

THE PREDICTION OF SELECTION RESPONSE IN A SELF-FERTILIZING SPECIES

II.* FAMILY SELECTION†

By D. G. PEDERSON‡

[Manuscript received January 9, 1969]

Summary

The response to between- and within-family selection is considered for self-fertilizing populations. For a general genetic model involving m alleles at each of k loci the expected responses are written as the sum of (1) the natural change in mean due to decreasing heterozygosity, and (2) a component which is a function of the genetic and environmental variances within an unselected population. A single-locus model is then considered and the expected responses are written in terms of parameters of the base population. General expressions are given for the ultimate response following relaxation of selection and these are interpreted in terms of the single-locus model. In the final section the efficiencies of the various methods are investigated. In general, within-family selection is relatively inefficient while between-family selection is usually to be preferred over the selection of individuals or the more complex method of combined selection.

I. INTRODUCTION

In Part I of this series (Pederson 1969) prediction equations were developed for the response to individual selection in a self-fertilizing species. Competition effects were assumed to be absent, as will be done in the present paper, and individuals were considered to be selected on the basis of phenotypic value with no regard being given to familial or ancestral data.

In practice a number of different selection methods are used for autogamous plants, among them being pedigree selection, the bulk-population method, and various systems of pure-line breeding (e.g. see Allard 1960). Each of these makes use of family data and it is the purpose of this paper to derive the relevant prediction equations. Both between- and within-family selection are considered, and brief attention is given to a method of combined selection.

Equations are initially developed for a general genetic model involving k unlinked, interacting loci of small effect, and then these equations are interpreted

* Part I, *Aust. J. biol. Sci.*, 1969, **22**, 117–29.

† Published with the approval of the Director of Research as Paper No. 2795 of the Journal Series of the North Carolina Agricultural Experiment Station, Raleigh, North Carolina. This investigation was supported by Public Health Service Research Grant GM 11546 from the Division of General Medical Sciences.

‡ Department of Genetics, North Carolina State University, Raleigh; present address: Department of Agricultural Botany, University of Sydney, N.S.W. 2006.

in terms of a single-locus model. The consequences of relaxation are considered and the relative efficiencies of the various selection methods are discussed.

II. GENERALIZED SELECTION RESPONSE

The notation of this paper will follow that of Pederson (1969). For the most general case attention will be given to the diploid genotypes generated by m alleles at each of k loci. These k loci, each of small individual effect, make up one of the many subsets of loci controlling the character undergoing selection. The effects of subsets are assumed to be additive.

Within the subset, the genotype of an arbitrary individual with alleles $A_{i_s}^s$ and $A_{j_s}^s$ at the s th locus may be written as

$$\prod_{s=1}^k A_{i_s}^s A_{j_s}^s,$$

with corresponding genotypic value $Y_{\Pi_{i_s j_s}}$. In this and other expressions, unless it is otherwise specified, the set of values $s = 1, \dots, k$ are to be assumed for the product operator. The genotypic model is completely general with respect to between- and within-locus interaction but does not allow for linkage between loci.

As in the earlier paper, the expression $S^t Y_{\Pi_{i_s j_s}}$ will be used to represent the array of genotypic values obtained from genotype $\Pi A_{i_s}^s A_{j_s}^s$ by t generations of self-fertilization.

Let the frequency of genotype $\Pi A_{i_s}^s A_{j_s}^s$ in the base population be ${}_0 f_{\Pi_{i_s j_s}}$, where the zero subscript indicates generation number. Then the symbol ${}_0 f_{\cdot i_s j_s \cdot}$, in which summation is over all of the i 's and j 's not specified in the subscript, represents the marginal genotypic frequencies at the s th locus. After t generations of self-fertilization the marginal frequencies become

$${}_t f_{\cdot i_s j_s \cdot}^* = \left(\frac{1}{2}\right)^t {}_0 f_{\cdot i_s j_s \cdot} + \beta_{i_s j_s} [1 - \left(\frac{1}{2}\right)^t] {}_0 f_{\cdot i_s \cdot}, \quad (1)$$

where

$$\beta_{i_s j_s} = \begin{cases} 0 & \text{if } i_s \neq j_s \\ 1 & \text{if } i_s = j_s \end{cases},$$

and ${}_0 f_{\cdot i_s \cdot}$ is the marginal frequency of allele $A_{i_s}^s$. The expression $[1 - (\frac{1}{2})^t]$ may alternatively be written as F_t , the inbreeding coefficient for the t th generation of selfing. The frequency in equation (1) is marked with an asterisk to indicate that it is the result of self-fertilization without selection. The k loci are unlinked, so that

$${}_t f_{\Pi_{i_s j_s}}^* = \prod {}_t f_{\cdot i_s j_s \cdot}^*. \quad (2)$$

The initial population mean, with respect to the k loci, is given by

$${}_0 \mu = \sum_{\text{all } i, j} {}_0 f_{\Pi_{i_s j_s}} Y_{\Pi_{i_s j_s}}.$$

After t generations of self-fertilization this becomes

$$\begin{aligned} {}_t \mu^* &= \sum {}_t f_{\Pi_{i_s j_s}}^* Y_{\Pi_{i_s j_s}} \\ &= \sum {}_0 f_{\Pi_{i_s j_s}} (S^t Y_{\Pi_{i_s j_s}}). \end{aligned} \quad (3)$$

(a) *Response from (n-1) Cycles of Family Selection*

Suppose that the initial population is allowed to self-fertilize for one cycle without selection, giving a set of families with overall mean ${}_1\mu^*$. The selection scheme will be considered in which each family with mean exceeding a certain arbitrary value is selected and its members allowed to contribute to the next generation. A new set of self-families is therefore generated, with overall mean ${}_2\mu$, and the process of selection and self-fertilization is then repeated. Families are selected or rejected solely on the basis of their mean performances with no regard being given to ancestral data.

