32

Selection and Crossbreeding II: Advanced Topics

Draft 26 June 2014

In plant breeding, one may use crosses to form a **synethic** population, which is then propagated by random mating. For example, with the first generation would consist of all $n(n-1)/2$ crosses among the n founding lines, while the second and subsequent generations are propagated through random mating with no control over mating. A type of synthetic is a **composite variety** (**varitey** usually used in place of line when the population is outbred), wherein a large number of individuals are used to found the synthetic population. Synthetic represent a comprise between the advantage of an F_1 hybrid and the ease of using random mating to propagate a population.

RECIPROCAL RECURRENT SELECTION (RRS)

The Pure-Line-Crossbred Covariance

The **purebred-crossbred covariance** is the covariance between a sire's purebred and crossbred offspring (usually half-sibs). Covariance with respect to a tester (test-cross progeny)

Bowman (1960) showed that when a trait is controlled by a single locus, a negative purebred-crossbred covariance can only arise if overdominance occurs. McNew and Bell (1971) showed that when epistasis is present, a negative covariance can arise in the absence of overdominance. While a negative covariance is often taken as a strong signal to use some sort of crossbreeding program,

---

Example 12.2. Consider the following example given by McNew and Bell (1971). Suppose two unlinked diallelic loci (A/a, B/b) underlie the character of interest, where the mean character values for the different two-locus genotypes are given

by

	BB	Bb	bb
AA	8	2	4
Aa	2	4	6
aa	4	6	0

Assume linkage equilibrium and suppose the allele frequencies are $p_A = 0.2$, $p_B = 0.4$ in the purebred population and $p_A = 0.7, p_B = 0.5$. in the testor population. With these frequencies, the purebred-crossbred genetic covariance is negative. Further, the mean of the selected (i.e., purebred) population is 4.00, the mean of the testor 3.81, and the mean of crossbred offspring is 4.05. Hence, it would appear from the negative covariance and higher mean in the crosses that a selection scheme involving crossbreeding would be favored over one involving only the purebred population. However, McNew and Bell found that under selection and crossbreeding the mean approaches a limit of 4.8, while under purebred selection, the AABB genotype is fixed, and the mean is 8.0. In this case selection restricted to purebreds produces a larger long-term response than one involving crossbreeding, even though the initial signs (negative purebred-crossbred covariance, higher crossbred mean) would suggest otherwise.

Selecting for both GCA and SCA – reciprocal recurrent selection

egg production, maize **Krosigk et al 1973 refered Facloner**

Does not stack up well against other methods **Calhoon and Bohern 1974, refed in Facloner**

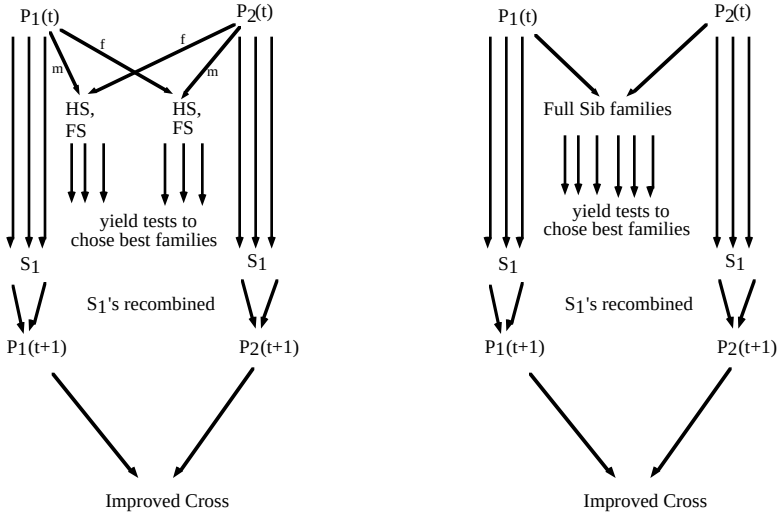


Figure 12.1. Half-sib (left) and full-sib reciprocal recurrent selection. Plants used as male (pollen) parents, denoted by m are selfed to create S_1 seed.

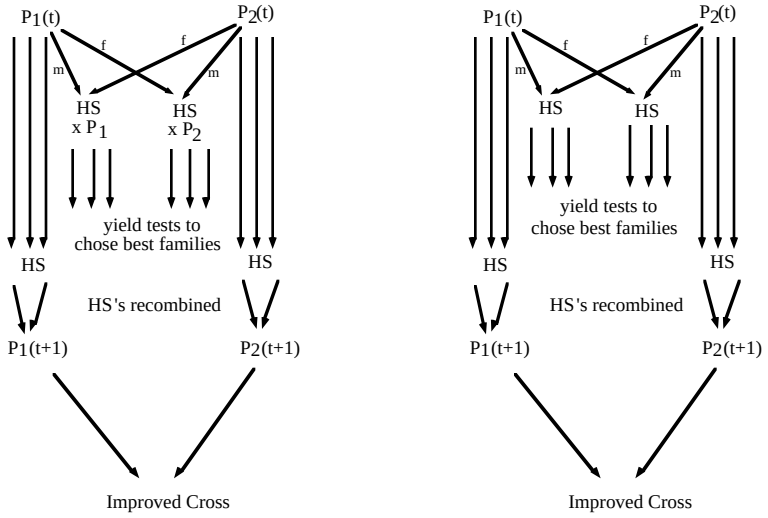


Figure 12.2. Two modified half-sib reciprocal recurrent selection scheme, proposed b y Paterniani (1967).

Table 12.1. Summary of several reciprocal recurrent selection schemes

Scheme	Selection Unit	Recombination Unit	Improved population
Half-Sib RRS Comstock et al. (1949)	HS, FS	S_1 families	$S_1 \times S_1$ progenies
Full-Sib RRS Hallauer and Eberhart (1970)	FS	S_1 families	$S_1 \times S_1$ progenies
Modified half-sib RRS-1 Paterniani (1967)	HS	pooled HS families	HS $\times$ HS
Modified half-sib RRS-2 Paterniani (1967)	HS	HS families	HS $\times$ HS

Selection for GCA and SCA

McNew and Bell (1976)

RRS selection (half-sib families, select on fathers, i)

$$\Delta\mu_P = \frac{1}{2}\,\frac{\sigma_{i,i}(A_C, A_P)}{\sigma_i^2(z_C)}\,S_{Ci} \qquad (12.xxa)$$

$$\Delta\mu_C = \frac{1}{4}\,\left(\frac{\sigma_i^2(A_C)}{\sigma_i^2(z_C)}\,S_{Ci} + \frac{\sigma_j^2(A_C)}{\sigma_j^2(z_C)}\,S_{Cj}\right) \qquad (12.xxb)$$

Within-line selection, n offspring per purebreed (full) sib family, d dams/male

$$\Delta\mu_P = \frac{\sigma_i^2(A_P)}{\sigma_i^2(z_P)}\,\left(\frac{nd+n+2}{4nd}\right)\,S_{Ci} \qquad (12.xxa)$$

$$\Delta\mu_C = \left(\frac{nd+n+2}{4nd}\right)\,\left(\frac{\sigma_{i,i}(A_C, A_P)}{\sigma_i^2(z_P)}\,S_{Pi} + \frac{\sigma_{j,j}(A_C, A_P)}{\sigma_j^2(z_P)}\,S_{Pj}\right) \qquad (12.xxb)$$

RRS Using Half-sibs Families

$$R = \overline{i}_1\,\frac{(1/4)\sigma^2_{A_{12}}}{\sigma(\,\overline{z}_{12}\,)} + \overline{i}_2\,\frac{(1/4)\sigma^2_{A_{21}}}{\sigma(\,\overline{z}_{21}\,)} \qquad (12.\text{xx})$$

RRS Using Full-sibs Families

$$R = \overline{i}\,\frac{(1/4)(\sigma^2_{A_{12}} + \sigma^2_{A_{21}})}{\sigma(\,\overline{z}\,)} \qquad (12.\text{xx})$$

For HS-RRS1

$$R = \overline{i}_1\,\frac{(1/16)\sigma^2_{A_{12}}}{\sigma(\,\overline{z}_{12}\,)} + \overline{i}_2\,\frac{(1/16)\sigma^2_{A_{21}}}{\sigma(\,\overline{z}_{21}\,)} \qquad (12.\text{xx})$$

For HS-RRS2

$$R = \overline{i}_1\,\frac{(1/8)\sigma^2_{A_{12}}}{\sigma(\,\overline{z}_{12}\,)} + \overline{i}_2\,\frac{(1/8)\sigma^2_{A_{21}}}{\sigma(\,\overline{z}_{21}\,)} \qquad (12.\text{xx})$$

Diallele Selection Schemes

SYNTHETICS

Balancing the often tremendous heterotic advantage of F_1 hybrids is the (often quite considerable) effort to continually generate these crosses each generation.

While one could use the F_2 and subsequent generations generated by random mating, as we have seen above the expected heterotic advantage over the parents, $\mu_{F_1} - \mu_{\overline{P}}$, is halved each generation for single crosses. Further, in many species (for example, alfalfa, *Medicago sativa*), the flowers are simply to small and/or too numerous to make controlled crosses practical.

Synthetic varieties, first suggested by Hayes and Garber (1919), offer one solution for making at least partial use of the heterotic potential available when crossing a series of lines. Generation of a synthetic variety requires involves three phases: choosing the parents or parental lines, controlled intercrossing these (usually making all $n(n-1)/2$ pairwise crosses) to generate a **Syn 1** population, and finally allowing subsequent generations to be generated under open pollination (i.e., no active pollination control, which may be mean some of the mating is by selfing when plants are not self-incompatible). The collection of initial lines/parents is denoted as the **Syn 0** population, and the population after $k-1$ generations of random mating is the **Syn k** population. Ideally, the total population size (and hence the total amount of seed) increases each generation. Usually Syn 4 gives the size required to generate a sufficient amount of seed for cultivar production.

The Syn 1 population is maximally heterozygous and is thus expected to have the best heterotic performance. Because of inbreeding in the Syn-2 and later generations (due to a limited number of founders), a reduction from the Syn-1 to the Syn-2 is expected, reflecting a decrease in the amount of heterozygosity among plants. For a diploid population with no selfing and no epistasis, the population mean reaches its equilibrium value in the Syn 2 generation. If there is only additive gene action, then Syn-1 and Syn-2 are expected to be identical. Linkage plays no role in the absence of epistasis, as the mean is still determined by frequencies of the individual alleles.

