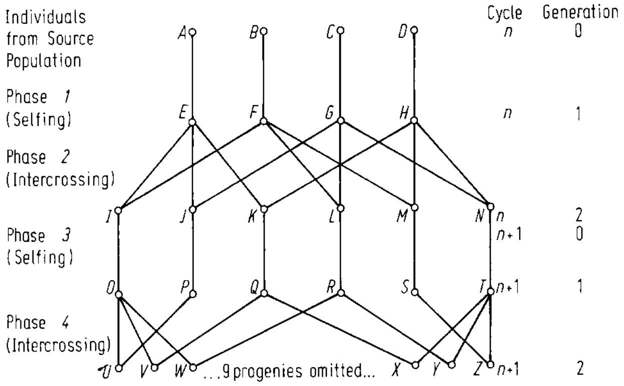

Fig. 1. Basic mating pattern for all four types of recurrent selection, showing only one offspring (of an assumed equal number of offspring) per mating

a closed loop. Note that in the *RS* system a closed loop encompasses three generations.

The very development of the *RS* procedure as one which minimizes inbreeding in a selection program indicates the plant breeder's concern regarding the effect of inbreeding on selection progress. In some cases in practice, pedigree information is available; in others, maintenance of pedigree records is not feasible. In all cases, continuation of selection effectiveness depends upon the degree of heterozygosity in the breeding population; thus the breeder is concerned with measuring not just the degree of inbreeding but instead the overall homozygosity in a population under *RS*.

Kempthorne (1960) comments: "The calculation of inbreeding coefficients may be misleading in a population under selection;" and Falconer (1960) observes that: "the coefficient of inbreeding is a measure of the state of dispersion [of gene frequencies] only in the absence of selection." Falconer goes on to explain that there must be a distinction made between the state of dispersion of gene frequencies (affecting additive variance, inbreeding depression, etc.) and the coefficient of inbreeding as computed from the population size or pedigree relationships, since under selection the actual dispersion will be less than that indicated by $F$. Falconer is referring to selection which favors the heterozygotes, such as natural selection operating during inbreeding depression. However, similar comments apply in the case of intense directional selection.

For example consider a special case in which gene fixation occurs under selection simultaneously with an inbreeding coefficient of zero. In simple recurrent selection if the selected trait depends on only one locus, and the frequency of the favorable allele is 50% or more, then in the absence of complete dominance fixation would occur with the first generation for a selection intensity of 25% or less. Note, however, that the inbreeding coefficient is only one-half upon fixation, assuming the open-pollinated source population that is usually the case. However the complete *RS* selection cycle is self-select-intercross, so that though upon selfing the source population, $F = 1/2$, still after intercrossing to complete the cycle, $F = 0$ again, even in the presence of genetic fixation. In this special case, selection has — on the basis of the parent plants' observed genotypes — eliminated all heterozygotes and all recessive homozygotes and even though $F = 1/2$ in the selected selfed seeds, genetic fixation with $F = 0$ results upon intercrossing at the close of the cycle.

In general, inbreeding promotes homozygosis; likewise directional selection promotes homozygosis in the presence of partial or complete dominance. However, when combined, each inhibits the action of the other in a recurrent program; directional selection seeks to fix the favorable allele, while in the selfing phase 25% of the heterozygotes segregate as homozygous for the unfavorable allele; inbreeding increases the proportion of homozygous loci only to have it decreased as selection removes part or all of the recessive homozygotes (unless, of course, the selection intensity is such that a large proportion of the heterozygotes also are discarded). Thus one must approach the interpretation of the inbreeding coefficient with caution when selection is involved; and if the point of interest is the degree of homozygosity of the population, then a homozygosity index should be obtained rather than any measure of degree of inbreeding.

Furthermore, if selection is effective, then the superior individuals in the source population will be represented more often in the ancestry of successive selected progenies than will be the inferior individuals; and also the various lines selected in any given cycle will tend to be more or less inbred according to the number of exceptional performers in their ancestry. This tendency of genetically superior individuals to be most inbred results in a reduction of effective population size under intense selection; see Robertson (1961) for further details. As a measure of inbreeding in a given cycle, what is needed is an average inbreeding coefficient for the intercross population (since this marks the termination of a breeding cycle) of a given cycle of *RS*.

This paper presents, for the one-locus case, the derivation of recurrence formulae for an average inbreeding coefficient and an average coefficient of parentage in the $n$th cycle of an *RS* program, and develops an index with which to measure the total degree of homozygosity in a population under recurrent selection. For the case of m independently segregating loci, to obtain the average inbreeding coefficient, the coefficient of parentage, or the index of total homozygosity for a given cycle, one may simply raise the appropriate one-locus coefficient to the power m. Extension of the result to m loci with linkage will appear in a sequel paper.

## 2. Assumptions and Definitions

It is assumed that the individuals in the breeding populations involved are diploid with only two alleles per locus, are capable of self fertilization, and breed in nonoverlapping generations. Any possibility of differential viability is ignored and a constant number of offspring per mating (typified by a single line) is assumed.

The inbreeding coefficient $F$ is defined as the probability that two genes at one randomly chosen locus of a diploid individual are identical by descent; and the coefficient of parentage $r$ is defined as the probability that two genes, drawn from the same locus of two different randomly chosen individuals, are identical by descent (see Malécot 1948).

Thus, given an individual $X$ with genes $x_1$ and $x_2$ at a given locus, $F_X = P\ (x_1 = x_2)$; similarly $r_{XY} = 1/4\ [P\ (x_1 = y_1) + P\ (x_1 = y_2) + P\ (x_2 = y_1) + P\ (x_2 = y_2)]$. Further, if $X$ is the offspring of a cross of the individuals $A$ and $B$ and $Y$ is the offspring of a cross between individuals $C$ and $D$ then $r_{XY}$ may be denoted $r_{A\times B, C\times D}$ and if $Z$ is an offspring of a cross between $X$ and $Y$, then $F_Z = r_{XY}$. $F^{(j,k)}$ and $r^{(j,k)}$, where $j = 0, 1, 2$ and $k = 1, 2, \ldots, n$, will be used to denote the coefficients of inbreeding and parentage, respectively, in the jth generation of the kth recurrent cycle. $F^{(k)}$ and $r^{(k)}$, where $k = 1, 2, \ldots, n$, will be used to denote the degree of inbreeding and parentage in the terminal (i.e., the second) generation of the kth recurrent cycle, and will be used interchangeably with $F^{(2,k)}$ whenever emphasis of the generation involved is unnecessary (See Section 6 for further discussion of the relationship between these single-superscripted and double-superscripted quantities).

