

THE PREDICTION OF SELECTION RESPONSE IN A SELF-FERTILIZING SPECIES*

I. INDIVIDUAL SELECTION

By D. G. PEDERSON†

[*Manuscript received July 26, 1968*]

Summary

The response to truncation selection is considered for populations which reproduce by self-fertilization. Family structure is disregarded in the selection process, and individuals are selected or rejected on the basis of phenotypic value. For k unlinked loci of small effect, the expected response from n cycles of selection is the sum of

- (1) the change in mean due to inbreeding, and
- (2) a component which is a function of the selection intensity and the variances and covariances of an unselected population.

For a single-locus model the expected response is interpreted in terms of the parameters of the base population. Predictions may be readily made for random-mating populations or for populations with two alleles per locus.

A general expression is derived for the expected change in mean following selection and reversion to homozygosity. This expression is investigated for the single-locus model.

I. INTRODUCTION

Procedures for estimating the genetic components of variance in an autogamous species are well documented. The pioneering study was that of Fisher, Immer, and Tedin (1932), who gave expressions for the variance of an F_2 generation, and of subsequent generations obtained by both selfing and intercrossing, in terms of the additive and dominance effects of single loci. The methods of estimation were considered in greater detail by Mather (1949), and the theory was extended to indicate the possible effects of linkage bias on the estimates. In a later paper, Hayman and Mather (1955) proposed an extension of the genetic model to include epistatic effects. But although estimation procedures suitable for an autogamous species have been considered in some detail, there is little information on how these estimates should be used to predict the response from artificial selection.

In a series of connected studies, Sakai (1954, 1955, 1956) considered a number of problems, including how to discriminate the most desirable hybrid combinations in early generations of hybrid bulks, and how to construct a selection index for

* Published with the approval of the Director of Research as Paper No. 2683 of the Journal Series of the North Carolina Agricultural Experiment Station, Raleigh, North Carolina. This investigation was supported in part by Public Health Service Research Grant GM 11546 from the Division of General Medical Sciences.

† Department of Genetics, North Carolina State University, Raleigh, North Carolina 27607; present address: North West Wheat Research Institute, Narrabri, N.S.W. 2390.

individual plants on the basis of individual measurements, line means, and line group means. It was shown that heritabilities are invariably low for the early generations following a cross of two lines, and the usefulness of the pedigree method of plant selection was questioned on the basis of this result. However, all of these investigations were based on a single-locus model, and the only populations considered were the segregating generations following the cross of two inbred lines. Panse (1940*a*, 1940*b*) made use of moment-generating functions in his investigation of the response to selection in an F_2 population. Several models of gene action were considered for a single locus, and values were derived for various characteristics of the F_3 population resulting from the selection and self-fertilization of F_2 individuals. A more comprehensive set of genetic models were later considered by Panse and Bokil (1953) and it was suggested that the results would be of use to a breeder in predicting qualitative responses to artificial selection. No formulae were given which could be used to predict the absolute magnitude of a selection response.

In the present study prediction formulae are therefore derived for the response to truncation selection in an infinitely large population of self-fertilizing individuals. Only the simplest situation is considered, in which the selection process is based on individual performance and takes no account of family structure within each generation. As such the results are applicable only to species and culture methods which permit the ascertainment of individual values free from competition effects. Results are presented for the generalized response due to a number of unlinked loci, each with many alleles, and then for a more restrictive model the response is expressed in terms of components of genetic variance. In the final section the consequences of relaxation of selection are investigated.

II. GENERALIZED SELECTION RESPONSE

Attention will be given to the diploid genotypes generated by m alleles at each of k unlinked loci. It will be assumed that these k loci represent only a small subset of the total number of loci controlling the character undergoing selection, and that the effects of subsets are additive.

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 values are completely general and permit all classes of between- and within-locus interaction.

The operation of self-fertilization will be denoted by the letter S . In general terms, the expression $(S^t Y_{\Pi i_s j_s})$ represents 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 ${}_0f_{\Pi i_s j_s}$, where the zero subscript indicates generation number. The symbol ${}_0f_{\cdot i_s j_s}$ will be used to represent the marginal genotypic frequencies at the s th locus. Summation is

assumed to be over all of the i 's and j 's not specified in the subscript. The initial population mean, with respect to the k loci under consideration, is given by

$$_0\mu = \sum_{\text{all } i,j} _0f_{\Pi_{i_s j_s}} Y_{\Pi_{i_s j_s}}.$$

Now consider the result of t generations of self-fertilization. The marginal genotypic frequencies at the s th locus for the t th generation are

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

where $_0 f_{i_s}$ is the frequency of allele $A_{i_s}^s$, and

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

The frequency in equation (1) is marked with an asterisk to indicate that it is the result of self-fertilization without selection. Since the k loci are unlinked, the overall genotypic frequency may be found as the product of marginal genotypic frequencies, i.e.

