

23

Selection Under Inbreeding

Either inbreeding or selection, never both at the same time — R. A. Fisher

Draft Version 26 May 2013

Why inbreeding and selection? One may have little choice. For many species, such as the autogamous crops that provide much of our food, the extra work required to insure outcrossing is considerable. Second, in many cases, creative use of inbreeding can increase selection response. Finally, many natural populations are highly inbred. Inbreeding redistributes the genetic variance in a population, reducing or removing it within an inbred line and increasing it between a collection of lines (Chapter 10). Inbreeding also (generally) increases the covariance between relatives, as relatives become increasingly more genetically similar under inbreeding (Chapter 10). As we will see, all these actions have important consequences for selection response.

When inbreeding occurs and nonadditive genetic action (dominance and/or epistasis) is present, the standard genetic variance components (σ_A^2 , σ_D^2 , σ_{AA}^2 , etc.) are no longer sufficient to predict response because even the simplest covariances between relatives depend on additional parameters. As discussed in Chapter 10, at least three additional components (ι^* , σ_{DI}^2 , and σ_{ADI}) are required to describe the covariance between inbred relatives when dominance is present (Table 10.1), and hence to specify short-term response. A further complication is inbreeding depression (LW Chapter 10), which changes the mean even in the absence of selection. Unless otherwise mentioned, we assume throughout gametic-phase equilibrium and no epistasis, genotype \times environment interactions or correlations. The complications these introduce for selection response with inbreeding remain largely unexplored.

Our examination of response under inbreeding begins with a general overview of the machinery and concepts of joint inbreeding and selection. This is followed by a discussion of family selection when the parents and/or tested progeny are inbred, extending the results of Chapter 21. These first two sections form the basics of inbreeding and selection response. The remainder of the chapter examines a number of special (but important) cases in more detail, such as selfing and partial selfing. Additional aspects of selection and inbreeding are covered in other chapters, with the interaction and selection and drift examined in Chapter 26 and the generation and selection among pure lines (briefly covered here) examined more extensively in Chapter 30.

BASIC ISSUES IN RESPONSE UNDER INBREEDING

Before we delve into the gory details for particular systems of inbreeding, a number of general comments on the response under inbreeding are in order for several reasons. First, for many readers, the discussion in this section may be sufficient unless/until they need specific details for a particular scheme. Second, there are a number of issues common to selection response any system of inbreeding. With the basic concepts and machinery in hand, one can skip from this section directly to the details for a particular system of interest without having to work

through a number of other system that may be of less interest.

Accounting for Inbreeding Depression

Even in the absence of selection, changes in the population level of inbreeding can cause changes in the mean due to inbreeding depression (LW Chapter 10). Let ΔI denote the change in mean from inbreeding depression. If dominance is the only nonadditive genetic effect, then the change from inbreeding at time t , $\Delta I_t = bf_t$, is a linear function of the inbreeding coefficient. Here b is the difference in character value between a completely inbred ($f = 1$) and outbred ($f = 0$) population. If epistasis is present, $\Delta I = bf + cf^2 + \dots$, the order of polynomial in f depends on the type of epistatic interactions (LW Chapter 10). To distinguish between the change due to inbreeding depression and the additional change due to selection, we decompose the total change in the population mean after t generations as

$$\Delta_\mu(t) = \mu_t - \mu_0 = R(t) + \Delta I_t \quad (23.1a)$$

with a component due to the response from selection $R(t)$ and one due to inbreeding depression ΔI_t . When computing the response to selection, we ignore the change from inbreeding depression, so that

$$R(t) = \Delta_\mu(t) - \Delta I_t \quad (23.1b)$$

If the population is randomly-mated, the inbreeding depression term becomes zero, exposing the true genetic response.

Response Under Small Amounts of Inbreeding

When the amount of inbreeding is small enough that changes in the covariances between relatives are negligible, its main effect is inbreeding depression. Consider a population of modest size undergoing random mating, where the amount of inbreeding generated by genetic drift at generation t is $f_t \simeq t/(2N_e)$, provided $t \ll N_e$ and $f(0) = 0$. If no epistasis is present,

$$\Delta I_t = \frac{bt}{2N_e} \quad (23.2a)$$

giving the expected response with a small amount of inbreeding (Nordskog and Hardiman 1980, Hill 1986) as approximately

$$R(t) \simeq t \cdot \bar{i} \cdot h^2 \cdot \sigma_z - \frac{bt}{2N_e} = t \cdot \bar{i} \cdot \left(h^2 \sigma_z - \frac{b}{2N_e \cdot \bar{i}} \right) \quad (23.2b)$$

This not an unreasonable approximation for small amounts of inbreeding (say $f < 0.05 - 0.1$). For larger amounts (i.e., $f = t/(2N_e) > 0.1$), the genetic variances change from their base population value and this must be taken into account. Chapter 26 examines long-term response in finite populations.

There is a rich literature on maximizing selection response under either constrained levels of inbreeding or under the minimization of inbreeding (Quinton et al. 1992; Quinton and Smith 1995; Grundy et al. 1994, 1998, 2000; Villanueva et al. 1994; Brisbane and Gibson 1995; Luo et al. 1995; Meuwissne 1997; Meuwissne and Sonesson 1998; Meszaros et al. 1999; Sonesson et al. 2000), and this is covered in Volume 3.

Using Ancestral Regressions to Predict Response

The simplicity of Equation 23.2 follows from the assumption that a small amount of inbreeding does not greatly change genetic variances. With larger amounts of inbreeding, variances and covariances can change each generation. Fortunately, with regular systems of inbreeding

the change in genetic variances (and hence the covariances between relatives) in the absence of selection is usually rather predictable. Motivated by this, we make the key assumption throughout the chapter that selection does not substantially alter these variances and covariances from their expected values in the absence of selection. *Provided* this assumption holds and that all regressions are linear and homoscedastic, the method of ancestral regressions (Chapter 15) offers a powerful approach for predicting short-term response.

Recall that under ancestral regression, the cumulative response is expressed as a series of regression coefficients (covariance divided by variance) of the contribution to the current total response from selection in a previous generation t , giving the expected response after T generations of selection and inbreeding as

$$R(T) = \sum_{t=0}^{T-1} S_t \frac{\sigma_G(T, t)}{\sigma^2(z_t)} = \sum_{t=0}^{T-1} \bar{r}_t \frac{\sigma_G(T, t)}{\sigma(z_t)} \quad (23.3)$$

Here $\sigma_G(T, t)$ is the covariance between a relative in generation t and one in the current generation $T \geq t$, while $\sigma^2(z_t)$ is the phenotypic variance of relatives in generation t . Note that this is just the regression of z_T on z_t , which has slope $\sigma_z(T, t)/\sigma^2(z_t) = \sigma_G(T, t)/\sigma^2(z_t)$ in the absence of environmental correlations between generations T and t . Under complicated systems of inbreeding, a number of relatives with different degrees of inbreeding must be simultaneously followed, leading to additional indices in the covariance, such as $\sigma_G(T, \tau, t)$ or $\sigma_G(T, \tau, t, k)$.

Throughout this chapter, the convention is that generation zero is the first generation of inbreeding from a random mating population. Selection in generation zero thus implies that parents from an outbred base population are selected and then inbred. If our first selection is on the first inbred progeny, then selection starts in generation one.

Equation 23.3 is based on the infinitesimal model, since it assumes selection-induced changes in allele frequencies are negligible. While *genotypic* frequencies change due to inbreeding (homozygotes increasing, heterozygotes decreasing), we assume that within the population *allele* frequencies do not significantly change. Hence, if p_i is the frequency of allele A_i in the base population, the frequency of lines eventually fixed for the $A_i A_i$ genotype is assumed to essentially remain p_i , despite selection. Formally, it is changed to $p_i + \epsilon_i$, where ϵ_i is a very small amount. With a very large number of loci all those very small values of ϵ_i can add up to a considerable change in the mean, while still resulting in little change in genetic variances (Chapter 24)

Since the covariance function also gives the genetic variance in generation t , as $\sigma_G^2(t) = \sigma_G(t, t)$, with the covariance function for our particular system of inbreeding in hand, we can immediately write the response as

$$R(T) = \sum_{t=0}^{T-1} S_t \frac{\sigma_G(T, t)}{\sigma_G(t, t) + \sigma_e^2} = \sum_{t=0}^{T-1} \bar{r}_t \frac{\sigma_G(T, t)}{\sqrt{\sigma_G(t, t) + \sigma_e^2}} \quad (23.4)$$

For example, the response after two generations of inbreeding and selection is

$$R(2) = \bar{r}_0 \frac{\sigma_G(2, 0)}{\sigma(z_0)} + \bar{r}_1 \frac{\sigma_G(2, 1)}{\sigma(z_1)} \quad (23.5a)$$

The first term represents the response that carries over to the second generation from selection in generation zero, while the second term is the response to selection from generation one. If we stop selection after two generations, but continue to inbred the population to complete homozygosity, the permanent response (after correcting for any inbreeding depression) is

$$R(2) = \bar{r}_0 \frac{\sigma_G(\infty, 0)}{\sigma(z_0)} + \bar{r}_1 \frac{\sigma_G(\infty, 1)}{\sigma(z_1)} \quad (23.5b)$$

Inspection of Equation 23.5a and 23.5b points out a key feature of response with inbreeding. In most cases, these covariances change, so that it is generally the case that $\sigma_G(i, t) \neq \sigma_G(j, t)$. The relative contribution to response from selection in any particular generation t can thus change over time, so that there is both a transient and permanent component to response. The permanent response for our simple example of selection in the first two generations of inbreeding is given by Equation 23.5b.

The Covariance Between Inbred Relatives

To apply ancestral regressions, we must obtain the covariance between relatives under the particular scheme of inbreeding of interest. These covariances were discussed in detail in Chapter 3, and here we just remind the reader of a few of the key concepts. Recalling Equation 3.14, which gives the genetic covariance between individuals x and y under general inbreeding, but assuming no linkage effects or epistasis, as

$$\begin{aligned} \sigma_G(x, y) = & 2\Theta_{xy}\sigma_A^2 + \Delta_{xy,7}\sigma_D^2 + \Delta_{xy,1}\sigma_{DI}^2 \\ & + (4\Delta_{xy,1} + \Delta_{xy,3} + \Delta_{xy,5})\sigma_{ADI} + (\Delta_{xy,2} - f_x f_y)\iota^* \end{aligned} \quad (23.6a)$$

where

$$2\Theta_{xy} = 2\Delta_{xy,1} + \Delta_{xy,3} + \Delta_{xy,5} + \Delta_{xy,7} + \frac{1}{2}\Delta_{xy,8} \quad (23.6b)$$

For reading ease, we henceforth suppress the xy subscript on Δ for those cases where the two relatives being considered are obvious.

The nine possible Δ_i coefficients of relatedness between two (diploid) individuals are defined in Figure 2 of LW Chapter 7, while the composite genetic parameters (the familiar additive and dominance variances σ_A^2 and σ_D^2 and the less-familiar quadratic components σ_{DI}^2 , ι^* , and the covariance σ_{ADI}) are defined from the standpoint of the non-inbred base population. While σ_{DI}^2 and ι^2 are non-negative (by construction), σ_{ADI} is a covariance and hence can be positive or negative.

Example 23.1. Consider a population with a single locus with genotypic values of $\mathbf{G}_1 \mathbf{G}_1 = 0$, $\mathbf{G}_1 \mathbf{G}_2 = 1.67$, and $\mathbf{G}_2 \mathbf{G}_2 = 2$. What are the quadratic components when $p = \text{freq}(\mathbf{G}_1) = 0.8$? We first find that the standard random-mating parameters are

α_1	α_2	δ_{11}	δ_{22}	σ_A^2	σ_D^2
-0.2804	1.1216	-0.0536	-0.8576	0.628993	0.045967

Since there are only two alleles, $\iota^* = \sigma_D^2$ (Cockerham and Matzinger 1985), and this is confirmed as

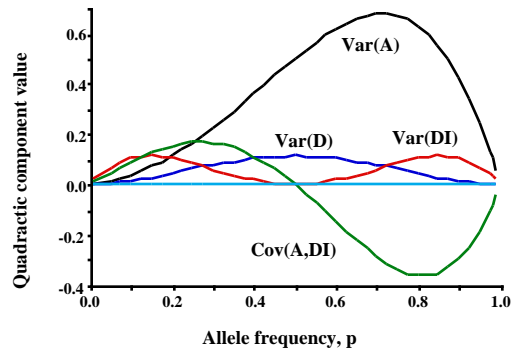
$$\iota^* = (p_1 \delta_{11} + p_2 \delta_{22})^2 = [0.8(-0.0536) + 0.2(-0.8576)]^2 = 0.045967 = \sigma_D^2$$

As for the other two components,

$$\begin{aligned} \sigma_{DI}^2 &= p_1 \delta_{11}^2 + p_2 \delta_{22}^2 - \iota^* \\ &= 0.8(-0.0536)^2 + 0.2(-0.8576)^2 - 0.045967 \\ &= 0.103427 \end{aligned}$$

$$\begin{aligned} \sigma_{ADI} &= 2(p_1 \alpha_1 \delta_{11} + p_2 \alpha_2 \delta_{22}) \\ &= 2[0.8(-0.2804)(-0.0536) + 0.2(1.1216)(-0.8576)] \\ &= -0.360707 \end{aligned}$$

The variance components for other allele frequencies are graphed below.



The nature of the identity by descent (ibd) measures Δ_i provide some insight into which components contribute to the transient, as opposed to the permanent, component to response. If we inbred to complete homozygosity, then both alleles in an individual y in this generation are identical by descent. Only four of the Δ_i measures are this state ($\Delta_1, \Delta_2, \Delta_5,$ and Δ_6). In such cases, σ_{DI}^2 and σ_{ADI}^2 can contribute to the permanent response, while σ_D^2 cannot. Likewise note that if $\Delta_1 = \Delta_3 = \Delta_5$ and $\Delta_2 = f_x f_y$, then σ_A^2 and σ_D^2 are sufficient to describe the covariance between relatives. Finally, as one reads the literature, no consistent notation is found for these additional genetic components required under inbreeding (Table 23.1).

Table 23.1. Some of the alternative notations used for the quadratic genetic components required under inbreeding. Gallis = Goldringer et al. (1996), Cornelius= Cornelius (1975), Van Sanford = Cornelius and Van Sanford (1988), Jinks = Mather and Jinks. Modified from Goldringer et al. (1996).

