Theor Appl Genet (1988) 76:209–216

Choice of method for identifying germplasm with superior alleles

1. Theoretical results*

J. E. Gerloff[1] ** **and O. S. Smith[2]**

[1] Seneca Associates, 84 Hammond Street, Cambridge, MA 02138, USA
[2] Pioneer Hi-Bred International, Johnston, IA 50131-0085, USA

Received November 10, 1986; Accepted November 20, 1987
Communicated by A. Hallauer

Summary. Elite, adapted germplasm is not likely to contain all the favorable alleles available in a species. Three statistics were evaluated for screening populations for their ability to contribute favorable dominant alleles not available in an elite single cross: (1) a statistic proposed by Dudley (SD) = [(P × I1 − I1) (I1 × I2 − I2) − (P × I2 − I2) (I1 × I2 − I1)]/[2(I1 − I2)]; (2) the upper bound (UBND) = minimum (P × I1 − I1, P × I2 − I2); and (3) the testcross to the single cross [TC (SC)] = P × (I1 × I2), where P is the population to be evaluated and I1 and I2 are homozygous parents of the elite single cross I1 × I2. A superiority measure for a population was defined as the product of frequencies of favorable alleles and effects summed over loci where I1 × I2 is homozygous unfavorable. Of the statistics considered, TC (SC) should have the highest genetic correlation with the superiority measure under the assumptions made, require the fewest testing resources and have the smallest standard error. Methods considered for screening inbreds were: (1) SD_I proposed by Dudley = [(I1 × I_W) + (I2 × I_W) − I1 − I2 − I_W − (I1 × I2)]/4; (2) TC (SC) = I_W × (I1 × I2); and (3) UBND = minimum (I_W × I1 − I1, I_W × I2 − I2) where I_W is the inbred to be evaluated. The superiority measure of an inbred I_W was defined as the relative number of loci where I1 and I2 are unfavorable and I_W is favorable. The genetic correlation with the superiority measure should be highest for SD_I. The larger number of measurements used in calculation, the necessity of evaluating potentially unadapted inbreds and larger testing resources required for SD_I suggest further research should be done to evaluate these statistics.

Key words: Exotic germplasm – Germplasm choice – Testcross – Upper bound

Introduction

Choice of germplasm for use in a breeding program is one of the most important decisions a breeder must make. Besides the elite germplasm available in a crop species, there is a vast amount of germplasm at various levels of immediate usefulness and adaptedness [e.g., 77,000 accessions of maize (*Zea mays* L.) in germplasm banks, Plucknett et al. 1983]. It has been suggested that benefits in yields could be realized by broadening the genetic base of important crop species (for example, Brown 1965). The usefulness of some unadapted germplasm and wild species as potential sources of alleles contributing disease and pest resistance has already been clearly demonstrated (e.g., Harlan 1976). Non-elite germplasm (including that from other species) also has the potential to expand the range of adaptation of a crop, improve cold and heat tolerance and improve quality factors such as protein content (Frey 1983; Harlan 1984). It is unlikely that the germplasm currently used in any crop contains all the desirable alleles available controlling a particular quantitatively inherited trait. Unfortunately, immediate utilization of non-elite or unadapted germplasm may be hindered by overall low mean performance, photoperiod sensitivity or undesirable agronomic traits. As Duvick (1981) stated, "We don't need diversity of deleterious genes; we do need to learn how to identify useful gene combinations in exotic materials, and how to transfer them efficiently and quickly." Brown (1983) stated, "In simplest terms, the breeder is interested in introducing useful alleles which are different from those present in

* Joint contribution: USDA-ARS and Journal Paper No. J-12420 of the Iowa Agric. and Home Econ. Exp. Stn., Ames, IA 50011. Project No. 2778. Part of a dissertation submitted by senior author in partial fulfillment of PhD requirements
** To whom correspondence should be addressed

populations now in use. With present methodology there are no completely satisfactory ways of identifying new alleles of most of the genes which make up a species."