The phenotypic variance of family means after t generations of selection and self-fertilization will be written as ${}_t\sigma_b^2$. This variance is composed of the genotypic variance generated by all subsets of loci plus the variance due to error effects. It is convenient at this point to define

$${}_tZ_{\Pi_{i_s j_s}} = (SY_{\Pi_{i_s j_s}} - {}_t\mu) / {}_t\sigma_b, \quad (4)$$

and

$${}_tZ_{\Pi_{i_s j_s}}^* = (SY_{\Pi_{i_s j_s}} - {}_t\mu^*) / {}_t\sigma_b. \quad (5)$$

The k loci are individually of small effect relative to the phenotypic standard deviation and it is therefore reasonable to assume that the squares and products of Z values are of negligible magnitude.

The consequences of truncation selection will now be considered in detail. Initially genotype $\Pi A_{i_s}^s A_{j_s}^s$ is present with frequency ${}_0f_{\Pi_{i_s j_s}}$, and self-fertilization of this genotype gives a self-family with mean $(SY_{\Pi_{i_s j_s}})$. Let the standardized selection differential, for selection among the self-families, have the value $\bar{i}$. Then it follows from Kimura (1958) that the family with mean $(SY_{\Pi_{i_s j_s}})$ occurs among the selected families with frequency

$${}_0g_{\Pi_{i_s j_s}} = {}_0f_{\Pi_{i_s j_s}}(1 + \bar{i} {}_1Z_{\Pi_{i_s j_s}}^*).$$

The individuals in the selected families are allowed to self to produce the next generation, which therefore has a mean of

$$\begin{aligned} {}_2\mu &= \sum_{\text{all } i, j} {}_0g_{\Pi_{i_s j_s}} (S^2 Y_{\Pi_{i_s j_s}}) \\ &= {}_2\mu^* + \bar{i} \sum {}_0f_{\Pi_{i_s j_s}} {}_1Z_{\Pi_{i_s j_s}}^* (S^2 Y_{\Pi_{i_s j_s}}). \end{aligned}$$

Within the first set of selected families, genotype $\Pi A_{i_s}^s A_{j_s}^s$ is present with an overall frequency of ${}_1f_{\Pi_{i_s j_s}}$. From (1), this is given by

$${}_1f_{\Pi_{i_s j_s}} = \Pi (\frac{1}{2} {}_0g_{\cdot i_s j_s} + \frac{1}{2} \beta_{i_s j_s} {}_0g_{\cdot i_s}).$$

Among the families upon which the second cycle of selection is to be carried out, the family with mean $(SY_{\Pi_{i_s j_s}})$ is therefore present with frequency ${}_1f_{\Pi_{i_s j_s}}$. Selection results in a new frequency

$${}_1g_{\Pi_{i_s j_s}} = {}_1f_{\Pi_{i_s j_s}}(1 + \bar{i} {}_2Z_{\Pi_{i_s j_s}}).$$

Self-fertilization of these families gives a population with mean

$$\begin{aligned} 3\mu &= \sum_{\text{all } i,j} 1f_{\Pi i_s j_s} (1 + i 2Z_{\Pi i_s j_s}) (S^2 Y_{\Pi i_s j_s}) \\ &= \sum 1f_{\Pi i_s j_s} (S^2 Y_{\Pi i_s j_s}) + i \sum 1f_{\Pi i_s j_s} 2Z_{\Pi i_s j_s} (S^2 Y_{\Pi i_s j_s}) \\ &= \sum 0g_{\Pi i_s j_s} (S^3 Y_{\Pi i_s j_s}) + i \sum 1f_{\Pi i_s j_s} 2Z_{\Pi i_s j_s} (S^2 Y_{\Pi i_s j_s}) \\ &= 3\mu^* + i \sum 0f_{\Pi i_s j_s} 1Z_{\Pi i_s j_s}^* (S^3 Y_{\Pi i_s j_s}) + i \sum 1f_{\Pi i_s j_s} 2Z_{\Pi i_s j_s} (S^2 Y_{\Pi i_s j_s}). \end{aligned}$$

In general it is found that

$$n\mu = n\mu^* + i \sum_{t=1}^{n-1} \left[\sum_{\text{all } i,j} t-1f_{\Pi i_s j_s} tZ_{\Pi i_s j_s} (S^{n-t+1} Y_{\Pi i_s j_s}) \right]. \quad (6)$$

Now from (4) and (5)

$$\begin{aligned} tZ_{\Pi i_s j_s} &= tZ_{\Pi i_s j_s}^* - (t\mu - t\mu^*) / t\sigma_b \\ &= tZ_{\Pi i_s j_s}^* - (i/t\sigma_b) \sum_{u=1}^{t-1} \left[\sum_{\text{all } i,j} u-1f_{\Pi i_s j_s} uZ_{\Pi i_s j_s} (S^{t-u+1} Y_{\Pi i_s j_s}) \right]. \end{aligned}$$