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

The population mean after t generations of selfing may be found either as

$$_t \mu^* = \sum_{\text{all } i,j} _t f_{\Pi_{i_s j_s}}^* Y_{\Pi_{i_s j_s}},$$

or as

$$_t \mu^* = \sum_{\text{all } i,j} _0 f_{\Pi_{i_s j_s}} (S^t Y_{\Pi_{i_s j_s}}).$$

(a) Covariances in the Absence of Selection

For selection with random mating, the expected response can be written in terms of covariances between related individuals (e.g. see Griffing 1960). It is therefore of interest in the present case to derive similar expressions for the covariances between relatives, and in this section particular attention will be given to the expected covariance between an individual in the t th generation and its offspring array in the n th generation of self-fertilization ($n > t$).

The frequency of $\Pi A_{i_s}^s A_{j_s}^s$ after t generations of self-fertilization without selection is $_t f_{\Pi_{i_s j_s}}^*$. In the n th generation this genotype will have produced an array of genotypes for which the genotypic values may be written $S^{n-t} Y_{\Pi_{i_s j_s}}$. The desired covariance in an unselected population is therefore

$$\begin{aligned} C(t,n) &= \sum_{\text{all } i,j} _t f_{\Pi_{i_s j_s}}^* Y_{\Pi_{i_s j_s}} (S^{n-t} Y_{\Pi_{i_s j_s}}) - _t \mu^* _n \mu^* \\ &= \sum _t f_{\Pi_{i_s j_s}}^* (Y_{\Pi_{i_s j_s}} - _t \mu^*) (S^{n-t} Y_{\Pi_{i_s j_s}}). \end{aligned} \quad (3)$$

(b) Response from n Cycles of Selection

Suppose that the population phenotypic variance, which takes into account all sets of loci, has a value of ${}_t \sigma^2$ after t generations of selection. It is convenient at this point to define

$$_t Z_{\Pi_{i_s j_s}} = (Y_{\Pi_{i_s j_s}} - _t \mu) / {}_t \sigma$$

and

$$_t Z_{\Pi_{i_s j_s}}^* = (Y_{\Pi_{i_s j_s}} - _t \mu^*) / {}_t \sigma.$$

The assumption will be made that the loci within the A subset are all of small effect, so that the deviation of any genotypic value from the mean is always small relative to the phenotypic standard deviation. It is then reasonable to assume that the squares and products of Z values are of negligible magnitude.

The selection scheme will now be considered in which only the individuals with phenotypic value exceeding a certain arbitrary value are chosen to contribute to the next generation. Let the standardized selection differential have a value \bar{i} . Then it follows from Kimura (1958) that the frequency of $\Pi A_{i_s}^s A_{j_s}^s$ in the set of selected individuals is given by

$${}_0g_{\Pi i_s j_s} = {}_0f_{\Pi i_s j_s}(1 + \bar{i} {}_0Z_{\Pi i_s j_s}).$$

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

$$\begin{aligned} {}_1\mu &= \sum_{\text{all } i, j} {}_0g_{\Pi i_s j_s} (S Y_{\Pi i_s j_s}) \\ &= \sum {}_0f_{\Pi i_s j_s} (1 + \bar{i} {}_0Z_{\Pi i_s j_s}) (S Y_{\Pi i_s j_s}) \\ &= {}_1\mu^* + \bar{i} \sum {}_0f_{\Pi i_s j_s} {}_0Z_{\Pi i_s j_s} (S Y_{\Pi i_s j_s}). \end{aligned} \quad (4)$$

The genotypic frequencies following one cycle of selection and self-fertilization will be denoted ${}_1f_{\Pi i_s j_s}$.

From equations (3) and (4) it follows that

$${}_1\mu = {}_1\mu^* + \bar{i} C(0, 1)/{}_0\sigma.$$

This result may be compared with that of Griffing (1960), who found that the mean of the progeny population which results from random mating selected parents from any arbitrary population is equal to the sum of (1) the mean of the population after it has been allowed to mate at random for one generation in the absence of selection, and (2) the increment $2\bar{i}\text{Cov}(\text{PO})/{}_0\sigma$, where $\text{Cov}(\text{PO})$ is the covariance of parents and offspring in the base population.