Given that the Syn 2 (and hence subsequent generations) show reduced performance relative to the Syn 1, just what is the advantage of synthetics? First, even though an equilibrium synthetic line shows a reduction from the maximal Syn 1 performance, it is still expected to have better performance than the average performance of all the founder lines. Further, the reduction from the Syn 1 becomes less dramatic as the number of founding lines increased. As we discuss below, the simplest version of Wright's predictor states (in the absence of epistasis) that when n founder lines are used, the decline from the Syn 1 to the Syn 2 (and hence the Syn ∞ equilibrium population) is just $1/n$ of the difference between the Syn 1 and Syn 0 means. Thus, but choosing a sufficient number of parents, the final synthetic population has a mean very close to the Syn 1 mean.

A second important use of synthetics is as reservoirs of desirable germ plasm (Sprague and Jenkins 1943) from which lines may be selected. The **Iowa Stiff Stalk Synthetic** in maize is an example the power of this approach (STUFF ABOUYT THIS LINE). Hill (1971) notes that the genetic variation between synthetics (whose parents are drawn from the same basic population) increases as fewer parents are

used for each synthetic variety. Hence, if the goal is to create a series of synthetic lines for subsequent between-line selection, then as few parents as possible should be used to form each line.

Wright's Predictor and Powers-Kinman-Sprague Extension

Wright (1922) noted that "a random-bred stock derived from n inbred families will have $1/n$th less superiority over its inbred ancestry than the first cross or a random-bred stock from which the inbred families might have been derived without selection". This verbal statement is the basis for the expression for the predicted F_2 value given the F_1 and average parental values $\overline{P}$,

$$F_2 = F_1 - \frac{F_1 - \overline{P}}{n} \tag{12.y1}$$

where F_x is the average performance across all F_x lines, $\overline{P}$ the average of all parental lines and n the number of lines. Equation 12.y1 appears to have been first presented by Kinman and Sprague (1945), who referred to it as "Wright's formula". We will use the term **Wright's predictor** to avoid confusion with Wright's (other) formula for the rate of change under selection (Equation XX.xx) which is very widely used in population genetics.

The logic behind Equation 12.y1 is as follows (Gilmore 1969): a random individual in the F_1 is a cross of $\ell_i \times \ell_j$ where $i \neq j$ are genes from different lines. In the F_1, the mean value is

$$\mu_{F_1} = \frac{1}{n(n-1)} \sum_{i \neq j} (\ell_i \times \ell_j)$$

as there are $n(n-1)$ crosses between the different lines (including reciprocals), and each occurs with the same frequency, so a random individuals if from a particular cross with probability $1/[\,n(n-1)\,]$. In the F_2, alleles derived form each line are present at frequency $1/n$,

$$\begin{aligned} F_2 &= \frac{1}{n^2} \sum_{i}^{n} \sum_{j}^{n} (\ell_i \times \ell_j) \\ &= \frac{1}{n^2} \sum_{i}^{n} (\ell_i \times \ell_i) + \frac{1}{n^2} \sum_{i \neq j} (\ell_i \times \ell_j) \\ &= \frac{1}{n} \overline{P} + \frac{n(n-1)}{n^2} F_1 = F_1 - \frac{F_1 - \overline{P}}{n} \end{aligned} \tag{12.zz}$$

Notice that the above derivation places no restrictions on the degree of inbreeding of the parents (Gilmore 1969), but does assume no epistasis. In the absence of epistasis, linkage has no effect on mean values, which is just a function of the

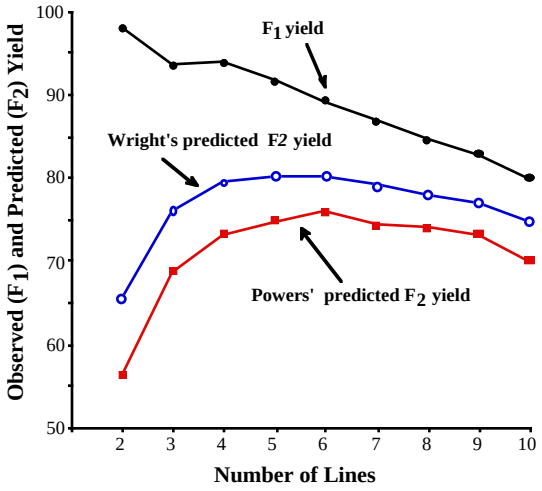
individual allele frequencies. This derivation also highlights the origin of the decline in performance in the F_2, namely some of the individuals have both alleles from the same founding population, while all of the individuals in the F_1 are formed by joining alleles from different lines/parents.

A few authors have also consider the case where gene action is entirely multiplicative (following Powers 1941). In particular, Kinman and Sprague (1945) suggested the expression

$$F_2 = \exp\left(\frac{2}{n^2} \sum_{i<j} \ln(\overline{z}_{ij}) + \frac{1}{n^2} \sum_{i}^{n} \ln(\overline{z}_{i}) \right) \qquad (12.y1)$$

this follows directly from Equation 12.xx and the assumption that gene action is entirely multiplcative. Although Equation 12.xx is occasionally credited to Powers in the literature, it first appeared in Kinman and Sprague, and is thus best referred to as the **Powers-Kinman-Sprague predictor**.

Example 12.x. Kinman and Sprague (1945) used data from 10 inbred lines and their 45 possible (non-reciprocal) single crosses to predict the expected yield of various synthetics formed by using from 2 to 10 inbred lines. They used both Wright's and Powers' expression to see if the different assumptions on gene action had a significant effect.



Lines were entered in order of their estimated general combining ability, so with $n = 4$, the top four lines in terms of GCA are used. As expected, as more lines are added, the overall F_1 yield decreases because the best pairs are chosen first. Countering this, the expected differential between the F_1 and F_2 means is a decreasing

function of n under both Wright's and Powers' expression. For this data, $n = 6$ is the optimal number of lines under both the Wright and Powers expressions.

Inbreeding in Synthetics

Synthetic lines, although random-mating populations, derive from a set of n parents and hence are inbred. The behavior of the inbreeding coefficient is of critical importance in predicting the mean of a synthetic, which (assuming no epistasis) is a function of the parental allele freuqnecies and the amoount of inbreeding in the synthetic population. In the simplest of settings (a non-selfing diploid population), the equilibrium level of inbreeding is just $1/n$ and is reached in one generation (the Syn-2). Busbice (1969) examined the expected inbreeding under more general settings. For the case of no selfing (i.e., individuals in the synthetic population are self-incompatible), then for a species with ploidy level $2k$ (k =1 for a diploid, 2 for a tetraploid, etc.), the amount of inbreeding f_j in the Syn j population is

$$f_j = \frac{kr_{j-1} + (k-1)f_{j-1}}{2k-1} \qquad \text{for} \quad j \geq 1 \tag{12.xxa}$$

where r_j is the probability that two random alleles from the parents of an individual are ibd. Note immediately that if the founding parents are unrelated ($r_0 = 0$), then $f_1 = 0$ for a diploid, independent of the level of inbreeding in the parents, and hence the Syn 1 is not inbred. For a tetraploid, if the parents are unrelated, then $f_1 = f_0/3$, or $1/3$ the average inbreeding of the parents, and here the Syn 1 is inbred. Inbreeding values in later generations are obtained by joint iteration of both f and r. Busbice argued at after the Syn 1 generation, r_t essentially reaches it equilibrium value, so that

$$r_e \simeq r_1 = \frac{1}{n}\left(\frac{1+(2k-1)f_0}{2k}\right) + \left(\frac{n-1}{n}\right) r_0 \tag{12.xxb}$$

This result and Equation 12.xxa gives the inbreeding in Syn generation $t > 1$ as

$$f_t = r_1\left[1 - \left(\frac{k-1}{2k-1}\right)^{t-1}\right] + f_1\left(\frac{k-1}{2k-1}\right)^{t-1} \tag{12.xxc}$$

Hence, the equilibrium level of inbreeding approaches r_1, or

$$f_e = r_e = \frac{1}{n}\left(\frac{1+(2k-1)f_0}{2k}\right) + \left(\frac{n-1}{n}\right) r_0 \tag{12.xxd}$$

Note for a diploid ($k = 1$) that f_e is reached after one generation of mating (i.e., in the Syn 2). For a tetraploid, the equilibrium level of inbreeding is approached

asymptotically (but is essentially at f_e after 4-5 generations). In the most common setting, the founding parents are unrelated, so that $r_0 = 0$, and the equilibrium inbreeding reduces to

$$f_e = \begin{cases} 1/(2kn)\,[\,1 + (2k-1)f_0\,] & \\ 1/2kn & \text{for } f_0 = 0 \\ 1/n & \text{for } f_0 = 1 \end{cases} \qquad (12.\text{xxe})$$

Hence, if the parents are unrelated and completely inbred, the equilibrium inbreeding is simply one over the number of parents (independent of the ploidy), while if the parents are not inbred, the equilibrium inbreeding is a function of both n and ploidy level.