But

$$(1/t\sigma_b) \sum_{\text{all } i,j} u-1f_{\Pi i_s j_s} uZ_{\Pi i_s j_s} (S^{t-u+1} Y_{\Pi i_s j_s})$$

may be rewritten as

$$\sum u-1f_{\Pi i_s j_s} uZ_{\Pi i_s j_s} (S^{t-u+1} Y_{\Pi i_s j_s} - t\mu) / t\sigma_b,$$

and this expression is of negligible magnitude since it effectively involves the product of two Z values. Equation (6) may therefore be rewritten as

$$n\mu = n\mu^* + i \sum_{t=1}^{n-1} \left[\sum_{\text{all } i,j} t-1f_{\Pi i_s j_s} tZ_{\Pi i_s j_s}^* (S^{n-t+1} Y_{\Pi i_s j_s}) \right]. \quad (7)$$

Further, it may be shown that after $(n-1)$ cycles of selection

$$\begin{aligned} n f_{\Pi i_s j_s} &= n f_{\Pi i_s j_s}^* \\ &+ i \sum_{t=1}^{n-1} \left\{ \left[\left(\frac{1}{2} \right)^{n-t} \right]^k \left[t-1f_{\Pi i_s j_s} tZ_{\Pi i_s j_s} \right. \right. \\ &+ (2^{n-t} - 1) \sum_{u=1}^k (\beta_{i_u j_u} \sum_{j_u} t-1f_{\Pi i_s j_s} tZ_{\Pi i_s j_s}) \\ &+ \frac{1}{2!} (2^{n-t} - 1)^2 \sum_{u \neq u'} (\beta_{i_u j_u} \beta_{i_{u'} j_{u'}} \sum_{j_u, j_{u'}} t-1f_{\Pi i_s j_s} tZ_{\Pi i_s j_s}) \\ &\left. \left. + \dots \right] \right\}. \quad (8) \end{aligned}$$

Equation (7) therefore becomes

$$n\mu = n\mu^* + i \sum_{t=1}^{n-1} \left[\sum_{\text{all } i,j} t-1f_{\Pi i_s j_s}^* tZ_{\Pi i_s j_s}^* (S^{n-t+1} Y_{\Pi i_s j_s}) \right]. \quad (9)$$

The right-hand side of equation (9) involves only the parameters of an unselected population, an important characteristic of which is the covariance between a family mean in the t th generation and the mean of its offspring array in the n th generation of self-fertilization. An expression for this covariance may be found in the following manner. After t cycles of selfing there is obtained a set of families such that the family with mean $(S Y_{\Pi i_s j_s})$ occurs with frequency $t-1f_{\Pi i_s j_s}^*$. The mean of this family

will change to $(S^{n-t+1}Y_{\Pi_{isj_s}})$ after a further $(n-t)$ generations of selfing without selection so that the required covariance may be found as

$$\begin{aligned} C_b(t, n) &= \sum_{\text{all } i, j} {}_{t-1}f_{\Pi_{isj_s}}^*(SY_{\Pi_{isj_s}})(S^{n-t+1}Y_{\Pi_{isj_s}}) - {}_t\mu^* {}_n\mu^* \\ &= \sum {}_{t-1}f_{\Pi_{isj_s}}^*(SY_{\Pi_{isj_s}} - {}_t\mu^*)(S^{n-t+1}Y_{\Pi_{isj_s}}). \end{aligned} \quad (10)$$

From (9) and (10) it follows that

$${}_n\mu = {}_n\mu^* + \bar{i} \sum_{t=1}^{n-1} [C_b(t, n)/{}_t\sigma_b]. \quad (11)$$

Although the parameters of this equation explicitly refer to the A subset of loci, the equation may equally well predict the population phenotypic response since the effects of subsets are additive.

Equation (6) could have been written in a form similar to (11) but the covariance term would have referred to the population undergoing selection. The foregoing theory therefore demonstrates that the covariances in an unselected and a selected population are effectively identical under the present assumptions. Similarly the genotypic, and hence phenotypic, variances generated by subsets of loci are the same for the unselected and selected populations and the symbol ${}_t\sigma_b^2$ will therefore be used throughout.

(b) *Response from (n-1) Cycles of Within-family Selection*

In the earlier paper of this series the expected response to individual selection under self-fertilization was given as

$${}_n\mu = {}_n\mu^* + \bar{i} \sum_{t=0}^{n-1} [C(t, n)/{}_t\sigma],$$

where $C(t, n)$ is the expected covariance between an individual of the t th generation and the mean of its offspring array in the n th generation of self-fertilization without selection. The variance ${}_t\sigma^2$ is the phenotypic variance of individuals in the t th generation.

Now consider the selection scheme in which the best individuals are chosen from within a number of families and allowed to produce the next generation of self-families. The intensity of selection is assumed to be the same within each family, such that the common standardized selection differential has a value of $\bar{i}$. Then by analogy with the results for individual and between-family selection it will be stated without proof that the expected response from $(n-1)$ cycles of the within-family selection is given by

$${}_n\mu = {}_n\mu^* + \bar{i} \sum_{t=1}^{n-1} [C_w(t, n)/{}_t\sigma_w], \quad (12)$$

where ${}_t\sigma_w^2$ is the mean phenotypic variance within families of the t th generation and $C_w(t, n)$ is the expected within-family covariance between an individual of the t th generation and the mean of its unselected offspring array in the n th generation ($n \geq t$). The algebraic proof of this equation is similar to that for individual and between-family selection but it is lengthy and thought to be unnecessary for the present purposes.