A second cycle of selection of the same intensity gives

$${}_1g_{\Pi i_s j_s} = {}_1f_{\Pi i_s j_s}(1 + \bar{i} {}_1Z_{\Pi i_s j_s}),$$

and subsequent self-fertilization gives a population with mean

$$\begin{aligned} {}_2\mu &= \sum_{\text{all } i, j} {}_1g_{\Pi i_s j_s} (S Y_{\Pi i_s j_s}) \\ &= \sum ({}_1f_{\Pi i_s j_s} + \bar{i} {}_1f_{\Pi i_s j_s} {}_1Z_{\Pi i_s j_s}) (S Y_{\Pi i_s j_s}). \end{aligned}$$

But the frequencies ${}_1f_{\Pi i_s j_s}$ are the result of self-fertilization in a population with frequencies ${}_0g_{\Pi i_s j_s}$, and hence

$$\begin{aligned} {}_2\mu &= \sum {}_0g_{\Pi i_s j_s} (S^2 Y_{\Pi i_s j_s}) + \bar{i} \sum {}_1f_{\Pi i_s j_s} {}_1Z_{\Pi i_s j_s} (S Y_{\Pi i_s j_s}) \\ &= {}_2\mu^* + \bar{i} \sum {}_0f_{\Pi i_s j_s} {}_0Z_{\Pi i_s j_s} (S^2 Y_{\Pi i_s j_s}) + \bar{i} \sum {}_1f_{\Pi i_s j_s} {}_1Z_{\Pi i_s j_s} (S Y_{\Pi i_s j_s}). \end{aligned}$$

In general, after n cycles of truncation selection the population mean is found to be

$${}_n\mu = {}_n\mu^* + \bar{i} \sum_{t=0}^{n-1} \left[\sum_{\text{all } i, j} {}_t f_{\Pi i_s j_s} {}_t Z_{\Pi i_s j_s} (S^{n-t} Y_{\Pi i_s j_s}) \right]. \quad (5)$$

Equation (5) is written in terms of the parameters of the selected population. However, by definition

$$\begin{aligned} {}_tZ_{\Pi_{i_s j_s}} &= (Y_{\Pi_{i_s j_s}} - {}_t\mu^*) / {}_t\sigma - ({}_t\mu - {}_t\mu^*) / {}_t\sigma \\ &= {}_tZ_{\Pi_{i_s j_s}}^* - (i / {}_t\sigma) \sum_{u=0}^{t-1} \left[\sum_{\text{all } i, j} {}_u f_{\Pi_{i_s j_s}} {}_u Z_{\Pi_{i_s j_s}} (S^{n-u} Y_{\Pi_{i_s j_s}}) \right]. \end{aligned}$$

The second term on the right-hand side of this equation is effectively zero since the product of two Z 's is a negligible quantity. Equation (5) therefore becomes

$${}_n\mu = {}_n\mu^* + i \sum_{t=0}^{n-1} \left[\sum_{\text{all } i, j} {}_t f_{\Pi_{i_s j_s}} {}_t Z_{\Pi_{i_s j_s}}^* (S^{n-t} Y_{\Pi_{i_s j_s}}) \right]. \quad (6)$$

Further, by a method similar to that used to derive equation (5), with repeated application of equations (1) and (2), it may be shown that

$$\begin{aligned} {}_n f_{\Pi_{i_s j_s}} &= {}_n f_{\Pi_{i_s j_s}}^* + i \sum_{t=0}^{n-1} \left\{ \left[\left(\frac{1}{2} \right)^{n-t} \right]^k [{}_t f_{\Pi_{i_s j_s}} {}_t Z_{\Pi_{i_s j_s}} \right. \\ &\quad + (2^{n-t} - 1) \sum_{u=1}^k (\beta_{i_u j_u} \sum_{j_u} {}_t f_{\Pi_{i_s j_s}} {}_t Z_{\Pi_{i_s j_s}}) \\ &\quad + (1/2!)(2^{n-t} - 1)^2 \sum_{u \neq u'} (\beta_{i_u j_u} \beta_{i_{u'} j_{u'}} \sum_{j_u, j_{u'}} {}_t f_{\Pi_{i_s j_s}} {}_t Z_{\Pi_{i_s j_s}}) \\ &\quad \left. + \dots \dots \dots \right\}. \end{aligned}$$

The second term on the right-hand side is a function of Z values, and substitution into equation (6) therefore gives the final result that

$${}_n\mu = {}_n\mu^* + i \sum_{t=0}^{n-1} \left[\sum_{\text{all } i, j} {}_t f_{\Pi_{i_s j_s}}^* {}_t Z_{\Pi_{i_s j_s}}^* (S^{n-t} Y_{\Pi_{i_s j_s}}) \right]. \quad (7)$$

From equation (3) it is clear that this expression may be rewritten as

$${}_n\mu = {}_n\mu^* + i \sum_{t=0}^{n-1} [C(t, n) / {}_t\sigma]. \quad (8)$$

Equation (5) could have been written in this form but the covariance term would have referred to the population undergoing selection. This equality of the covariance terms follows from the assumption that loci are of small effect relative to the phenotypic standard deviation. In the same way the overall genotypic, and hence phenotypic, variances are identical for the selected and unselected populations, and only the symbol ${}_t\sigma^2$ will therefore be used.