Suggestions for identifying useful germplasm sources have varied with breeding goals. For simply inherited traits, evaluation of the germplasm per se in any environment that allows expression of that trait will be adequate. For quantitatively inherited traits, crosses between high-performing parents have been reported to result in the best chance of finding high-performing progeny (Hayes and Johnson 1939; Green 1948; Busch et al. 1974). Dudley (1984a) suggested estimating the relative number of loci where an inbred (I_W) contained the more favorable of two alleles and an elite, adapted single cross (I1 × I2) was homozygous unfavorable. The estimator proposed was $[(I_W \times I2)+(I_W \times I1)-I_W-I2-I1-(I1 \times I2)]/4$. Other methods of choosing populations include: evaluation per se (Spencer 1980; Burton and Davies 1984), crossing to other populations in diallels (Lonnquist and Gardner 1961; Eberhart 1971; Josephson 1982), crossing to an elite single cross (Kramer and Ullstrup 1959; Stuber 1978) and crossing to elite inbreds (Burton and Davies 1984). The amount of genetic variance in a population has also been suggested as a criterion for choice (Hallauer and Miranda 1981).

In screening unadapted or non-elite germplasm, the breeder is likely to be most interested in finding favorable alleles not available in elite sources. Dudley (1984b) proposed estimating the relative frequencies of the favorable allele in a population (P) at loci for which an elite, adapted single cross (I1 × I2) is homozygous unfavorable. This estimator was: $[(P \times I1-I1)\ (I1 \times I2-I2)-(P \times I2-I2)\ (I1 \times I2-I1)]/[2(I1-I2)]$.

Per se or diallel cross performance of unadapted germplasm may be impossible to assess in the target environment because of photoperiod sensitivity, response to temperature and other environmental factors (Stuber 1978; Goodman 1985). Estimation of genetic variances and diallel crossing involve extensive testing and would be unsuitable for screening large numbers of germplasm sources. The objectives of this research were to consider three relatively simple methods of screening both populations and inbreds for their potential to contribute alleles superior to those available in elite, adapted germplasm. These methods will be compared by using a genetic model of two alleles per locus, complete dominance and no epistasis, but unrestricted in terms of alleles effects at each locus and average frequencies of the favorable allele at different classes of loci.

Discussion

Screening populations

Theory. Two models of gene action for a quantitatively inherited trait will be described. Both models assume regular Mendelian inheritance in a diploid species, two alleles per locus, complete dominance of the favorable allele and no epistasis. The frequencies of the favorable allele in the homozygous inbred parents of a single cross can be used to group the *N* loci affecting the trait into classes (Dudley 1984b). Four classes of loci are considered (Table 1). For class i loci, both inbreds (I1 and I2) are homozygous favorable. For class j loci, I1 is favorable and I2 is unfavorable, and I2 is favorable and I1 is unfavorable at class k loci. Both inbreds are fixed for the unfavorable allele at loci in class l. The letters i, j, k and l identify the classes and the number of loci in their respective classes. The two models differ in the assumptions made about gene frequencies, u, which is defined as half the difference between the values of the two homozygotes (Comstock and Robinson 1948), and z, the value of the unfavorable homozygote. The + +, + − and − − genotypes are assumed to have the values $z+2u$, $z+2u$, and z, respectively. In Model I, as used by Dudley (1984b) and generalized here, the average frequency of the favorable allele, $\bar{p}$, in the population to be tested (P) is equal for loci in classes i, j and k. The average frequency of the favorable allele for loci in class l is $\bar{p}_l$. The values of u and z are assumed to be equal for all loci. In Model II, average gene frequencies in a population vary from class to class, and u and z vary from locus to locus. Let p_i, u_i and z_i be the frequency of the favorable allele in P, the value of u, and

Table 1. Frequencies of the favorable allele, half the difference between the two homozygotes and the value of the unfavorable homozygote under Models I and II