## 3. Derivation of Recurrence Formulae for a Finite Population under Random Selection

Consider a population which is reproducing under the mating pattern of $RS$, as illustrated by Figure 1. By the assumption of equal numbers of progeny and no differential viability, there is no loss of generality in tracing the progress of the population assuming one offspring (or line) per mating.

From Figure 1, it is apparent that even upon assuming one offspring per mating, only for the unrealistic source population sizes of $N = 2$ and $N = 3$ will the population size remain within manageable limits under no selection. A source population of size $N$, in generation 2 has expanded to size $N(N-1)/2$, in generation 4 has $M(M-1)/2$, where $M = N(N-1)/2$, individuals, etc.

Thus for purposes of investigating the change in the inbreeding coefficient, consider the progress of this population assuming that a population size of $N$ (the size of the initial source population) is maintained through an arbitrary number of generations by random selection of $N$ out of the $N(N-1)/2$ offspring in each cycle.

The assumption of random selection from a population of all possible intercrosses is equivalent to the assumption of random mating. Thus, the calculation of $F$ in a given generation is similar to that given by Wright (1921) and Malécot (1948) and summarized by Kempthorne (1957), except that in this case the mating pattern alternates selfing and intercrossing in successive generations.

Given any diploid individual $X$ in a certain cycle of a recurrent selection mating system, and given that $X$ has genes $x_1$ and $x_2$, if $x_1$ and $x_2$ are identical by descent, then either they resulted by descent from one out of a possible $N$ ancestors who was a common parent to half sibs whose selfed offspring subsequently crossed to produce $X$, or else they came separately from the parents not common to the half sibs whose selfed offspring crossed to produce $X$ and were identical by descent in those two parents. As an illustration, consider any individual (say $X$ in Figure 1) in generation 2 of cycle $n+1$ from the source population: tracing the closed loop involved (from $X$ to $Q-T$ to $K-N$ to $H$), one observes that both genes of this individual ($X$) could be traced to one common ancestor ($H$) in generation 1 of cycle $n$ with probability $1/N$ (1/4, in this case). On the other hand, if the two genes did not come from a common ancestor, an event that has probability $1 - 1/N$ (i.e., 3/4), then they must have descended simultaneously from two separate ancestors (i.e., one only from each $E$, $G$, or $H$). Thus, for a mating pattern similar to those illustrated by Figure 1, in which alternate generations have been produced by intercrossing, if one assumes $N$ lines in the source generation of cycle $n$, then

$$F^{(2,n+1)} = r^{(1,n+1)} = r^{(2,n)} = \frac{1}{N}\left[\frac{1+F^{(1,n)}}{2}\right] + \left(1-\frac{1}{N}\right)F^{(2,n)} \tag{3.1}$$

reflecting the effect of finite population size $N$, and

$$F^{(1,n+1)} = \frac{1}{2}\left(1 + F^{(2,n)}\right), \tag{3.2}$$

independent of the value of $N$.

In terms of the panmictic index $P = 1 - F$,

$$P^{(2,n+1)} = \left(1-\frac{1}{N}\right)P^{(2,n)} + \frac{1}{2N}P^{(1,n)}, \tag{3.3}$$

and

$$P^{(1,n+1)} = \frac{1}{2}P^{(2,n)}, \tag{3.4}$$

which when combined with (3.3) yields

$$P^{(2,n+1)} = \left(1-\frac{1}{N}\right)P^{(2,n)} + \frac{1}{4N}P^{(2,n-1)}, \tag{3.5}$$

or

$$P^{(n+1)} = \left(1-\frac{1}{N}\right)P^{(n)} + \frac{1}{4N}P^{(n-1)}. \tag{3.6}$$

Again it should be emphasized that generations 0 and 2 are defined to be those produced by intercrossing. As shown by (3.4), the panmictic index of

Table 1. *Progress of the panmictic index ($P^{(n)}$) through 15 cycles of recurrent selection for varying population sizes ($N$) and varying values of the panmictic index in the source population ($P^{(0)}$)*