Use will be made of the relationships

$${}_t\sigma^2 = {}_t\sigma_b^2 + {}_t\sigma_w^2, \quad (13)$$

$$C(t, n) = C_b(t, n) + C_w(t, n). \quad (14)$$

III. THE RESPONSE IN TERMS OF GENETIC PARAMETERS

In practice, equations (11) and (12) would be of little predictive value since separate covariance and variance estimates are required for each cycle of prediction. A single-locus model will therefore be considered and the covariances and variances will be written as functions of the genetic and environmental variances in the base population. The expected response to continuous selection may then be readily determined in any population for which the initial variances are known. It will be assumed that loci do not interact, so that their individual effects may be summed.

The single-locus model has been given in the previous paper and will only be repeated in outline. It will be assumed that genotypes are ordered and that ${}_0f_{ij} = {}_0f_{ji}$ for all i and j .

The subscripts of Section II are simplified by putting $i_1 = i$, $j_1 = j$ and omitting all others. The genotypic value of A_iA_j is written as the sum of additive and dominance effects, viz.

$$Y_{ij} = {}_0\mu + \alpha_i + \alpha_j + \delta_{ij} \quad (i, j = 1, \dots, m).$$

The genetic effects are defined by the set of least-squares equations

$${}_0f_i \alpha_i + \sum_j {}_0f_{ij} \alpha_j = \sum_j {}_0f_{ij} (Y_{ij} - {}_0\mu) \quad (i = 1, \dots, m).$$

For the present purposes it is necessary that all ${}_0f_{ii}$ be non-zero at segregating loci so that dominance effects are defined.

The additive and dominance variances generated by the locus in the base population are, respectively,

$$\sigma_A^2 = 2 \sum_i {}_0f_i \alpha_i^2 + 2 \sum_{ij} {}_0f_{ij} \alpha_i \alpha_j,$$

and

$$\sigma_D^2 = \sum_{ij} {}_0f_{ij} \delta_{ij}^2.$$

The "dot" notation for gene frequencies indicates summation over the unspecified subscript. A third parameter will be defined, namely

$$\sigma_{AD} = \frac{1}{2} \sum_{ij} {}_0f_{ij} (\alpha_i + \alpha_j) \delta_{ij}.$$

For a single locus, equation (10) reduces to

$$\begin{aligned} C_b(t, n) &= \sum_{ij} {}_{t-1}f_{ij}^* (S Y_{ij} - {}_t\mu^*) (S^{n-t+1} Y_{ij}) \\ &= \left(\frac{1}{2}\right)^{n-t+1} \sum_{ij} {}_{t-1}f_{ij}^* (S Y_{ij} - {}_t\mu^*) [Y_{ij} + (2^{n-t+1} - 1) Y_{ii}]. \end{aligned}$$

Both the frequency ${}_{t-1}f_{ij}^*$ and the mean ${}_t\mu^*$ may be written in terms of the base population frequencies and effects, and there is obtained

$$\begin{aligned} C_b(t, n) &= \left(\frac{1}{2}\right)^n \sum_{ij} {}_0f_{ij} (S Y_{ij} - {}_0\mu) [Y_{ij} + (2^{n-t+1} - 1) Y_{ii}] \\ &\quad - \left(\frac{1}{2}\right)^n [1 - \left(\frac{1}{2}\right)^t] \sum_i {}_0f_i \delta_{ii} \sum_{ij} {}_0f_{ij} [Y_{ij} + (2^{n-t+1} - 1) Y_{ii}] \\ &\quad + [1 - \left(\frac{1}{2}\right)^{t-1}] \sum_i {}_0f_i (Y_{ii} - {}_0\mu) Y_{ii} \\ &\quad - [1 - \left(\frac{1}{2}\right)^t] [1 - \left(\frac{1}{2}\right)^{t-1}] \sum_i {}_0f_i \delta_{ii} \sum_i {}_0f_i Y_{ii}. \end{aligned}$$

Substitution of gene effects for genotypic values gives

$$\begin{aligned}
 C_b(t, n) = & 4[1 - (\tfrac{1}{2})^t] \sum_i 0f_i \alpha_i^2 + 4(\tfrac{1}{2})^t \sum_{ij} 0f_{ij} \alpha_i \alpha_j \\
 & + [3(\tfrac{1}{2})^t - (\tfrac{1}{2})^n] \sum_{ij} 0f_{ij} (\alpha_i + \alpha_j) \delta_{ii} + 4[1 - 2(\tfrac{1}{2})^t] \sum_i 0f_i \alpha_i \delta_{ii} \\
 & + [(\tfrac{1}{2})^{t+1} - (\tfrac{1}{2})^{n+2}] \sum_{ij} 0f_{ij} \delta_{ii} (\delta_{ii} + \delta_{jj}) \\
 & + [1 - 2(\tfrac{1}{2})^t] \sum_i 0f_i \delta_{ii}^2 - [1 - (\tfrac{1}{2})^t][1 - (\tfrac{1}{2})^n] (\sum_i 0f_i \delta_{ii})^2 \\
 & + (\tfrac{1}{2})^{n+1} \sum_{ij} 0f_{ij} \delta_{ij}^2.
 \end{aligned} \tag{15}$$