Locus class and no. of loci	Frequency of the favorable allele					One-half difference between the two homozygotes		Value of the unfavourable homozygote	
	I1	I2	I1 × I2	P					
				Model I	Model II	Model I	Model II	Model I	Model II
i	1.0	1.0	1.0	$\bar{p}$	p_i	u	u_i	z	z_i
j	1.0	0.0	0.5	$\bar{p}$	p_j	u	u_j	z	z_j
k	0.0	1.0	0.5	$\bar{p}$	p_k	u	u_k	z	z_k
l	0.0	0.0	0.0	$\bar{p}_l$	p_l	u	u_l	z	z_l

the value of z, respectively, for the i-th locus in class i. The definitions of p_j, u_j, z_j, p_k, u_k, z_k, p_l, u_l and z_l are similar (Table 1). The substitution of $i\bar{p}_i\bar{u}_i$ could be made for $\sum_i p_i u_i$ in the following discussion. Similarly, $j\bar{p}_j\bar{u}_j$ can be substituted for $\sum_j p_j u_j$, $k\bar{p}_k\bar{u}_k$ for $\sum_k p_k u_k$ and $l\bar{p}_l\bar{u}_l$ for $\sum_l p_l u_l$. These are analogous, but not necessarily equal, to $i\bar{p}u$, $j\bar{p}u$, $k\bar{p}u$ and $l\bar{p}_l u$ under Model I. Measurements for a trait of interest replace the symbols P, I1, I2, and their crosses, which allows for estimation of the genotypic values of the respective entries under Models I and II (Tables 2 and 3, respectively). The assumption of random-mating equilibrium is necessary for the genotypic value of the population to be as shown.

An elite, adapted single cross was considered to contain the highest concentration of favorable alleles available in adapted germplasm (Dudley 1984a, b). Classes i, j, k and l are properties of this elite reference single cross I1 × I2; the u's and z's are a property of the species. Under Model I, the relative contribution a particular population with average frequency $\bar{p}_l$ can make to the single cross at loci where the favorable allele is not present is given by the term $l\bar{p}_l u$, because lu will not vary from population to population (Dudley 1984b). This term can be defined as a superiority measure of the population. Under Model II, the analogous term for the definition of the superiority measure is $\sum_l p_l u_l$. Dudley (1984b) proposed the statistic $[(P \times I1 - I1)\,(I1 \times I2 - I2) - (P \times I2 - I2)\,(I1 \times I2 - I1)]/[2(I1 - I2)]$ as the estimator of the superiority measure inasmuch as this has expectation under Model I of $l\bar{p}_l u$. We will denote this statistic by SD.

The testcross to the single cross, $TC(SC) = [P \times (I1 \times I2)]$, has expectation under Model II of:

$$\sum_i u_i + \sum_j p_j u_j + \sum_k p_k u_k + 2\sum_l p_l u_l - \sum_l u_l + \sum_N (u_N + z_N).$$

The difference between the testcrosses to I1 × I2 for two populations, P and P′, has expectation:

$$\sum_j (p_j - p'_j)\,u_j + \sum_k (p_k - p'_k)\,u_k + 2\sum_l (p_l - p'_l)\,u_l.$$

As Dudley (1984b) emphasized, this compares the two populations for their differences in allelic frequencies at loci where the favorable allele is unavailable in the elite germplasm (class l) but also at loci where the favorable allele is available in one of the elite inbreds (classes j and k). A population P with high frequencies of favorable alleles at the j and k classes but low frequencies at the l-th class might exhibit higher testcross performance than a population P′ with low frequencies at classes j and k but high frequencies at class l. P would have fewer favorable alleles not already available in the elite single cross than P′.

Table 2. Expectations of genotypic values of I1, I2, P and their hybrids under the assumptions of Model I [taken from Dudley (1984b) with minor changes]

I1	$= (i+j-k-l)\,u + N(u+z)$
I2	$= (i-j+k-l)\,u + N(u+z)$
P	$= (i+j+k)(4\bar{p}-2\bar{p}^2-1)\,u + l(4\bar{p}_l - 2\bar{p}_l^2 - 1)\,u + N(u+z)$
I1 × I2	$= (i+j+k-l)\,u + N(u+z)$
P × I1	$= (i+j+2k\bar{p}-k+2l\bar{p}_l-l)\,u + N(u+z)$
P × I2	$= (i+2j\bar{p}-j+k+2l\bar{p}_l-l)\,u + N(u+z)$
P × (I1 × I2)	$= (i+j\bar{p}+k\bar{p}+2l\bar{p}_l-l)\,u + N(u+z)$