| $N$ | Roots of Quadratic | | $P(0)$ | Recurrent Cycle Number ($n$) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $r_1$ | $r_2$ | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 4 | +.8257 | −.0757 | 1.0 | .81 | .67 | .55 | .46 | .38 | .31 | .26 | .21 | .18 | .15 | .12 | .10 | .08 | .07 |
| | | | 0.8 | .65 | .54 | .44 | .37 | .30 | .25 | .21 | .17 | .14 | .12 | .10 | .08 | .06 | .05 |
| | | | 0.6 | .49 | .40 | .33 | .27 | .23 | .19 | .15 | .13 | .11 | .08 | .07 | 0.6 | .05 | .04 |
| | | | 0.4 | .32 | .27 | .22 | .18 | .15 | .12 | .10 | .08 | .07 | .06 | 0.5 | .04 | .03 | .03 |
| | | | 0.2 | .16 | .13 | .11 | .09 | .08 | .06 | .05 | .04 | .04 | .03 | .02 | .02 | .02 | .01 |
| 5 | +.8583 | −.0583 | 1.0 | .85 | .73 | .63 | .54 | .46 | .40 | .34 | .29 | .25 | .22 | .18 | .16 | .14 | .12 |
| | | | 0.8 | .68 | .58 | .50 | .43 | .37 | .32 | .27 | .23 | .20 | .17 | .15 | .13 | .11 | .09 |
| | | | 0.6 | .51 | .44 | .36 | .32 | .28 | .24 | .20 | .18 | .15 | .13 | .11 | .09 | .08 | .07 |
| | | | 0.4 | .34 | .29 | .25 | .22 | .18 | .16 | .14 | .12 | .10 | .09 | .07 | .06 | .05 | .05 |
| | | | 0.2 | .17 | .15 | .13 | .11 | .09 | .08 | .07 | .06 | .05 | .04 | .04 | .03 | .02 | .02 |
| 10 | +.9270 | −.0270 | 1.0 | .93 | .86 | .79 | .74 | .68 | .63 | .59 | .54 | .50 | .47 | .43 | .40 | .37 | .35 |
| | | | 0.8 | .74 | .69 | .64 | .59 | .55 | .51 | .47 | .44 | .40 | .37 | .35 | .32 | .30 | .28 |
| | | | 0.6 | .56 | .51 | .48 | .44 | .41 | .38 | .35 | .33 | .30 | .28 | .26 | .24 | .22 | .21 |
| | | | 0.4 | .37 | .34 | .32 | .29 | .27 | .25 | .23 | .22 | .20 | .19 | .17 | .16 | .15 | .14 |
| | | | 0.2 | .19 | .17 | .16 | .15 | .14 | .13 | .12 | .11 | .10 | .09 | .09 | .08 | .07 | .07 |
| 25 | +.9703 | −.0103 | 1.0 | .97 | .94 | .91 | .89 | .86 | .83 | .81 | .79 | .76 | .74 | .72 | .70 | .68 | .66 |
| | | | 0.8 | .78 | .75 | .73 | .71 | .69 | .67 | .65 | .63 | .61 | .59 | .57 | .56 | .54 | .52 |
| | | | 0.6 | .58 | .56 | .54 | .53 | .52 | .50 | .49 | .47 | .46 | .44 | .43 | .42 | .41 | .39 |
| | | | 0.4 | .39 | .38 | .37 | .35 | .34 | .33 | .32 | .31 | .30 | .30 | .29 | .28 | .27 | .26 |
| | | | 0.2 | .19 | .19 | .18 | .18 | .17 | .17 | .16 | .16 | .15 | .15 | .14 | .14 | .14 | .13 |
| 100 | +.9925 | −.0025 | 1.0 | .99 | .99 | .98 | .97 | .96 | .96 | .95 | .94 | .93 | .92 | .92 | .91 | .91 | .90 |
| | | | 0.8 | .79 | .79 | .78 | .78 | .77 | .76 | .76 | .75 | .75 | .74 | .74 | .73 | .73 | .72 |
| | | | 0.6 | .60 | .59 | .59 | .58 | .58 | .57 | .57 | .56 | .56 | .56 | .55 | .55 | .54 | .54 |
| | | | 0.4 | .40 | .39 | .39 | .39 | .39 | .38 | .38 | .38 | .37 | .37 | .37 | .37 | .36 | .36 |
| | | | 0.2 | .20 | .20 | .20 | .19 | .19 | .19 | .19 | .19 | .19 | .19 | .18 | .18 | .18 | .18 |

any generation produced by selfing is half that of the parent generation, regardless of population size. However, the course of the panmictic index over time in terms of either generations or $RS$ cycles is reflected entirely by (3.5) or (3.6) respectively, since these expressions combine (3.3) and (3.4).

Thus for $n > 1$,

$$P^{(n)} = a_1 r_1^n + a_2 r_2^n$$

where $r_1$ and $r_2$ are roots of the quadratic

$$x^2 - \left(1 - \frac{1}{N}\right) x - \frac{1}{4N} = 0 \qquad (3.7)$$

obtained from (3.6) following the usual procedure for solving difference equations, and $a_1$ and $a_2$ are determined by the parameters in the source population.

Table 1 indicates the roots of the quadratic equation for varying values of $N$ and then, given values of $P^{(0)}$ varying from 0.2 through 1.0, traces the progress of $P^{(n)}$ from the second through the fifteenth recurrent cycles (through thirty breeding generations). Since it is immediately preceded by intercrossing, $P^{(1)} = P^{(0)}$. Only in or following the generation which terminates closed loops in the pedigree is it possible to reduce the degree of panmixia: i.e., no sooner than $P^{(2)}$.

The steady progress of $P^{(n)}$ toward an eventual value of zero is quite evident for small $N$. These data give an indication of the effect of finite population size and also the effect of degree of panmixia in the source population of a recurrent breeding program, assuming a single gene model and no artificial selection, over a period of 15 recurrent cycles. Figure 2 illustrates the effect of population size upon the progress of the population for $P^{(0)} = .5$ and $P^{(0)} = .8$.

## 4. Derivation of Recurrence Formulae for $F^{(2)}$ in a Finite Population under Effective Directional Selection

Consider Figure 3, which gives a possible selection pattern. Referring to the definitions of the four basic methods of $RS$ given in Section 1, one may observe that if the selection of lines $E$, $F$, $G$, and $H$ (or $O$, $P$, $Q$, and $R$ in the second cycle) in Figure 3 is based upon observation of phenotypes, then $SRS$ is involved; if their selection is based upon observation of the phenotypes of the progeny of $A$, $B$, $C$, and $D$ (or of $I$, $J$, $L$, and M in the second cycle) when crossed with a homozygous tester stock (heterozygous tester stock) then $RSSCA$ ($RSGCA$) is involved. With $RRS$, for each of the two populations, a mating series is conducted such as illustrated by Figure 3; and selection of, say individuals $E$, $F$, $G$, and $H$ (or $O$, $P$, $Q$, and $R$ in the second cycle), in the first population is based upon the observed phenotypes of progenies resulting from the outcrossing of $A$, $B$, $C$, and $D$ (or $I$, $J$, $L$, and $M$) with a random sample of individuals from the second population involved. Since, with $RRS$, the selection of lines is based on observation of the offspring of an outcrossing of these lines with a genetically unrelated population, the tendency to select the more inbred lines in either of the two populations is reduced. Thus, with *RRS*, one may expect the degree of inbreeding to be less, even though the procedure for calculating it is identical to that used in the case of the other three types of *RS*.