Our	Gallais	Cornelius	Cockerham	Van Sanford	Jinks
σ_A^2	σ_A^2	σ_A^2	σ_A^2	σ_A^2	$D/2$
σ_D^2	σ_D^2	σ_D^2	σ_D^2	σ_D^2	$H/4$
σ_{ADI}	σ_{ADO}	$C - 2\sigma_A^2$	$2D_1$	$\frac{Q_{yy} - Q_{xx}}{2} - \sigma_A^2$	
σ_{DI}^2	$\sigma_{D_o}^2$	$\sigma_\infty^2 - 2C + 2\sigma_A^2$	D_2^*	Q_{xx}	
l^*	$\sum D o^2$	μ_∞	H^*	$\frac{Q_{yy} + Q_{xx}}{2} - Q_{xy} - \sigma_A^2$	

For full-sib mating, the coefficients required for Equation 23.6 can be obtained by iteration of the Cockerham (1971) transition matrix given by Equation 10.18. These can then be used in conjunction with Equation 23.3 to predict response. Cornelius and Dudley (1975) present numerical tables for these coefficients for the first eight generations for both full-sib and uncle (aunt) – niece (nephew) systems of inbreeding.

Limitations

The major limitation with the ancestral regression approach is the assumption that selection does not significantly alter the covariances between relatives over what is expected under the system of inbreeding (in the absence of selection). Clearly, if there are major alleles, selection will chose individuals carrying these, further increasing the amount of inbreeding in the population. Thus, this general approach is best thought of as a weak selection approximation if there are alleles present of even modest effect. Even in the absence of major alleles, the

effect of selection is to generally make individuals more inbred than expected by the particular system of inbreeding. In such cases, the covariances between relatives are also affected.

FAMILY SELECTION WITH INBREEDING AND RANDOM MATING

A widespread use of inbreeding in selection is in conjunction with random mating in any number of family-based selection schemes. As detailed in Chapter 8, the approach of family-based selection is to use the family mean to provide a better estimate of the breeding value of the parents for the next cycle of selection. This section extends some of the results from Chapter 8 to families with inbred parents and/or sibs. In the terminology introduced in Chapter 8, the selection unit we will use throughout this section is the mean of the sibs, while the recombination unit will either be a sib or the parent of the measured sibs.

Two minor complications commonly arise in family-based selection that account for some of the variety of response equations found in the literature. First, if strict family selection is used, then a measured sib will also be a parent for the next generation. In this case, the covariance between the sib mean \bar{z} and an offspring y starting the next cycle of selection has two components. If z_1 denotes the measured sib used as a parent of y , then with n measured sibs in a family,

$$\sigma(\bar{z}, y) = \frac{1}{n} \sum_{i=1}^n \sigma(z_i, y) = \left(\frac{1}{n}\right) \sigma(z_1, y) + \left(1 - \frac{1}{n}\right) \sigma(z_2, y) \quad (23.7)$$

The first covariance, $\sigma(z_1, y)$, is between parent and offspring, while the second, $\sigma(z_2, y)$, is that between an individual z_2 and the offspring y of its sib z_1 . When sib selection occurs, (such as through the use of remnant seed), the sib used in the recombination is *not* one of the sibs measured for the selection unit, and $\sigma(\bar{z}, y) = \sigma(z, y)$. To simplify our results, we assume only sib or parental selection (progeny testing, where parents are chosen based on the performance of their offspring). For moderate to large family size, the difference between sib and family selection is expected to be very small.

The second issue relates to the variance of the recombination unit, $\sigma^2(\bar{z})$. From ANOVA theory, the variance in observed family means is the between-group variance σ_b^2 plus the error in estimating the true mean μ_i from \bar{z}_i , which is σ_w^2/n . Here $\sigma_w^2 = \sigma_T^2 - \sigma_b^2$ is the within-group variance, the difference between the total and between-group variances. Hence,

$$\begin{aligned} \sigma^2(\bar{z}) &= \sigma_b^2 + \sigma_w^2/n \\ &= \sigma_z(\text{sibs}) + \frac{\sigma_z^2 - \sigma_z(\text{sibs})}{n} \\ &= \left(1 - \frac{1}{n}\right) \sigma_z(\text{sibs}) + \frac{\sigma^2(\bar{z})}{n} \\ &= \left(1 - \frac{1}{n}\right) \left(\sigma_G(\text{sibs}) + \sigma_{E(wf)}^2\right) + \frac{\sigma^2(\bar{z})}{n} \end{aligned} \quad (23.8)$$

where $\sigma_{E(wf)}^2$ is the within-family environmental variance and σ_E^2 the total environmental variance. It is often assumed that n is sufficiently large that the second term can be ignored (as we do in this chapter), leaving the variance of selection units as the covariance between sibs plus and within-family environmental variance. Replication of family members provides a fair amount of control over the error variance. We ignore this error structure here, but it is examined in detail in Chapters 19 and <CHECK >.

Family Selection Using Inbred Parents

One scheme for increasing the response to family selection is to inbred one (or both) parents, and then score the resulting half- or full-sib progeny as the family unit. This has two effects on response, one positive (increasing the covariance between relatives) and one negative (increasing the variance of the selection unit).

The genetic covariance among half-sibs where the common parent is inbred (to amount f) is

$$\sigma_G(HS) = \left(\frac{1+f}{4}\right)\sigma_A^2 + \left(\frac{1+f}{4}\right)^2\sigma_{AA}^2 + \dots + \left(\frac{1+f}{4}\right)^k\sigma_{A^k}^2 \quad (23.9a)$$

For full sibs, if $\bar{f} = (f_1 + f_2)/2$ is the average inbreeding coefficient for the parents, then

$$\sigma_G(FS) = \left(\frac{1+\bar{f}}{2}\right)\sigma_A^2 + \left(\frac{(1+f_1)(1+f_2)}{4}\right)\sigma_D^2 + \left(\frac{1+\bar{f}}{2}\right)^2\sigma_{AA}^2 + \dots \quad (23.9b)$$

This inflation of the between-sib covariances relative to random mating increases the variance of the selection unit. For the reader wondering why the inbreeding components ($\sigma_{ADI}^2, \sigma_{DI}^2$, etc.) do not appear in Equation 23.9, its because the parents of the sibs, while being inbred, are *unrelated*. Hence, alleles within the sibs are not identical by descent, and this implies Δ_1 to Δ_6 are zero. This also applies to the selection unit-offspring covariances (Equation 23.10).

Table 23.2. The response to family selection when parents are inbred. $\sigma_{E(HS)}^2$ and $\sigma_{E(FS)}^2$ are the environmental variances within half- and full-sib families, respectively. Half- vs. full-sibs refer to the family unit being measured, while the parents for the next generation are either remnant seed (sib selection) or the parent of the selection unit itself (progeny testing). For comparison purposes, selection on both parents is assumed. Response is halved if only a single parent has been chosen by family selection. The effects of epistasis are ignored. Additive \times additive epistasis inflates the immediate response, but its contribution decays with recombination. Epistasis inflates the selection unit variance over the values given here, reducing response.

Selection Scheme	$R/(\sigma_A^2 \bar{v})$
Half-sibs, remnant seed	$\frac{(1+f)/4}{\sigma(\bar{z}_{HS,f})} = \frac{(1/2)\sqrt{1+f}}{\sqrt{\sigma_A^2 + 4\sigma_{E(HS)}^2/(1+f)}}$
Half-sibs, Parental	$\frac{(1+f)/2}{\sigma(\bar{z}_{HS,f})} = \frac{\sqrt{1+f}}{\sqrt{\sigma_A^2 + 4\sigma_{E(HS)}^2/(1+f)}}$
Full-sibs, remnant seed	$\frac{(1+f)/2}{\sigma(\bar{z}_{FS,f})} = \frac{\sqrt{(1+f)/2}}{\sqrt{\sigma_A^2 + (1+f)\sigma_D^2/2 + 2\sigma_{E(FS)}^2/(1+f)}}$

Turning to the selection unit-offspring covariances, we will ignore the effects of additive epistasis, as this contributes to the transient, rather than permanent, component of response (as random mating breaks up linkage associations). Here the covariances between

the selection unit and a *single* parent \mathcal{R} of y are

$$\sigma_G(x, y | \mathcal{R} = P \text{ of } x) = \left(\frac{1+f}{4}\right) \sigma_A^2 \quad (23.10a)$$

$$\sigma_G(x, y | \mathcal{R} = HS \text{ of } x) = \left(\frac{1+f}{8}\right) \sigma_A^2 \quad (23.10b)$$

$$\sigma_G(x, y | \mathcal{R} = FS \text{ of } x) = \left(\frac{1+f}{4}\right) \sigma_A^2 \quad (23.10c)$$

with P , HS , and FS implying that the parent \mathcal{R} of y is related to the measured sibs as either a parent, a half-sib, or a full-sib (respectively). Substitution of these results into Equation 21.1 gives the response to a single cycle of selection under various schemes, and these are summarized in Table 23.2.

As Table 23.2 shows, for half-sibs the response under inbreeding is greater than under random mating. This is also generally true for full-sibs, but random mating can give a larger response if the dominance variance is sufficiently large.

Example 23.2. Consider a population with $\sigma_A^2 = 50$, $\sigma_D^2 = 25$, and $\sigma_E^2 = 50$. The response to half-sib, full-sib, and parental selection for various values of parent inbreeding f are given below as a fraction of the response for $f = 0$. The values in the $f = 0$ column are R/\bar{v} , while the values for $f > 0$ represent the ratio of response for that f value relative to the response under random-mating.

Selection	$f=0$	Fraction of random-mating response				
		$f=1/8$	$f=1/4$	$f=1/2$	$f=3/4$	$f=1$
Half-sib						
remanent seed	1.581	1.111	1.220	1.430	1.632	1.826
Parental	3.162	1.111	1.220	1.430	1.632	1.826
Full-sib	2.774	1.093	1.181	1.342	1.485	1.612

Progeny Testing Using Inbred Offspring

Toro (1993) proposes that sire progeny testing be performed using inbred offspring (by crossing the sire to full-sib or half-sib sisters to generate the family), with superior sires then outcrossed (Figure 23.1). This suggestion takes advantage of improved accuracy for testing using inbred sibs while still having an outcrossed population. To quantify the advantage of testing inbred progeny, consider a sire crossed to full-sibs. The selection unit -offspring covariance is that of an outcrossed sib from the sire and an inbred sib from the sire. The probability that the sire allele in the inbred and outcrossed offspring is identical by descent (ibd) is $1/2$. This sire allele can also be transmitted through the dam to the inbred offspring, generating a Δ_3 ibd state (both copies of an allele in x are ibd, and these are also ibd with the sire allele in y). If the sire and dam are full-sibs, the probability of the dam transmitting this allele is $1/4$, while if the sire and dam are half-sibs, the probability is $1/8$. Hence, when the sire and dams are full-sibs (SDFS),

$$\Delta_3 = (1/2)(1/4) = 1/8, \quad \Delta_8 = (1/2)(1 - 1/4) = 3/8 \quad (23.11a)$$

Here state Δ_8 corresponds to a single allele in x and y being ibd. For half-sib sire and dam (SDHS),

$$\Delta_3 = (1/2)(1/8) = 1/16, \quad \Delta_8 = (1/2)(1 - 1/8) = 7/16 \quad (23.11b)$$

Substituting into Equation 23.6 gives the resulting covariance between inbred and outcrossed sibs for full-sib sire and dam as

$$\sigma_G(I, O | SDFS) = (5/16) \sigma_A^2 + (1/8) \sigma_{ADI} \quad (23.12a)$$

while for half-sib sire and dam,

$$\sigma_G(I, O | SDHS) = (9/32) \sigma_A^2 + (1/16) \sigma_{ADI} \quad (23.12b)$$

By comparison, if the dam and sire are unrelated, the covariance is just that between half-sibs, $(1/4) \sigma_A^2$. Thus, in the absence of dominance, the covariance under SDFS is 125% that of half-sibs, and for SDHS the covariance is 112% of half-sibs.

Figure for Progeny testing with inbred test offspring

Figure 23.1. Progeny testing with inbred offspring in test unit

S_1 , S_2 , and $S_{i,j}$ Family Selection

Another scheme for family selection using inbreeding is **S_1 family selection**, where an organism (usually, but not restricted to, a plant) is selfed, and the family mean of the selfed progeny used for selection decisions. Remnant seed for the initial selfing is used for outcrossing and starting the next cycle of selection. This scheme takes two generations — the selfed seed must be grown for scoring families and then in the next generation the S_1 seeds from superior families grown and crossed. Note that S_1 family selection is different from S_1 seed selection, discussed in Chapter 21. While seed selection also use remnant S_1 seeds as the recombination unit, the tested family is a half-sib, rather than an S_1 .

Selection can also be based on S_2 families. Under classical **S_2 family selection**, an individual is selfed to form an S_1 , a single plant of which is then selfed again to form the S_2 family for testing. Remnant seed from the S_1 is used as the recombination unit, with seed from superior families grown and crossed at random to start the next cycle of selection. There is the potential for ambiguity with S_2 selection, as one could use remnant seed from the S_0 instead of the S_1 . Because of this ambiguity, we use a modification of the notation suggested by Wricke and Weber (1986), and consider **$S_{i,j}$ family selection** (Wricke and Weber use $I_{i,j}$). Here, the S_j family is tested, while remnant seed is used from a parent in generation S_i , where $i < j$. Hence, S_1 family selection becomes $S_{0,1}$ selection, classical S_2 is $S_{1,2}$, while the variant of keeping S_0 seed but testing S_2 families is $S_{0,2}$ selection, or **bulk S_2 family selection**.

Expressions for the response to S_1 selection in the literature (e.g., Hallauer and Miranda 1981, Choo and Kannenberg 1981, Bradshaw 1983) are based on the derivation by Empig et al (1972), which gives the covariance between an individual x in the selection unit and the offspring y of its selfed sib as

$$\sigma(x, y) = \sigma_A^2 + \beta_1 \quad \text{where} \quad \beta = \sum 2pq(p - 1/2)d[a + (q - p)d]$$

Similar expressions exist for the response to $S_{1,2}$ selection (Hallauer and Miranda 1981) and for $S_{0,j}$ selection (Wricke and Weber 1986). However, it is fairly easily to obtain a variance-components expression for response to general $S_{i,j}$ family selection. Since a member of the recombination unit \mathcal{R} is outbred, it passes on only single alleles to its offspring. This situation

excludes all of the identity states except for Δ_3 (both alleles in x are ibd, and one is passed on to the offspring y through \mathcal{R}), Δ_8 (the alleles in x are unrelated and one is passed onto y via \mathcal{R}), and Δ_9 (the alleles in x are unrelated to those in y). As a result, Equation 23.6 implies that the selection unit-offspring covariance only depends on σ_A^2 and σ_{ADI} . As shown in Example 23.3 (which can be skipped by the causal reader), these can be obtained by some simple book-keeping, and are

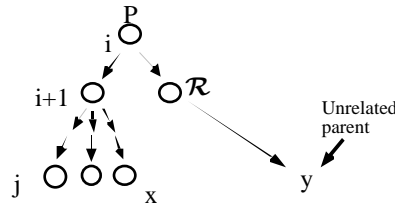
$$\Delta_3 = f_i + (1 - f_i) \left(\frac{1 - 2^{-(j-i)}}{2} \right) = 1 - \frac{1}{2} \left(\frac{1}{2^i} + \frac{1}{2^j} \right) \quad (23.13a)$$

$$\Delta_8 = (1 - f_i) 2^{-(j-i)} = 2^{-j} \quad (23.13b)$$

giving

$$2\Theta_{xy} = \Delta_3 + \Delta_8/2 = 1 - \frac{1}{2^{i+1}} \quad (23.13c)$$

Example 23.3. To compute the probabilities of the ibd states Δ_3 , Δ_8 , and Δ_9 , first recall the various relatives involved. P is the parent that has undergone i generations of selfing and it generates both remant seed \mathcal{R} (that will be grown and crossed with an unrelated individual to form the offspring y) and a collection of S_j families that will be scored.