Table 3. Expectations of genotypic values of I1, I2, P and their hybrids under the assumptions of Model II

I1	$= \sum_i u_i + \sum_j u_j - \sum_k u_k - \sum_l u_l + \sum_N (u_N + z_N)$
I2	$= \sum_i u_i - \sum_j u_j + \sum_k u_k - \sum_l u_l + \sum_N (u_N + z_N)$
P	$= \sum_i u_i(4p_i - 2p_i^2 - 1) + \sum_j u_j(4p_j - 2p_j^2 - 1) + \sum_k u_k(4p_k - 2p_k^2 - 1) + \sum_l u_l(4p_l - 2p_l^2 - 1) + \sum_N (u_N + z_N)$
I1 × I2	$= \sum_i u_i + \sum_j u_j + \sum_k u_k - \sum_l u_l + \sum_N (u_N + z_N)$
P × I1	$= \sum_i u_i + \sum_j u_j + 2\sum_k p_k u_k - \sum_k u_k + 2\sum_l p_l u_l - \sum_l u_l + \sum_N (u_N + z_N)$
P × I2	$= \sum_i u_i + 2\sum_j p_j u_j - \sum_j u_j + \sum_k u_k + 2\sum_l p_l u_l - \sum_l u_l + \sum_N (u_N + z_N)$
P × (I1 × I2)	$= \sum_i u_i + \sum_j p_j u_j + \sum_k p_k u_k + 2\sum_l p_l u_l - \sum_l u_l + \sum_N (u_N + z_N)$

An upper bound (UBND) can be placed on $2\sum_l p_l u_l$ by noting that the expectation of P × I1 − I1 under Model II is $2\sum_l p_l u_l + 2\sum_k p_k u_k$ and of P × I2 − I2 is $2\sum_l p_l u_l + 2\sum_j p_j u_j$. By taking the minimum of the difference between the testcross to the inbreds and the respective inbred parent, the contribution of the terms $\sum_j p_j u_j$ or $\sum_k p_k u_k$ to the estimator is minimized. Both the UBND and the TC(SC), when divided by two, are biased estimators of the superiority measure.

The frequencies of the favorable alleles in hypothetical inbreds I1 and I2, I1 × I2 and five populations (P1 through P5) can be used to illustrate the calculation of SD, TC(SC)/2 and the UBND/2. The true $\sum_l p_l u_l$ is also given (Table 4). In this example, u was assumed to equal one for all loci, z was equal to zero, and $p_i = \bar{p}_i$, $p_j = \bar{p}_j$, $p_k = \bar{p}_k$ and $p_l = \bar{p}_l$. The SD was the only estimator that correctly estimated $\sum_l p_l u_l$ for any of the hypothetical

Table 4. Frequencies of the favorable allele at loci in a class in hypothetical inbreds, single cross and five populations (P1–P5) and parameters and statistics for the populations when dominance is complete, all u's are equal to one and all z's are equal to zero

No. of loci	Class	Entry							
		I1	I2	I1 × I2	P1	P2	P3	P4	P5
10	i	1.0	1.0	1.0	0.2	0.5	0.2	0.4	0.3
8	j	1.0	0.0	0.5	0.7	0.3	0.4	0.4	0.5
7	k	0.0	1.0	0.5	0.2	0.6	0.7	0.4	0.6
5	l	0.0	0.0	0.0	0.3	0.5	0.2	0.3	0.0
Parameters and statistics									
$\sum_l p_l u_l$					1.5	2.5	1.0	1.5	0.0
I1					36.0	36.0	36.0	36.0	36.0
I2					34.0	34.0	34.0	34.0	34.0
I1 × I2					50.0	50.0	50.0	50.0	50.0
P × I1					41.8	49.4	47.8	44.6	44.4
P × I1 − I1					5.8	13.4	11.8	8.6	8.4
P × I2					48.2	43.8	42.4	43.4	42.0
P × I2 − I2					14.2	9.8	8.4	9.4	8.0
UBND/2					2.9	4.9	4.2	4.3	4.0
TC(SC)/2					22.5	23.3	22.55	22.0	21.6
SD					−26.5	19.3	17.8	1.5	5.6
UBND/2 − $\sum_l p_l u_l$					1.4	2.4	3.2	2.8	4.0
TC(SC)/2 − $\sum_l p_l u_l$					21.0	20.8	21.55	20.5	21.6
SD − $\sum_l p_l u_l$					−28.0	16.8	16.8	0.0	5.6