Busbice also consider the most general case, where individuals can also self (with constant probability s). This is certainly a reasonable model for many crop species. Busbice found that $r_e \simeq r_1$ is independent of s (this follows because all parents contribute equally to the next generation), and hence is given by Equation 12.xxb. The equilibrium inbreeding is given by

$$f_e = r_1 + \frac{s(1-r_1)}{s + 2k(1-s)} \qquad (12.\text{xxg})$$

This reduces to r_1 when $s = 0$ (no selfing) and to 1 when $s = 1$ (complete selfing). The expected inbreeding in the Syn 1 generation under partial selfing becomes

$$f_1 = \frac{k(1-s)r_0 + (s/2)(1 + (2k-1)f_0) + (k-1)f_0}{2k-1} \qquad (12.\text{xxh})$$

For a diploid with unrelated parents ($r_0 = 0$), this reduces to $f_1 = (s/2)(1 + f_0)$. Finally, the expected inbreeding in generation $t > 1$ is

$$f_t = f_e \left[1 - \left(\frac{s}{2} + \frac{k-1}{2k-1} \right)^{t-1} \right] + f_1 \left(\frac{s}{2} + \frac{k-1}{2k-1} \right)^{t-1} \qquad (12.\text{xxi})$$

Busbice's Generalization of Wright's Predictor

In the absence of epistasis, the mean of the equilibrium synthetic population is simply a function of the allele frequencies in the founding population and the amount of inbreeding in the equilibrium population. Busbice (1970) used his general inbreeding expression to obtain a very general extension of Wright's predictor, allowing for the effect of ploidy, partial inbreeding in the parents, selfing among the Syn progeny, and relatedness of founding parents. Busbice stars with the assumption that yield Y_t in the Syn t generation is a linear function of inbreeding, with

$$Y_t = A + (1 - f_t)B \qquad (12.\text{meee})$$

where f_t is the inbreeding in Syn t. This is just a restatement of the assumption that yield increases linearly with the amount of heterozygosity, which is reduced by inbreeding. The relationship between yield and percent heterozygosity (which we will call H here) has been examined by a number of authors and generally found to be linear, or at a minimum have a very significant linear trend. For example, Sentz et al. used inbred line crosses to measure trait values in the F_1 (100% heterozygosity), the F_2 and simple backcrosses ($H = 50\%$), the parental lines ($H = 0\%$), the second backcross $B_2 = P_i \times (F_1 \times P_i)$ where $H = 25\%$ and the double backcross $P_1 \times (F_1 \times P_2)$ with $H = 75\%$. While the regression of traits on H was generally curvilinear (i.e. significant quadratic or higher order terms), the bulk of variation was still accounted for by the linear part of the regression. Stringfield (1950) also found that the regression of yield on H was essentially linear.

Busbice's expression for Y_t is a simple linear system of equations and hence if we know the yield values for two generations (and the corresponding level of inbreeding), we can solve for A and B. Using the yields Y_0 and Y_1 for Syn 0 and Syn 1 generations,

$$Y_0 = A + (1 + f_0)B, \quad \text{and} \quad Y_1 = A + (1 + f_1)B$$

yielding

$$A = Y_0 + \frac{f_0 - 1}{f_0 - f_1}\,(Y_1 - Y_0)\,, \quad B = \frac{Y_1 - Y_0}{f_0 - f_1}, \quad \text{for } f_0 \neq f_1$$

Hence, the general expression for the performance in Syn generation t is given by

$$Y_t = Y_1 - \frac{f_t - f_1}{f_0 - f_1}\,(Y_1 - Y_0) \tag{12.xx}$$

and in particular equilibrium mean in the synthetic populations becomes

$$Y_\infty = Y_1 - \frac{f_e - f_1}{f_0 - f_1}\,(Y_1 - Y_0) \tag{12.xx}$$

Thus the effect of the number of founders (n), how they are related (r_0), ploidy levels (k), and partial selfing (s) on the mean of the synthetic all enter through the ratio of inbreeding coefficients. Table 12.3 gives values of this ratio under several common situations. Note that the above derivaition requires that $f_0 \neq f_1$. If parents are unrelated ($r_0 = 0$ and noninbred $f_0 = 0$), then $f_0 = f_1$ and the Syn 1 population is immediately in Hardy-Weinberg equilibrium.

Table 12.3. The reduction $(f_1 - f_e)/(f_0 - f_1)$ of the initial heterotic advantage, $Y_1 - Y_0$ under different assumptions about ploidy ($2k$), parental inbreeding (f_0) and levels of selfing (s).

Parents completely inbred ($f_0 = 1$)

$$\frac{f_1 - f_e}{f_0 - f_1} = \begin{cases} \dfrac{2 - ns}{n(2-s)} = \dfrac{1}{n} \text{ (for no selfing, } s = 0) & \text{Diploids} \\ \dfrac{6 - n(2+3s)}{n(4-3s)} = \dfrac{3-n}{2n} \text{ (for } s = 0) & \text{Tetraploids} \\ \dfrac{10 - n(4+5s)}{n(6-5s)} = \dfrac{5-2n}{3n} \text{ (for } s = 0) & \text{Hexaploids} \end{cases}$$

No selfing ($s = 0$), parents unrelated ($r_0 = 0$)

$$\frac{f_1 - f_e}{f_0 - f_1} = \begin{cases} \dfrac{1+f_0}{2nf_0} & \text{Diploids} \\ \dfrac{3 + f_0(3-2n)}{4nf_0} & \text{Tetraploids} \\ \dfrac{5 + f_0(25-12n)}{18nf_0} & \text{Hexaploids} \end{cases}$$

Neighter the Wright or Busbice predictors are valid when epistasis is present. When significnat epistatic interactions are present, the mean of the synthetic population will likely not be at equlbrium in the Syn-2 generation, even for a non-selfing diploid. While Hardy-Weinberg equilibrium is reached after a singel generation of random mating, linkage disequilibrium is halfed each generation for unlinke dloci, but changes much more slowly for linked loci. Since the decay of linkage diseuqilbrium (which would be generated by the crossing for loci that differ in allele frequnecies between lines) changes the gamete frequencies, an equilibrium in the mean value is not reached until linkage equilibrium between contributign epistatic loci is reached. Gallais (1975a) provides an expression that allows for additive by additive epistasis, which is as follows. Consider a synethic constructed from n lines, whose equilibrium value we denote by $Y_\infty(n)$. Denote the mean equilbrium value over all synthetics involving $m < n$ lines from the original n lines as $\overline{Y}_\infty(m)$ and likewise let $\overline{Y}_\infty(k)$ denote the mean involving $k < m$ lines. Gallais' predictor is

$$\overline{Y}_\infty(n) = \frac{m(n-k)\overline{Y}_\infty(m) - k(n-m)\overline{Y}_\infty(k)}{n(m-k)} \qquad (12.\text{xxa})$$

Making all these requires crosses invovles a considerable effort, and this still only corrects for additive by additive epistasis only.

Synthetics Based on the Offspring of Selected Parents

The n parents chosen to found a synthetic may themselves be related. A common situation is where n_f unrelated families are chosen, each of which contributions n_o offspring, where $n = n_f\, n_o$. For diploids, if we let the parents of the families be potentially inbred with coefficient f_0, then with probability $1/n_f$ two randomly-chosen individuals are from the same family, so that $r_o = 2\Theta/n_f$ where 2Θ is the coefficient of coancestry for the family in question. In particular,

$$2\Theta = \begin{cases} (1+f_0)/4 & \text{half-sibs} \\ (1+f_0)/2 & \text{full-sibs} \\ (1+f_0)/2 & S_1 \text{ sibs} \\ 1-2^{-k}(1-f_0) & \text{sibs from an } S_k \text{ line} \\ 1 & \text{fully inbred lines} \end{cases} \tag{12.xx}$$

Substitution of the appropriate r_0 value allows us to compute the correct inbreeding coefficient and estimate the mean of the synthetic line.

Example 12.xx: Suppose one has n parents and considers two possible ways to form a synthetic. The Syn 0 could consist of the n parents, and we denote the equilibrium level of inbreeding by f_e. One potential drawback with this particular Syn 0 is that the total amount of seed in the Syn 1 may be small. If the species can be selfed, one could instead from the Syn 0 by taking the collection of selfed progeny (S_1 families) from these parents. The resulting Syn 0 population should be roughly the size of the Syn 1 population using just the parents. Denoting the equilibrium level of inbreeding using S_1 families are f'_e, how comparable are equilibrium levels of inbreeding under these two different synthetics? We show here that the levels of inbreeding are identical ($f_e = f'_e$) for the special case of a diploid with no selfing and the initial n parents are unrelated ($k = 1, s = 0, r_0 = 0$) that they are equal. Buschice (1970) shows this more generally for arbitrary ploidy, selfing, and allowing the initial parents to potentially be related ($r_0 \neq 0$).

For the synthetic where the Syn 0 consist of the n parents, Equation 12.xxg gives

$$f_e = r_1 = \frac{1+f_0}{2n}$$

where f_0 is the (average) level of inbreeding among the founding parents.

Now consider the situation where the Syn 0 consists of the S_1 progeny of these n parents. Let f'_0 and r'_0 denote the initial inbreeding and the degree of relatedness between the Syn 0 individuals. Suppose each selfing produce m offspring, implying that the Syn 0 has $n' = n\,m$ individuals. With probability $1/n$ two random Syn 0 individuals come from the same S_1 family and are related, otherwise they come from different families (and are unrelated). The coefficient of coancestry

for sibs from an S_1 family is $(1 + f_p)/2$. Since the average inbreeding of the S_1 parents is $f_p = f_0$, the relationship among two random individuals in the Syn 0 formed from S_1 families becomes

$$r'_0 = \frac{1 + f_0}{2n}$$

Note from Equation 12.xxb that if n' is large (as would be the case using S_1 families to form the Syn 0), then $r'_1 \simeq r'_0$, and hence $f'_e = r'_1 \simeq r'_0$. Noting the $r_1 = r'_0$ shows that $f'_e = f_e$, so that the equilibrium level of inbreeding the same whether the Syn 0 consists of n parents or the n selfed (S_1) families descant from them.

The mindset of the early corn breeders was often to focus on inbred lines. If one wishes to construct a synthetic from a subset of individuals in an open pollinated population, considerable time and effort must be expended if we first construct fully inbred lines from this population before constructing the synthetic. A short-cut was suggested by Jenkins (1940), who advocated the use S_1 individuals (selfed) individuals, as opposed to their completely-inbred relatives, to form a synthetic. His idea was that S_1 progeny had similar combining ability to their inbred relatives (and hence similar Y_1 values when used in crosses), but required far less effort to generate. Here one first selfs a number of plants, and then top-crosses their S_1 offspring to find those individuals with superior general combining ability. Remnant S_1 seed are then used to form the synthetic. Given that Jenkins (1935) and Sprague (1946) found that topcross performance is relatively constant after the S_1 generation, the GCA values from S_1 lines is expected to be similar to that for a collection of fully-inbred lines derived from them. Lonquist (1949) used this approach to construct synthetics that out-yielded the open-pollinated variety (OPV) from which the S_1 lines were extracted. Appropriately 200 plants from the open pollinated variety were selfed, from which 36 plants were selected for future work. A fraction of seeds from each saved S_1 ear were planted in rows, detassleted (to prevent selfing) and top-crossed using the OPV at the pollen parent. GCA values for each plant were estimate from the bulk of all seeds from a row, and the best plants chosen, and their remnant seed used to construct the synthetic.