The following derivation is illustrated by Figure 3, though it is not restricted by the fact that Figure 3 specifies a particular case. Let $a_i$, $i = 1, \ldots, N$, be the number of times each line saved in cycle 1 is represented in the parents of progeny saved in cycle 2, where $N$ is the total number of lines saved. Let $b_i = a_i$ if $a_i \geq 2$, let $\sum' b_i$ denote the sum of all $b_i$, and let $\sum'' a_i$ denote the sum of all $a_i = 1$. Then $2N = \sum_{i=1}^{N} a_i = \sum' b_i + \sum'' a_i$. In the derivation, the inbreeding coefficient in the population is developed in terms of a weighted average of the inbreeding coefficients of the individuals comprising the population; because of the regularity of the breeding pattern, some population results may be derived in terms of corresponding results based on individual pedigrees.

In developing the results of this section, the following notation and definitions will be observed (in addition to those specified in Section 2):

$F_X$ denotes the inbreeding coefficient for the individual $X$;

$r_{XY}$ denotes the coefficient of parentage between $X$ and $Y$;

$K_1$ denotes the inbreeding coefficient of an individual which results from a closed loop pedigree;

$K_2$ denotes the inbreeding coefficient of an individual which results from a single line pedigree (not a closed loop);

$K_1'$ denotes the coefficient of relationship between two individuals who have a common parent; and

$K_2'$ denotes the coefficient of relationship between two individuals who do not have a common parent.

If $r = 0$ in the source population, then the only possible contribution to the degree of inbreeding in generation 2 of cycle 2, $F^{(2,2)}$, is that made by lines represented two or more times in the first generation of any cycle. If $r = 0$ in the source population, then any line represented only once, such as $H$ in Figure 3, cannot contribute to $F^{(2,2)}$. Notice in Figure 3 that in the second generation of the second cycle, $W$ is the only offspring that does not terminate a closed loop since the ancestor J results from the cross of $E$ and $G$ and the ancestor $M$ is $F \times H$. The offspring $S$, $T$, $U$, $V$, and $X$ can be traced to loop ancestors $E$, $F$, $F$, $G$, and $F$, respectively, reflecting the fact that $E$ is represented twice in generation 2 of cycle 1, $F$ is represented three times, and $G$ twice. By the above definitions, the value of the contribution to $F^{(2,2)}$ of individuals who are loop ancestors is $K_1$ since the size of the inbreeding coefficient of the closed loop offspring results from the possibility of transmission by the loop ancestor of identical genes through both sides of the loop. Similarly the contribution of $H$, as measured by $F_W$, is denoted $K_2$.

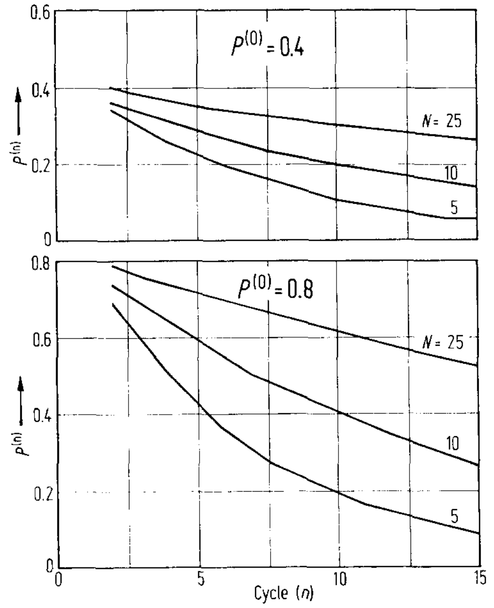

Fig. 2. Comparison of $P^{(n)}$ values for varying initial population sizes and fixed initial values of the panmictic index $P^{(0)}$

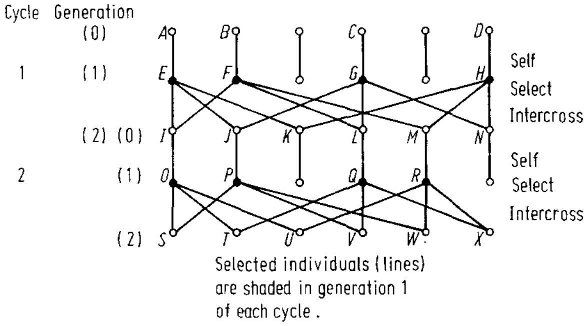

Fig. 3. A possible selection pattern illustrating derivation of the average inbreeding coefficient in the second cycle of recurrent selection

All lines represented exactly once in the first generation of any cycle (such as $H$ in Figure 3) will contribute $K_2 \sum'' a_i$ to $F^{(2,2)}$. Since any two lines in the selected population of the second cycle can have at most one common ancestor in the first cycle, any line represented twice (such as $E$) supplies exactly one closed loop and contributes $1 \times K_1$ to the inbreeding coefficient; any line represented exactly three times (such as $F$) supplies exactly three closed loops and contributes $3 \times K_1$; and, in general, one represented $a_i \geq 2$ times out of the total of $N\,(N-1)/2$ possible offspring contributes $K_1\, b_i\, (b_i - 1)/2$. Thus the total contribution to $F^{(2,2)}$ of individuals resulting from closed loop pedigrees is $K_1\, b_i\, (b_i - 1)/2$ and of those resulting from single line pedigrees is

$$K_2\left[\frac{N\,(N-1)}{2} - \sum \frac{b_i\,(b_i - 1)}{2}\right],$$

where

$$\frac{N\,(N-1)}{2} - \sum \frac{b_i\,(b_i-1)}{2} = \sum'' a_i\,.$$

By the regularity of the mating pattern, $K_1$ may be calculated as $F_S$ and thus

$$K_1 = F_S = r_{OP} = r_{IJ} = r_{E\times F,\,E\times G} = \\ = \frac{1}{4}\,(r_{EE} + r_{EF} + r_{EG} + r_{FG})\,,$$

from Figure 3; it follows that, in general,

$$K_1 = \frac{3}{4}\,r^{(1,1)} + \frac{1}{8}\,(1 + F^{(1,1)}) = \\ = \frac{3}{4}\,r^{(1,0)} + \frac{1}{8}\left[1 + \frac{1}{2}\,(1 + F^{(1,0)})\right],$$

and thus

$$K_1 = \frac{3}{4}\,r^{(1,0)} + \frac{1}{16}\,F^{(1,0)} + \frac{3}{16}. \tag{4.1}$$