Consider a random locus. If the alleles at this locus are ibd in P (which occurs with probability f_i), then the only ibd state between x and y is Δ_3 , as both alleles in x are ibd and this allele is also passed onto \mathcal{R} and hence to y . Otherwise (with probability $1 - f_i$) a locus in P is not ibd, and we denote the two alleles as **A** and **a** (these states are only for the purpose of following ibd alleles, and in fact the two alleles can in reality be identical in state). There are then three possible genotypes for x , **AA**, **Aa**, and **aa**, while y can receive either **A** or **a** from \mathcal{R} . The ibd states for these different combinations are as follows:

Genotype of x	allele in y for \mathcal{R}	
	A	a
AA	Δ_3	Δ_9
Aa	Δ_8	Δ_8
aa	Δ_9	Δ_3

Hence, the probability of Δ_8 equals the probability that a locus in P is not ibd ($1 - f_i$) times the probability that the locus is still not ibd by generation j (i.e., it is in state **Aa**). Since the probability a locus is not converted to an ibd state is $1/2$ for each generation of selfing,

$$\Pr(\Delta_8) = (1 - f_i) 2^{-(j-i)}$$

The probability of state Δ_3 is the probability that P is ibd times twice the probability x is genotype **AA** and y gets allele **A**. The factor of two arises because of symmetry, as the case of **aa** and **a** has equal probability. The probability y gets **A** from \mathcal{R} is just $1/2$, while the probability

x is **AA** (given it was not ibd in generation i) is $(1 - 2^{-(j-i)})(1/2)$, the probability of the locus becoming ibd (i.e., not staying as non-ibd) times $(1/2)$ for randomly fixing allele **A**. Hence

$$\begin{aligned} \Pr(\Delta_3) &= f_i + 2[(1 - f_i)(1/2)(1 - 2^{-(j-i)})](1/2) \\ &= f_i + (1 - f_i)(1/2)(1 - 2^{-(j-i)}) \end{aligned}$$

Since both parents of y come from superior families, we double the covariance to give the total (i.e., accounting for both parents of y) selection unit-offspring covariance under $S_{i,j}$ family selection as

$$\begin{aligned} 2\sigma_G(x, y) &= 4\theta_{x,y}\sigma_A^2 + 2\Delta_3\sigma_{ADI} \\ &= 2\sigma_A^2 \left(1 - \frac{1}{2^{i+1}}\right) + 2\sigma_{ADI} \left[1 - \frac{1}{2^{(i+1)}} - \frac{1}{2^{(j+1)}}\right] \end{aligned} \tag{23.14}$$

Numerical values for these coefficients are given in Table 23.3.

Table 23.3. Coefficients for Equation 23.14, the selection unit-offspring covariance under $S_{i,j}$ family selection. The column under σ_A^2 gives the coefficient for the additive variance (which is a function only of i), while the σ_{ADI} coefficient is also a function of j and is given in the remaining columns.

i	σ_A^2	Coefficient on σ_{ADI} for $j =$					
		$i + 1$	$i + 2$	$i + 3$	$i + 4$	$i + 5$	∞
0	1.00	0.50	0.75	0.88	0.94	0.97	1.00
1	1.50	1.25	1.38	1.44	1.47	1.48	1.50
2	1.75	1.63	1.69	1.72	1.73	1.74	1.75
3	1.88	1.81	1.84	1.86	1.87	1.87	1.88
4	1.94	1.91	1.92	1.93	1.93	1.94	1.94
5	1.97	1.95	1.96	1.96	1.97	1.97	1.97
6	1.98	1.98	1.98	1.98	1.98	1.98	1.98
7	1.99	1.99	1.99	1.99	1.99	1.99	1.99
8	2.00	1.99	2.00	2.00	2.00	2.00	2.00

Finally, from (?? CHECK) Example 23.8, the genetic variance between $S_{i,j}$ families is

$$\begin{aligned} \sigma_G^2(S_{i,j}) &= (2 - 2^i)\sigma_A^2 + 2^{-(2j-i)}\sigma_D^2 + (2 - 2^{-i} - 2^{-j})\sigma_{ADI}^2 \\ &+ \left(1 + 2^{-(2j+1-i)} - 2^{-j} - 2^{-(i+1)}\right)\sigma_{DI}^2 + 2^{-(2j-i)}(1 - 2^{-i})\iota^* \end{aligned} \tag{23.15}$$

Thus, the general expression for response to $S_{i,j}$ family selection is

$$R_{S_{i,j}} = \bar{i} \frac{2\sigma_A^2(1 - 2^{-(i+1)}) + 2\sigma_{ADI}(1 - 2^{-(i+1)} - 2^{-(j+1)})}{\sqrt{\sigma_G^2(S_{i,j}) + \sigma_{E(S_{i,j})}^2}} \tag{23.16}$$

In particular, the response to S_1 family selection is

$$R_{S_{0,1}} = \bar{i} \frac{\sigma_A^2 + (1/2)\sigma_{ADI}}{\sqrt{\sigma_A^2 + (1/4)\sigma_D^2 + \sigma_{ADI}/2 + (1/8)\sigma_{DI}^2 + \sigma_{E(S_{0,1})}^2}} \tag{23.17a}$$

The response to “classic” S_2 family selection is

$$R_{S_{1,2}} = \bar{i} \frac{(3/2)\sigma_A^2 + (5/4)\sigma_{ADI}}{\sqrt{(3/2)\sigma_A^2 + (1/8)\sigma_D^2 + (5/4)\sigma_{ADI} + (9/16)\sigma_{DI}^2 + (1/16)\iota^* + \sigma_{E(S_{1,2})}^2}} \quad (23.17b)$$

while the response to bulk S_2 family selection is

$$R_{S_{0,2}} = \bar{i} \frac{\sigma_A^2 + (3/4)\sigma_{ADI}}{\sqrt{\sigma_A^2 + (1/16)\sigma_D^2 + (3/4)\sigma_{ADI} + (9/32)\sigma_{DI}^2 + \sigma_{E(S_{0,2})}^2}} \quad (23.17c)$$

Finally, for a population with two equally frequent alleles (such as occurs with a pure-line cross), we saw in Chapter 10 that $\sigma_{DI}^2 = \sigma_{ADI} = 0$ and $\iota^* = \sigma_D^2$. In this case, the response reduces to

$$R_{S_1} = \bar{i} \frac{2\sigma_A^2(1 - 2^{-(i+1)})}{\sqrt{(2 - 2^i)\sigma_A^2 + 2^{-(2j-i-1)}(1 - 2^{-(i+1)})\sigma_D^2 + \sigma_{E(S_1)}^2}} \quad (23.18)$$

How do the various schemes using selfed families compare with other types of between-family selection? If one assumes that most of the between-family variance is due to non-genetic effects (i.e., the error variance dominates), then an appropriate comparison of different family-based schemes is to examine the response divided by the standard deviation of family mean as (to a first approximation) the error variances will be roughly similar. Further, since different schemes take different number of generations, the scaled response ratio, $R/[\bar{i}\sigma_A^2\sigma(\bar{z})]$ should be expressed in terms of response per generation. We also need to adjust for whether one or both parents have been chosen from superior families. Table 23.4 gives the response per cycle accounting for all these factors under the assumption of no dominance.

Table 23.4. Comparison of different types of family-based selection, under the assumption of no dominance. $R^* = R/[\sigma_A^2 \bar{i} \sigma(\bar{z})]$ is the scaled selection response per cycle per selected parent (using the contribution to the selection unit-offspring covariance from a *single* parent), g is the number of generations per cycle, and c is the number of parents under selection. The response per generation is given by the final column, ctR^*/g .

Type	R^*	g	c	ctR^*/g
S_1	1/2	2	2	1/2
$S_{1,2}$	3/4	3	2	1/2
$S_{0,2}$	1/2	3	2	1/3
Full Sibs	1/4	2	2	1/4
HS, S_1 seed	1/4	2	2	1/4
HS, remant seed	1/8	2	2	1/8
HS, Parent	1/4	2	1	1/8

Table 23.4 shows that S_1 and $S_{1,2}$ selection are superior to other approaches (under the assumption of no dominance and roughly equal family variances). While we have not included comparisons with methods using inbred parents, these are easily obtained by multiplying the scaled response per generation by $1 + f$ (see Table 23.2). While $S_{1,2}$ selection gives a larger response per cycle, this is countered by increased cycle time. Note that S_2 bulk family selection ($S_{0,2}$) is not as efficient as S_1 or $S_{1,2}$. For other types of $S_{i,j}$ selection,

the tradeoff between an increase in additive variance $2(1 - 2^{-(i+1)})$ versus the increase in generation time is such that the scaled response per generation is under 1/2 for $i > 2$ and hence not as efficient as either S_1 or $S_{1,2}$ selection.

Choo and Kannenberg (1979a,b) compared (via computer simulations allowing for dominance) the relative efficiencies of S_1 family, mass, and modified (Lonnquist) ear-to-row selection (Chapter 21). S_1 selection was found to be superior, with the largest advantage occurring at low heritabilities. Similar conclusions were offered by Eberhart (1972). Favorable allele frequencies changed most rapidly under S_1 , but the loss of genetic variance was also fastest under this method. Both of these are likely consequences of the smaller effective population size associated with S_1 selection (which is about 1/3 of a comparable mass or ear-to-row scheme), and indeed Choo and Kannenberg (1979b) observed that this method had the highest loss of favorable alleles, as did Bradshaw (1984), while this was not seen by Wright (1980), who used higher starting allele frequencies. Choo and Kannenberg also noted that linkage can slow down S_1 response, as recombination occurs only every other year (as opposed to every year under mass and ear-to-row selection)

Consistent with these theoretical predictions, several workers have demonstrated that S_1 recurrent selection is better than testcross (half-sib) selection for increased yield in maize (Duclos and Crane 1968, Burton et al. 1971, Carangal et al. 1971, Geneter 1973, Moll and Smith 1981, Tanner and Smith 1987) and Sorghum (Doggett 1972). Likewise, Moll and Smith (1981) reported that S_1 selection for yield in maize gave a roughly 50% greater response than full-sib selection. S_1 lines, however, do show an increased loss of genetic variation (Mulamba et al. 1983, Tanner and Smith 1987).

However, S_1 lines can show greater genotype \times environment interaction (Lonnquist and Lindsay 1964, Wricke 1976, Jan-orn et al. 1976). Caution is thus in order for the general superior of S_1 or $S_{1,2}$ selection over other family-based approaches. The results in Table 23.4 made two major assumptions: no dominance and equal between-family variance. For the later, both between-family genetic differences as well as $G \times E$ and other error terms need be considered. One can easily imagine situations where the difference in error variance more than compensates for the difference in covariances. For example, half-sib selection may generate far more family members for testing than an S_1 , greatly reducing the error variance. Likewise, σ_{ADI} can be negative, reducing the expected advantage of S_1 and $S_{1,2}$ selection. Indeed, Jan-orn et al. (1976) estimated that $(\sigma_A^2 + \beta)/\sigma_A^2 = 1 + \sigma_{ADI}/\sigma_A^2$ was in the range of 0.5-0.55 for many traits in sorghum, suggesting that σ_{ADI} is both negative and substantial.

Other family selection schemes involving inbreeding have been proposed, such as the **selfed half-sib** and **selfed full-sib families** (SHS and SFS) methods of Burton and Carver (1993). Here, progeny from either a half- or full-sib family are selfed and it is the selfed progeny that is used as the family mean for selection decisions. The advantage of this approach is a large increase in the amount of seed (and hence the ability to more greatly replicate a family, reducing the error variance) — if there are M initial sibs, each of which is crossed to obtain N selfed offspring, there are a total of MN offspring per family. Burton and Carver suggest that this approach can be at least as efficient as S_1 family selection, largely due to the decreased variance in the selection unit compared to S_1 families. Another variant is joint **half-sib, S_1 family selection**, proposed by Goulas and Lonnquist (1976) for maize. On prolific (multiple eared) plants, the lower ear is selfed, the upper ear open pollinated. Both the HS and S_1 progenies are jointly evaluated and the best families chosen, using the remnant HS seed. Dhillon (1991b) proposes a scheme of **alternate recurrent selection** of S_1 and half-sib families, involving alternate cycles of S_1 selection and either ear-to-row or half-sib selection. The idea is to take advantage of breeding situations that involve a trail field season and a winter nursery for creating and/or recombining new families. Under the right settings, this approach can exceed the per-generation response of S_1 selection.

Cycles of Inbreeding and Outcrossing

Dickerson (1973) and Dickerson and Lindhé (1977) have suggested that in some cases the response to selection with random mating alternating every other generation with full-sib mating enhances short-term response. Their logic is that a generation (or two) of inbreeding increases the between-group variance and this can be exploited by selection. However, given the extra generations used for inbreeding (instead of selection), the conditions for such a cyclic inbreeding-selection system to give a larger response than mass selection are stringent. Dickerson and Lindhé show that the ratio of response under cyclic inbreeding (R_I) versus mass selection (R_m) is approximately

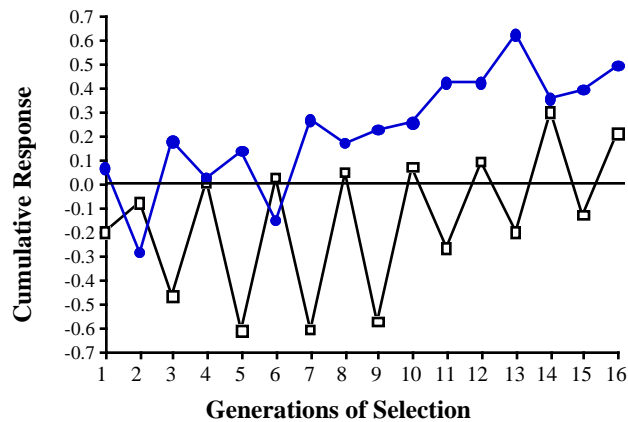
$$\frac{R_I}{R_m} \simeq \left(\frac{\bar{v}_m g_I}{\bar{v}_I g_m} \right) \sqrt{\frac{(1+f)r_f}{h^2}} \quad (23.19)$$

where g is the generation time per cycle (typically $g_m = 1$, $g_I = 2$) and r_f is the genetic correlation among the inbred line members. For example, if one crosses full-sibs and then crosses and selects on inbred families in alternate years, $f = 0.25$, $r_f = 0.6$, and $g_I/g_m = 1/2$, implying

$$\frac{R_I}{R_m} \simeq \left(\frac{\bar{v}_m}{\bar{v}_I} \right) \sqrt{\frac{0.1875}{h^2}}$$

or (under equal selection intensities) that $h^2 < 0.1875$ for cyclic inbreeding to exceed mass selection (Dickerson and Lindhé 1977).