populations (i.e., P4). When u is equal for all loci, SD will estimate the superiority measure accurately only when $\bar{p}_j = \bar{p}_k$. There was not a perfect correlation between the ranking of the populations based on the superiority measure and the ranking based on SD, UBND/2 or TC(SC)/2. A negative estimate of the superiority measure was obtained with SD for P1. The difference between UBND/2 and the superiority measure was always smaller than the difference between TC(SC)/2 and the superiority measure. The magnitude of the difference between SD and the measure of superiority was sometimes higher and sometimes lower than these same differences for the UBND/2 and TC(SC)/2.

Interpretation. An estimator for the superiority measure that involves more measurements than SD would be unlikely to be usable in practice because of the resources needed to evaluate each population and the higher standard error of an estimator when more measurements are involved. A simple screening method is more desirable. Because of the difficulties in evaluating unadapted germplasm, per se performance may not be a good criterion for screening. Hence, only the TC(SC), UBND and SD will be considered as possible screening methods. Under Model I, TC(SC)/2 and UBND/2 are biased estimators of $l\bar{p}_l u$, but SD is not. All three are biased estimators of the superiority ($\sum_l p_l u_l$) under Model II. The magnitude of the bias varies among populations depending on allelic frequencies.

Under Model II, SD has expectation $\sum_l p_l u_l + \beta_{SD}$, where β_{SD} is equal to $[(\sum_k p_k u_k)\sum_j u_j - (\sum_j p_j u_j)\sum_k u_k]/(\sum_j u_j - \sum_k u_k)$. This bias, β_{SD}, may be positive or negative. If a population has nothing to contribute toward improving elite germplasm (i.e., if $\sum_l p_l u_l = 0$), SD will be positive if $\beta_{SD} > 0$. This will occur when I1 > I2 (i.e., $\sum_j u_j > \sum_k u_k$) if $\sum_j u_j / \sum_k u_k > \sum_j p_j u_j / \sum_k p_k u_k$. For example, if $\sum_j p_j u_j = \sum_k p_k u_k \neq 0$ but I1 > I2, then $\beta_{SD} = \sum_j p_j u_j$, which is greater than zero. In the example shown in Table 4, SD was positive for P5 even though $\bar{p}_l$ was zero. The quotient $\sum_j u_j / \sum_k u_k$ was 1.14 for this population and $\sum_j p_j u_j / \sum_k p_k u_k$ was 0.95. If $\sum_j u_j - \sum_k u_k > 0$, β_{SD} will be negative when $\sum_j u_j / \sum_k u_k < \sum_j p_j u_j / \sum_k p_k u_k$. If $|\beta_{SD}| > \sum_l p_l u_l$ and $\beta_{SD} < 0$, SD will be negative even though $\sum_l p_l u_l$ may be greater than zero. This occurred for P1 in the example. The quotient $\sum_j p_j u_j / \sum_k p_k u_k$ was 4.0 for P1. This was greater than $\sum_j u_j / \sum_k u_k$, which had the value of 1.14. False negative and positive values of SD are thus possible. The bias term when half the testcross to I1 × I2 is used to estimate $\sum_l p_l u_l$, $\beta_{1/2\,TC(SC)}$, is $[\sum_i u_i + \sum_j p_j u_j + \sum_k p_k u_k$