Kinman and Sprague (1945) and other authors (e.g., Allard 1960) have adocated a second reason to use S_1, namely that they can increase the yield of a synthetic. Their reasoning that the mean mean yield of S_1 parents should be greater than the mean yield of fully-inbred lines deived from them (which would suffer greater inbreeding depression). However, the equilibrium mean of a synthetic (assuming no epistasis) depends only on the allele frequencies in the founding population and the equilibrium level of inbreeding. Allele frequencies are unchanged in S_1's and their fully inbred descendant lines (assuming no selection).

Autopolyploids

Synthetics formed with diploids have two important features. First, provided selfing does not occur, the equilibrium level of inbreeding is reached in one generation by randomly pollinating the Syn-1 population. Provided epistatic contributions are mimial, the mean of the Syn-2 population is also at its equilibrium value. Second, the amount of heterosis declines from its maximal vlaue in the Syn-1 to its equilbrium value in the Syn-2. When dealing with autopolyploids, for example autotraploids such as potatoes and the forgae crops alfalfa (*Medicago sativa* L.) and birdsfoot trefiol (*Lotus corniculatus* L.), populations take many generations to reach their equilibrum inbreeding value and the mean of the population can *increase* from its value in the Syn-1 generation. This later point arises because the inbreeding coefficient in the Syn-1 generation equals

$$f_1 = \frac{kr_0 + (k-1)f_0}{2k-1}$$

which can be greater than the equilibrium level of inbreeding (Dudley 1964, Dessureaux and Gallais 1969, Busbice 1969). For example, for a tetraploid with completely unrelated ($r_0 = 0$) but fully inbred ($f_0 = 1$) parents, sign$(f_1 > f_e) =$ sign$(3 - n)$. Thus for $n = 2$, the $f_1 > f_e$ and the mean is expected to decline from the Syn-1 value, while for $n \geq 4$, the mean is expected to increase each generation as the equilbrium level of inbreeding (which is less than the syn-1 level) is approached.

However, there is another complication with autoploids, namely the the inbreeding coefficient itself can be insufficent to fully describe inbreeding (Busbice and Wilsie 1966, Gallais 1967). Thus occurs because while a diploid locus has only two possible genetic states (homozygotes and heterzygotes), an autopolyploid locus can have up to five different genetic states:

State	Representive genotype	Num. of alleles
Monogenic	$A_iA_iA_iA_i$	1
Digenic simplex	$A_iA_iA_iA_j$	2
Digenic duplex	$A_iA_iA_jA_j$	2
Trigenic	$A_iA_iA_jA_k$	3
Tetragenic	$A_iA_jA_kA_\ell$	4

Demarly (1963) to suggested that the interaction between multiple alleles in the tri- and tetragenic states can potentially contribute further heterosis, so that the greatest effect would occur in tetragenic individuals. Thus, while autotetraploid have a slower approach to complete homozygosity than to diploids (Demarly 1963, Busbice 1969), they usually show greater inbreeding depression, most likely due to the decay of multialleic heterozygosity.

Dunbier and Bingham (1975) essentially confirmed the maximum heterozygosity hypothesis for yield in alfalfa through the clever use of haploid-derived autotetraploids (denoted by HD4x). A number of such lines are developed and

used to generate single and double crosses. Single crosses would be expected to carry no more than two alleles, while double crosses potentially could involve four alleles. Dunbier and Bingham observed an average performance ranking of DC > SC > HD4x, as expected from the maximum heterozygosity hypothesis. They note that this has important implications for breeding, comparing the expected frequencies of the different allelic classes tow two different situations involving four unrelated duplex clones (A, B, C, D): an equilibrium synthetic of all four versus the double cross AB $\times$ CD

State	freq in Syn	freq in DC
Monogenic	0.2%	0.0%
Digenic	9.6%	1.2%
Trigenic	49.2%	19.8%
Tetragenic	41.0%	79.0%

If tetragenic interactions are important congtribtuors to heterosis, we would expect a significant decrease between the double cross and syntethic.

Composites

A synthetic population formed using individuals from open-pollinated populations (as opposed to inbred lines) is usually referred to as either a **composite** or a **composite vareity**. Often Syn 0 generation of a composite consists of equal mixture of seed from a number of parents/varietal lines. For a diploid population, the Syn 1 generation, formed by randomly mating among the Syn 0 seed is immediately in Hardy Weinberg equilibrium, and (in the absence of epistasis) the mean also reaches its equilibrium value. Assuming fairly large numbers of individuals are used for each line, then if there is no inbreeding within each line, there is no inbreeding in the Syn 0 or Syn 1 population, and Busbice's predictor of the mean of a synthetic cannot be used. However, if C denotes the mean yield over a set of controlled crosses between the lines and Y_0 the mean yield over all the parential lines, then the equilibrium yield in the composite using n equally-frequent founding populations is just

$$Y_e = C - \frac{C - Y_0}{n} \tag{12.xx}$$

This was first suggested by Eberhart et al. (1967). While Equation 12.xx looks very much like Wright's formula, its derivation does not depend on inbreeding. Busbice's argument to obtain this expression is as follows: at equilibrium, $1/n$ of the individuals result from crosses between individuals from the same population, while the rest $(1-1/n)$ of the individuals result from crosses involving individuals from two different populations. Hence

$$Y_e = C\left(1 - \frac{1}{n}\right) + \frac{Y_0}{n} = C - \frac{C - Y_0}{n} \tag{12.xx}$$

Now suppose that their are unequal contributions among lines. Among the n lines, the π_i denote the fractional contribution from population i. Assuming all starting populations are in Hardy-Weinber equilibrium and that we can ignore epistasis, then the logic used in Equation 12.xx (WRIGHAT'S predictor) suggest the following predictor for the equilibrium value of a composite

$$\widehat{\mu} = \sum_{i,j}^{n} \pi_i \, \pi_j \, \overline{z}_{ij} \tag{12.xVx}$$

where $\overline{z}_{ij}$ is the observed mean for a cross involving line i as the male and line j as the female parents.

Example 12.Vx: Consider the following diallel data set, where the diagonal elements are the line means and the ijth diagonal element the $i \times j$ cross (ignoring parental order):

	Line A	Line B	line C
Line A	50	70	65
Line B		60	65
Line C			55

Considering all possible two-line synthetics, which two lines, and at which frequnecies, generates the composite with the largest mean? Consider A and B first, and let π denote the fraction of A (and hence $1 - \pi$ is the fraction of B. The result line mean is

$$\mu_{A,B}(\pi) = \pi^2 \, 50 + 2\pi(1-\pi) \, 70 + (1-\pi)^2 \, 60$$

Since

$$\frac{d\mu_{A,B}(\pi)}{d\pi} = 20 - 60\pi$$

the maximum value is give by $\pi_{max} = 20/60 = 1/3$ and the result composite has an equilibrium mean of 63.33. In a similiar fashion, we find for A and C that $\pi_{max} = 2/5$ with corresponding mean 59.0, while B and C, $\pi_{max} = 2/3$ with mean 61.67. Hence, the best two-line composite is given by a founder population of $1/3$ A and $2/3$ B.

We can use this same approach to compute the expected value for a cross between two composites. For example, consider the cross is between C_1 and C_2, where the composite C_j consists of n_j varities, the ith of which, P_{ji}, makes up a fraction π_{ji} . Assume that all the founding populations, and both composites, are

in Hardy-Weinberg equilibrium, and that we can ignore the effects of epistasis. The expected mean of the cross $C_1 \times C_2$ is given by

$$\mu_{C_1 C_2} = \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} \pi_{1j}\, s\pi_{1k}\, \mu_{jk} \tag{12.xx}$$

where μ_{ij} si the mean of a cross between lines i and j. Thus from a diallele analysis invovlign all lines of interest, we can compute the expected means of the F_1 cross between any two composites contructions from these lines. Likewise, by the same agrument, the mean in the F_2 (and subsequent) generations is given by

$$\mu = \sum_{j=1}^{n} \sum_{k=1}^{n} \pi_i\, \pi_j\, \mu_{ij} \tag{12.xx}$$

where there are a total of $n \leq n_1 + n_2$ lines, the ith of which contributions a fraction π_i to the final composite.

Example 12.Vx: Suppose a fourth line, D, was examined in a diallel ivolves lines $A - C$ in the presiovus example, with

	Line D	Line A	Line B	line C
Line D	60	55	85	70

Suppose were from two composiutes, the first with equal mixtures of lines A and C, the second with equal mixtures of lines B and D. The equilibrium means of these composites are

$$\mu_{AC} = \frac{\mu_A + \mu_C + 2\mu_{AC}}{4} = \frac{50 + 55 + 2 \cdot 65}{4} = 58.75$$

$$\mu_{BD} = \frac{\mu_B + \mu_D + 2\mu_{BD}}{4} = \frac{70 + 60 + 2 \cdot 85}{4} = 75$$

The expected F_1 in the cross between these two composites is

$$\mu_{F_1} = \left(\frac{A}{2} + \frac{C}{2}\right) * \left(\frac{B}{2} + \frac{D}{2}\right) = \frac{\mu_{AB} + \mu_{AD} + \mu_{CB} + \mu_{CD}}{4}$$

$$= \frac{70 + 55 + 65 + 70}{4} = 65$$

The expected F_2 (and equilibrium) mean is given by

$$\mu_{F_2} = \left(\frac{A}{4} + \frac{B}{4} + \frac{C}{4} + \frac{D}{4}\right)\left(\frac{A}{4} + \frac{B}{4} + \frac{C}{4} + \frac{D}{4}\right)$$

$$= \frac{\mu_A + \mu_B + \mu_C + \mu_D}{16} + \frac{\mu_{AB} + \mu_{AC} + \mu_{AD} + \mu_{BC} + \mu_{BD} + \mu_{CD}}{8}$$

$$= 65.3125$$

In this case, the mean increases from the F_1 to the F_2.