III. RESPONSE IN TERMS OF GENETIC PARAMETERS

In the prediction of selection response under random mating the parameter of overriding importance is the covariance between parents and offspring in the base population (e.g. Griffing 1960). From equation (8) it is clear that a less simple situation exists for selection with self-fertilization, in that the number of covariances to be estimated is equal to the number of cycles for which a predicted response is required. In addition, while the phenotypic variance remains constant for a population mating at random, this is not necessarily so when propagation is by self-

fertilization. For the particular case of a population with a high level of heterozygosity and no overdominant loci the phenotypic variance will tend to increase as selfing proceeds.

In an attempt to deal with these difficulties a single locus model will be considered and the covariance $C(t, n)$ will be expressed in terms of genetic parameters of the base population, that is the 0th generation. When this is done, it is sufficient to estimate the genetic parameters to determine $C(t, n)$ at any stage of self-fertilization. Although the theory is based on a single locus, it will again be assumed that the effects of many independent loci may be summed. Explicit expressions will not be given for the population phenotypic variance since the genotypic variance for the single locus may always be found by substituting $n = t$ in the appropriate expression for $C(t, n)$, and then

$$t\sigma^2 = \sum_{\text{loci}} C(t, t) + \sigma_E^2,$$

where σ_E^2 is the error variance, assumed to be constant.

The subscripts of the previous section will be simplified by putting $i_1 = i$ and $j_1 = j$, and omitting all other subscripts. The "dot" notation will be used, e.g.

$${}_0f_{i.} = \sum_j {}_0f_{ij}.$$

A completely general model will be considered, such that the value of genotype $A_i A_j$ is given by

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

In this equation, α_i and α_j are the additive effects of the i th and j th alleles respectively, and δ_{ij} is the dominance effect associated with genotype $A_i A_j$. The α 's may be determined by a least-squares procedure, which yields the normal equations

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

Summation over this set of equations gives

$$\sum_i {}_0f_{i.}\alpha_i = 0. \quad (9)$$

It can be further shown that

$$\sum_i {}_0f_{ij}\delta_{ij} = \sum_j {}_0f_{ij}\delta_{ij} = 0. \quad (10)$$

Following Kempthorne (1957), the additive variance in the base population is given by

$$\begin{aligned} \sigma_A^2 &= 2 \sum_{ij} {}_0f_{ij}\alpha_i(Y_{ij} - {}_0\mu) \\ &= 2 \sum_i {}_0f_{i.}\alpha_i^2 + 2 \sum_{ij} {}_0f_{ij}\alpha_i\alpha_j. \end{aligned} \quad (11)$$

The remainder of the variance is attributable to dominance effects, and is given by

$$\sigma_D^2 = \sum_{ij} {}_0f_{ij}\delta_{ij}^2. \quad (12)$$

For a single locus, equation (3) reduces to

$$C(t, n) = \left(\frac{1}{2}\right)^{n-t} \sum_{ij} {}_t f_{ij}^* (Y_{ij} - {}_t\mu^*) [Y_{ij} + (2^{n-t} - 1)Y_{ii}]. \quad (13)$$

Similarly, from equation (1)

$${}_i f_{ij}^* = (\frac{1}{2})^t {}_0 f_{ij} + \beta_{ij} [1 - (\frac{1}{2})^t] {}_0 f_{i.},$$

so that

$$C(t, n) = (\frac{1}{2})^n \sum_{ij} {}_0 f_{ij} (Y_{ij} - {}_t \mu^*) [Y_{ij} + (2^{n-t} - 1) Y_{ii}] + [1 - (\frac{1}{2})^t] \sum_i {}_0 f_{i.} Y_{ii} (Y_{ii} - {}_t \mu^*).$$

But

$$\begin{aligned} {}_t \mu^* &= \sum_{ij} {}_i f_{ij}^* Y_{ij} \\ &= (\frac{1}{2})^t {}_0 \mu + [1 - (\frac{1}{2})^t] \sum_i {}_0 f_{i.} Y_{ii} \\ &= {}_0 \mu + [1 - (\frac{1}{2})^t] \sum_i {}_0 f_{i.} \delta_{ii}, \end{aligned} \quad (14)$$

and therefore

$$\begin{aligned} C(t, n) &= (\frac{1}{2})^n \sum_{ij} {}_0 f_{ij} (Y_{ij} - {}_0 \mu) [Y_{ij} + (2^{n-t} - 1) Y_{ii}] + [1 - (\frac{1}{2})^t] \sum_i {}_0 f_{i.} Y_{ii} (Y_{ii} - {}_0 \mu) \\ &\quad - (\frac{1}{2})^n [1 - (\frac{1}{2})^t] \sum_i {}_0 f_{i.} \delta_{ii} \sum_{ij} {}_0 f_{ij} [Y_{ij} + (2^{n-t} - 1) Y_{ii}] \\ &\quad - [1 - (\frac{1}{2})^t]^2 \sum_i {}_0 f_{i.} \delta_{ii} \sum_i {}_0 f_{i.} Y_{ii}. \end{aligned}$$