$-\sum_l u_l + \sum_N (u_N + z_N)]/2$. This will always be greater than or equal to zero (unless $-\sum_l u_l + \sum_N (u_N + z_N)$ is relatively large and negative), and may be greater, less than or equal to β_{SD}. The bias of half the upper bound as an estimator of $\sum_l p_l u_l$ ($\beta_{1/2\,UBND}$) is the minimum of $\sum_j p_j u_j$ or $\sum_k p_k u_k$. This will be greater than $\beta_{1/2\,TC(SC)}$ if $\sum_l u_l > \sum_i u_i + \text{maximum}(\sum_j p_j u_j, \sum_k p_k u_k) - \text{minimum}(\sum_j p_j u_j, \sum_k p_k u_k) + \sum_N (u_N + z_N)$. This occurs when: (1) there are many loci (or with large values of u) where the elite single cross is fixed for the unfavorable allele compared to where it is fixed for the favorable allele; (2) $\sum_j p_j u_j$ is nearly equal to $\sum_k p_k u_k$; or (3) there are large negative values of z_N relative to u_N. At least one of these conditions must be met, and, depending on magnitude, they might all be necessary in order for $\beta_{1/2\,UBND}$ to be greater than $\beta_{1/2\,TC(SC)}$. It seems likely, however, that $\beta_{1/2\,UBND}$ will be less than $\beta_{1/2\,TC(SC)}$. The relationship between $\beta_{1/2\,UBND}$ and β_{SD} is likely to be variable.

One would like to correctly rank populations in terms of their relative superiority even if the superiority measure cannot be estimated exactly. The most important consideration is the correlation between the estimators and the superiority measure. The correlation, r, between any statistic T and the superiority measure for a sample of populations to be screened will be:

$$\text{cov}(\sum_l p_l u_l, T)/\{[\text{var}(\sum_l p_l u_l)][\text{var}(T)]\}^{1/2}.$$

If the statistic is a function of testcross measurements, it will contain some or all of the terms: $\sum_i u_i$, $\sum_j u_j$, $\sum_k u_k$, $\sum_l u_l$, $\sum_i p_i u_i$, $\sum_j p_j u_j$, $\sum_k p_k u_k$, $\sum_l p_l u_l$, and $\sum_N (u_N + z_N)$. Terms with summations not involving p will be the same for each population and will not change the correlation between T and $\sum_l p_l u_l$. Such terms will not contribute to the genetic variance of T. If epistasis exists and/or the populations are not in equilibrium, there will be positive or negative correlations between $\sum_j p_j u_j$, $\sum_k p_k u_k$, and $\sum_l p_l u_l$, with sign and magnitude varying among populations and classes of loci. Over all populations, these terms at the different classes of loci may still be uncorrelated unless there are consistent linkage disequilibrium relationships among populations and/or systematic epistatic relationships among loci in certain classes. However, in any sample of populations, correlations between frequencies and effects at the different classes will occur. These correlations are impossible to predict. For the general case, the covariance of any terms involving p and u for different classes of loci is assumed to be zero. These terms will contribute to the variance of T, however. Also note that the correlation of any multiple of T with the superiority measure is equal to the correlation between the superiority measure and T itself.

The correlation of UBND with $\sum_l p_l u_l$ under the above assumptions is:

$$\text{var}(\sum_l p_l u_l)/\{[\text{var}(\sum_l p_l u_l)][\text{var}(\sum_l p_l u_l) + \text{var}(\text{minimum of } \sum_j p_j u_j,\ \sum_k p_k u_k)]\}^{1/2}.$$

The correlation between TC(SC) and $\sum_l p_l u_l$ is:

$$\text{var}(\sum_l p_l u_l)/\{[\text{var}(\sum_l p_l u_l)][\text{var}(\sum_l p_l u_l) + (1/4)\,\text{var}(\sum_j p_j u_j) + (1/4)\,\text{var}(\sum_k p_k u_k)]\}^{1/2}.$$