Synthetic Values and Syntheizing Ability

The prediction formulae for the value of a synthetic presented above require controlled crosses, for example the vaue of the first generation syntetic. If we have a number of lines from which a subset will be draw as paretns for a syntethic line, it is generally not feasibale to make all appropriate crosses to find the best combination of lines. One approach, as we have seen before with hybrids, is to use the general combining ability of the lines to reduce set of potential parents. The GCA for any line can easily be estimated for by using a topcross with all the other lines, with $GCA_i = \overline{T}_i - \overline{\overline{T}}$. However, as we have noted above, syntehtics are inbred, while the general combining ability is a measure in an outbred population. Wright and Gallais have provided a correction, the **synthetic value** SV_i and the related **genreal syntheizing ability** GSA_i. Let I_i denote the mean value in the progeny formed by selfing parents from line i, and let $v_i = I_i - \overline{I}$. For a synthetic diploid population at equilibrium formed from equal contrbutions of n lines, $1/n$ of the crosses involve gametes from the same populations, while the rest invovle gametes from two different populations. Using this, the syntehtic value SV_i of line i in a synthetic if size n is given by

$$SV_i^n = \frac{v_i + 2(n-1)GCA_i}{n} = \frac{v_i}{n} + 2\,GCA_i\left(1 - \frac{1}{n}\right) \qquad (12.\text{xxa})$$

Note that for large n, the synthetic value is essentially twice the general combining ability. A closely realted measure is the general syntehzing ability (also refered to as the **general varietal ability**) which is the mean value of all synthetics of size n that have i as one of their parents, where

$$GSA_I^n = \frac{SV_i}{k} = \frac{v_i + 2(n-1)GCA_i}{n^2} = \frac{v_i}{n^2} + \frac{2(1-1/n)GCA_i}{n} \qquad (12.\text{xxb})$$

The mean of a syntethic value can be predicted by the sum of the GSA's,

$$Y_\infty(n) = \sum_{i=1}^{n} GSA_I^n = \frac{\sum_i v_i + 2(n-1)\sum_i GCA_i}{n^2}$$

Gallais' Test and Varietal Values

Gallais (1978, 1979, 1990) the **general varietal value** of a genotype is the expected value for all varieties of a given type that can be extracted from it. Likewise, **specific varietal values** can be obtained for multiparental varieties such as synthetics and hybrids. Clone value

line value

$$L(A_iA_j) = \mu^L + \alpha_i^L + \alpha_i^L \tag{12.xx}$$

where $\mu_l = \mu + \sum p_i\delta_{ii}$ as $\sum\ p_i\,\alpha_i = 0$

$$\alpha_i^L = (1/2)(2\alpha_i + \delta_{ii}) = \alpha_i + \delta'_{ii}/2 \tag{12.xxb}$$

where $\delta'_{ii} = \delta_{ii} - \sum p_i\delta_{ii}$ as the genotypic value of A_iA_i is $2\alpha_i + \delta_{ii}$.

$$\sigma^2_{A_L} = 2\sigma^2_A + \sigma_{ADI} + \sigma^2_{DI} \tag{12.xx}$$

Synthesizing ability

DETECTION OF LINES CARRYING NEW ELITE ALLELES

The goal of breeders is the continued accumulation of favorable (or **elite**) alleles in a line or population. Suppose I_1 and I_2 are two such pure (completely homogenous) elite lines. While breeding and selection has concerted elite (favorable) alleles in these lines, there are certainly other favorable alleles not present in these lines. If I_x denotes another pure line, or more generally P_x another population (potentially segregating), we are very interested in determining whether this candidate line/population contains additional favorable alleles. Likewise, given a collection of lines/populations, we would like to determine which contains the most new favorable alleles. Further, given we have identifies such lines, what is the best breeding strategy to introgress their new favorable alleles into the existing elite lines. This problem has been examined by Dudley (1982, 1984a, 1984b, 1984c, 1987a, 1987b) and others. As our discussion should make clear, there is clearly a need for further improvement in these methods, and this is an important area of future research. The use of molecular markers obviously facilitates identification and introgression of elite alleles, and we review this in Chapter 26. Our focus here is on using line-cross means to identifying lines carrying additional elite alleles.

There are two different, but related, goals in improving a pure line. For hybrid breeding, the goal is to improve one (or both) of the elite lines through the introgression of new alleles such that the hybrid between these improve lines exceeds the performance of the hybrid between the original elite lines. Suppose I_1 is the elite parent to be improved by introgressing alleles for I_x. The goal is to extract a line (or lines), say I_{1x^*}, from $I_1 \times I_x$ cross such that $I_{1x^*} \times I_2$ has better performance than $I_1 \times I_2$. We will use the notation here that I_2 is the **tester** line, and the various lines extracted from the $I_1 \times I_x$ cross which are them crossed to

the tester generate the **testcross population**. Note that the extracted lines may result from one (or more) generations of either selfing or backcrossing the F_1 population to generate the I_{1x^*} lines to be tested. A related goal would be to improve the resulting pure lines extracted from the $I_{1x^*} \times I_2$ cross.

Both pure lines (I_x) and segregating populations (P_x) can be candidates for new elite alleles. If the new elite alleles have large qualitative effects (such as disease resistance), the choice of lines may be obvious. More generally, with a complex trait (such as yield), choosing the best line from a collection of candidates is not obvious. One approach is to chose among the lines based solely on their means $\overline{z}_i$ (often called **per se selection**, or selection of performance per se). There is certainly no guarantee that per se selection chooses the line with the most new elite alleles. For example, the best performing candidate line may contain only elite alleles that are already present in I_1 and/or I_2. Clearly, line cross information is needed to make an informed decision, and a number of different estimators have been proposed.

Triple Cross Estimators

Given that our goal is to find those lines that given high performance to trip-cross hybrids (lines derived from a cross between one elite line and a candidate are crossed to the other elite line), it is not surprusing the estimators based on the triple cross have been suggested to identify candidate lines.

Two simple line-cross measures have been proposed, the **predicted triple cross** (denoted PTC) estimator of Sprague and Eberhart (1977) which requires two single-crosses ($I_1 \times I_x$ and $I_2 \times I_x$) and by the observed value of the triple cross ($[\, I_1 \times I_2 \,] \times I_x$), which is also referred to as the **test cross to a single cross**, as proposed by Gerloff and Smith (1988a, b). The predicted triple test cross value is given as the average of the single crosses of I_1 and I_2 to I_x,

$$\text{PTC} = \frac{\overline{z}_{1x} + \overline{z}_{2x}}{2} \tag{12.A1}$$

where $\overline{z}_{ij}$ is the observed mean in a cross between I_i and I_j. With n candidate lines, this involves making and measuring $2n$ crosses. The line with the best PTC value with I_1 and I_2 is chosen as containing the most elite alleles not present in the two elite lines. The test cross to a single cross involves the triple cross formed from crossing I_x to the F_1 from $I_1 \times I_2$,

$$\text{OTC} = \overline{z}_{x \cdot 12} \tag{12.A2}$$

where $\overline{z}_{x \cdot 12}$ denotes the observed mean value of this triple cross (this has been denoted as both TC(SC) and TCSC, for test cross to a single cross, in the literature). Again, the line with the largest value is used. Here, to test n lines, as few as $n + 1$ crosses are required, provided that the $I_1 \times I_2$ cross can provide enough parents for crosses to the I_x.

Dudley's μG and $\ell\,\mu\,\overline{p}_\ell$ Estimators

Following Dudley (1984a, b), a more direct approach is to partition the loci for I_1, I_2 and I_x into eight classes, denoted A through H (Table 12.A1). Loci in Dudley's **Class G** contain favorable alleles in I_x that are missing from both elite lines (I_1 and I_2). The number of loci Class G thus provides a direct measure for choosing between candidate lines. Other classes that are of interest are Class C (I_1 and I_x contain favorable alleles missing in I_2) and Class E (I_2 and I_x contain favorable alleles missing in I_1). Finally, Classes D and F refer to loci with favorable alleles only found in I_1 and I_2, respectively. The joint classes of **D + E** and **C + F** also appear in the literature. Loci in classes D + E have the same states in I_2 and I_x and different states in 1_1, as Class D loci are both less favorable ($-$) in I_2 and I_x but more favorable ($+$) in I_1, while Class E loci are both favorable in I_2 and I_x while being less favorable in I_1. Similarly, C + F is a measure of alleles common to I_1 and I_x, but not I_2.

Table 12.A1. Genotypes at a given locus for the two elite pure lines I_1 and I_2 and a pure line I_x being tested for additional favorable alleles. Here $+$ denotes a favorable allele, $-$ a less favorable allele.

Class of loci	I_1	I_2	I_x
A	$++$	$++$	$++$
B	$++$	$++$	$--$
C	$++$	$--$	$++$
D	$++$	$--$	$--$
E	$--$	$++$	$++$
F	$--$	$++$	$--$
G	$--$	$--$	$++$
H	$--$	$--$	$--$

To estimate the contribution from each class, additional assumptions are required. Dudley (1984a,b) initially assumed that the values of the genotypes are the same at each locus, with (in our standard notation) $++ = c+2a$, $+- = c+ad$, $-- = c$. Note that this is a different notation from that used in Dudley (and related papers), who use the notation $++ = z + 2\mu$, $+- = z + a\mu$, $-- = z$ (so that our c corresponds to their z, our a to their μ and our d to their a). We use the notation n_x to denote the number of loci in class x, while Dudley (and related papers) use (say) G interchangeably for both the locus class and the number of loci in that class. Estimators are obtained by first expressing the various line-cross means in terms of the n_x and the genotype parameters (μ, a, k), and then solving for the desired n_x. For example, the mean of line I_x can be expressed as

$$\mu_x = 2a(n_A + n_C + n_E + n_G) + \mu(n_A + n_B + n_C + n_D + n_E + n_F + n_G)$$

while the mean for $I_x \times I_2$ can be expressed in this format by noting in the F_1 that all Class A and C loci are $++$, all Class B, D, E, and G loci are $+-$, and all class F and H loci are $--$, giving

$$\mu_{1x} = a(n_A + n_C) + ad(n_B + n_D + n_E + n_G) + \mu n$$

where $n = n_A + n_B + n_C + n_D + n_E + n_F + n_G + n_H$ is the number of loci influencing the trait.