The corresponding expression for $C(t, n)$, which ignores family structure, has been given in the first paper of this series, and from (14) there is obtained

$$\begin{aligned}
 C_w(t, n) = & 2(\tfrac{1}{2})^t \sum_i 0f_i \alpha_i^2 - 2(\tfrac{1}{2})^t \sum_{ij} 0f_{ij} \alpha_i \alpha_j \\
 & + 2(\tfrac{1}{2})^t \sum_{ij} 0f_{ij} (\alpha_i - \alpha_j) \delta_{ii} \\
 & + (\tfrac{1}{2})^t \sum_i 0f_i \delta_{ii}^2 - [(\tfrac{1}{2})^{t+1} - (\tfrac{1}{2})^{n+2}] \sum_{ij} 0f_{ij} \delta_{ii} (\delta_{ii} + \delta_{jj}) \\
 & + (\tfrac{1}{2})^{n+1} \sum_{ij} 0f_{ij} \delta_{ij}^2.
 \end{aligned} \tag{16}$$

The equations for predicting $n\mu$ are functions of both variances and covariances. However, explicit expressions will not be given for the variances since the genetic variance generated by a single locus in the t th generation may always be found by substituting $n = t$ in the appropriate covariance, either $C_b(t, n)$ or $C_w(t, n)$. The assumption is then made that the total phenotypic variance is the sum of an error variance and the genetic variance for the many loci involved.

As presented above, each of the covariance equations involves a large number of unknown terms. However, they are of a much simpler form for populations which have arisen by a random-mating procedure and for populations which have only two alleles present at each segregating locus. These two classes will now be considered in turn.

(a) *Randomly Mated Base Population*

A feature of the single-locus model is that

$$\sum_i 0f_i \alpha_i = 0.$$

If genotypic frequencies may be expressed as the products of gene frequencies, as in a random-mating population, the terms $\sum_{ij} 0f_{ij} \alpha_i \alpha_j$ and $\sum_{ij} 0f_{ij} \alpha_j \delta_{ii}$ are therefore zero. Equations (15) and (16) correspondingly reduce to

$$\begin{aligned}
 C_b(t, n) = & 4[1 - (\tfrac{1}{2})^t] \sum_i 0f_i \alpha_i^2 + [4 - 5(\tfrac{1}{2})^t - (\tfrac{1}{2})^n] \sum_i 0f_i \alpha_i \delta_{ii} \\
 & + [1 - 3(\tfrac{1}{2})^{t+1} - (\tfrac{1}{2})^{n+2}] \sum_i 0f_i \delta_{ii}^2 \\
 & - [1 - 3(\tfrac{1}{2})^{t+1} - 3(\tfrac{1}{2})^{n+2} + (\tfrac{1}{2})^{n+t}] (\sum_i 0f_i \delta_{ii})^2 \\
 & + (\tfrac{1}{2})^{n+1} \sum_{ij} 0f_{ij} \delta_{ij}^2,
 \end{aligned} \tag{17}$$

and

$$\begin{aligned}
 C_w(t, n) &= 2(\tfrac{1}{2})^t \sum_i {}_0f_{i.} \alpha_i^2 + 2(\tfrac{1}{2})^t \sum_i {}_0f_{i.} \alpha_i \delta_{ii} \\
 &\quad + [(\tfrac{1}{2})^{t+1} + (\tfrac{1}{2})^{n+2}] \sum_i {}_0f_{i.} \delta_{ii}^2 \\
 &\quad - [(\tfrac{1}{2})^{t+1} - (\tfrac{1}{2})^{n+2}] (\sum_i {}_0f_{i.} \delta_{ii})^2 + (\tfrac{1}{2})^{n+1} \sum_{ij} {}_0f_{ij} \delta_{ij}^2.
 \end{aligned} \tag{18}$$

To predict a selection response it would be necessary to estimate the unknown functions in these equations. The problem is made much simpler if all dominance effects are assumed to be zero. The covariances then become

$$\begin{aligned}
 C_b(t, n) &= 4[1 - (\tfrac{1}{2})^t] \sum_i {}_0f_{i.} \alpha_i^2 \\
 &= 2[1 - (\tfrac{1}{2})^t] \sigma_A^2, \text{ and} \\
 C_w(t, n) &= 2(\tfrac{1}{2})^t \sum_i {}_0f_{i.} \alpha_i^2 \\
 &= (\tfrac{1}{2})^t \sigma_A^2.
 \end{aligned}$$