The correlation of SD with $\sum_l p_l u_l$ is:

$$\text{var}(\sum_l p_l u_l)/\{[\text{var}(\sum_l p_l u_l)][\text{var}(\sum_l p_l u_l) + (1/(\sum_j u_j - \sum_k u_k)^2)[(\sum_j u_j)^2\,\text{var}(\sum_k p_k u_k) + (\sum_k u_k)^2\,\text{var}(\sum_j p_j u_j)]]\}^{1/2}.$$

Because the covariances of $\sum p u$ for different classes of loci are assumed to be zero, the superiority measure and the bias are uncorrelated. Differences among the correlations of the statistics with the superiority measure are due to the variances of the biases. As the variance of the bias increases, the correlation decreases. There is no reason to expect that the $\text{var}(\sum_j p_j u_j)$ will be different from $\text{var}(\sum_k p_k u_k)$, so let $\text{var}(\sum_j p_j u_j) = \text{var}(\sum_k p_k u_k)$. There are two extremes for the distributions of $\sum_j p_j u_j$ and $\sum_k p_k u_k$: they may be non-overlapping or they may have the same mean. If they are non-overlapping, the variance of the minimum will be equal to $\text{var}(\sum_j p_j u_j)$ because values from the distribution with the lower mean will always be chosen. If they have the same means and are assumed to have the same distributions, the minimum $(\sum_j p_j u_j, \sum_k p_k u_k)$ is the minimum of a random sample of size two from the same distribution because the two values for a particular population are assumed to be uncorrelated. If this distribution is assumed to be normal, then the variance of the minimum is $0.68\,[\text{var}(\sum_j p_j u_j)]$ (Beyer 1968, p. 333). Under the same assumption of equal variances for $\sum_j p_j u_j$ and $\sum_k p_k u_k$, the correlation of TC(SC) with the superiority measure contains the term $(1/2)\,[\text{var}(\sum_j p_j u_j)]$. The correlation of SD has the term $\{[(\sum_j u_j)^2 + (\sum_k u_k)^2]/(\sum_j u_j - \sum_k u_k)^2\}\,[\text{var}(\sum_j p_j u_j)]$. The coefficient on $\text{var}(\sum_j p_j u_j)$ for the SD is greater than one because the denominator of the coefficient is equal to the numerator

minus $2\sum_j u_j \sum_k u_k$, a positive term under the assumption of directional dominance. The TC (SC) is then expected to be most highly correlated with the superiority measure, and the UBND is expected to be more highly correlated than the SD.

When gene action is additive among and within loci, any evaluation using testcross information results in equal correlations between the test statistic and the superiority measure. The TC (SC), UBND and SD are then equally correlated with $\sum_l p_l u_l$.

With multi-allelic loci, the loci can be grouped into classes analogous to the two allele case. In the i-th class, I1 and I2 are homozygous for the best allele. In the j-th class, I1 has the best allele but I2 has some other allele, and vice versa for the k-th class. Neither I1 nor I2 has the best allele at loci in the l-th class. If the superiority measure is defined as the product of the relative effect and frequency of the alleles in the population that are better than the best in the single cross, the three estimators estimate this with biases analogous to the two allele case. However, these biases also include terms involving the effects and frequencies of alleles in the population at class l loci that are better than one of the alleles in the single cross but not better than the best allele in the single cross. Because the relationships between the coefficients on the bias terms are analogous to those in the two allele case, multiple alleles are not expected to change the relative correlations between TC (SC), UBND and SD and the superiority measure.

In biological populations, measurements are made with error, which reduces the correlation between the superiority measure and its estimator. The greatest reduction under field conditions in the correlation is expected for SD because it is a function of the greatest number of measurements. Reduction in the correlation is expected to be less for the TC (SC) than the UBND because the TC (SC) involves fewer measurements (only one per population). Use of the single-cross tester has an advantage over the UBND and SD because of its ease of use in field experiments. Crossing may be easier with a more vigorous pollen or ear parent single cross as compared with an inbred line. Both the UBND and SD require crossing to two different inbred testers as well as growing the inbred parents per se. At least twice as many testing resources are required compared with the TC (SC). In addition, there may be problems estimating the performance of inbreds because of their reduced vigor and potentially different interactions with density and fertility levels when compared with testcrosses. If inbreds I1 and I2 are close in genetic value but are measured with sufficient error to reverse the sign of their difference, this will have a profound effect on the ranking of populations based on SD inasmuch as I1 – I2 serves as a scaling factor. The testcross to the reference single cross is expected to be more highly correlated with the superiority measure of a population as defined by $\sum_l p_l u_l$ and is more efficient in terms of testing resources required than either SD or UBND. The TC (SC) would seem to be the most practical choice for screening populations for their relative potential to contribute favorable alleles not available in elite, adapted germplasm.