By making three additional assumptions, namely that: (i) $c = 0$, (ii) $n_A = n_H$, and (iii) complete dominance ($d = 1$), Dudley was able to solve for $a\,n_G$ (the number of unique favorable loci in a candidate times their effect) in terms of the line cross means. The resulting estimator is named μG (which in our notation translates to $a\,n_G$, the genotypic value a times the number of loci n_G), where

$$\widehat{\mu G} = \frac{\overline{z}_{1x} + \overline{z}_{2x} - \overline{z}_x - \overline{z}_1 - \overline{z}_2 - \overline{z}_{12}}{4} \tag{12.A3a}$$

If the F_1 do not produce sufficient seed for suitable replication, then F_2s can be used instead, in which case Dudley (1984a) gives the estimator as

$$\widehat{\mu G}_{F_2} = \frac{2\,\overline{z}_{1x\cdot 1x} + 2\,\overline{z}_{2x\cdot 2x} - 2\,\overline{z}_x - \overline{z}_1 - \overline{z}_2 - 2\,\overline{z}_{12\cdot 12}}{4} \tag{12.A3b}$$

where $\overline{z}_{ij\cdot ij}$ is the observed F_2 mean of the $I_i \times I_j$ cross, i.e., the cross $(I_i \times I_j) \times (I_i \times I_j)$. Table 12.A2 gives similar estimates for the other locus classes (based on Dudley's 1984a assumptions). Dudley suggests that approximate variances on the estimator can be obtained by ignoring any potential covariances between the sample mean and using the variance of a sum. The sampling variance for $\widehat{\mu G}$ (Equation 12.A3a) is thus

$$\text{Var}\left(\widehat{\mu G}\right) = \frac{\text{Var}(\overline{z}_{1x}) + \text{Var}(\overline{z}_{2x}) + \text{Var}(\overline{z}_x) + \text{Var}(\overline{z}_1) + \text{Var}(\overline{z}_2) + \text{Var}(\overline{z}_{12})}{16} \tag{12.A3c}$$

Note that this is also the sampling variance for the other estimators listed in Table 12.A2.

As a point of comparison, under the genetic assumptions leading the Equation, the expected values of the two triple-cross based estimators (PTC and TC[SC]) are identical and equal to

$$E[\,PTC\,] = \mu_{x\cdot 12} = \mu n_G + \mu(n + n_B + n_C + n_E)$$

so that both these estimators overestimate μn_G under these assumptions.

Table 12.A2. Dudley's (1984a) estimator for the number of loci in Classes A through F. The assumptions used are the same as those leading to Equation 12.A3. Expressions using the F_2 in place of the F_1 are given in Dudley (1984a).

$$
\begin{aligned}
4\,\widehat{\mu B} &= 4\,\widehat{a\,n_B} = \overline{z}_{1x} + \overline{z}_{2x} - \overline{z}_{12} + \overline{z}_1 + \overline{z}_2 - \overline{z}_x \\
4\,\widehat{\mu C} &= 4\,\widehat{a\,n_C} = -\overline{z}_{1x} + \overline{z}_{2x} + \overline{z}_{12} + \overline{z}_1 - \overline{z}_2 + \overline{z}_x \\
4\,\widehat{\mu D} &= 4\,\widehat{a\,n_D} = \overline{z}_{1x} - \overline{z}_{2x} + \overline{z}_{12} - \overline{z}_1 - \overline{z}_2 - \overline{z}_x \\
4\,\widehat{\mu E} &= 4\,\widehat{a\,n_E} = \overline{z}_{1x} - \overline{z}_{2x} + \overline{z}_{12} - \overline{z}_1 + \overline{z}_2 + \overline{z}_x \\
4\,\widehat{\mu F} &= 4\,\widehat{a\,n_F} = -\overline{z}_{1x} + \overline{z}_{2x} + \overline{z}_{12} - \overline{z}_1 - \overline{z}_2 - \overline{z}_x
\end{aligned}
$$

As mentioned, these estimators are biased if there is not complete dominance. If our goal is to isolate completely recessive alleles, we can still use these equations, by treating $-$ alleles as favorable Zanoni and Dudley 1989). In this case, our goal is to find the line with the largest value for Class B (dominant $+$ alleles in both elite lines, recessive $-$ alleles in the candidate). To express these alleles in a hybrid (say by crossing the F_1 from $I_1 \times I_x$ to I_2), one would chose the candidate with the largest value of Class D, as the recessive allele is present in both I_2 and I_x, but not I_1. In the resulting $(I_1 \times I_x) \times I_2$ hybrid, loci in this class would be half $--$ and half $+-$.

To deal with arbitrary levels of dominance (but still keeping the assumptions $n_A = n_H$ and $c = 0$), Dudley (1984a) showed that the expected value of Equation 12.A3a, is

$$E[\,\widehat{\mu G}\,] = (a/2)(\,n_G\,[\,1+d\,] - n_B\,[\,1-d\,]\,)$$

so that $\widehat{\mu G}$ underestimates $a\,n_G$ if $0 < d < 1$, while with overdominance ($d > 1$), $\widehat{\mu G}$ overestimates $a\,n_G$.

The above estimators are for testing candidate *inbred lines*. However, we may also wish to test different candidate *populations* for favorable alleles. In this case, while both elite inbred lines (I_1 and I_2) are fixed for alleles, a candidate population P_x is potentially segregating for faovrable and less favorable alleles at any given loci (Dudley 1984b). In such cases, Dudley's eight allelic classes (Table 12.A1) reduce to four classes (Table 12.A3), which he denotes $i = A + B$ (loci fixed for favorable alleles in both elite lines), $j = C + D$ (favorable alleles fixed in I_1, absent in I_2), $k = E+G$ (favorable alleles fixed in I_2, absent in I_2), and $\ell = G+H$ (loci fixed for unfavorable alleles in both elite lines). Assuming all loci have genotypic values of $c + 2a\ :\ c + 2a\ :\ c$ (i.e., complete dominance) and assuming that $\overline{p}_i = \overline{p}_j = \overline{p}_k$, then an estimator of $n_\ell\,a\,\overline{p}_\ell$, is

$$\widehat{n_\ell\,a\,\overline{p}_\ell} = \frac{(\,\overline{z}_{1x} - \overline{z}_1\,)(\,\overline{z}_{12} - \overline{z}_2\,) - (\,\overline{z}_{2x} - \overline{z}_2\,)(\,\overline{z}_{12} - \overline{z}_1\,)}{2\,(\,\overline{z}_1 - \overline{z}_2\,)} \qquad (12.A4)$$

Notice that this estimator does not require the means from each of the candidate populations, $\overline{z}_x$, and hence requires only growing $2n + 3$ populations to test n populations.

Table 12.A3. Dudley's (1984b) loci classes when testing two elite inbred lines (I_1 and I_2) against a (potentially segregating) population P_x. Loci in Class i are fixed for favorable alleles in both elite lines, Class j fixed in elite line one and absent in elite line 2, the reverse for Class k, while Class ℓ are loci that are fixed for less favorable alleles in both elite lines. $\overline{p}_y$ is the frequency of favorable alleles in class y in the population (P_x) of interest. The last column (I_x) gives the corresponding p values for the four classes in terms of the number of loci in each of Dudley's eight classes (Table 12.A1) when the population being tested is completely inbred.

	Allele Frequencies			
Class of loci	I_1	I_2	P_x	I_x
i	1.0	1.0	$\overline{p}_i$	$= n_A/(n_A + n_B)$
j	1.0	0.0	$\overline{p}_j$	$= n_C/(n_C + n_D)$
k	0.0	1.0	$\overline{p}_k$	$= n_E/(n_E + n_F)$
ℓ	0.0	0.0	$\overline{p}_\ell$	$= n_G/(n_G + n_H)$

What is the connection between μG (the measure for testing a completely inbred line) and the more general population measure $n_\ell\, a\, \overline{p}_\ell$. Comparing Tables 12.A1 and 12.A3, Class ℓ for the population measure corresponds to Classes G + H of the inbred line measure. Since Class H refers to loci with no favorable alleles in any of the lines, $\overline{p}_\ell$ denotes the frequency of favorable alleles in class ℓ (both elite inbred lines lack favorable alleles),

$$\overline{p}_\ell = \frac{n_G}{n_G + n_H} = \frac{n_G}{n_\ell} \tag{12.A5a}$$

so that

$$n_\ell\, \overline{p}_\ell = n_G, \quad \text{implying} \quad n_\ell\, a\, \overline{p}_\ell = a\, n_G \tag{12.A5b}$$

Thus, the general population measure also estimates the desired value (an_G) for an inbred line.

Crossing Schemes for Optimal Introgression

Suppose that we have identified one (or more) candidate lines I_x containing favorable alleles not present in either elite line (I_1 and I_2). What is the best crossing scheme to introgress these alleles? First, we need to determine which elite line is crossed to the candidate, i.e, should we use $I_1 \times I_x$ or $I_2 \times I_x$? Second, given we have identified the elite line to be crossed to the chosen candidate, (say I_1), how should the resting hybrid be crossed to the other elite line I_2. For example, one might cross F_2's from $I_1 \times I_x$ to I_2. However, one might also backcross the F_1 to one of the parents (I_1 or I_x) for one (or more) generations before crossing to I_2.

Dudley (1982, 1984a, c) notes that the estimated number of loci in the various classes (A through G) provide a guide on both which elite line to cross to the

candidate and whether the resulting hyrbid should be backcrossed or selfed before proceeding to cross it to the other elite parental line. For a chosen candidate line, if $n_C + n_F > n_D + n_E$, then the candidate line I_x shares more genes in common with I_1 than with I_2. If the inequality if reversed, I_x shared more genes with I_2. Dudley (1984c) suggests the candidate be first crossed to the line it shares more genes in common with, so that if $n_C + n_F > n_D + n_E$ the first cross (which we will call the F_1) should be $I_x \times I_1$, else the cross is $I_x \times I_2$. Whether the F_1 should be backcrossed to one of its parents or selfed to an F_2 before crossing with the other elite parent depends on whether the elite line or the candidate contains more favorable alleles (Dudley 1984). In particular, if both lines contain similar amounts of favorable alleles, a generation of selfing (to faciluate recombination) should be allowed before crossing to the other elite line. However, if one of the parents contains a disportotatine number of favorable alleles, the F_1 should be backcrossed to the parent containing the most favorable alleles before the hybrid is crossed to the other elite line (Reddy and Comstock 1976; Bailey 1977; Ho and Comstock 1980; Dudley 1982, 1984c).