Example 23.4. MacNeil et al. (1984) examined two populations of Japanese Quail (*Coturnix coturnix japonica*) subjected to alternative cycles of full-sib and random mating, selecting on an index of total egg mass divided by female weight. The results for two replicated populations are given below. While both populations ultimately responded to selection, there was significant inbreeding depression, especially in population one (open squares). The cyclic mating scheme produced inbred individuals every odd generation, resulting in population one showing dramatic shifts between positive selection response and significant inbreeding depression countering any positive response. While population two also showed negative response during some of the inbred generations, the effect was not nearly as dramatic.



While random mating did have a slightly higher response than cyclic mating, this was largely due to reduced selection differentials in the cyclically-selected lines. The realized heritabilities were similar.

Given these stringent conditions, it is perhaps not surprising that experimental support for the advantage of cyclic inbreeding is lacking. Dion and Minvielle (1985) used 15 generations of cyclic full-sib / random mating to select for increased pupal weight in *Tribolium castaneum*, finding no differences in the response or realized heritabilities relative to random mating. Similar results were observed in Japanese Quail (Example 23.4). While López-Fanjul and Villaverde (1989) observed that one generation of full-sib mating resulted in a four-fold increase in the realized heritability of egg to pupal viability in *Drosophila melanogaster*, this was more than offset by inbreeding depression.

Another cyclic scheme, **S₁ mass selection**, was proposed by Dhillon (1991a). Here, individuals are crossed and the resulting offspring selfed. The S₁ are then evaluated by individual selection, and superior individuals outcrossed to start the cycle again. The covariance between an S₁ and its outbred offspring is obtained as follows. With probability one, an S₁ individual passes on a single allele to its outbred offspring, so $\Delta_3 + \Delta_8 = 1$. With probability 1/2, the S₁ individual has both alleles ibd at a locus (due to the generation of selfing), giving $\Delta_3 = \Delta_8 = 1/2$. More generally, if k generations of selfing are used before random mating, then $\Delta_3 = f_k = 1 - 2^{-k}$, $\Delta_8 = 1 - f_k = 2^{-k}$, $2\theta = \Delta_3 + \Delta_8/2 = (1/2)(2 - 2^{-k})$, and the S_k - offspring covariance becomes

$$\sigma(S_k, y) = (1/2)(2 - 2^{-k})\sigma_A^2 + (1 - 2^{-k})\sigma_{ADI} \quad (23.20a)$$

Assuming selection on both parents, the response per generation is then

$$R_{S_k} = \left(\frac{\bar{i}}{k+1} \right) \frac{(2 - 2^{-k})\sigma_A^2 + 2(1 - 2^{-k})\sigma_{ADI}}{\sqrt{\sigma_g^2(S_k) + \sigma_E^2(S_k)}} \quad (23.20b)$$

The factor of $1/(k+1)$ arises because there are k generations of selfing for each single generation of selection. The genetic variance $\sigma_g^2(S_k)$ among S_k individuals can be obtained from Equation 23.23, and $\sigma_E^2(S_k)$ is the error variance for single S_k individuals, which is expected to be considerably larger than the error variance for families as no replication is involved. For strict additivity, the ratio of per generation response under S_k mass selection relative to mass selection is

$$\frac{R_{S_k}}{R_M} = \left(\frac{2 - 2^{-k}}{k+1} \right) \sqrt{\frac{\sigma_A^2 + \sigma_E^2(M)}{(2 - 2^{-k})\sigma_A^2 + \sigma_E^2(S_k)}} \quad (23.20c)$$

Note that $R_{S_k} < R_M$ for all values of k (assuming the error variances are roughly equal). Dhillon assumes that a greenhouse can be used for the S₁, giving one cycle per field generation. In such cases, the ratio is roughly $R_{S_k}/R_M = 3/2$ (assuming error variances dominate the term in the square root and are roughly equal). However, a major biological limitation in the assumptions behind obtaining this ratio is that the selected traits must be expressed before reproduction. For traits expressed during or after reproduction, only a single sex has been selected upon (as presumably the S₁ is outcrossed to random individuals). In such cases the response ratio is reduced to $(1/2)(3/2) = 3/4$ of mass selection.

INDIVIDUAL SELECTION UNDER PURE SELFING

Under pure selfing, one starts with a collection of individuals and continually selfs each to form a series of inbred lines. Let S_k denote such a line after k generations of selfing, with the

S_0 being the collection of individuals that are initially selfed to start the lines, and the S_∞ the completely inbred lines. A variety of options exist for generating the initial collection of lines. The simplest is to use a random sample of individuals from an outbred population. Another common situation is the **pure line cross**, where one crosses two completely inbred (pure) lines, and continually selfs starting with the F_1 . In this case, the initial cross produces a number of F_1 individuals, and even though these are selfed to create a series of F_2 lines, the first generation of selfing is formally defined as the F_2 . The reason is that all the F_1 s in this case are genetically identical, being heterozygous at every loci at which the two lines differ. Such a population of only heterozygotes is not in Hardy-Weinberg equilibrium, but the F_2 are (for diploid autosomal loci). Hence, it is the F_2 that we take as the base for starting to count generations of selfing, so that $S_0 = F_2, S_1 = F_3$, etc. If loci are unlinked, then linkage disequilibrium (which is maximal in the F_1) is zero in the F_2 's. If loci are linked, it may take several rounds of random mating to mitigate the effects of the F_1 disequilibrium on tightly-linked loci.

Several other line-cross situations may also form the foundation population from which individuals are drawn for selfing. If one intermates a collection of lines, the first generation will also not be in Hardy-Weinberg equilibrium unless allele frequencies are the same in each line. However, for diploid autosomal loci Hardy-Weinberg is reached with an additional generation of random mating (sex-linked loci and polyploids take several generations, see LW Chapter 4). Linkage disequilibrium is also created in such a cross, due to differences in the gamete frequencies across lines. Unlike the case for crossing two pure lines, the F_2 from a multiple-line cross is not necessarily in linkage equilibrium, even for unlinked loci. In this case, the disequilibrium decays as $(1/2)^r$ where r is the number of generations that the F_1 s are randomly mated. For linked loci, the decay is $(1 - c)^r$ where c is the recombination frequency. The instant achievement of linkage equilibrium in the F_2 from a pure line cross arises because all F_1 individuals are genetically identical and heterozygous at all segregating loci, which is not the situation for crosses of three (or more) lines. Other common types of line crosses involving pure (or otherwise) lines are **three-way hybrids**, $(L_1 \times L_2) \times L_3$ (the F_1 from an $L_1 \times L_2$ crossed to L_3), and **double-crosses** (or **four-way hybrids**) $(L_1 \times L_2) \times (L_3 \times L_4)$ which commonly arise in maize breeding. Again, it is often advisable to instead take such crosses through at least one additional round of random mating to achieve Hardy-Weinberg (so that our expressions for response are valid) and approach linkage equilibrium before starting inbreeding.

Response Under Pure Selfing

Suppose we indeed have a collection of individuals that are in Hardy-Weinberg and linkage equilibrium and we commence selfing. After all lines have become completely inbred, there is no response to selection *within* a line as there is no within-line genetic variation (in the absence of mutation). However, the response *between* lines involves the entire genotypic variance as selection is essentially between clones.

We first consider one extreme, inbreeding each line entirely to fixation and then selecting among the lines. At this point, a parent and its (selfed) offspring are genetically identical, and the resulting parent-offspring covariance equals the total genetic variance in the population. Inbreeding alters the total genetic variation from its random-mating value of $\sigma_A^2 + \sigma_D^2$ to a new value $\tilde{\sigma}_G^2$ over the entire collection of pure lines. The resulting parent-offspring covariance among these lines is

$$\sigma(z_p, z_o) = \tilde{\sigma}_G^2 = 2\sigma_A^2 + 2\sigma_{ADI} + \sigma_{DI}^2 \quad (23.21)$$

When k th order additive epistasis is present, $2^k \sigma_{A^k}^2$ is added (e.g., $4\sigma_{AA}^2, 8\sigma_{AAA}^2$, etc.). Assuming linearity, the response to a generation of selection among these inbred lines produces

an expected response of

$$R = S \frac{\tilde{\sigma}_G^2}{\tilde{\sigma}_G^2 + \sigma_e^2} \tag{23.22}$$

Even if selection is moderate, a single generation is likely to significantly alter the distribution of remaining genotypes (and hence change the genetic variance), and thus the validity of Equation 23.22 over more than a few generations is very doubtful. There are a number of subtleties with trying to select the best pure line from a collection, which are examined in Chapter 30.

Instead of waiting for inbreeding to be complete, suppose that we select among individuals while inbreeding is still occurring. The response in generation T from selection in generation t is then a function of the cross-generation covariance $\sigma_G(T, t)$. For strict selfing, Equation 10.16 gives the covariance between a relative and generation T and its ancestor in generation $t < T$, as

$$\sigma_G(T, t) = (1 + f_t)\sigma_A^2 + (1 - f_T)(\sigma_D^2 + f_t t^*) + \frac{f_T + 3f_t}{2} \sigma_{ADI} + f_t \sigma_{DI}^2 \tag{23.23}$$

where $f_t = 1 - (\frac{1}{2})^t$ is the amount of inbreeding in generation t . The phenotypic variance in generation t is $\sigma^2(z_t) = \sigma_G(t, t) + \sigma_e^2$, where $\sigma_G(t, t)$ is obtained from Equation 23.23 by setting $T = t$. Equation 23.21 also follows, as $f_\infty = 1$. Recall (LW Chapter 6) that in some cases, the environmental variance σ_e^2 may increase with inbreeding, and thus we may need to account for this as well. Cockerham and Matzinger (1985) extend Equation 23.23 to include additive by additive epistasis (but still assuming gametic-phase equilibrium). If additive epistasis up to order k is present, extra terms are added to the covariance given by Equation 23.23,

$$(1 + f_t)^2 \sigma_{AA}^2 + \dots + (1 + f_t)^k \sigma_{A^k}^2$$

When all possible types of pairwise epistasis (e.g., $A \times A, A \times D, D \times D$) occur, 12 variance components are required to describe $\sigma_G(T, t)$ under selfing (Wright 1987, 1988), but we will ignore this level of complication.

Substitution of Equation 23.23 into Equation 23.4 gives the response to selection while the line is being inbred. For complete additivity, $\sigma_G(T, t) = (1 + f_t)\sigma_A^2 = (2 - 2^{-t})\sigma_A^2$, giving the response as

$$R(T) = \sum_{t=0}^{T-1} S_t \frac{(2 - 2^{-t})\sigma_A^2}{(2 - 2^{-t})\sigma_A^2 + \sigma_e^2} \tag{23.24}$$

as obtained by Brim and Cockerham (1961) and under much more general conditions by Pederson (1969a). If dominance is present, the selection response under selfing has both a transient and a permanent component. When selection is relaxed, the mean potentially changes as the transient component decays. The expected total change in the mean after n generations, the first T of which were under selection (generations 0 to $T - 1$), is given by

$$R(n | T) = \sum_{t=0}^{T-1} S_t \frac{\sigma_G(n, t)}{\sigma_G(t, t) + \sigma_e^2} = \sum_{t=0}^{T-1} \bar{i}_t \frac{\sigma_G(n, t)}{\sqrt{\sigma_G(t, t) + \sigma_e^2}} \tag{23.25}$$

The permanent response to T generations of selection, $\tilde{R}(T)$, is given by

$$\tilde{R}(T) = R(\infty | T) = \sum_{t=0}^{T-1} S_t \frac{\sigma_G(\infty, t)}{\sigma_G(t, t) + \sigma_e^2} = \sum_{t=0}^{T-1} \bar{i}_t \frac{\sigma_G(\infty, t)}{\sqrt{\sigma_G(t, t) + \sigma_e^2}} \tag{23.26}$$

Since $f_\infty = 1$, Equation 23.23 (extended for additive \times additive epistasis) gives the covariance between an individual in generation t and a completely inbred (F_∞) line descended from it as

$$\sigma_G(\infty, t) = \left(2 - \frac{1}{2^t}\right) \sigma_A^2 + \left(2 - \frac{3}{2^{t+1}}\right) \sigma_{ADI} + \left(1 - \frac{1}{2^t}\right) \sigma_{DI}^2 \quad (23.27)$$

which is essentially $\tilde{\sigma}_G^2$ for $t > 5$. Additive variance contributes to the permanent response, while σ_D^2 and t^* contribute to the transient, but not the permanent, response. However, dominance does make a contribution to the permanent response through σ_{DI}^2 and σ_{ADI}^2 . To see why, consider the case when inbreeding is complete. Here the only genotypes are of the form $\mathbf{A}_i\mathbf{A}_i$ and have genotypic decomposition $2\alpha_i + \delta_{ii}$. The frequency of such genotypes (in the collection of completely inbred lines) is p_i , assuming no change in the population allele frequencies. The resulting genetic variance between lines is thus

$$\tilde{\sigma}_G^2 = \sigma^2(2\alpha_i + \delta_{ii}) = 4\sigma^2(\alpha_i) + 2 \times 2\sigma(\alpha_i, \delta_{ii}) + \sigma^2(\delta_{ii}) = 2\sigma_A^2 + 2\sigma_{AD} + \sigma_{DI}^2$$

The contribution from standard dominance variance, $\sigma_D^2 = \sigma^2(\delta_{ij})$, decays as $\mathbf{A}_i\mathbf{A}_j$ heterozygotes are lost due to inbreeding.

Example 23.5. Using the genetic parameter values from Example 23.1, what fraction of the response from various generations of selection is ultimately passed on to the completely inbred line? We assume $\sigma_e^2 = 1$ (which corresponds to a random-mating heritability of $h^2 = \sigma_A^2 / (\sigma_A^2 + \sigma_D^2 + \sigma_e^2) = 0.38$). For selection in various generations (t), the fifth column gives the expected single-generation response, the fourth column gives the eventual contribution, while the final column gives the percentage of the response from a particular generation that is translated into the final response.

t	$\sigma_G(\infty, t)$	$\sigma_G^2(t, t)$	$\frac{\sigma_G(\infty, t)}{\sigma_G^2(t, t)}$	$\frac{\sigma_G(t+1, t)}{\sigma_G^2(t, t)}$	%
0	0.449	0.675	0.268	0.363	73.82
1	0.544	0.669	0.326	0.377	86.45
2	0.592	0.657	0.357	0.384	93.07
3	0.616	0.649	0.374	0.387	96.49
4	0.628	0.645	0.382	0.389	98.23
5	0.634	0.642	0.386	0.389	99.11
7	0.639	0.641	0.389	0.390	99.78

The largest reduction came in the first generation of selection, where selected individuals are selfed for the first time. As selfing progressed, the genotypic values between ancestors and descendants become much more similar, and hence they retain almost all of their initial response. While more dramatic changes are expected under some particular values of σ_{ADI} , σ_{DI}^2 , and t^* , after the first six generations, inbreeding is largely completely, with offspring being almost genetically identical to their parents.