(b) *Two Alleles per Locus in the Base Population*

Two new variables will be defined, namely

$$S = {}_0f_{22} + {}_0f_{11},$$

and

$$D = {}_0f_{22} - {}_0f_{11}.$$

As a consequence of the restrictions

$$\sum_i {}_0f_{i.} \alpha_i = 0$$

and

$$\sum_i {}_0f_{ij} \delta_{ij} = \sum_j {}_0f_{ij} \delta_{ij} = 0,$$

the gene effect α_2 can be written in terms of α_1 and effects δ_{12} and δ_{22} can be written in terms of δ_{11} . Equations (15) and (16) become

$$\begin{aligned}
 C_b(t, n) &= (\alpha_1^2 / {}_0f_{2.}^2) [1 - 4(\tfrac{1}{2})^t {}_0f_{12} - D^2] \\
 &\quad + (D\alpha_1 \delta_{11} / 4{}_0f_{2.} {}_0f_{22}) \{ [4 - 2(\tfrac{1}{2})^t] (1 - D^2) - 12(\tfrac{1}{2})^t {}_0f_{12} - 2(\tfrac{1}{2})^n (S - D^2) \} \\
 &\quad + (D^2 \delta_{11}^2 / 4{}_0f_{22}^2) \{ [1 - (\tfrac{1}{2})^t] (1 - D^2) - 2(\tfrac{1}{2})^t {}_0f_{12} \} \\
 &\quad + [(\tfrac{1}{2})^{n+2} \delta_{11}^2 (S - D^2) / 4{}_0f_{12} {}_0f_{22}^2] \{ [1 - 4(\tfrac{1}{2})^t {}_0f_{12}] (S - D^2) - 4D^2 {}_0f_{12} \},
 \end{aligned}$$

and

$$\begin{aligned}
 C_w(t, n) &= (\alpha_1^2 / {}_0f_{2.}^2) 2(\tfrac{1}{2})^t {}_0f_{12} + (D\alpha_1 \delta_{11} / {}_0f_{2.} {}_0f_{22}) 2(\tfrac{1}{2})^t {}_0f_{12} \\
 &\quad + (D^2 \delta_{11}^2 / {}_0f_{22}^2) (\tfrac{1}{2})^{t+1} {}_0f_{12} + (\tfrac{1}{2})^{n+2} \delta_{11}^2 (S - D^2)^2 / 4{}_0f_{12} {}_0f_{22}^2.
 \end{aligned}$$

But in addition

$$\sigma_A^2 = (\alpha_1^2 / {}_0f_{2.}^2) (S - D^2),$$

$$\sigma_D^2 = ({}_0f_{11} \delta_{11}^2 / 2{}_0f_{12} {}_0f_{22}) (S - D^2),$$

and

$$\sigma_{AD} = (D\alpha_1 \delta_{11} / 4{}_0f_{2.} {}_0f_{22}) (S - D^2).$$

It follows that

$$\begin{aligned}
 C_b(t, n) = & \left[\frac{1 - 4(\frac{1}{2})^t {}_0f_{12} - D^2}{S - D^2} \right] \sigma_A^2 \\
 & + 2 \left\{ \frac{[2 - (\frac{1}{2})^t](1 - D^2) - 6(\frac{1}{2})^t {}_0f_{12} - (\frac{1}{2})^n (S - D^2)}{S - D^2} \right\} \sigma_{AD} \\
 & + 2D^2 {}_0f_{12} \left\{ \frac{[1 - (\frac{1}{2})^t](1 - D^2) - 2(\frac{1}{2})^t {}_0f_{12}}{(S - D^2)(S^2 - D^2)} \right\} \sigma_D^2 \\
 & + (\frac{1}{2})^{n+1} \left\{ \frac{[1 - 4(\frac{1}{2})^t {}_0f_{12}](S - D^2) - 4D^2 {}_0f_{12}}{S^2 - D^2} \right\} \sigma_D^2,
 \end{aligned} \tag{19}$$

and

$$\begin{aligned}
 C_w(t, n) = & \left[\frac{2(\frac{1}{2})^t {}_0f_{12}}{S - D^2} \right] \sigma_A^2 + \left[\frac{8(\frac{1}{2})^t {}_0f_{12}}{S - D^2} \right] \sigma_{AD} \\
 & + \left[\frac{4(\frac{1}{2})^t D^2 {}_0f_{12}}{(S - D^2)(S^2 - D^2)} \right] \sigma_D^2 + \left[\frac{(\frac{1}{2})^{n+1}(S - D^2)}{S^2 - D^2} \right] \sigma_D^2.
 \end{aligned} \tag{20}$$

For the special case of a base population with gene frequencies of 0.5 σ_{AD} is zero by definition so that

$$C_b(t, n) = \left[\frac{1 - 4(\frac{1}{2})^t {}_0f_{12}}{S} \right] \left[\sigma_A^2 + (\frac{1}{2})^{n+1} \sigma_D^2 \right], \tag{21}$$

and

$$C_w(t, n) = \left[\frac{2(\frac{1}{2})^t {}_0f_{12}}{S} \right] \sigma_A^2 + \left[\frac{(\frac{1}{2})^{n+1}}{S} \right] \sigma_D^2. \tag{22}$$

IV. RELAXATION OF SELECTION

Homozygosity is the natural state for a completely self-fertilizing organism. When a population is subjected to $(n-1)$ cycles of between-family selection and then allowed to revert to homozygosity, its expected mean is given by

$${}_n\mu_H = \sum_{\text{all } i, j} {}_0f_{\Pi i_s j_s} Y_{\Pi i_s i_s} + i \sum_{t=1}^{n-1} \left[\frac{C_b(t, \infty)}{t\sigma_b} \right]. \tag{23}$$

The subscript H indicates that the mean refers to a population in the homozygous condition. The proof of this equation is similar to that for the case of individual selection (Pederson 1969). When the $(n-1)$ cycles of selection are within families

$${}_n\mu_H = \sum_{\text{all } i, j} {}_0f_{\Pi i_s j_s} Y_{\Pi i_s i_s} + i \sum_{t=1}^{n-1} \left[\frac{C_w(t, \infty)}{t\sigma_w} \right]. \tag{24}$$

The parameters of interest are therefore the covariances $C_b(t, \infty)$ and $C_w(t, \infty)$, and the general expressions for these may be found by taking the limit as $n \rightarrow \infty$ in equations (15) and (16).