Similarly

$$K_2 = F_W = r_{PR} = r_{JM} = r_{E\times G,\,F\times H} = \\ = \frac{1}{4}\,(r_{EF} + r_{EH} + r_{FG} + r_{GH})\,,$$

or, in general,

$$K_2 = r^{(1,0)}\,. \tag{4.2}$$

Thus, using (4.1) and (4.2), the average degree of inbreeding at the termination of cycle 2 is

$$F^{(2)} = F^{(2,2)} = \frac{2}{N\,(N-1)}\left\{\left[\frac{N\,(N-1)}{2} - \sum\frac{b_i\,(b_i-1)}{2}\right]\right. \\ \left.\times\, r^{(1,0)} + \left[\sum\frac{b_i\,(b_i-1)}{2}\right]\left[\frac{3}{4}r^{(1,0)} + \frac{1}{16}F^{(1,0)} + \frac{3}{16}\right]\right\}. \tag{4.3}$$

One may as well add the restriction $N > 3$ at this point, since for $N = 2$ no selection would be possible; and for $N = 3$, selection would either terminate the population or result in a population of size 2 with the same difficulties just mentioned.

The minimum inbreeding would occur if all lines were represented equally (i.e. twice) in which case $\frac{b_i\,(b_i-1)}{2} = 1$ for all $i$, so that

$$F^{(2)} = K_1\left[\frac{2}{N-1}\right] + K_2\left[\frac{N-3}{N-1}\right]. \tag{4.4}$$

## 5. Relationship to Derivation given by Sprague, et al. (1952)

If, in the derivation of (4.3) one were to delete the generation of selfing which is assumed to initiate the recurrent series (and which is almost always employed in practice) and if it were assumed that $F = r = 0$ in the source population, then $K_1 = 1/8$, and instead of (4.3),

$$F = \frac{1}{8}\left[\frac{\dfrac{b_1\,(b_1-1)}{2} + \cdots + \dfrac{b_N\,(b_N-1)}{2}}{\dfrac{N\,(N-1)}{2}}\right], \tag{5.1}$$

given by Sprague, et al. (1952). Thus this expression is valid for only one cycleand only under the restrictions noted above.

## 6. General Recurrence Formulae for $F^{(n+2)}$ and $r^{(n+2)}$

The derivation in Section 4 was made in terms of the first and second recurrent cycles in order to emphasize its relationship to the well-known formula given by Sprague, et al.; it is obvious, however, that the argument holds if 1 and 2 are replaced by $n$ and $n + 1$ to indicate a general cycle in the selection procedure. Also, since $r^{(n,0)}$ and $F^{(n,0)}$ represent the coefficients of parentage and inbreeding, respectively, in the source population of the $n$th cycle of recurrent selection, they could be expressed as $r^{(n-1)}$ and $F^{(n-1)}$, inasmuch as the source population of any given cycle is the population produced by intercrossing the selected population (in other words, the terminal generation) of the preceding cycle and it is the intercrossed population for which an inbreeding coefficient is desired. Thus, for $N > 3$,

$$F^{(n+2)} = \frac{2}{N\,(N-1)}\left\{\left[\frac{N\,(N-1)}{2} - \sum\frac{b_i\,(b_i-1)}{2}\right]r^{(n)} + \right. \\ \left. + \sum\frac{b_i\,(b_i-1)}{2}\left[\frac{3}{4}r^{(n)} + \frac{1}{16}F^{(n)} + \frac{3}{16}\right]\right\}. \tag{6.1}$$

To apply this formula to successive generations, one must be supplied the values of $r^{(0)}$ and $F^{(0)}$, recognize that $F^{(1)} = r^{(0)}$, and also have a recurrence formula for $r^{(n+1)}$.

A recurrence formula for $r^{(n+1)}$ may be obtained by observing that the contribution of any two individuals to the value of $r^{(n+1)}$ will be

$$K_1' = \frac{3}{4}\,r^{(n)} + \frac{1}{16}\,F^{(n)} + \frac{3}{16}$$

if the individuals have a common parent (by definition of the mating system, any two individuals in the $(n + 1)$st cycle may have at most one common parent); and the contribution will be

$$K_2' = r^{(n)}$$

if they do not have a common parent. Any single individual has two parents, of which one is a common parent with a set of $N - 2$ other individuals and the second is a common parent with a disjoint set of $N - 2$ other individuals. Furthermore, the two individuals in question are chosen from the $N\,(N - 1)/2$ offspring of the parent generation of size $N > 3$. Thus, given any two individuals in the $(n + 1)$st cycle of a recurrent selection program (the intercrossed generation), the probability that they have a parent in common is given by

$$1 - \frac{\dfrac{N\,(N-1)}{2} - 2\,(N-2) - 1}{\dfrac{N\,(N-1)}{2} - 1} = \frac{4}{N+1}$$

so that

$$r^{(n+1)} = \frac{4}{N+1}\,K_1' + \left(1 - \frac{4}{(N+1)}\right)K_2' = \\ = \frac{N}{N+1}\,r^{(n)} + \frac{F^{(n)} + 3}{4\,(N+1)}. \tag{6.2}$$

Note that the selection intensity, per se, has no effect on the inbreeding coefficient, though intense selection will increase the total homozygosity; however the number of lines (or individuals) saved has a profound effect on the size of the inbreeding coefficient. Also, since the breeding pattern calls for all possible intercrosses of the selected generation, the effect of selection upon the coefficient of parentage occurs only through the contribution of $F$ to $K_1'$.

## 7. The Homozygosity Index

Let $H$ denote the probability that two genes at a given locus in an individual are homozygous (either by descent or in state). Then $H$ is an index of total homozygosity and $1 - H$ an index of total heterozygosity within an individual. Let $t$ denote the probability that two genes randomly chosen from a given locus in two individuals are identical either by descent or in state. For a given individual or in a given population, these quantities may be calculated (or estimated) from observation of the genotype or from known (or estimated) gene frequencies.