It is important to again stress that these results for expected response are based on infinitesimal model approximations. Clearly, selection with a small number of loci can change allele frequencies, violating the assumptions leading to Equations 23.24-23.27. Likewise, with

a small number of lines and/or strong selection, these results are also biased. Another, more subtle, violation of this basic model occurs if some lines are disproportionately chosen over other (as one might expect). In such cases, the covariances that are now appropriate are not for the population as a whole, but rather those for *within* particular sublines. The unstated assumption of Equation 23.23 is that when individuals are being compared for selection, they most recent ancestors are those drawn from the base population. If their most recent ancestor is more current, then the covariances are incorrect, and the estimated response is biased.

Response When Inbreeding Pure Line Crosses

Considerable simplification occurs when two pure lines are crossed. In this case, each locus has only two alleles segregating (one from each line) each with frequency 1/2, and as a result $t^* = \sigma_D^2$ and $\sigma_{DI}^2 = \sigma_{ADI} = 0$. Equation 23.23 reduces to

$$\sigma_G(T, t) = c_t \sigma_A^2 + 2^{-T} c_t \sigma_D^2 + c_t^2 \sigma_{AA}^2 + \dots + c_t^k \sigma_{A^k}^2, \quad \text{where } c_t = 2 - \frac{1}{2^t} \quad (23.28)$$

Starting selection on the F_2 's and denoting this as generation 0, Equation 23.25 simplifies (Pederson 1969a) to

$$R(n|T) = \sum_{t=0}^{T-1} \bar{i}_t \frac{(2 - 2^{-t})(\sigma_A^2 + 2^{-n} \sigma_D^2)}{\sqrt{(2 - 2^{-t})(\sigma_A^2 + 2^{-t} \sigma_D^2) + \sigma_e^2}} \quad (23.29)$$

Example 23.6. Suppose a cross between two inbred lines is subjected to truncation selection for the uppermost 20% in the first two generations of selfing. What is the cumulative response to selection at the n th generation of selfing? Here $\bar{i} = 1.402$ and we assume the (random mating) variance components of $\sigma_A^2 = 50$, $\sigma_D^2 = 25$, and $\sigma_E^2 = 50$ (giving a random-mating heritability of 0.4). The total phenotypic variance in the first two generations is $\sigma_z^2(0) = \sigma_G^2(0) + \sigma_e = 125$ and $\sigma_z^2(1) = \sigma_G^2(1) + \sigma_e = 93.75$. Letting $r(n, t)$ denote the response to selection in generation t that is present at generation n and $R(n)$ denote the cumulative response from selection present at generation n , we find that

n	t	$\sigma_G(n, t)$	$r(n, t)$	$R(n)$
1	0	62.5	7.837	7.837
2	0	56.3	7.054	
2	1	84.4	9.866	16.920
3	0	53.1	6.662	
3	1	79.7	9.318	15.980
4	0	51.6	6.466	
4	1	77.3	9.044	15.510
5	0	50.8	6.368	
5	1	76.2	8.907	15.275
10	0	50.0	6.273	
10	1	75.0	8.774	15.047
∞	0	50	6.270	
∞	1	75	8.770	15.040

The Bulmer Effect Under Selfing

The effect of linkage and number of loci has been examined for completely additive models (no dominance or epistasis) in small-scale simulation studies by Bliss and Gates (1968) and Stam (1977). Linkage reduces the rate of response, while (for fixed σ_A^2) the per generation response increases as the number of loci decreases.

The first theoretical investigation of the magnitude of Bulmer effect (reduction in the genetic variance from selection-induced negative gametic disequilibrium) during selfing was by Cornish (1990a, b). In Cornish's model, a single generation of selection occurs in the F_2 and the effect on the final (F_∞) lines was examined. It was found that the genetic variance in the offspring of the selected parents is reduced (relative to that under pure inbreeding) by $h^2 \bar{i} (\bar{i} - z) \sigma_A^2$. This is very reminiscent of the reduction in additive variance in an outbred population due to selection under the infinitesimal model (Chapters 5, 15), with \bar{i} being the selection intensity and z the corresponding unit normal for the fraction selected (see Chapter 5 for details). Truncation selection on a normal distribution of phenotypes reduces the phenotypic variance of selected parents by $\bar{i} (\bar{i} - z) \sigma_z^2$, and only a fraction of this is passed onto the offspring. In a random mating population, the reduction in variance rapidly approaches an equilibrium value (depending on the strength of selection \bar{i} and the heritability h^2). In a selfing population, since there is no random mating to recover variation, the reduction in variance over that under pure inbreeding is permanent. By constant, upon relaxation of selection under random mating, the variance eventually returns to its pre-selection value under the infinitesimal model. Thus, while inbreeding in a random mating population reduces the rate of response, with selection in a selfing population, it not only reduces the rate, but also the final selection limit.

A far more detailed investigation of the Bulmer effect under the infinitesimal model was given by Hayashi and Ukai (1994). The authors assumed that the effect of any given locus scaled as $n^{-1/2}$ (standard for an infinitesimal model, see Chapter 15). Using this assumption, they produce recursion equations for the changes in variance and covariance for a pure line cross. They assumed truncation selection starts F_2 generation and remains constant for t generations.

If only additive variance is present, Hayashi and Ukai find that

$$\sigma_A^2(t+1) = \sigma_{A^o}^2(t+1) - \sigma_{A^o}^2(t) + \left(1 - \bar{i}(\bar{i} - z) \frac{2\sigma_A^2(t)}{2\sigma_A^2(t) + \sigma_e^2}\right) \sigma_A^2(t) \quad (23.30)$$

where $\sigma_{A^o}^2(t) = (2 - 2^t) \sigma_A^2$ is the additive variance in the population of lines under strict selfing and t is the generation of selfing. Here $\sigma_{A^o}^2(t+1) - \sigma_{A^o}^2(t)$ can be regarded as the within-family variance, which is unaffected by selection. The remaining component in Equation 23.30 represent the change in the between-family variance (the variance in progeny means), which is reduced by selection.

If both additive and dominance effects are present, they have correlated changes and the recursion equation is a little more complex. Letting $\sigma_G^o(T, t)$ denote the cross-generational covariance under pure selfing, Hayashi and Ukai show that

$$\sigma_G(T, t) = \sigma_G^o(T, t) - \bar{i}(\bar{i} - z) \sum_{k=0}^{t-1} \frac{\sigma_G(t, k) \sigma_G(T, k)}{\sigma_G(k, k) + \sigma_e^2} \quad (23.31a)$$

where

$$\sigma_G^o(T, t) = (2 - 2^t) (\sigma_A^2 + 2^{-T} \sigma_D^2) \quad (23.31b)$$

This equation is solved by iteration, starting with

$$\sigma_G(T, 0) = \sigma_G^o(T, 0) = \sigma_A^2 + 2^{-T} \sigma_D^2 \quad (23.31c)$$

Example 23.7. As an application of the Hayashi-Ukai variance correction, let us reconsider Example 23.6. Here $z = 0.84$ (Example 2 in Chapter 5), so that $\bar{i}(\bar{i} - z) = 0.788$. From Equation 23.31c, the genetic covariance between a selected F_2 individual (generation $t = 0$) and a relative after T generations of selfing

$$\sigma_G(T, 0) = \sigma_G^o(T, 0) = \sigma_A^2 + 2^{-T} \sigma_D^2 = 50 + 25/2^T$$

Note that this covariance is the same as with pure-selfing. The covariance between an individual selected in the next generation and its relative in generation T of selfing does, however, show a reduction, with

$$\sigma_G(T, 1) = \sigma_G^o(T, 1) - \bar{i}(\bar{i} - z) \frac{\sigma_G(1, 0) \sigma_G(T, 0)}{\sigma_G(0, 0) + \sigma_E^2}$$

The first term is the pure-selfing covariance, the second the reduction due to selection. To obtain the value of the latter, first note that $\sigma_G(1, 0) = 50 + 25/2 = 62.5$, while $\sigma_G(0, 0) + \sigma_E^2 = 125$ and $\sigma_G^o(T, 1) = (3/2)(50 + 25/2^T)$, giving

$$\begin{aligned} \sigma_G(T, 1) &= (3/2) (50 + 25/2^T) - 0.788 \frac{62.5 \cdot (50 + 25/2^T)}{125} \\ &= (50 + 25/2^T) \left(3/2 - 0.788 \frac{62.5}{125} \right) \end{aligned}$$

Since the first quantity is proportional to $\sigma_G^o(T, 1)$, the ratio $\sigma_G(T, 1)/\sigma_G^o(T, 1) = 0.737$ is a constant independent of T . Similarly, the genetic variance in generation one becomes

$$\begin{aligned} \sigma_G(1, 1) &= \sigma_G^o(1, 1) - \bar{i}(\bar{i} - z) \frac{\sigma_G(1, 0) \sigma_G(1, 0)}{\sigma_G(0, 0) + \sigma_E^2} \\ &= (3/2) (50 + 25/2) - 0.788 \frac{(62.5)^2}{125} = 69.125 \end{aligned}$$

Again, the first quantity is the pure-inbreeding value, the second the correction for linkage disequilibrium. In this case, the correct genetic variance is only 0.737 of the pure-inbreeding value. Since neither the genetic variance or covariance for generation 0 are effected by selection ($\sigma_G(0, 0) = \sigma_G^o(0)$, and $\sigma_G(T, 0) = \sigma_G^o(T, 0)$), the response to selection in generation 0 is unaffected by linkage disequilibrium (this is also the case for a random mating population, see Chapter 5). The response in generation two, however, is effected by selection. Since both the covariance and genetic variance are reduced by the same fraction (0.737) relative to strict inbreeding, the ratio of response at generation T from selection in generation one to its predicted value ignoring disequilibrium is

$$\frac{r(T, 1)}{r^0(T, 1)} = \left(\frac{\sigma_G(T, 1)}{\sigma_G^o(T, 1)} \right) \left(\frac{\sqrt{\sigma_G^o(1, 1)}}{\sqrt{\sigma_G(1, 1)}} \right) = \frac{0.737}{\sqrt{0.737}} = 0.859$$

The presence of gametic disequilibrium thus reduces response by 14%.

FAMILY SELECTION UNDER PURE SELFING

Predicting response to family selection — using the selfed offspring (perhaps several generations worth) to choose lines — requires first a consideration of the hierarchical structure among the selfed lines in a population (Figure 23.2). The collection of lines descended from a parent at time τ (which we can think of as this individual's extended family) are expected to show less within-line variation than a collection of lines from an early ancestor of τ at time $t < \tau$. With family selection, our goal is to predict the response given that we select individuals from generation t on the basis of the performance of their offspring in generation $\tau > t$. We may then wish to know what fraction of this response is around at some future generation $T > \tau$. For example, we may select the best lines in generation t based on the performance of their selfed offspring, using remnant seed from the selected parents to form the next generation. In this case, $\tau = t + 1$ (see Figure 23.3). If individual plants do not produce sufficient seed for family testing, two generations of selfing may be used to generate sufficiently large family, in which case $\tau = t + 2$.

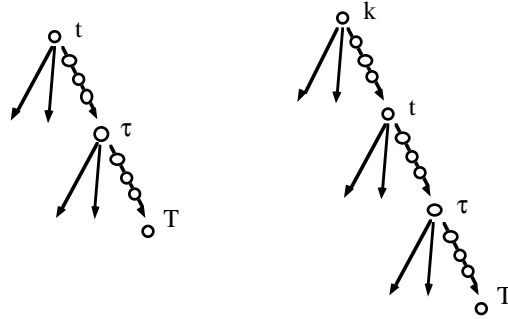


Figure 23.2. The hierarchical structuring of selfed populations. **A:** (Left) Often we select using a parent in generation t by scoring its offspring in generation τ , and we wish to know the covariance between τ and some future generation T , given that both share that common parent in generation t . Here, the individual at τ is a **subline** of t , while T is a subline of τ , and hence a **sub-subline** of t . **B** (Right): Another level of hierarchical structuring of selfed populations. When selecting within a substructure of the selfing pedigree, we may be interested in the response using parents in generation t whose offspring are scored in generation τ and the response is across those families in the pedigree sharing the deeper common parent k .

The Covariance Between Relatives in a Structured Selfing Population

Given the need to account for the structure in a selfing population, Cockerham (1983) and Cockerham and Martzinger (1985), building on concepts from Horner (1952) and Gates et al. (1957), generalized the covariance given by Equation 23.23 to $\sigma_G(T, \tau, t)$, the covariance between a relative in generation T and another relative in generation $\tau \leq T$ when the last common relative is in generation $t \leq \tau$. This is given by

$$\begin{aligned} \sigma_G(T, \tau, t) &= (1 + f_t)\sigma_A^2 + \left(\frac{(1 - f_T)(1 - f_\tau)}{1 - f_t} \right) \sigma_D^2 + \left(f_t + \frac{f_T + f_\tau}{2} \right) \sigma_{ADI} \\ &+ \left(f_t + \frac{(f_T - f_t)(f_\tau - f_t)}{2(1 - f_t)} \right) \sigma_{DI}^2 + \left(\frac{f_t(1 - f_T)(1 - f_\tau)}{1 - f_t} \right) \iota^* \\ &+ (1 + f_t)^2 \sigma_{AA}^2 + (1 + f_t)^3 \sigma_{AAA}^2 + \cdots + (1 + f_t)^k \sigma_{A^k}^2 \end{aligned} \quad (23.32)$$

Notice that Equation 23.32 reduces to Equation 23.23 when $\tau = t$ (parents are the selection unit), as $\sigma_G(T, t, t) = \sigma_G(T, t)$. The epistatic terms are often ignored, and the model does

not account for non-additive epistatic terms and also assumes linkage equilibrium. For cross-generational covariances indexed by two or more relatives, such as $\sigma_G(T, t)$ and $\sigma_G(T, \tau, t)$, we use the indexing convention that the right-most index (t in this cases) references the oldest (earliest generation) individual, while the leftmost (T in this case) references the youngest (latest generation). Thus as one proceeds right-to-left in the index, more recent relatives are being considered.

For the special case of a pure line cross, Equation 23.32 reduces considerably to

$$\begin{aligned} \sigma_G(T, \tau, t) &= (1 + f_t)\sigma_A^2 + \left(\frac{(1 + f_t)(1 - f_T)(1 - f_\tau)}{1 - f_t} \right) \sigma_D^2 + (1 + f_t)^2 \sigma_{AA}^2 + \dots \\ &= \left(2 - \frac{1}{2^t} \right) \left(\sigma_A^2 + \frac{\sigma_D^2}{2^{T+\tau-t}} + \left(2 - \frac{1}{2^t} \right) \sigma_{AA}^2 + \dots \right) \end{aligned} \tag{23.33}$$

The permanent selection response is given by the covariance between a completely inbred F_∞ line ($T = \infty$) and a relative (for example, from the selection unit) in generation τ if they last both shared a relative in generation t (as would occur if remnant seed from t is used to form the new lines). Here, Equation 23.32 reduces to

$$\begin{aligned} \sigma_G(\infty, \tau, t) &= (1 + f_t)\sigma_A^2 + \frac{1 + 2f_t + f_\tau}{2} \sigma_{ADI} + \frac{f_t + f_\tau}{2} \sigma_{DI}^2 \\ &\quad + (1 + f_t)^2 \sigma_{AA}^2 + (1 + f_t)^3 \sigma_{AAA}^2 + \dots + (1 + f_t)^k \sigma_{A^k}^2 \end{aligned} \tag{23.34}$$

Similarly, the covariance between a parent in generation t and an offspring in generation T follows by noting that here $t = \tau$, and Equation 23.32 reduces to Equation 23.23.