Two cases are of interest from a practical point of view, the first being that of a random-mating population with only additive gene effects operating. The

appropriate expressions are then

$$C_b(t, \infty) = 2[1 - (\frac{1}{2})^t] \sigma_A^2$$

and

$$C_w(t, \infty) = (\frac{1}{2})^t \sigma_A^2.$$

Secondly, for a population derived by selfing from a cross of two homozygous parents

$$C_b(t, \infty) = (1/S)[1 - 4(\frac{1}{2})^t {}_0f_{12}] \sigma_A^2$$

and

$$C_w(t, \infty) = (1/S)2(\frac{1}{2})^t {}_0f_{12} \sigma_A^2.$$

It may be noted that dominance effects do not enter into these expressions.

V. EFFICIENCY OF FAMILY SELECTION

The efficiency of family selection has been discussed by Lush (1947) and Falconer (1960), and it is not often recognized that the results are directly applicable to self-fertilizing populations. For the sake of completeness the efficiency equations will be written in terms of the present notation.

It will be assumed that in the t th generation there are a large number of families each with N_t individuals. The intra-class phenotypic correlation between members of families is

$$T_t = {}_t\sigma_b^2 / {}_t\sigma^2.$$

Lush used the symbol t for this correlation. A similar parameter, analogous to Lush's r , is defined as

$$R(t, n) = C_b(t, n) / C(t, n).$$

Suppose that a single cycle of individual selection is carried out in the t th generation and that in the n th generation the population mean is measured relative to the mean of a completely unselected population. The observed gain will be taken as the unit response, and in relative measure the expected gains from between- and within-family selection are

$$\Delta\mu_b(t, n) = [1 + (N_t - 1)R(t, n)] / \{N_t[1 + (N_t - 1)T_t]\}^{\frac{1}{2}}, \quad (25)$$

and

$$\Delta\mu_w(t, n) = [1 - R(t, n)] \{(N_t - 1) / [N_t(1 - T_t)]\}^{\frac{1}{2}}, \quad (26)$$

Family and individual data may be combined to give a selection index. Lush has shown that the best estimate of an individual's genetic worth is given by

$$I = P + \left[\frac{R(t, n) - T_t}{1 - R(t, n)} \cdot \frac{N_t}{1 + (N_t - 1)T_t} \right] \bar{P},$$

where P is the individual's phenotypic value and $\bar{P}$ is the mean of the family of which it is a member. For this system of combined selection the expected relative gain is

$$\Delta\mu_c(t, n) = \left\{ 1 + \frac{[R(t, n) - T_t]^2}{1 - T_t} \cdot \frac{N_t - 1}{1 + (N_t - 1)T_t} \right\}^{\frac{1}{2}}. \quad (27)$$

Equations (25), (26), and (27) may therefore be used to determine the efficiencies of the various selection methods relative to individual selection. It is convenient that only the three variables $R(t, n)$, T_t , and N_t need be known in each case. The correlation T_t is a function of both genetic and environmental variances, so that

a certain amount of control is possible over its value, but the value of $R(t,n)$ depends only on the genetic system operating and is fixed for a given t and n .

Selection from a set of F_3 -generation families will be used as an example, with the F_2 generation being taken as the reference population for measurement of genetic variances. If only additive genetic effects are operating then $R(1,2)$ has the value $\frac{2}{3}$. This is also the value of $R(1,\infty)$ for the model including both additive and dominance effects. Equations (25), (26), and (27) are therefore functions of N_t and T_t and the relative efficiencies may be depicted graphically as has been done by Lush.

A simple result is also obtained when the population undergoing selection is a mixture of homozygous genotypes, in which case $R(t,n) = 1$ and the efficiency equations become

$$\begin{aligned}\Delta\mu_b(t,n) &= \Delta\mu_c(t,n) \\ &= \{N_t/[1+(N_t-1)T_t]\}^{\frac{1}{2}},\end{aligned}$$

and

$$\Delta\mu_w(t,n) = 0.$$

These functions are plotted in Figure 1. Since most populations are effectively homozygous after several generations of selfing, the curves are of general usefulness for determining required family sizes.

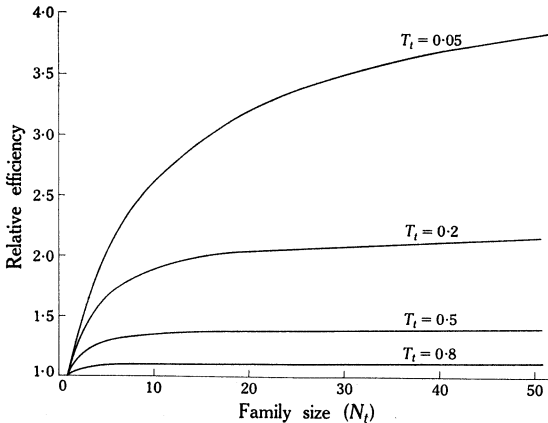


Fig. 1.—Selection within a completely homozygous population. Effect of family size on the relative responses due to between-family and individual selection.