Substitution of gene effects into this equation yields the final result

$$\begin{aligned} C(t, n) &= [4 - 2(\frac{1}{2})^t] \sum_i {}_0 f_{i.} \alpha_i^2 + 2(\frac{1}{2})^t \sum_{ij} {}_0 f_{ij} \alpha_i \alpha_j \\ &\quad + [(\frac{1}{2})^t - (\frac{1}{2})^n] \sum_{ij} {}_0 f_{ij} (\alpha_i + \alpha_j) \delta_{ii} + 4[1 - (\frac{1}{2})^t] \sum_i {}_0 f_{i.} \alpha_i \delta_{ii} \\ &\quad + [1 - (\frac{1}{2})^t] \sum_i {}_0 f_{i.} \delta_{ii}^2 - [1 - (\frac{1}{2})^t][1 - (\frac{1}{2})^n] (\sum_i {}_0 f_{i.} \delta_{ii})^2 + (\frac{1}{2})^n \sum_{ij} {}_0 f_{ij} \delta_{ij}^2. \end{aligned} \quad (15)$$

This equation is valid only if the α 's and δ 's are defined in the base population, and for this to be so all of the ${}_0 f_{ij}$ must be non-zero. This somewhat limiting restriction will apply in all of the discussion which follows.

Although each of the unknown functions in equation (15) is estimable, it is clear that the equation is of little practical value. However, there are two classes of population for which a simplified result may be obtained, namely a population which has arisen by a random-mating procedure and a population which has only two alleles present per locus. These two cases will now be considered in turn.

(a) *Random-mating Population*

The significant feature of a random-mating population is that, for all i and j ,

$${}_0 f_{ij} = ({}_0 f_{i.})({}_0 f_{.j}).$$

The terms involving the products $\alpha_i \alpha_j$ and $\alpha_j \delta_{ii}$ in equation (15) are therefore zero. However, the number of unknown functions in the equation for $C(t, n)$ is still prohibitively large. The assumption will therefore be made that all dominance effects are zero, in which case the mean value of the character is not expected to change

with self-fertilization in the absence of selection. Substitution of the expressions for $C(t, n)$ and σ^2 into equation (8) therefore gives

$$n\mu = i \sum_{t=0}^{n-1} \{[\theta(t) \cdot \sigma_A^2]/[\theta(t) \cdot \sigma_A^2 + \sigma_E^2]^{\frac{1}{2}}\},$$

where

$$\theta(t) = 2 - (\frac{1}{2})^t.$$

This is equivalent to the expression presented by Brim and Cockerham (1961).

(b) *Populations with two Alleles per Locus*

The restriction (9) may be written as

$$\alpha_2 = -({}_0f_{1.}\alpha_1)/{}_0f_{2.},$$

so that equation (11) for the additive variance in the base population becomes

$$\sigma_A^2 = (\alpha_1^2/{}_0f_{2.}^2)(S - D^2), \quad (16)$$

where

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

and

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

Similarly, δ_{12} and δ_{22} may be written in terms of δ_{11} , and there results

$$\sigma_D^2 = ({}_0f_{11}\delta_{11}^2/2{}_0f_{12}{}_0f_{22})(S - D^2). \quad (17)$$

Finally, the covariance σ_{AD} will be defined as

$$\begin{aligned} \sigma_{AD} &= \frac{1}{2} \sum_{ij} {}_0f_{ij} (\alpha_i + \alpha_j) \delta_{ii} \\ &= (D \alpha_1 \delta_{11}/4{}_0f_{2.}{}_0f_{22})(S - D^2). \end{aligned} \quad (18)$$

With this definition,

$$C(0, 1) = \sigma_A^2 + \sigma_{AD} + \frac{1}{2}\sigma_D^2.$$

There exist a number of useful relationships involving the genotypic and genic frequencies, namely:

$$\begin{aligned} {}_0f_{1.} &= \frac{1}{2} (1 - D), \\ {}_0f_{2.} &= \frac{1}{2} (1 + D), \\ {}_0f_{22} + {}_0f_{11} &= \frac{1}{2}(S^2 + D^2), \end{aligned}$$

and

$${}_0f_{11}{}_0f_{22} = \frac{1}{4}(S^2 - D^2).$$

These relationships, and those between the α 's and the δ 's, are used to write equation (15) in the form

$$\begin{aligned} C(t, n) &= (\alpha_1^2/{}_0f_{2.}^2)[1 - 2(\frac{1}{2})^t{}_0f_{12} - D^2] \\ &\quad + (D\alpha_1\delta_{11}/2{}_0f_{2.}{}_0f_{22})\{[(\frac{1}{2})^t - (\frac{1}{2})^n](S - D^2) + 2[1 - (\frac{1}{2})^t](1 - D^2)\} \\ &\quad + \{[1 - (\frac{1}{2})^t]D^2(1 - D^2)\delta_{11}^2/4{}_0f_{22}^2\} \\ &\quad + [(\frac{1}{2})^n(S - D^2)\delta_{11}^2/8{}_0f_{12}{}_0f_{22}^2]\{(S - D^2)[1 - 2(\frac{1}{2})^t{}_0f_{12}] - 2D^2{}_0f_{12}\}. \end{aligned}$$