Some of the most useful cases of Equation 23.32 relate to genetic variances. For the sake of a clearer exposition, we will ignore additive epistasis in what follows (although its inclusion is trivial). First, the total genetic variance in generation T is given by

$$\sigma_G(T, T, T) = (1 + f_T)\sigma_A^2 + (1 - f_T)\sigma_D^2 + 2f_T\sigma_{ADI} + f_T\sigma_{DI}^2 + f_T(1 - f_T)\iota^* \tag{23.35}$$

This is the genetic variance across the entire population (across *all* the lines present in generation T). We also require the genetic variance in generation T among the subset of lines that descend from a single individual in generation t . Here, $\tau = T$ and the variance becomes

$$\begin{aligned} \sigma_G(T, T, t) &= (1 + f_t)\sigma_A^2 + \frac{(1 - f_T)^2}{1 - f_t} \sigma_D^2 + (f_t + f_T) \sigma_{ADI} \\ &\quad + \left(f_t + \frac{(f_T - f_t)^2}{2(1 - f_t)} \right) \sigma_{DI}^2 + \frac{f_t(1 - f_T)^2}{1 - f_t} \iota^* \end{aligned} \tag{23.36}$$

An example of this would be the genetic variance across the collection of F_3 or F_4 **bulk families** from a single F_2 parent. For an F_3 family this is $\sigma_G(1, 1, 0)$, as the F_2 represents generation zero of selfing, while across the entire collection of F_4 families that trace back to this F_2 individual has variance $\sigma_G(2, 2, 0)$.

Example 23.8. What is the between-family genetic variance for $S_{i,j}$ families, the bulk collection of S_j families from a single individual in generation i ? From Equation 23.36, $\sigma_G^2(S_{i,j}) = \sigma_G(j, j, i)$. Substitution of $f_t = 1 - 2^{-t}$ and simplifying gives

$$\sigma_G^2(j, i) = (2 - 2^i)\sigma_A^2 + 2^{-(2j-i)}\sigma_D^2 + (2 - 2^{-i} - 2^{-j})\sigma_{ADI}^2$$

$$+ \left(1 + 2^{-(2j+1-i)} - 2^{-j} - 2^{-(i+1)}\right) \sigma_{DI}^2 + 2^{-(2j-i)} (1 - 2^{-i}) t^*$$

Some particular values of interest are

$$\begin{aligned}\sigma_G(S_{0,1}) &= \sigma_A^2 + (1/4)\sigma_D^2 + \sigma_{ADI}/2 + (1/8)\sigma_{DI}^2 \\ \sigma_G(S_{0,2}) &= \sigma_A^2 + (1/16)\sigma_D^2 + (3/4)\sigma_{ADI} + (9/32)\sigma_{DI}^2 \\ \sigma_G(S_{1,2}) &= (3/2)\sigma_A^2 + (1/8)\sigma_D^2 + (5/4)\sigma_{ADI} + (9/16)\sigma_{DI}^2 + (1/16)t^*\end{aligned}$$

Finally, it will prove useful to decompose the total genetic covariance $\sigma_G(T, t)$ into a within- and between-family covariance, $\sigma_{Gw}(T, t)$ and $\sigma_{Gb}(T, t)$, where

$$\sigma_G(T, t) = \sigma_{Gw}(T, t) + \sigma_{Gb}(T, t) \quad (23.37a)$$

The between family covariance in generation t is the covariance between sibs from a parent in generation $t - 1$,

$$\sigma_{Gb}(T, t) = \sigma_G(T, t, t - 1) \quad (23.37b)$$

The within-family genetic covariance follows as

$$\begin{aligned}\sigma_{Gw}(T, t) &= \sigma_G(T, t) - \sigma_{Gb}(T, t) \\ &= \sigma_G(T, t, t) - \sigma_G(T, t, t - 1)\end{aligned} \quad (23.37c)$$

For more general families, $t - 1$ is replaced by $t - j$ when the last common ancestor to the family was j generations before the collection of families is scored. Note that the within- and between-family genetic variances in generation t are given by $\sigma_{Gw}(t, t)$ and $\sigma_{Gb}(t, t)$, respectively. Recalling Equation 23.8, this implies a phenotypic variance for the between-family means of

$$\begin{aligned}\sigma_z(t, t) &= \sigma_{Gb}(t, t) + E_w + \frac{\sigma_{Gw}(t, t) + E - E_w}{n} \\ &= \left(1 - \frac{1}{n}\right) (\sigma_G(t, t, t - 1) + E_w) + \frac{\sigma_G(t, t, t) + E}{n}\end{aligned} \quad (23.37d)$$

when n sibs are examined.

Example 23.9. Consider the within and between family genetic variances for an $S_{j-1,j}$ family (the offspring from a single S_{j-1} individual). Here $T = \tau = j$ and $t = j - 1$, and from Equation 23.37b, the between-family genetic variance is

$$\sigma_{Gb}(j, j) = \sigma_G(j, j, j - 1)$$

For a pure line cross, Equation 23.33 (ignoring epistasis) gives

$$\sigma_{Gb}(j, j) = \sigma_G(j, j, j - 1) = \left(2 - \frac{1}{2^{j-1}}\right) \left(\sigma_A^2 + \frac{\sigma_D^2}{2^{j+1}}\right)$$

The within-family variance is, from Equation 23.37c,

$$\sigma_{Gw}^2(j, j) = \sigma_G(j, j, j) - \sigma_G(j, j, j - 1)$$

For a pure-line cross, the genetic variation (ignoring epistasis) in the population is

$$\sigma_G(j, j, j) = \left(2 - \frac{1}{2^j}\right) \left(\sigma_A^2 + \frac{\sigma_D^2}{2^j}\right)$$

giving the within-family variance as

$$\sigma_{Gw}(S_j) = \sigma_G(j, j, j) - \sigma_G(j, j, j - 1) = \left(\frac{1}{2^j}\right) \sigma_A^2 + \left(\frac{3 - 2^{j+1}}{4^j}\right) \sigma_D^2$$

As Figure 23.2B illustrates, we can consider yet ever-deeper hierarchical levels of population structure with selfing. Suppose we are interested in the response in generation T due to selection among parents in generation t chosen on the basis of their relatives in generation τ , but that we are considering only the response among the subpopulation that descended from a common ancestor in generation k . For example, among all the descendants from a particular S_3 , what is the response to selection on their S_4 's if we base selection on the S_6 family means? Here T is the generation of interest for the response ($T = \infty$ for the permanent response), $k = 3$, $t = 4$, and $\tau = 6$.

Defining $\sigma_G(T, \tau, t, k)$ as the covariance between T and τ given they shared an ancestor in generation t from subpopulation k , Wright and Cockerham (1986), following Gates (1954), show that

$$\sigma_G(T, \tau, t, k) = \sigma_G(T, \tau, t) - \sigma_G(T, \tau, k) \quad \text{for } t > k \tag{23.38}$$

Hence, Equation 23.32 can be used to compute these covariances.

Example 23.10. As an example of the difference between $\sigma_G(T, \tau, t, k)$ and $\sigma_G(T, \tau, t)$, consider the following situation. Suppose we randomly chose F_2 s from a pure line cross and self a large collection of these, generating a total of 1000 lines. We have a number of options for advancing the lines, the two extremes are to advance single lines from each of the original F_2 's or to advance 1000 lines from a single F_2 . Lacking any other information on the lines, it is obvious that keeping all the original lines is likely the better strategy (indeed, this was proposed by Compton 1968), but just how much better is it? For both situations, consider the contribution to response in a completely inbred individual ($T = \infty$) from an individual in generation t chosen by evaluation of relatives in generation $\tau \geq t$.

Since within each line there are only two alleles, the covariances simplify (Wright and Cockerham 1986), and we are left for the first strategy with

$$\sigma_G(\infty, \tau, t) = \left(1 - \frac{1}{2^{t+1}}\right) (2\sigma_A^2 + 2\sigma_{ADI} + \sigma_{DI}^2) + \left(\frac{1}{2^{\tau+1}}\right) (\sigma_{ADI} + \sigma_{DI}^2)$$

If selection is delayed until at least the fifth generation of selfing, this is essentially $2\sigma_A^2 + 2\sigma_{ADI} + \sigma_{DI}^2$.

Conversely, the covariance for the second strategy (advancing all the lines from a single F_2) is given by $\sigma_G(\infty, \tau, t, 0)$, as all individuals trace back to a single individual in generation zero. Here

$$\begin{aligned}\sigma_G(\infty, \tau, t, 0) &= \sigma_G(\infty, \tau, t) - \sigma_G(\infty, \tau, 0) \\ &= \left(1 - \frac{1}{2^{t+1}}\right) (2\sigma_A^2 + 2\sigma_{ADI} + \sigma_{DI}^2) + \left(\frac{1}{2^{\tau+1}}\right) (\sigma_{ADI} + \sigma_{DI}^2) \\ &\quad - \left[\left(1 - \frac{1}{2}\right) (2\sigma_A^2 + 2\sigma_{ADI} + \sigma_{DI}^2) + \left(\frac{1}{2^{\tau+1}}\right) (\sigma_{ADI} + \sigma_{DI}^2) \right] \\ &= \left(\frac{1}{2} - \frac{1}{2^{t+1}}\right) (2\sigma_A^2 + 2\sigma_{ADI} + \sigma_{DI}^2)\end{aligned}$$

If selection is not started until rather late, this is one half the covariance as that for using the entire collection of lines. Considering only the additive variance, this makes sense, as under the infinitesimal model, in an outbred population (such as the collection of F_2 s), half the additive variance is between individuals and half is generated by segregation within individuals. The first covariance, $\sigma_G(\infty, \tau, t)$, considers only not the variance within a particular line, but also the variance among the initial line founders, while the second focuses solely on the variance within a particular line.

Response to Family Selection

Our earlier discussions of selection with selfing (Equations 23.23-23.28) assumed that the selection unit was the parent (individual selection), so that $\tau = t$. More generally, consider a parent in generation t where we save selfed seed from this individual for the recombination unit, and test the parent using the mean of its bulked selfed offspring in generation τ , an $S_{t,\tau}$ family (Figure 23.3). For such cases, the response in generation T from selection among parents in generation t is

$$r(T, \tau, t) = \bar{i} \frac{\sigma_G(T, \tau, t)}{\sqrt{\sigma_G(\tau, \tau, t) + \sigma_e^2}} \quad (23.39)$$

This follows since the genetic variance of the selection unit is $\sigma_G(\tau, \tau, t)$, while the covariance between the selection unit (τ) and an offspring of the recombination unit measured in generation T have their last common parent in generation t (Figure 23.3).

For the cases in Figure 23.3, where families are selected and remnant seed from those families used to form the next generation, then for selection based on $S_{t,t+1}$ families,

$$\sigma_G(\infty, t+1, t) = (1 + f_t)\sigma_A^2 + \frac{3 + 5f_t}{4} \sigma_{ADI} + \frac{1 + 3f_t}{4} \sigma_{DI}^2 + (1 + f_t)^2 \sigma_{AA}^2 \quad (23.40a)$$

while for selection based on $S_{t,t+2}$ families,

$$\sigma_G(\infty, t+2, t) = (1 + f_t)\sigma_A^2 + \frac{7 + 9f_t}{8} \sigma_{ADI} + \frac{3 + 5f_t}{8} \sigma_{DI}^2 + (1 + f_t)^2 \sigma_{AA}^2 \quad (23.40b)$$

As pointed out by Cockerham and Matzinger (1985), the long-term response under these two schemes differs only in the dominance-related terms, σ_{DI}^2 and σ_{ADI} . If instead of remnant seed, families are chosen and then one chooses one (or more) members from the selected families for selfing (strict family selection, as opposed to sib, = remnant seed, selection, the expressions become a little more complex. If n family members are scored, the appropriate covariance for selected based upon families from one generation of selfing is

$$\frac{1}{n} \sigma_G(\infty, t+1, t+1) + \frac{n-1}{n} \sigma_G(\infty, t+1, t) \quad (23.41a)$$

This arises by recalling Equation 23.7. Since selection is based on the family mean, $1/n$ is the weight on the individual in the family that is also the recombination unit (hence, the common ancestor is just this individual, which is in generation $t + 1$) and the remainder $(1 - 1/n)$ are sibs of this individual (and the common ancestor is the previous generation back). Similarly, if family selection is based on two generations of selfing, then the appropriate covariance becomes

$$\frac{1}{n^2} \sigma_G(\infty, t + 2, t + 2) + \frac{n - 1}{n^2} \sigma_G(\infty, t + 2, t + 1) + \frac{n - 1}{n} \sigma_G(\infty, t + 2, t) \quad (23.41b)$$

If the number n of family members tested is large, the corrections given by Equation 23.41 can be ignored and the simpler version (Equation 23.40) used instead.

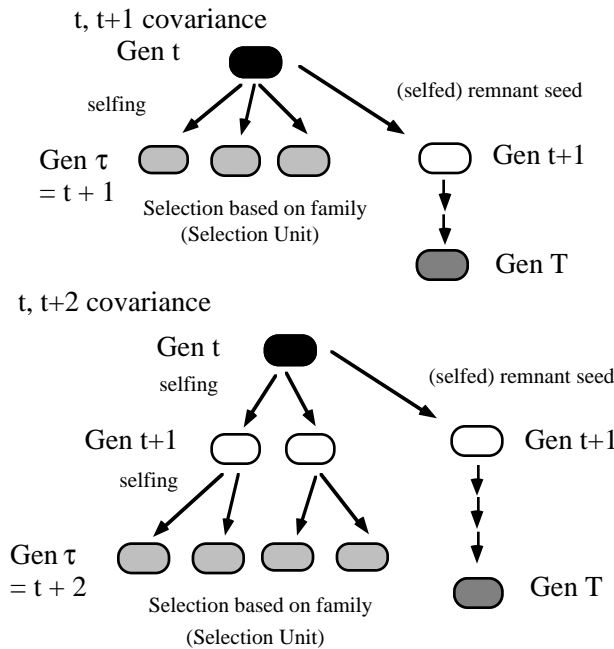


Figure 23.3. Examples of family selection in selfed lines. **Top:** The selection unit here is the selfed offspring of a parent, and the recombination unit is a remnant seed from this parent, so that selection is based on $S_{t,t+1}$ families. Here, the covariance of interest is between an individual in the selection unit (generation $\tau = t + 1$) and a descendant of the recombination unit (measured in generation T), which have a common parent in generation t . **Bottom:** In species with low seed set, a single plant may not generate sufficient seed for family testing. In this case, additional seed can be generated by another round of selfing. If selection is based upon such $S_{t,t+2}$ families and we use a remnant (selfed) seed from the parent as the recombination unit, the common parent is two generations removed from the selection unit ($\tau = t + 2$).