VI. DISCUSSION

Many systems of artificial selection are based on family performance. In this paper, between-family selection implies the selection of complete families on the basis of their mean performances, while within-family selection implies the selection of superior individuals from within each family. In each case, the selected individuals are pooled and allowed to give rise to a new set of self-families, and these form the basis of the subsequent selection cycle.

The expected population mean after $(n-1)$ cycles of between-family selection is

$${}_n\mu = {}_1\mu^* + i \sum_{t=1}^{n-1} [C_b(t,n)/{}_t\sigma_b].$$

For selection within families

$${}_n\mu = {}_n\mu^* + i \sum_{t=1}^{n-1} [C_w(t,n)/{}_t\sigma_w].$$

In each case the natural change in mean due to decreasing heterozygosity is a component of the observed response. When allowance is made for this factor, the heritability for the t th cycle of between-family selection is

$$h_b^2(t,n) = C_b(t,n)/{}_t\sigma_b^2,$$

and the corresponding expression for selection within families is

$$h_w^2(t,n) = C_w(t,n)/{}_t\sigma_w^2.$$

These are the appropriate expressions when selection is carried out in the t th generation and the response is measured in the n th generation.

Estimates of genetic and environmental variance may be substituted into the prediction equations to determine which of a number of populations would respond most to artificial selection. Moll and Robinson (1967) have shown that this would be a practicable procedure in corn, but they have noted that a great deal of data would be necessary to distinguish between populations with genetic variances of comparable magnitude. The difficulty for an autogamous species is to obtain estimates of variance which are free from competition effects and which are realistic for a derived population grown at standard planting rates.

From a practical viewpoint the main interest is in base populations which derive from the cross of two homozygous parents, and Sakai (1954) has given expressions for the expected heritabilities in such populations. In terms of the present notation these expressions are comparable to $C(t,\infty)/{}_t\sigma^2$, for individual selection, and $C_b(t,\infty)/{}_t\sigma_b^2$, for selection between families. Sakai notes the superiority of family selection over individual selection as indicated by the respective heritabilities, but it may be commented that these parameters are a misleading guide to the potential of a selection method. In very general terms, the expected gain from one cycle of selection may be written as

$$\Delta G = i h^2 \sigma.$$

Thus, for example, even though the heritability may be higher for between-family selection than for individual selection, the lower phenotypic standard deviation in the former case would tend to balance out this effect.

In fact, for the example of selection from an F_3 population it is found that individual selection may be more efficient than between-family selection if the heritability is high or if there is an appreciable effect of common environment. Further, the use of a selection index is probably never warranted since between-family selection is almost as efficient as combined selection for low values of T_t , while for medium values the selection of individuals is a relatively efficient method. In describing pedigree selection most authors (e.g. Poehlman 1959; Briggs and Knowles 1967) recommend an F_3 family size of 25-30, and from the figure given by Lush (1947) it is seen that this number ensures a high efficiency of selection for most values of T_t .

Only very rarely is within-family selection for a quantitative character expected to be the most efficient method. The rare cases occur in highly heterozygous populations when there is either a large variance due to common environment or the heritability is high. Although rigorous between-family selection will invariably be the preferred method, it must be remembered that in an autogamous population each self-family descends from a single individual of the previous generation. The selection of complete families may therefore cause a rapid decrease in genetic variability, this being particularly so for small populations. Consideration should be given to a method such as that suggested by Palmer (1953), in which selection is carried out simultaneously in a number of subpopulations and the resulting lines are then either mixed or intercrossed to initiate a new cycle of selection.

The present paper takes no account of linkage between loci, although this factor is doubtless of some importance when linkage disequilibrium is at a high value, as in the early generations following a cross of two homozygous lines. It is also noted that actual selection responses must necessarily tend towards a limit so that the prediction equations are of doubtful reliability beyond three or four cycles of selection.

VII. REFERENCES

- ALLARD, R. W. (1960).—"Principles of Plant Breeding." (John Wiley and Sons, Inc.: New York.)
- BRIGGS, F. N., and KNOWLES, P. F. (1967).—"Introduction to Plant Breeding." (Reinhold Publ. Corp.: New York.)
- FALCONER, D. S. (1960).—"Introduction to Quantitative Genetics." (Ronald Press Co.: New York.)
- KIMURA, M. (1958).—On the change of population fitness by natural selection. *Heredity, Lond.* **12**, 145-67.
- LUSH, J. L. (1947).—Family merit and individual merit as bases for selection. Pt. I. *Am. Nat.* **81**, 241-61.
- MOLL, R. H., and ROBINSON, H. F. (1967).—Quantitative genetic investigations of yield in maize. *Züchter* **37**, 192-9.
- PALMER, T. P. (1953).—Progressive improvement in self-fertilized crops. *Heredity, Lond.* **7**, 127-9.
- PEDERSON, D. G. (1969).—The prediction of selection response in a self-fertilizing species. I. Individual selection. *Aust. J. biol. Sci.* **22**, 117-29.
- POEHLMAN, J. M. (1959).—"Breeding Field Crops." (Henry Holt and Co., Inc.: New York.)
- SAKAI, K. (1954).—Theoretical studies on the technique of plant breeding. I. Value of heritability in bulks and plant progenies of successive hybrid generations in autogamous plants. *Jap. J. Breed.* **4**, 145-8. (In Japanese.)