If $n_G - n_D < 0$, at least one generation of backcrossing the $I_1 \times I_x$ F_1 back to I_1 is recommended selfing to obtain new lines to cross with the testor (I_2). If I_1 serves as the testor (i.e., I_2 is the elite line to be improved), then $n_G - n_F < 0$ suggests at least one generation of the backcrossing $I_2 \times I_x$ F_1 to $I_2$is recommended. If the appropiate inequailty (given the testor) is not significantly greater than zero, then no backcrossing is needed, and the F_1 can be directly selfed. If the appropriate inequality is significantly greater than zero, then the elite parent is backcrossed to the candidate line. Logic: If I_1 is the elite line being improved, then $n_D > n_G$ implies that the number of favorable loci unique to I_1 exceeds the number of favorable loci unique to I_x. Likewise, $n_F > n_G$ implies that I_2 contains more unique favorable loci than I_x. In these settings, backcrossing increases the chance that more favorable loci are maintained during selfing.

Modification of Dudley's Original Estimators

Even with the assumption of complete dominance holds, Dubley's estimator for an inbred line (μG) and a population ($n_\ell \, a \, \overline{p}_\ell$) are still biased if the other genetic assumptions leading to these estimator fail. Working with the population estimator, Dudley (1987a) showed that the bias from assuming $\overline{p}_i = \overline{p}_j = \overline{p}_k$ is

$$\text{bias}\left(\widehat{n_\ell \, a \, \overline{p}_\ell}\right) = \frac{a \, n_j \, n_k \, (\, \overline{q}_j - \overline{q}_k \,)}{n_j - n_k} \tag{12A.5a}$$

where $\overline{q}_y = 1 - \overline{p}_y$. Hence, Equation 12A.4 is a biased estimator unless $\overline{q}_j = \overline{q}_k$. Dudley further shows that $a \, n_j$ and $a \, n_k$ can be estimated by $(\, \overline{z}_{12} - \overline{z}_2)/2$ and $(\, \overline{z}_{12} - \overline{z}_1)/2$, respectively, implying that this bias term can thus be estimated as

$$\widehat{\text{bias}}\left(\widehat{n_\ell \, a \, \overline{p}_\ell}\right) = \frac{(\, \overline{z}_{12} - \overline{z}_1 \,)(\, \overline{z}_{12} - \overline{z}_2 \,)(\, \overline{q}_j - \overline{q}_k \,)}{2(\, \overline{z}_1 - \overline{z}_2 \,)} \tag{12A.5b}$$

Given an estimate of $(\overline{q}_j - \overline{q}_k)$, the bias arising from assuming $\overline{q}_j = \overline{q}_k$ can be accounted for. Dudley proposes to estimate upper and lower bounds on the $\overline{q}_y$ and then substitute the average of these for $\overline{q}_y$ to estimate the bias. The bounds for the $\overline{q}_y$ are obtained as follows. It can be shown (again by equating the expected means with observed values, Dudley 1987a) that

$$\overline{z}_{1x} - \overline{z}_{2x} = (\overline{z}_{12} - \overline{z}_2)\,\overline{q}_j - (\overline{z}_{12} - \overline{z}_1)\,\overline{q}_k \tag{12.A6}$$

Hence, a lower bound for $\overline{q}_j$ occurs by first setting $\overline{q}_k = 0$ and then solving to give $\overline{q}_j = (\overline{z}_{1x} - \overline{z}_{2x})/(\overline{z}_{12} - \overline{z}_2)$. Dudley denotes this lower bound by $\overline{q}_{k0}$, i.e., by setting $\overline{q}_k = 0$. For this bound to yield an admissible solution for $\overline{q}_j$, we require $\overline{z}_{1x} - \overline{z}_{2x} \leq \overline{z}_{12} - \overline{z}_2$, otherwise the lower bound for $\overline{q}_j$ is greater than one. Likewise, the $\overline{q}_{k1}$ upper bound (obtained by setting $\overline{q}_k = 1$) is

$$\overline{q}_j = \frac{\overline{z}_{1x} - \overline{z}_{2x} + \overline{z}_{12} - \overline{z}_1}{\overline{z}_{12} - \overline{z}_2}$$

Table 12.A5. Various bounds on the $\overline{q}_y$ for Dudley's population estimator. Using the admissible bounds given the data, the average of the upper and lower bound is substituted for $\overline{q}_y$ to estimate the bias (Equation 12A.5a) generated by the assumption $\overline{p}_i = \overline{p}_j = \overline{p}_k$.

Bound	$\overline{q}_j$	$\overline{q}_k$
$\overline{q}_{j0}$	0	$-\dfrac{\overline{z}_{1x} - \overline{z}_{2x}}{\overline{z}_{12} - \overline{z}_1}$
$\overline{q}_{k0}$	$\dfrac{\overline{z}_{1x} - \overline{z}_{2x}}{\overline{z}_{12} - \overline{z}_2}$	0
$\overline{q}_{j1}$	1	$-\dfrac{\overline{z}_{1x} - \overline{z}_{2x} - (\overline{z}_{12} - \overline{z}_2)}{\overline{z}_{12} - \overline{z}_1}$
$\overline{q}_{k1}$	$\dfrac{\overline{z}_{1x} - \overline{z}_{2x} + (\overline{z}_{12} - \overline{z}_1)}{\overline{z}_{12} - \overline{z}_2}$	1

Table 12.A5 summaries the various bounds derived in the same fashion. The bounds that one uses for any particular data set are those that are admissible (i.e., lower bound is less than one, upper bound is greater than zero), see Example A1. Substitution in the appropriate bounds gives **Dudley's** (1987a) **modified estimator**, $n_\ell\, a\, \overline{p}^*_\ell$, which becomes

$$4\, n_\ell\, a\, \overline{p}^*_\ell = \begin{cases} \overline{z}_{1x} + \overline{z}_{2x} - \overline{z}_{12} - \overline{z}_1, & \text{using } \overline{q}_{j0},\, \overline{q}_{k1} \\ \overline{z}_{1x} + \overline{z}_{2x} - \overline{z}_{12} - \overline{z}_2, & \text{using } \overline{q}_{j1},\, \overline{q}_{k0} \\ 2\,\overline{z}_{2x} - \overline{z}_{12} - \overline{z}_2, & \text{using } \overline{q}_{j0},\, \overline{q}_{j1} \\ 2\,\overline{z}_{1x} - \overline{z}_{12} - \overline{z}_1, & \text{using } \overline{q}_{k0},\, \overline{q}_{k1} \end{cases} \tag{12.A7}$$

While this improved estimator largely corrects for the bias generated by assuming $\overline{p}_i = \overline{p}_j = \overline{p}_k$, it does *not* correct for other sources of bias (such as the assumption of complete dominance or that all loci have the same effect). Note that with the improved estimator, we do not need $\overline{z}_x$, reducing the number of populations to be measured to $2n + 3$ (Table 12.A6).

As Equation 12.A5b points out, this modified estimator equals the inbred line (as opposed to general population) estimator μG under the assumptions leading to Equation 12.A7. Dudley (1987b) notes that this modified estimator is free of the assumptions $c = 0$ and $n_A = n_H$, so that Equation 12.A7 is also am improved estimator for testing among inbred candidates. Similar modified estimators for μA through μF can be found in Dudley (1987b), while Zanoni and Dudley (1989b) gives expressions for the modified estimators using F_2 (as opposed to F_1) crosses. An important observation (Dudley 1987b) that that $\mu C + \mu F$ and $\mu D + \mu E$ give identical values using either the original or modified estimator. From Table 12.A2,

$$a(n_C + n_F) = 2\,(\,-\overline{z}_{1x} + \overline{z}_{2x} + \overline{z}_{12} - \overline{z}_2\,)$$

and

$$a(n_D + n_E) = 2\,(\,\overline{z}_{1x} - \overline{z}_{2x} + \overline{z}_{12} - \overline{z}_1\,)$$

Example A1. Moreno-Gonzalez (1978, given in Dudley 1984a) reports the following yield data for maize inbred lines and their crosses:

Mean	M017	B73	N28	B37	C103	Oh43
Pure Line	55	95	38	27	41	86
× M017	—	241	218	208	108	183
× B73	241	—	183	190	233	220

Taking M017 (I_1) and B73 (I_2) as the elite lines (their F_1's having the highest value, 241), which of the remaining lines contains the largest number of favorable alleles not present in M017 and/or B73?

Using predicted test cross values (Equation 12.A1), find the following PTC $= (\,\overline{z}_{1x} + \overline{z}_{2x}\,)/2$ values:

Line	N28	B37	C103	Oh43
PTC	200.5	199.0	170.5	201.5

suggesting that Oh43, N28, and B37 are nearly equal candidates.