Within Family Selection Under Selfing

Our results for selfing thus far have only been concerned with selection between (or among) lines. If selection is practiced entirely *within* a single selfed lineage (e.g., within the progeny of a single selfed individual), genetic variation is quickly removed and the selection response rapidly ceases. Pederson (1969b) gives the response remaining in generation T from within-

family selection in generation t as

$$r_w(T, n) = \bar{i}_t \frac{\sigma_{Gw}(T, t)}{\sqrt{\sigma_{Gw}(t, t) + \sigma_{Ew}^2}} \quad (23.42a)$$

Applying Equation 23.37c, this becomes

$$r_w(T, n) = \bar{i}_t \frac{\sigma_G(T, t, t) - \sigma_G(T, t, t-1)}{\sqrt{\sigma_G(t, t, t) - \sigma_G(t, t, t-1) + \sigma_{Ew}^2}} \quad (23.42b)$$

Recalling Equation 23.34, for a pure line cross, we have

$$\begin{aligned} \sigma_G(T, t, t) - \sigma_G(T, t, t-1) = \\ 2^{-t} \sigma_A^2 + 2^{-T} \sigma_D^2 + \sum_{k=2} \left(\left(2 - \frac{1}{2^t} \right)^k - \left(2 - \frac{1}{2^{t-1}} \right)^{k-1} \right) \sigma_{A^k}^2 \end{aligned}$$

Ignoring epistasis, the cumulative response to k generations of selection at generation $T > k$ is

$$R(T | k) = \sum_{t=1}^k \bar{i}_t \frac{2^{-t} \sigma_A^2 + 2^{-n} \sigma_D^2}{\sqrt{2^{-t} (\sigma_A^2 + \sigma_D^2) + \sigma_e^2}} \quad (23.43)$$

(Pederson 1969b). Note that the sum here starts at selfing generation one, as selection starts within the S_1 inbred family. Since any within-family genetic variation rapidly decays, after a few generations the response essentially stops, with the amount of permanent response due to selection in generation t depending on $\sigma_A^2 / 2^t$. Since

$$\sigma_e^2 \leq 2^{-t} (\sigma_A^2 + \sigma_D^2) + \sigma_e^2 \leq \sigma_z^2$$

it follows that

$$\bar{i} \frac{\sigma_A^2}{\sigma_E} \sum_{t=1}^{\infty} 2^{-t} \leq \bar{i} \sigma_A^2 \sum_{t=1}^{\infty} \frac{2^{-t}}{\sqrt{2^{-t} (\sigma_A^2 + \sigma_D^2) + \sigma_e^2}} \leq \bar{i} \frac{\sigma_A^2}{\sigma_z} \sum_{t=1}^{\infty} 2^{-t}$$

Since the power series sums to one, the total permanent response $R(\infty | \infty)$ under continued within-family selection (ignoring new mutation) is

$$\bar{i} \sigma_A^2 / \sigma_E \leq R(\infty | \infty) \leq \bar{i} h \sigma_A$$

which is no better than twice the response of the first generation ($\bar{i} h \sigma_A / 2$).

Combined Selection

Suppose n_2 F_2 individuals are collected and selfed to create F_3 families of size n_3 for each of the F_2 founding lines. In trying to choose the best F_3 line, we might consider **combined selection**, using both its individual value z_{ij} and also the average value of the entire F_2 family from which it derives,

$$\bar{z}_i = \frac{1}{n_3} \sum_{k=1}^{n_3} z_{ij}$$

By analogy with the family index (Chapter 8), we can select lines using the index

$$I_{ij} = b \bar{z}_i + z_{ij} \quad (23.44)$$

As with the standard family index, a large b value places more weight on family average, a small b places more weight on the individual line value. Lines chosen are selfed to complete fixation. The expected response in generation T of selfing follows from a slight modification the generalized Breeders' equation (8.3) and is

$$R_y(T) = \bar{i} \frac{\sigma(I_{ij}, y_T)}{\sigma_I} \quad (23.45)$$

where y_T is a direct descent of the line z_{ij} in generation T of selfing. The permanent response is given by considering $T = \infty$.

It will be useful at this point to introduce two variance components that will (somewhat) simplify our results. First define

$$\sigma_c^2(T) = \sigma_G(T, 1, 0) \quad (23.46a)$$

as the variance between crosses (the different F_2 s) and

$$\sigma_l^2(T) = \sigma_G(T, 1, 1) - \sigma_G(T, 1, 0) \quad (23.46b)$$

as the variance in lines within crosses (the variance in deviations from the mean cross effect). For a pure line cross, these variances are

$$\sigma_c^2(T) = \sigma_A^2 + 2^{-(T+1)}\sigma_D^2 + \sigma_{AA}^2 + \cdots + \sigma_{A^k}^2 \quad (23.47a)$$

$$\sigma_l^2(T) = (1/2)\sigma_A^2 + 2^{-T}\sigma_D^2 + (5/4)\sigma_{AA}^2 + \cdots + \left(\left(\frac{3}{2}\right)^k - 1\right)\sigma_{A^k}^2 \quad (23.47b)$$

With these variance components in hand, consider the numerator of Equation 23.45. First note that

$$\sigma(I_{ij}, y_T) = b\sigma(\bar{z}_i, y_T) + \sigma(z_{ij}, y_T)$$

Since y_T is a direct descendant of line ij , $\tau = t = 1$, hence

$$\sigma(z_{ij}, y_T) = \sigma_G(T, 1, 1) \quad (23.48a)$$

Likewise,

$$\sigma(\bar{z}_i, y_T) = \frac{1}{n_3} \sum_{k=1}^{n_3} \sigma(z_{ik}, y_T) = \frac{\sigma(z_{ij}, y_T)}{n_3} + \left(1 - \frac{1}{n_3}\right) \sigma(z_{ik}, y_T)$$

The first term is given by Equation 23.48a, while the second is the covariance between a sib z_{ik} of line ij . Here, the common parent between line ik and y is the F_2 , so that $t = 0$, $\tau = 1$, and this covariance is $\sigma_G(T, 1, 0)$. Putting these together,

$$\sigma(\bar{z}_i, y_T) = \sigma_G(T, 1, 0) + \frac{\sigma_G(T, 1, 1) - \sigma_G(T, 1, 0)}{n_3} = \sigma_c^2(T) + \frac{\sigma_l^2(T)}{n_3} \quad (23.48b)$$

The numerator covariance in Equation 23.45 thus becomes

$$\sigma(I_{ij}, y_T) = b \left(\sigma_c^2(T) + \frac{\sigma_l^2(T)}{n_3} \right) + \sigma_c^2(T) + \sigma_l^2(T) \quad (23.49)$$

Now turning to the variance of the selection index,

$$\sigma^2(I_{ij}) = \sigma^2(b\bar{z}_{i.} + z_{ij}) = b^2\sigma^2(\bar{z}_{i.}) + \sigma^2(z_{ij}) + 2b\sigma(\bar{z}_{i.}, z_{ij}) \quad (23.50a)$$

First, consider the variance of a random F_3 line,

$$\sigma^2(z_{ij}) = \sigma_G(1, 1, 1) + \sigma_\epsilon^2 = \sigma_c^2(1) + \sigma_l^2(1) + \sigma_\epsilon^2 \quad (23.50b)$$

where σ_ϵ^2 is the within-line error variance. Next, note that

$$\sigma^2(\bar{z}_{i.}) = \frac{n_3\sigma^2(z_{ij})}{n_3^2} + \frac{n_3(n_3-1)\sigma(z_{ij}, z_{ik})}{n_3^2} = \frac{\sigma^2(z_{ij})}{n_3} + \left(1 - \frac{1}{n_3}\right)\sigma(z_{ij}, z_{ik})$$

The first covariance is given by Equation 23.50b, while the second is the covariance between sibs, $\sigma(1, 1, 0)$, giving

$$\sigma^2(\bar{z}_{i.}) = \frac{\sigma(1, 1, 1) + \sigma_\epsilon^2}{n_3} + \left(1 - \frac{1}{n_3}\right)\sigma(1, 1, 0) = \sigma_c^2(1) + \frac{\sigma_l^2(1) + \sigma_\epsilon^2}{n_3} \quad (23.50c)$$

Finally, we can show that $\sigma(\bar{z}_{i.}, z_{ij}) = \sigma^2(\bar{z}_{i.})$. Putting all these together gives

$$\sigma^2(I_{ij}) = b(b+2)\left(\sigma_c^2(1) + \frac{\sigma_l^2(1) + \sigma_\epsilon^2}{n_3}\right) + (\sigma_c^2(1) + \sigma_l^2(1) + \sigma_\epsilon^2) \quad (23.51)$$

Substitution of Equations 23.49 and 23.51 into 23.45 gives the expected response for arbitrary b and $T > 1$.

An obvious question is what value of b maximizes response? Taking the derivative of Equation 23.45 with respect to b and solving for zero gives the optimal value as

$$b = \frac{n_3 [\sigma_c^2(T)\sigma_E^2 - \sigma_c^2(1)\sigma_l^2(1) + \sigma_c^2(T)\sigma_l^2(1)]}{\sigma_l^2(T)(n_3\sigma_c^2(1) + \sigma_l^2(1) + \sigma_E^2)} \quad (23.52)$$

As with family index selection, the selection intensity is reduced by the fact that F_3 lines from the same F_2 are related and hence correlated, and we must correct for this correlation. Letting ρ be the correlation between the index scores from lines in the same cross, then if we chose the best K from $N = n_2n_3$ total lines (n_3 lines from each of n_2 F_2 's), then Equation 8.54b gives the selection intensity as

$$\bar{i}_{n_2n_3, K}(\rho) = \bar{i}_p - \frac{1-p}{2\bar{i}_p p(n_2n_3(1-\rho) + n_2\rho + 1)} \quad (23.53)$$

where $p = K/N$ is the fraction saved and \bar{i}_p the infinite-population size selection intensity. To compute the correlation ρ between index scores, first note that

$$\begin{aligned} \sigma(I_{ij}, I_{ik}) &= \sigma(b\bar{z}_{i.} + z_{ij}, b\bar{z}_{i.} + z_{ik}) \\ &= b^2\sigma^2(\bar{z}_{i.}) + 2b\sigma(\bar{z}_{i.}, z_{ij}) + \sigma(z_{ij}, z_{ik}) \\ &= [b^2\sigma^2(\bar{z}_{i.}) + 2b\sigma(\bar{z}_{i.}, z_{ij}) + \sigma^2(z_{ij})] - \sigma^2(z_{ij}) + \sigma(z_{ij}, z_{ik}) \\ &= \sigma^2(I_{ij}) - (\sigma_c^2 + \sigma_l^2 + \sigma_\epsilon^2) + \sigma_c^2 \\ &= \sigma^2(I_{ij}) - (\sigma_l^2 + \sigma_\epsilon^2) \end{aligned} \quad (23.54a)$$

The third step follows upon recalling Equation 23.50a. The correlation between index scores among individuals from the same line thus becomes

$$\rho = \frac{\sigma(I_{ij}, I_{ik})}{\sigma^2(I)} = \frac{\sigma^2(I_{ij}) - (\sigma_l^2 + \sigma_e^2)}{\sigma^2(I_{ij})} = 1 - \frac{\sigma_l^2 + \sigma_e^2}{\sigma^2(I_{ij})} \quad (23.54b)$$

Substitution of this value into Equation 23.53 gives the appropriate selection intensity, corrected for both finite population size and correlation between lines.

For a fixed total number of lines to examine, there is a tradeoff between selection intensity and accuracy. The optimal design for known (or estimated) genetic variances can be obtained by numerically trying different combinations of the n_2/n_3 ratio (for fixed N) in Equation 23.45 (and its associated components, Equations 23.49 and 23.51-53). (Equation 23.52). Weber (1982, 1984, Wricke and Weber 1986) has extensively examined index selection under selfing, including both optimal design (in the absence of dominance) and the extension to additional generations of selfing. For example, Weber (1982) considers the more general situation where each F_2 family consists of n_3 F_3 families, each F_3 family consisting of n_4 F_4 families and so on to F_j families. In this case, the full index is

$$I = b_2(\bar{z}_2 - \bar{z}) + b_3(\bar{z}_3 - \bar{z}) + \dots + b_j(\bar{z}_j - \bar{z}_{j-1}) \quad (23.55a)$$

The weights b_k are chosen to maximize the correlation between the index I and the final genetic value of the completely inbred lines (which we denote g_∞), which (Weber 1982) implies

$$b_k = \frac{\sigma(\bar{z}_k - \bar{z}_{k-1}, g_\infty)}{\sigma^2(\bar{z}_k - \bar{z}_{k-1})} \quad (23.55b)$$

Consult Weber for further details.

RESPONSE UNDER PARTIAL SELFING

One of the most widespread systems of inbreeding is **partial selfing**, where each individual can either self or outcross. If η is the probability of selfing, then the population approaches a *mean* inbreeding value of $f = \eta/(2 - \eta)$, but there is a distribution of inbreeding values among individuals within the population. In particular, the probability an individual is inbred to level $f_i = 1 - 2^{-i}$ is $(1 - \eta)\eta^i$. This lack of uniformity in f greatly complicates the prediction of selection response. In particular, nonlinear parent-offspring regressions can occur (Wright and Cockerham 1985, Kelly 1999a). Our treatment of this difficult area first presents approximate results using covariances. We then consider a more careful treatment due to Kelly (1999a,b) that considers response *within* each group with a given level of inbreeding (i.e., all group members have the same f value).

An Approximate Treatment Using Covariances

Partial selfing has been examined by Wright and Cockerham (1985, 1986) and Wright (1988), who obtain the appropriate cross-generation covariances and predict response using the method on ancestral regression. They assume that the population is at the equilibrium mean inbreeding value, and that each individual has the same probability η of selfing (i.e., no genetic variation in selfing vs. outcrossing rates). Since a parent can either self or outcross, the parent-offspring needs to take this into account. Further, because of selfing, the covariances for selection before and after reproduction are slightly different. For selection before reproduction (BR), the single-generation response is given by

$$R_{BR} = \bar{i} \frac{\sigma_{G(BR)}(t + 1, t)}{\sqrt{\sigma_G^2 + \sigma_e^2}} \quad (23.56a)$$

with

$$\sigma_{G(BR)}(t + 1, t) = 2(1 - \eta)\sigma_{G_O}(P, O) + \eta\sigma_{G_S}(P, O) \quad (23.56b)$$

where $\sigma_{G(O)}(P, O)$ is the parent-offspring covariance under outcrossing and $\sigma_{G(S)}(P, O)$ the parent-offspring covariance under selfing. Since the population is assumed to be at its inbreeding equilibrium, the parent-offspring covariances and total genetic variation (σ_G^2) are assumed to constants, independent of t . Table 23.5 gives the coefficients for these covariances, as well as for σ_G^2 . Note that an additional quadratic component, $\iota^2 - \iota^*$, also appears in the covariance between relatives. As we will see, there is a significant transient component to the immediate response. For selection after reproduction (AR), an outcrossed individual has only one parent under selection (the pollen parent not being under selection), and the covariance in the numerator of Equation 23.56a is replaced by

$$\sigma_{G(AR)}(t + 1, t) = (1 - \eta)\sigma_{G_O}(P, O) + \eta\sigma_{G_S}(P, O) \quad (23.56c)$$

Table 23.5. Covariances required to predict response under partial selfing using Equation 23.56. The probability of selfing is η , and $f = \eta/(2 - \eta)$. After Wright and Cockerham (1985).