From equations (16), (17), and (18) it then follows that

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

For the special case of $D = 0$, equation (19) and the corresponding equation for the phenotypic variance reduce to

$$C(t, n) = (1/S)[1 - 2(\frac{1}{2})^t {}_0f_{12}][\sigma_A^2 + (\frac{1}{2})^n \sigma_D^2], \tag{20}$$

and

$$t\sigma^2 = (1/S)[1 - 2(\frac{1}{2})^t {}_0f_{12}][\sigma_A^2 + (\frac{1}{2})^t \sigma_D^2] + \sigma_E^2. \tag{21}$$

These would be the appropriate expressions for selection in a population derived by self-fertilization from a cross of two homozygous lines.

IV. RELAXATION OF SELECTION

Of particular interest for an autogamous species is the gain after selection has been relaxed, and the population has resumed its homozygous condition. For a population which has been selected for n cycles and then allowed to revert to homozygosity the expected mean will be written as ${}_n\mu_H$.

In the general notation, the frequency of genotype $\Pi A_{i_s}^s A_{j_s}^s$ in the first set of selected individuals is ${}_0g_{\Pi i_s j_s}$. If self-fertilization is then carried out to homozygosity, the expected frequency of $\Pi A_{i_s}^s A_{i_s}^s$ is given by

$$\sum_{\text{all } j} {}_0g_{\Pi i_s j_s},$$

since the k loci are unlinked. The proof of this relationship for the two-locus case is given by Griffing (1956). It follows that

$$\begin{aligned}
 {}_1\mu_H &= \sum_{\text{all } i} \left(\sum_{\text{all } j} {}_0g_{\Pi i_s j_s} \right) Y_{\Pi i_s i_s} \\
 &= \sum_{\text{all } i, j} {}_0f_{\Pi i_s j_s} Y_{\Pi i_s i_s} + \bar{i} \sum_{\text{all } i, j} {}_0f_{\Pi i_s j_s} {}_0Z_{\Pi i_s j_s} Y_{\Pi i_s i_s}.
 \end{aligned}$$

In a similar fashion

$$\begin{aligned}
 {}_2\mu_H &= \sum_{\text{all } i} \left(\sum_{\text{all } j} {}_1g_{\Pi i_s j_s} \right) Y_{\Pi i_s i_s} \\
 &= \sum_{\text{all } i, j} {}_1f_{\Pi i_s j_s} (1 + \bar{i} {}_1Z_{\Pi i_s j_s}) Y_{\Pi i_s i_s} \\
 &= \sum_{\text{all } i, j} {}_0g_{\Pi i_s j_s} Y_{\Pi i_s i_s} + \bar{i} \sum_{\text{all } i, j} {}_1f_{\Pi i_s j_s} {}_1Z_{\Pi i_s j_s} Y_{\Pi i_s i_s} \\
 &= \sum_{\text{all } i, j} {}_0f_{\Pi i_s j_s} Y_{\Pi i_s i_s} + \bar{i} \sum_{\text{all } i, j} {}_0f_{\Pi i_s j_s} {}_0Z_{\Pi i_s j_s} Y_{\Pi i_s i_s} \\
 &\quad + \bar{i} \sum_{\text{all } i, j} {}_1f_{\Pi i_s j_s} {}_1Z_{\Pi i_s j_s} Y_{\Pi i_s i_s}.
 \end{aligned}$$

As a general result, after n cycles of selection

$$n\mu_H = \sum_{\text{all } i,j} {}_0f_{\Pi i_s j_s} Y_{\Pi i_s t_s} + i \sum_{t=0}^{n-1} \left(\sum_{\text{all } i,j} {}_t f_{\Pi i_s j_s} {}_t Z_{\Pi i_s j_s} Y_{\Pi i_s t_s} \right).$$

By the same manner in which equation (7) was derived from equation (5), this expression may be transformed to give

$$n\mu_H = \sum_{\text{all } i,j} {}_0f_{\Pi i_s j_s} Y_{\Pi i_s t_s} + i \sum_{t=0}^{n-1} \left(\sum_{\text{all } i,j} {}_t f_{\Pi i_s j_s}^* {}_t Z_{\Pi i_s j_s}^* Y_{\Pi i_s t_s} \right).$$