Applying Dudley's μG statistic (Equation 12.A3a) gives

Line	N28	B37	C103	Oh43
μG	-7.0	-5.0	-22.75	-18.5

Hence, none of the lines appear to contain favorable alleles not present in the two elite lines. However, chosing the best line based on this statistic shows a clear difference between N28 and B37 versus Oh43. Note that the best line based on μG, B37, is the third worst line based on PTC values. Moving to Dudley's modified estimator, we find that

Line	$\overline{q}_{j0}$	$\overline{q}_{k0}$	$\overline{q}_{j1}$	$\overline{q}_{k1}$	Estimator	$n_\ell\, a\, \overline{p}^*_\ell$
N28	-0.19	**0.24**	**0.60**	1.51	$\overline{q}_{k0}, \overline{q}_{j1}$	16.25
B37	-0.10	**0.12**	**0.69**	1.40	$\overline{q}_{k0}, \overline{q}_{j1}$	15.50
C103	**0.67**	-0.86	1.46	**0.42**	$\overline{q}_{j0}, \overline{q}_{k1}$	11.25
Oh43	**0.20**	-0.25	**0.98**	1.02	$\overline{q}_{j0}, \overline{q}_{j1}$	26.00

With this estimator, OH43 show a very clear advantage over the other candidates. Using the estimators given in Table 12.A2, $\mu C + \mu F = 91.5$, while $\mu D + \mu E = 74.5$ (a significant difference), OH42 should thus be first crossed to I_1 = Mo17. Likewise, since $n_B + n_D = 12.0 + 36.5 = 48.5$ while $n_E + n_G = 38.0 + 26.0 = 64$ (a significant difference),

---

The Upper Bound (UBND) and Net Improvement (NI) Estimators

Three other estimators related to the number of new favorable alleles in a candidate line have been proposed. Gerloff and Smith (1988a,b) proposed the **upper bound statistic** for testing a (potentially segregating) population

$$\text{UBND} = \min\left(\, \overline{z}_{x1} - \overline{z}_1, \overline{z}_{x2} - \overline{z}_2 \,\right) \tag{12.A8}$$

Under the assumptions of the Dudley (1984) estimator, Gerloff and Smith find that

$$\frac{\mu_{x1} - \mu_1}{2} = n_\ell\, a\, \overline{p}_\ell + n_k\, a\, \overline{p}_k, \quad \text{and} \quad \frac{\mu_{x2} - \mu_1}{2} = n_\ell\, a\, \overline{p}_\ell + n_j\, a\, \overline{p}_j \tag{12.Ax}$$

When the population being tested is an inbred line, these reduce to $a(n_G + n_E)$ and $a(n_G + n_C)$, respectively.

Bernardo (1990a) notes that there are two sources contributing to improvement of the elite lines and their hybrid from a candidate line. As mentioned through, the candidate line may contain favorable alleles not found in either elite line. Balancing this is the risk of losing favorable alleles already present in the elite lines during inbred and selection to form new elite lines. Thus, net improvement of hybrid involving both elite lines and the candidate only occurs if more favorable alleles are gain than lost. If I_2 is the line crossed to the candidate, alleles can be gained from Class G, but can also be lost from Class F (advantageous alleles present in I_2 but lacking in both I_1 and I_x). Likewise, if the candidate is first crossed to I_1, alleles can be lost from Class D. Thus $n_G - n_F$ and $n_G - n_D$ are

measures for the net improvement, and these can be estimated since (under the assumption of complete dominance)

$$E[\,\overline{x}_{1x} - \overline{x}_{12}\,] = 2a(n_G - n_F), \quad \text{and} \quad E[\,\overline{x}_{2x} - \overline{x}_{12}\,] = 2a(n_G - n_D)$$

This motivates Bernardo's**net improvement statistic**,

$$\text{NI} = \max\left(\frac{\overline{z}_{x1} - \overline{z}_{12}}{2}, \frac{\overline{z}_{x2} - \overline{z}_{12}}{2}\right) \tag{12.A10}$$

The candidate line with the largest NI statistic is chosen, and if $\overline{z}_{1z} > \overline{z}_{2z}$, the candidate I_x is crossed to I_2 with I_1 the **tester**. If the inequality if reversed, then the candidate is crossed to I_1.

Table 12.A6. Number of crosses and populations needed to test n candidate populations for elite alleles.

Estimator	Populations needed	Total
Test cross to a single cross (TCSC)	$I_x \times (I_x \times I_2)$	$n+1$
Predicted triple cross (PTC)	$I_1 \times I_x, I_2 \times I_x$	$2n$
Net improvement statistic (NI)	$I_1 \times I_x, I_2 \times I_x$	$2n+1$
Upper bound statistic (UBND)	$I_1, I_2, I_1 \times I_x, I_2 \times I_x$	$2n+2$
Dudley's $n_\ell\, a\, \overline{p}_\ell$	$I_1, I_2, I_1 \times I_2, I_1 \times I_x, I_2 \times I_x$	$2n+3$
Dudley's μG	$I_1, I_2, I_1 \times I_2, I_x, I_1 \times I_x, I_2 \times I_x$	$3n+3$
Hohls et al. estimator	$I_1, I_2, I_1 \times I_2, I_x, I_1 \times I_x, I_2 \times I_x$	$3n+3$

As might be expected, the ranking of candidate lines using these different measures are generally fairly strongly correlated. Zanoni and Dudley (1989a) found in maize crosses for several characters that the modifed Dudley estimator (Equation 12.A7) was high correlated with the PTC and UBND estimators, but was very poorly correlated with the original Dudley estimator (Equation 12A.3a). Mišević (1989b), also working with maize, found strong correlations (in terms of the ranks of candidiate lines) between The modified Dudley, PTC, and UBND estimators. The correlation between the modified Dudley and PTC often exceeded 95%, leading Mišević the propose using this estimator (as it required the fewest lines be grown).

The various estimators have also been comparied using computer simulations and (in a few cases) actual data. Simulation studies by Gerloff and Smith found that their UBND statistic (Equation 12.A8) is more consistently correlated with the superiority of a candidate line ($n_\ell\, a\, \overline{p}_\ell$) that is Dudley's original (Equation 12.A3) estimator. However, they also found that the observed mean of a triple

cross (TCSC, Equation 12.A2) was also equally correlated with the superiority, and requires fewer total crosses (Table 12.A6). Bernardo (1990) also used simulations under a variety of genetic models, considering the mean $\overline{x}_{0.10}$ of the upper 10% from the F_2 of the cross to the tester. Overall, the NI statistic had the highest correlation (averaged over the various tested genetic models) with $\overline{x}_{0.10}$ ($r = 0.81$), followed by PTC (0.75), modified Dudley (0.73), and finally UBND (0.52).

Mišević (1989a) also examined the F_2 from the elite line to be improved with the candidate, testcrossed to the other elite line. Out of 22 tested candidate lines, three (ranked 1, 15, and 17th under modifed Dudley; 1, 18, 19 by PTC; and 1, 13, 19 by UBND) where chosen to have their F_2s testcrossed. The candidate ranked first by all three estimators had significantly more test crosses with higher yeilds than the traget hybrid (the cross of the original two elite lines) than testcrosses involving the other two (much mroe poorly ranked) lines.

Dudley (1987b) showed that his modified estiamtor generaly did better than UBND, TSCS, and the origianl estiamtor in prdicting the lines with the best superiority.

Perhaps the most direct test was Zanoni and Dudley (1989c), who examined the preformance of F_2's from crosses between the elite lines and candidate lines with low very high predicted superiority values. Candidate lines were ranked using both the original and modified Dudley estimator, as well as UBND, PTC, and per-se evaluation. The performance ranking of the resulting test crosses showed that only the modified Dudley and UBND were (somewhat) reliable predictors of performance, with the later two measures being able to predict whether a candidate line feel into the good vs. poor classes. However, neighter estimator correlated well with the ranking of the three top lines, as measured by either the proportion of testcrosses superior to the original hybrid or the number the exceed the original by at least one standard deviation.

Hogan and Dudley (1991) regression of estimates of $\ell_{|ell}\,\mu'$ on lines whose donor composition was known in advance. Regression accounted on the precentage of donor genotype accounted for 87 to 99% of the variation in $\ell_{|ell}\,\mu'$ over four traits.

A study by Pfarr and Lamkey (1992a) was less optimstic about the perfornace of Dudley's modified esitmator. These authors used seven lines formed by various combiantions of backcrossing from two original parental lines. hence, the relative propostion of each parent line is known. Five traits were examine (yield, ear and plant hieght, pollen and silking dates). Four other traits were also measured, but the resulting estimators for $\overline{q}$ were outside the presimailable (0,1) space and hence no conistent estimator for $\ell_{|ell}\,\mu'$ is aviable in these cases. Dudley's modified statistc was able to predict the line carrying the msot favorable alleles about 60-80% of the time. They express that Dudley's modified estimator may often be inable to idenfity favorable popuations when the favorable alleles are at low frequnecies.

Pfarr and Lamkey (1992b) compared but the original and modified Dudley estiamtes, UBND, TCSE, test crossing to an inbed line, and perfornance per see

(candiate line means) in populations of known genetic composition. Modified Dudlyey and UBDN had the higest correlation with kow genetic composition, and the rank correlations between these two approaches was on the order of 0.98. Both approaches were equall effectively, and hence an edge is given to UBND, given taht fewer popualtions need be evaluated.

Per se perofrnace was no correlated with precentage of the favorable parent, and Dudleyt's original measure also faired very poorly.

FOr some traits (such as early silking), none of the estiamtors corectly identified the most superior line. Failire of the genetic model (evidence of additivity

Hohls et al. (1995) unbiased estimator:

$$\mu\, a\,(B+G) = \frac{s_{1x} + s_{2x} - s_{12}}{2} \tag{12.A11}$$

where

$$s_{ij} = \overline{z}_{ij} - \frac{\overline{z}_{i\cdot} + \overline{z}_{j\cdot}}{2}$$

is the specific combining ability for the cross of $i \times j$. Similar unbiased estimators for other allelic classes are

$$\mu\, a\,(C+F) = \frac{s_{2x} + s_{12} - s_{1x}}{2} \tag{12.A12a}$$

C = present in I_1 and I_x, absent in I_2, F = only present in I_2. And

$$\mu\, a\,(D+E) = \frac{s_{12} + s_{1x} - s_{2x}}{2} \tag{12.A12b}$$

E = present in I_2 and I_x, absent in I_1, D = only present in I_1.

Example A2.

Applying Gerloff's (1985) Upper bound statistic (Equation 12.A9) gives

Line	N28	B37	C103	Oh43
UBND	**123**	113	13	88

Bernardo's (1990a) net improvement statistic (Equation 12.A10) gives

Line	N28	B37	C103	Oh43
NI	−11.5	−16.5	− 4.0	−10.5

Finally, using Hohls et al. (1995) unbiased estimator

Line	N28	B37	C103	Oh43
$\mu\, a\,(B+G)$	−241.5	**−17.65**	−44.90	−18.15

Experimental Checks of the Various Estimators

Stam (1977)

$$= \frac{c}{2} + \frac{1}{4}\, c(1-2c) \left(\frac{1 - ([1-2c]/2)^n}{2 - (1-2c)} \right) \qquad ()$$

$$\frac{2c}{2(1+2c)}$$