The degree of inbreeding at any given cycle in a population under selection may not reflect the degree of homozygosis because the extent of the homogeneity in the source population is unknown, because selection has accelerated the approach toward genetic fixation, or because inbreeding depression is working against the trend of artificial selection so that the "more desirable" phenotypes are perhaps sterile or less viable. Thus if the experimenter relies upon the degree of inbreeding to yield a measure of expected variability upon which to base estimates of future selection effectiveness, his experimental results may deviate considerably from his predictions. In other words (see Section 1) the calculation of inbreeding coefficients may well be misleading, as observed by Kempthorne and Falconer. Specifically, the measures needed are $H$, an index of total homozygosity within the individuals of the breeding population, and $t$, an index measuring the genetic similarity among individuals in a breeding population.

Given $H$ and $t$ in a source population, the change in these quantities from one generation to the next is calculable in exactly the same manner as $F$ and $r$ for any given breeding pattern. Thus all of the development of Sections 4 and 6, specifically (6.1) and (6.2), holds if $F$ is replaced by $H$ and $r$ by $t$, yielding measures of the homozygosity in a population under selection; of course $H \geq F$ and $t \geq r$ in all cases. It would be these indexes which would be of value to the plant breeder in evaluating the variability to be expected in any generation of a recurrent selection breeding program. $H^{(0)}$ and $t^{(0)}$ may be estimated from known gene frequencies in the source population; and in the case of an open-pollinated source population, prior to artificial selecton, $H^{(0)} = t^{(0)}$.

## 8. Application to Special Cases

In Table 2 are outlined the numerical results obtained from (6.1) and (6.2) if one assumes equal

Table 2. *Progress of populations of varying sizes, varying initial inbreeding coefficients, and varying initial coefficients of parentage through twenty-five cycles of recurrent selection, assuming minimum inbreeding*

| Initial Conditions | | | Values of the Panmictic Index ($P = 1 - F$) at the End of the Indicated Recurrent Cycle | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $r^{(0)}$ | $F^{(0)}$ | $N$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 10 | 15 | 20 | 25 |
| | | 10 | 1.00 | .96 | .89 | .83 | .78 | .73 | .68 | .64 | .55 | .39 | .28 | .20 |
| | | 15 | 1.00 | .97 | .93 | .89 | .84 | .81 | .77 | .73 | .67 | .53 | .42 | .33 |
| | | 25 | 1.00 | .98 | .96 | .93 | .90 | .88 | .85 | .83 | .78 | .68 | .59 | .51 |
| 0 | 0 | 100 | 1.00 | 1.00 | .99 | .98 | .97 | .97 | .96 | .95 | .94 | .90 | .87 | .84 |
| | | 10 | .75 | .72 | .68 | .63 | .59 | .55 | .51 | .48 | .42 | .30 | .21 | .15 |
| | | 15 | .75 | .73 | .70 | .67 | .64 | .61 | .58 | .55 | .50 | .40 | .32 | .25 |
| | | 25 | .75 | .74 | .72 | .70 | .68 | .66 | .64 | .62 | .59 | .51 | .44 | .38 |
| .25 | 0 | 100 | .75 | .75 | .74 | .74 | .73 | .73 | .72 | .72 | .70 | .68 | .65 | .63 |
| | | 10 | .25 | .25 | .24 | .22 | .21 | .19 | .18 | .17 | .15 | .11 | .08 | .05 |
| | | 15 | .25 | .25 | .24 | .23 | .22 | .21 | .20 | .19 | .18 | .14 | .11 | .09 |
| | | 25 | .25 | .25 | .25 | .24 | .23 | .23 | .22 | .21 | .20 | .17 | .15 | .13 |
| .75 | 0 | 100 | .25 | .25 | .25 | .25 | .25 | .24 | .24 | .24 | .24 | .23 | .22 | .21 |
| | | 10 | 1.00 | .95 | .89 | .83 | .78 | .72 | .68 | .63 | .55 | .39 | .28 | .20 |
| | | 15 | 1.00 | .97 | .92 | .88 | .84 | .80 | .77 | .73 | .64 | .53 | .42 | .33 |
| | | 25 | 1.00 | .98 | .95 | .93 | .90 | .87 | .85 | .83 | .78 | .67 | .58 | .51 |
| 0 | .25 | 100 | 1.00 | 1.00 | .99 | .98 | .97 | .97 | .96 | .95 | .94 | .90 | .87 | .84 |
| | | 10 | 1.00 | .95 | .88 | .82 | .77 | .72 | .67 | .62 | .54 | .39 | .20 | .20 |
| | | 15 | 1.00 | .97 | .92 | .87 | .83 | .80 | .76 | .73 | .66 | .52 | .41 | .33 |
| | | 25 | 1.00 | .98 | .95 | .92 | .90 | .87 | .85 | .82 | .78 | .67 | .58 | .50 |
| 0 | .75 | 100 | 1.00 | 1.00 | .99 | .98 | .97 | .97 | .96 | .95 | .94 | .90 | .87 | .84 |
| | | 10 | .50 | .48 | .45 | .42 | .39 | .36 | .34 | .32 | .28 | .20 | .14 | .10 |
| | | 15 | .50 | .49 | .46 | .44 | .42 | .40 | .38 | .37 | .33 | .26 | .21 | .17 |
| | | 25 | .50 | .49 | .48 | .46 | .45 | .44 | .43 | .41 | .39 | .34 | .29 | .25 |
| .50 | .50 | 100 | .50 | .50 | .49 | .49 | .49 | .48 | .48 | .48 | .47 | .45 | .44 | .42 |