It follows that

$$n\mu_H = \sum_{\text{all } i,j} {}_0f_{\Pi i_s j_s} Y_{\Pi i_s t_s} + i \sum_{t=0}^{n-1} [C(t, \infty)/t\sigma], \quad (22)$$

where

$$C(t, \infty) = \lim_{n \rightarrow \infty} C(t, n).$$

For the general one-locus case, from (15)

$$\begin{aligned} C(t, \infty) = & [4 - 2(\tfrac{1}{2})^t] \sum_i {}_0f_i. \alpha_i^2 + 2(\tfrac{1}{2})^t \sum_{ij} {}_0f_{ij} \alpha_i \alpha_j + (\tfrac{1}{2})^t \sum_{ij} {}_0f_{ij} (\alpha_i + \alpha_j) \delta_{ii} \\ & + 4[1 - (\tfrac{1}{2})^t] \sum_i {}_0f_i. \alpha_i \delta_{ii} + [1 - (\tfrac{1}{2})^t] \left[\sum_i {}_0f_i. \delta_{ii}^2 - \left(\sum_i {}_0f_i. \delta_{ii} \right)^2 \right]. \end{aligned} \quad (23)$$

When the genotypic frequencies in the base population are the product of gene frequencies, two of the product terms disappear and there remains

$$\begin{aligned} C(t, \infty) = & [4 - 2(\tfrac{1}{2})^t] \sum_i {}_0f_i. \alpha_i^2 + [4 - 3(\tfrac{1}{2})^t] \sum_i {}_0f_i. \alpha_i \delta_{ii} \\ & + [1 - (\tfrac{1}{2})^t] \left[\sum_i {}_0f_i. \delta_{ii}^2 - \left(\sum_i {}_0f_i. \delta_{ii} \right)^2 \right]. \end{aligned} \quad (24)$$

For the case of two alleles per locus, substitution of $n = \infty$ into equation (19) gives

$$\begin{aligned} C(t, \infty) = & \left[\frac{1 - 2(\tfrac{1}{2})^t {}_0f_{12} - D^2}{(S - D^2)} \right] \sigma_A^2 \\ & + \left\{ \frac{2(\tfrac{1}{2})^t (S - D^2) + 4[1 - (\tfrac{1}{2})^t] (1 - D^2)}{(S - D^2)} \right\} \sigma_{AD} \\ & + \left\{ \frac{2[1 - (\tfrac{1}{2})^t] D^2 (1 - D^2) {}_0f_{12}}{(S^2 - D^2)(S - D^2)} \right\} \sigma_D^2. \end{aligned} \quad (25)$$

As may be expected, when $D = 0$ the last two terms disappear so that the covariance is a function of the additive variance alone.

V. DISCUSSION

The genotypic frequencies of a large random-mating population remain constant from generation to generation if selective forces are not acting, and genetic variability is therefore maintained in a form readily utilizable by artificial selection. By way of contrast, a self-fertilizing population normally comprises a number of homozygous genotypes which, with continued selfing, offer limited potential for genetic improvement. Hybridization should therefore precede artificial selection for such a population, and it is with hybrid material constructed in this manner that the present study is concerned.

The key result of the study is the equation for the expected population mean after n cycles of individual selection, viz.

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

It is clear that the natural change in mean due to decreasing heterozygosity is a component of the observed response. If allowance is made for this factor, the heritability for the t th cycle of selection is

$$h^2(t,n) = C(t,n)/{}_t\sigma^2.$$

This is the relevant expression when selection is carried out in the t th generation and the response is measured in the n th generation.

There are several courses of action open to a person wishing to predict the selection response for a population of self-fertilizers. The first method would be to grow the population for several generations in the absence of selection and thus estimate the means, variances, and covariances necessary for prediction. The major assumption would be that of no linkage between the loci responding to selection. This method has little attraction, however, since it would be necessary to grow n generations of plants to predict the response for n cycles of selection. One of the less general methods may therefore be more suitable.

On the assumption that loci do not interact, equation (15) gives the required covariance in terms of single-locus effects. Although the number of unknown parameters is probably too great for the equation to be of any predictive value, there are two classes of population for which a simplified result is obtained. The first type has genotypic frequencies equal to the products of gene frequencies, for example the population resulting from a diallel intercross of several parents. Some of the problems of estimation have been considered by Harris (1964). There are few difficulties if no dominance is assumed, possibly a reasonable assumption for an autogamous species.

A simple result is also obtained if it is reasonable to assume only two alleles per segregating locus in the base population. Such a population would result from a hybridization scheme between two homozygous lines. However, it would be necessary to know the mating procedure by which the population was constructed, since genotypic frequencies enter into the prediction equation.

A common procedure for an autogamous species is to cross two inbred parents and then repeatedly self the hybrid material. The resulting populations have two alleles per locus and the homozygotes equally frequent, and, as was noted in the previous section, the covariance $C(t,n)$ and variance ${}_t\sigma^2$ are functions of the additive, dominance, and error variances alone. The covariance is a function of only the additive variance when the gain to be estimated is the ultimate gain following relaxation of selection.