Cov	σ_A^2	σ_D^2	σ_{ADI}	σ_{DI}^2	ι^*	$\iota^2 - \iota^*$
$\sigma_{G_O}(P, O)$	$\frac{1+f}{2}$	0	$\frac{f}{2}$	0	0	0
$\sigma_{G_S}(P, O)$	$1 + f$	$\frac{1-f}{2}$	$\frac{1+7f}{4}$	f	$\frac{f(1-f)}{2(2+f)}$	$\frac{f(1-f^2)}{2(2+f)}$
$\sigma_{G(AR)}$	$\frac{1+3f}{2}$	$\frac{f(1-f)}{1+f}$	$\frac{2f(1+3f)}{2(1+f)}$	$\frac{2f^2}{1+f}$	$\frac{f^2(1-f)}{(1+f)(2+f)}$	$\frac{f^2(1-f^2)}{(1+f)(2+f)}$
$\sigma_{G(BR)}$	$1 + f$	$\frac{f(1-f)}{1+f}$	$\frac{f(3+5f)}{2(1+f)}$	$\frac{2f^2}{1+f}$	$\frac{f^2(1-f)}{(1+f)(2+f)}$	$\frac{f^2(1-f^2)}{(1+f)(2+f)}$
σ_G^2	$1 + f$	$1 - f$	$2f$	f	$\frac{f(1-f)}{2+f}$	$\frac{f(1-f^2)}{2+f}$

The permanent response to selection is given by replacing the numerator covariance in Equation 23.56a with $\sigma_G(\infty, t)$. For selection before reproduction, Wright and Cockerham show this equals

$$\sigma_{G(BR)}(\infty, t) = \left(\frac{2}{2 - \eta}\right)\sigma_A^2 + \left(\frac{\eta(3 - \eta)}{(2 - \eta)^2}\right)\sigma_{ADI} + \left(\frac{\eta^2}{(2 - \eta)^2}\right)\sigma_{DI}^2 \quad (23.57a)$$

Further, the before and after covariances are related by

$$\sigma_{G(AR)}(\infty, t) = \left(\frac{1 + \eta}{2}\right)\sigma_{G(BR)}(\infty, t) \quad (23.57b)$$

While the coefficient on σ_A^2 is unchanged, all the other quadratic components have rather different coefficients relative to their single-generation values. Thus, the transient component of response is expected to be considerable. Wright and Cockerham find that

$$\sigma_{G(BR)}(t + T, t) = \sigma_{G(BR)}(\infty, t) + \left(\sigma_G^2 - \sigma_{G(BR)}(\infty, t)\right)\left(\frac{\eta}{2}\right)^T \quad (23.58)$$

with a similar expression for the covariance for selection after reproduction. Hence, the transient component decays rather rapidly (as the rate of decay is no slower than 2^{-T}).

Wright and Cockerham also show that the final change in the mean from a generation of selection can be predicted from the response in the first two generations after selection, with

$$\bar{z}_\infty = \frac{2\bar{z}_2 - \eta\bar{z}_1}{2 - \eta} \quad (23.59)$$

where \bar{z}_i is the mean i generations after a generation of selection.

Family selection is examined by Wright and Cockerham (1986). With partial-selfing, there are a number of potential families that one can consider for the selection unit: outcrossed half- and full-sibs, selfed, or naturally pollinated (a mixture of selfed and outcrossed progeny). Consult their paper for details on response under these different systems.

A major caveat with the covariance approach is that selected individuals may not be a random sample of the inbreeding classes. For example, with inbreeding depression, less inbred individuals are expected to be chosen by selection more often. This has the effect of creating a nonlinear parent-offspring regression (Figure 23.4). If the nonlinearity is significant, the covariance approach is inappropriate.

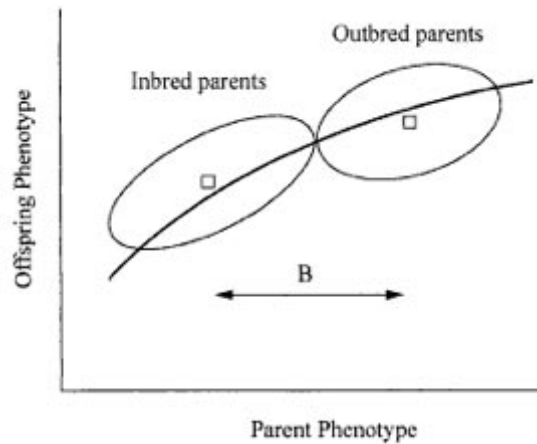


Figure 23.4. When inbreeding depression is present, the parent-offspring regression under partial selfing can be nonlinear. Suppose we have just two groups: $f = 0$ (outcrossed) and $f = 1$ (fully inbred). Within each group, parent-offspring regressions are linear. However, for the population as a whole (i.e., we examine individuals without knowledge of which group they belong to), the regression is non-linear. Here b denotes the inbreeding depression (the change in mean, the open squares, between the outcrossed and fully-inbred populations). After Kelly (1999a.)

A More Careful Treatment: Kelly's Structured Linear Model

Kelly (1999a,b) proposed an alternative approach for examining the selection response with partial selfing. By analogy with a geographically structured population, Kelly suggested that the basis for a more careful approach to partial selfing is to consider the population as a set of cohorts within which all individuals have the same level of inbreeding. Cohort zero are all individuals that arose through random mating in the previous generation, cohort one are those individual that arose from selfing in the previous generation, cohort i those individuals that have undergone i rounds of selfing since they last enjoyed random mating. He terms this approach a **structure linear model** (or nbfl SLM), as within each cohort, it is assumed that standard linear regressions hold. The response under selfing is then predicted by specifying both the within-cohort variances and covariances and the between-cohort transitions.

Besides properly accounting for the structured nature of partly selfing populations, Kelly's treatment also accommodates the effects of linkage disequilibrium (under the infinitesimal model assumptions).

Before examining the SLM approach, results from simulations reported in Kelly (1999a) reveal several important observations that we discuss first. If directional dominance is present (which implies inbreeding depression occurs), the distribution of the cohorts does not follow the geometric distribution predicted from the strict (i.e., no selection) partial selfing model. Rather, selection greatly skews this distribution. For example, when advantageous alleles are recessive, inbred individuals will be over-represented, while when advantageous alleles are dominant, inbreds are under-represented. In either setting, the covariance approach will give incorrect answers as it assumes the geometric distribution of inbreeding values. The deviation from this distribution caused by selection can be quite striking. For example, after 10 generations of selection when beneficial recessives are present, the cohort distribution is strongly bimodal, with peaks at 0 (outbreeding) and 10 (selfing for all ten generations). A second feature from the simulations was that linkage disequilibrium generated significant changes in the variances and covariances. Changes in f for the population directly affect the usable level of additive variance and selfing reduces the chance for recombination to break down non-random associations generated by selection. Such selection-induced changes in the covariances from their strict (i.e., no selection) partial selfing values also compromise the results from the covariance approach.

Turning now to the details of Kelly's approach, define cohort i as the collection of individuals that have undergone i generations of selfing since their last outcrossing. Let $\pi_i(t)$ denote the frequency of cohort i in generation t , under strict partial selfing $\pi_i(t) = (1 - \eta)\eta^i$. Kelly assumes selection acts before reproduction. The transition probabilities between cohorts are as follows. With probability $1 - \eta$, an individual outcrosses, so that

$$\pi_0(t+1) = 1 - \eta \quad (23.60a)$$

For an individual to enter cohort $i + 1$, it must have been a member of cohort i in the previous generation and it must self. Further, Kelly allows for selection against cohorts (for example, highly inbred individuals may have lower fitness). If $W(t)$ and $W_i(t)$ denote the average fitness of a random individual from the population and from cohort i (respectively) in generation t , then the fraction of the population that is in cohort i after selection (but before reproduction) is $\pi_i(t) W_i(t)/W(t)$. Putting all these together, the fraction of the population in cohort $i + 1$ in the next generation is

$$\pi_{i+1}(t+1) = \eta \pi_i(t) \frac{W_i(t)}{W(t)} \quad (23.60b)$$

Now consider the dynamics of the mean. Since a member of cohort i has $f = 1 - 2^{-i}$, the mean value of inbreeding for the population is

$$F(t) = \sum \pi_i f_i = \sum \pi_i (1 - 2^{-i}) \quad (23.61)$$

Define the weighted sum of average effects as a and the weighted sum of homozygous dominance deviations by b , so that

$$a = \sum_{ij} \alpha_{ij} p_{ij}, \quad b = \sum_{ij} \delta_{ijj} p_{ij} \quad (23.62)$$

where α_{ij} is the average effect for allele j from locus i , p_{ij} the allele frequency and δ_{ijj} the dominance deviations for homozygotes. If $A(t)$ and $B(t)$ denote the average values of a and b at generation t , then the mean population phenotype is

$$z(t) = \mu + A(t) + F(t)B(t) \quad (23.63a)$$

Letting $A_i(t)$ and $B_i(t)$ denote the average value of a and b in cohort i in generation t , it also follows that

$$A(t) = \sum \pi_i A_i(t), \quad \text{and} \quad B(t) = \sum \pi_i B_i(t) \quad (23.63b)$$

Kelly assumes that *within* each cohort, the parent-offspring regressions are linear, so that the value A' of A after selection in cohort i is

$$A'_i(t) = A_i(t) + \frac{\sigma_{az,i}(t)}{\sigma_{z,i}^2(t)} S_i(t) \quad (23.64a)$$

Since the mean additive value of selfed offspring equals that of their parents,

$$A_{i+1}(t+1) = A'_i(t) = A_i(t) + \frac{\sigma_{az,i}(t)}{\sigma_{z,i}^2(t)} S_i(t) \quad (23.64b)$$

Likewise, the mean additive value of outcrossed offspring equals the average additive values of their parents,

$$A_0(t+1) = \sum_i \pi_i(t) A'_i(t) = A(t) + \sum_i \pi_i(t) \frac{\sigma_{az,i}(t)}{\sigma_{z,i}^2(t)} S_i(t) \quad (23.64c)$$

Under linkage equilibrium, the covariance functions are given by

$$\sigma_{az,i}(t) = \sigma_A^2 + f_i(\sigma_A^2 + 2\sigma_{ADI}) \quad (23.65a)$$

Similar expressions exist for $B_i(t)$, with B replacing A and (under linkage equilibrium)

$$\sigma_{bz,i}(t) = \sigma_{ADI} + f_i(\sigma_{ADI} + 2\sigma_{DI}^2) \quad (23.65b)$$

Finally, again under linkage equilibrium, the phenotypic variance in cohort i is given by

$$\sigma_i^2(t) = \sigma_G^2(i, i) + \sigma_{E,i}^2 \quad (23.65c)$$

where $\sigma_G^2(i, i)$, the genetic variance after i generations of selfing, is given by Equation 23.23. Iteration of Equations 23.60, 23.61, and 23.64 and 23.65 gives the evolution of the $A_i(t)$ and $B_i(t)$ values, from which the population mean in generation t can be obtained by use of Equation 23.63b and c.

Kelly's simulations showed that selection-induced linkage disequilibrium can have a rather dramatic effect on the covariances, and suggested an iterative approach (along the lines of our results for random mating and of Hayashi and Ukai's results for pure selfing) for computing all the required covariances. As we did in Chapter 5, let

$$\kappa_i(t) = 1 - \frac{\sigma_{z,i}^{2*}(t)}{\sigma_{z,i}^2(t)} \quad (23.66)$$

denote the reduction in the phenotypic variance (here in cohort i) caused by selection, where $\sigma_{z,i}^{2*}(t)$ is the phenotypic variance after selection. Further, define the covariances

$$\sigma_{az,i}^*(t) = \sigma_{a,i}^2(t) + (1 - 2^{-i}) \sigma_{ab,i}(t) \quad (23.67a)$$

$$\sigma_{bz,i}^*(t) = \sigma_{ab,i}(t) + (1 - 2^{-i}) \sigma_{b,i}^2(t) \quad (23.67b)$$

Kelly shows that, under infinitesimal model assumptions that the recursion equations for the variances and covariances in the inbred cohorts ($i > /ge 1$) become

$$\sigma_{a,i+1}^2(t+1) = \sigma_{A,i}^2(t) - \kappa_i(t) \frac{(\sigma_{az,i}^*(t))^2}{\sigma_{z,i}^2(t)} + 2^{-(i+1)}\sigma_A^2 \quad (23.68a)$$

$$\sigma_{ab,i+1}(t+1) = \sigma_{ab,i}(t) - \kappa_i(t) \frac{\sigma_{az,i}^*(t) \sigma_{bz,i}^*(t) \sigma_{b,i}^2(t)}{\sigma_{z,i}^2(t)} + 2^{-(i+1)}\sigma_{ADI} \quad (23.68b)$$

$$\sigma_{b,i+1}^2(t+1) = \sigma_{b,i}^2(t) - \kappa_i(t) \frac{(\sigma_{bz,i}^*(t))^2}{\sigma_{z,i}^2(t)} + 2^{-(i+2)}\sigma_{DI}^2 \quad (23.68c)$$

For all of these expressions, the first term is the variance (covariance) before selection, the second term the reduction from selection, and the final term the contribution from segregation. The recursion equations for the outbred cohort are a little more involved, as parents are inbred to differing degrees. Letting $F'(t)$ denote the average value of inbreeding among selected parents,

$$\sigma_{A,o}^2(t+1) = \frac{1}{2} (\sigma_{a'}^2(t) + [1 - f'(t)]\sigma_A^2) \quad (23.69a)$$

$$\sigma_{ab,o}(t+1) = \frac{1}{2} (\sigma_{ab'}(t) + [1 - f'(t)]\sigma_{ADI}) \quad (23.69b)$$

$$\sigma_{b,o}^2(t+1) = \frac{1}{2} (\sigma_{b'}^2(t) + [1 - f'(t)]\sigma_{DI}^2) \quad (23.69c)$$

Where the variances (and covariances) are the population averages in the selected parents. Iteration of these equations accounts for the disequilibrium generated by selection.