Table 3. *Progress of populations of varying sizes, varying initial inbreeding coefficients, and varying initial coefficients of parentage through twenty-five cycles of current selection, assuming maximum inbreeding*

| Initial Conditions | | | Values of the Panmictic Index ($P = 1 - F$) at the End of the Indicated Recurrent Cycle | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $r^{(0)}$ | $F^{(0)}$ | $N$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 10 | 15 | 20 | 25 |
| 0 | 0 | 10 | 1.00 | .85 | .80 | .74 | .69 | .64 | .60 | .56 | .48 | .34 | .24 | .17 |
| | | 15 | 1.00 | .84 | .80 | .76 | .72 | .69 | .65 | .62 | .56 | .44 | .34 | .27 |
| | | 25 | 1.00 | .83 | .81 | .77 | .75 | .73 | .71 | .68 | .64 | .55 | .47 | .41 |
| | | 100 | 1.00 | .82 | .81 | .79 | .79 | .78 | .77 | .77 | .76 | .73 | .70 | .67 |
| .25 | 0 | 10 | .75 | .65 | .60 | .56 | .52 | .48 | .45 | .42 | .36 | .26 | .18 | .13 |
| | | 15 | .75 | .64 | .60 | .57 | .54 | .52 | .49 | .47 | .42 | .33 | .26 | .20 |
| | | 25 | .75 | .64 | .61 | .58 | .57 | .55 | .53 | .52 | .50 | .42 | .36 | .31 |
| | | 100 | .75 | .63 | .61 | .60 | .59 | .59 | .58 | .58 | .57 | .55 | .52 | .50 |
| .75 | 0 | 10 | .25 | .25 | .21 | .20 | .18 | .17 | .16 | .15 | .13 | .09 | .06 | .04 |
| | | 15 | .25 | .25 | .21 | .20 | .19 | .18 | .17 | .16 | .15 | .12 | .09 | .07 |
| | | 25 | .25 | .25 | .21 | .20 | .19 | .19 | .18 | .18 | .16 | .14 | .12 | .11 |
| | | 100 | .25 | .25 | .20 | .20 | .20 | .20 | .20 | .19 | .19 | .18 | .18 | .17 |
| 0 | .25 | 10 | 1.00 | .84 | .79 | .73 | .68 | .64 | .59 | .55 | .49 | .34 | .24 | .16 |
| | | 15 | 1.00 | .82 | .80 | .75 | .72 | .68 | .65 | .62 | .56 | .44 | .34 | .27 |
| | | 25 | 1.00 | .81 | .80 | .77 | .75 | .73 | .70 | .68 | .64 | .55 | .47 | .41 |
| | | 100 | 1.00 | .80 | .81 | .79 | .79 | .78 | .77 | .77 | .76 | .73 | .70 | .67 |
| 0 | .75 | 10 | 1.00 | .81 | .78 | .72 | .68 | .63 | .59 | .55 | .47 | .33 | .23 | .16 |
| | | 15 | 1.00 | .80 | .79 | .75 | .71 | .68 | .65 | .61 | .58 | .43 | .34 | .27 |
| | | 25 | 1.00 | .78 | .78 | .77 | .75 | .72 | .70 | .68 | .64 | .55 | .47 | .40 |
| | | 100 | 1.00 | .77 | .81 | .79 | .79 | .78 | .77 | .77 | .75 | .73 | .70 | .67 |
| .50 | .50 | 10 | .50 | .42 | .40 | .37 | .34 | .32 | .30 | .28 | .24 | .17 | .12 | .08 |
| | | 15 | .50 | .42 | .40 | .38 | .36 | .34 | .33 | .31 | .28 | .22 | .17 | .13 |
| | | 25 | .50 | .43 | .40 | .39 | .38 | .38 | .36 | .35 | .32 | .28 | .24 | .20 |
| | | 100 | .50 | .41 | .41 | .40 | .39 | .39 | .39 | .38 | .38 | .36 | .35 | .34 |

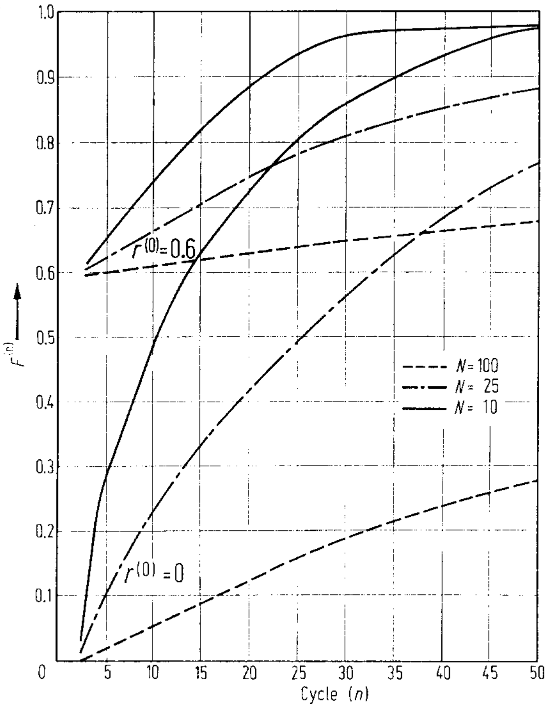

Fig. 4. Comparisons of the effects of varying population sizes ($N$) and varying initial coefficients of parentage ($r^{(0)}$) on the progress of a population under recurrent selection. $F^{(0)} = 0$ is assumed

representation of all lines (i.e. $b_i = 2$ for all $i$) and allows the parameters $N$, $r^{(0)}$, and $F^{(0)}$ to vary and then traces the progress of the degree of panmixia through twenty-five cycles (50 generations) of *RS*. Maintenance of a constant population size (by selection following the selfing phase) is assumed in the method by which Table 2 is generated. The dramatic effects of population size ($N$) and initial coefficient of parentage ($r^{(0)}$) are evident from Table 2 and of these two, $r^{(0)}$ is the more important; the size of the inbreeding coefficient in the source population ($F^{(0)}$) has little effect on the panmictic index at the ith recurrent cycle ($P^{(i)}$). These effects are seen in greater detail in the graphs of Figure 4, in which $F^{(0)}$ is assumed to be zero. Under the restriction that $b_i = 2$ for all $i$, (6.1) and (6.2) reduce to

$$F^{(n+2)} = \frac{2}{N-1}\left(\frac{3}{4}r^{(n)} + \frac{1}{16}F^{(n)} + \frac{3}{16}\right) + \frac{N-3}{N-1}r^{(n)},$$

and

$$r^{(n+1)} = \frac{N}{N+1}r^{(n)} + \frac{F^{(n)}+3}{4(N+1)},$$

which were used to obtain the results of Table 2 and Figure 4.