Selection in an F_2 population, which has ${}_0f_{11} = {}_0f_{12} = {}_0f_{22} = \frac{1}{4}$, will be used as an example. Direct application of equations (8), (20), and (21) gives

$${}_n\mu = {}_n\mu^* + \bar{i} \sum_{t=0}^{n-1} \frac{[2 - (\frac{1}{2})^t][\sigma_A^2 + (\frac{1}{2})^n\sigma_D^2]}{\{[2 - (\frac{1}{2})^t][\sigma_A^2 + (\frac{1}{2})^t\sigma_D^2] + \sigma_E^2\}^{\frac{1}{2}}}.$$

The equation for the ultimate gain following relaxation of selection is similar except that the dominance variance does not appear in the numerator.

From (14)

$$n\mu^* - n_{-1}\mu^* = \frac{1}{2}(n_{-1}\mu^* - n_{-2}\mu^*).$$

This equation could be used to estimate the change in mean in the absence of selection for the above example. In practice, it would be necessary to observe the unselected mean for at least two generations. It may be noted that if dominance of the "plus" genes is a significant factor in the genetic structure of a population, then the natural decrease in mean may be of such a high value that the overall response is negative for one or more cycles of selection. This result may be compared with the situation for a random-mating population, for which the expected selection response is always a positive function of the additive variance.

For the single-locus theory to be applicable it is necessary that the gene effects are defined in the base population. This is always so for a random-mated base population, but there is a restriction on the type of two-allele population which can be considered. For example, in a backcross population only one homozygote and the heterozygote are represented at each locus, so that dominance effects are undefined. One way of overcoming this difficulty would be to determine, from first principles, the expected genotypic frequencies after one cycle of selection and then treat this as the base population.

A basic assumption of the present study is that the squares and products of Z values are of negligible magnitude. In its original form (Kimura 1958) this assumption stated that the effects of individual loci are small relative to the total phenotypic standard deviation, but Griffing (1960) extended the concept to include the effects of subsets of loci. The present study follows this later development. No general theory exists to predict the response of loci of large effect, although their patterns of response may be investigated by methods such as that used by Latter (1965) for selection with random mating. In practice the effects of individual loci are always unknown, and so the present theory should not be applied for more than two or three cycles of selection.

VI. ACKNOWLEDGMENTS

The author wishes to express his gratitude to Drs. W. D. Hanson and C. C. Cockerham for critically reading the manuscript.

VII. REFERENCES

- BRIM, C. A., and COCKERHAM, C. C. (1961).—Inheritance of quantitative characters in soybeans. *Crop Sci.* **1**, 187–90.
- FISHER, R. A., IMMER, F. R., and TEDIN, O. (1932).—The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics, Princeton* **17**, 107–24.
- GRIFFING, B. (1956).—A generalized treatment of the use of diallel crosses in quantitative inheritance. *Heredity, Lond.* **10**, 31–50.
- GRIFFING, B. (1960).—Theoretical consequences of truncation selection based on the individual phenotype. *Aust. J. biol. Sci.* **13**, 307–43.
- HARRIS, D. L. (1964).—Genotypic covariances between inbred relatives. *Genetics, Princeton* **50**, 1319–48.

- HAYMAN, B. I., and MATHER, K. (1955).—The description of genic interactions in continuous variation. *Biometrics* **11**, 69–82.
- KEMPTHORNE, O. (1957).—“An Introduction to Genetic Statistics.” (John Wiley & Sons, Inc.: New York.)
- KIMURA, M. (1958).—On the change of population fitness by natural selection. *Heredity, Lond.* **12**, 145–67.
- LATTER, B. D. H. (1965).—The response to artificial selection due to autosomal genes of large effect. I. Changes in gene frequency at an additive locus. *Aust. J. biol. Sci.* **18**, 585–98.
- MATHER, K. (1949).—“Biometrical Genetics.” (Methuen & Co. Ltd.: London.)
- PANSE, V. G. (1940*a*).—A statistical study of quantitative inheritance. *Ann. Eugen.* **10**, 76–105.
- PANSE, V. G. (1940*b*).—Application of genetics to plant breeding. II. The inheritance of quantitative characters and plant breeding. *J. Genet.* **40**, 283–302.
- PANSE, V. G., and BOKIL, S. D. (1953).—Statistical properties of characters in quantitative inheritance with special reference to plant breeding. Proc. 9th Int. Conf. Genet., Bellagio (Como), 1953. *Caryologia* **6** (suppl.), 865–7.
- 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.
- SAKAI, K. (1955).—Theoretical studies on the technique of plant breeding. II. Theoretical basis for discriminating the most desirable hybrid combinations in early generations of hybrid bulks. *Jap. J. Breed.* **5**, 110–14.
- SAKAI, K. (1956).—Theoretical studies on the technique of plant breeding. III. Selection index for individual plants on the basis of individual measurements, line means and line-group means in autogamous plants. *Jap. J. Breed.* **6**, 175–9.