Maximum inbreeding results when, at each cycle, one line is represented $N-1$ times, $N-3$ lines are each represented only once, and two lines are represented twice. Table 3 shows the progress of the panmictic index under the same assumptions as in

Table 2 except that a minimum inbreeding selection pattern is replaced by one of maximum inbreeding. Tables 2 and 3 could be used to approximate inbreeding coefficients in *RS* systems for which their basic assumptions were applicable. The last section of the tables, for which $r^{(0)} = F^{(0)} = .5$, could be used to approximate the progress of $1 - H$ in a breeding population.

## 9. Conclusions

General recurrence formulae for the calculation of the inbreeding coefficient and the coefficient of parentage in the $(n + 2)$ nd cycle of a *RS* mating pattern are developed for the one-locus, two-allele case. These formulae involve no restrictions as to the initial values $r^{(0)}$ or $F^{(0)}$ or as to the number of times each line is represented in the progeny at any stage of reproduction; and they assume only a basic *RS* mating scheme.

For the special case in which

(i) $n = 0$

(ii) $r^{(1)} = F^{(1)} = 0$;

(iii) the usually-employed initial generation of selfing is ignored; so that the mating cycle involves only select-intercross-self-select-intercross;

(iv) all selected lines are represented equally in the progeny of the final intercrossed generation, the formula for $F^{(2)}$ reduces to that given by Sprague, *et al.* (1952).

From the derivation, it is evident that substituting random mating for intercrossing would result in decreases in panmixia at each successive cycle.

It is pointed out in Sections 1 and 7 that when selection is present, an overall index of homozygosity ($H$) or an overall index of heterozygosity ($1 - H$) would be more meaningful than $F$ to use as an indication of possible future selection progress and a method of calculation of this index is outlined, as well as that of calculating an index $t$ measuring the relationship between individuals in terms of the probability of genes identical by descent or in state at a given locus for two individuals.

From computer-simulated populations the progress of populations assuming both minimum and maximum inbreeding is investigated with varying initial coefficients of relationship, varying initial degrees of homozygosity by descent, and varying initial sample sizes; and it is observed that the contribution of $r^{(0)}$ to the degree of panmixia after several breeding cycles far outweighs the contribution of $F^{(0)}$. If only a few cycles of *RS* are anticipated (where one cycle represents from two to three years in a breeding program), the initial sample size actually makes very little difference in the degree of panmixia, expecially if $r^{(0)}$ is close to zero. The tables presented could be used to estimate minimum and maximum inbreeding coefficients in practice, given populations which conform to the given assumptions.

## Zusammenfassung

Für Selektionsprogramme, die durch aufeinanderfolgende Selbstungs-Selektions-Kreuzungs-Zyklen (wie z. B. rekurrente Selektion oder reziproke rekurrente Selektion) charakterisiert sind, werden allgemeine Rekurrenzformeln zur Berechnung von Inzucht- und Homozygotie-Koeffizienten in jedem Zyklus entwickelt.

Die Formel für den Inzuchtkoeffizienten stellt eine Verallgemeinerung eines von Sprague et al. (1952) erhaltenen Ergebnisses dar.

Es wird gezeigt, daß der "coefficient of parentage" der Ausgangspopulation ebenso wie die Populationsgröße einen nachhaltigen Einfluß auf den Inzuchtkoeffizienten der folgenden Zyklen haben. Die Beziehung beider Typen von Koeffizienten und ihre Bedeutung für die praktische Arbeit werden diskutiert.

## References

1. Comstock, R. E., Robinson, H. F., Harvey, P. H.: A Breeding Procedure Designed to Make Maximum Use of Both General and Specific Combining Ability. Agronomy Journal **41**, 360–367 (1949). — 2. Falconer, D. S.: Introduction to Quantitative Genetics. New York: The Ronald Press Co. 1960. — 3. Hull, F. H.: Recurrent Selection for Specific Combining Ability in Corn. J. Amer. Soc. Agron. **37** 134–145 (1945). — 4. Jenkins, M. T.: The Effect of Inbreeding and of Selection Within Inbred Lines of Maize upon the Hybrids Made After Successive Generations of Selfing. Iowa State Coll. J. Sci. **9**, 429–450 (1935). — 5. Jenkins, M. T.: The Segregation of Genes Affecting Yield of Grain in Maize. J. Amer. Soc. Agron. **32**, 55–63 (1940). — 6. Kempthorne, O.: An Introduction to Genetic Statistics. New York: John Wiley & Sons 1957. — 7. Kempthorne, O.: Biometrical Relations Beween Relatives and Selection Theory. The Proceedings of the 1958 International Symposium on Biometrical Genetics. London: Pergamon Press 1960. — 8. Malécot, G.: Les Mathématiques de l'Hérédité. Paris: Masson et Cie. 1948. — 9. Robertson, A.: Inbreeding in Artificial Selection Programmes. Genetical Research **2**, 189–194 (1961). — 10. Robinson, H. F., Comstock, R. E., Harvey, P. H.: Genotypic Variances in Open Pollinated Varieties of Corn. Genetics **40** (1), 45–60 (1955). — 11. Shikata, M.: Transformation of Generalized Inbreeding Coefficient in Components of Pedigree. J. Theor. Biol. **10**, 11–14 (1966). — 12. 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. Agro. J. **44**, 329–331 (1952). — 13. Wright, S.: Systems of Mating I–V. Genetics **6**, 111–178 (1921).

Received June 29, 1970
Communicated by W. Seyffert

Dr. R. L. Cain
Morehead State University
Morehead, Ky. 40351
Dr. K. Hinkelmann
Virginia Polytechnic Institute Blacksburg,
Virginia 24061 (USA)