Table 5. Summarization of presence (+) or absence (−) of the favorable allele in inbreds I1, I2 and I_W for the eight classes of loci (taken from Dudley [1984a] with minor notational changes)

Class of loci	Inbreds		
	I1	I2	I_W
A	+	+	+
B	+	+	−
C	+	−	+
D	+	−	−
E	−	+	+
F	−	+	−
G	−	−	+
H	−	−	−

Table 6. Expectation of genotypic value of inbreds, their F_1 and three-way hybrids and the upper bound (UBND) assuming complete dominance. Values for I1, I2, I_W and their single crosses taken from Dudley (1984a) with notational changes and without the assumption that A = H

I1	$= A+B+C+D-E-F-G-H+\sum_N (u_N+z_N)$
I2	$= A+B-C-D+E+F-G-H+\sum_N (u_N+z_N)$
I_W	$= A-B+C-D+E-F+G-H+\sum_N (u_N+z_N)$
I1 × I2	$= A+B+C+D+E+F-G-H+\sum_N (u_N+z_N)$
I1 × I_W	$= A+B+C+D+E-F+G-H+\sum_N (u_N+z_N)$
I2 × I_W	$= A+B+C-D+E+F+G-H+\sum_N (u_N+z_N)$
I_W × (I1 × I2)	$= A+B+C+E+G-H+\sum_N (u_N+z_N)$
UBND	$= \min(I_W \times I1 - I1,\ I_W \times I2 - I2)$ $= \min(2E+2G,\ 2C+2G) = 2G+\min(2E,\ 2C)$

Screening inbreds

Theory and interpretation. Dudley (1984a) grouped the loci affecting a quantitative trait into eight classes based on the presence or absence of the favorable allele in three different inbreds, assuming two alleles per locus (Table 5). I1 and I2 are the homozygous parents of the elite reference single cross; I_W is the homozygous line to be evaluated for the presence of favorable alleles not available in either I1 or I2 (at class G loci). The letters represent both the classes and number of loci in the class. The superiority measure of an inbred can be defined as G. Assuming

complete dominance and no epistasis, the genotypic value of the inbreds, their crosses, the three-way cross [TC(SC)] and the UBND (=minimum $[I_W \times I1 - I1$, $I_W \times I2 - I2]$) can be predicted (Table 6). If the u's vary among loci, the letters A through H represent the respective sums over the u's in the class. For example, $A = \sum_A u_A$. If there is no correlation between the sums of effects for the different classes of loci, the correlations between G and its estimators will be $var(G)/\{[var(G)][var(G) + var(\beta)]\}^{1/2}$, where β is the bias, because G and β will be uncorrelated. The estimator with the smallest variance of β will be most highly correlated with the superiority measure. If $A = H$, then $SD_I = [(I2 \times I_W) + (I1 \times I_W) - I_W - I2 - I1 - (I1 \times I2)]/4$ estimates G when the means of the crosses and inbreds replace their symbols (Dudley 1984a). If $A \neq H$, then the bias using SD_I as an estimator of G (β_{SD_I}) is $-(A-H)/2 - \sum_N (u_N + z_N)/2$. When the UBND is halved to estimate G, the bias ($\beta_{1/2\,UBND}$) is minimum (E, C). The bias to the TC(SC) when estimating G [$\beta_{TC(SC)}$] is $A - H + B + C + E + \sum_N (u_N + z_N)$. When these estimators are used to compare two inbreds, I_{W1} and I_{W2}, the differences between the estimators reflect differences in the superiority measures ($G_{W1} - G_{W2}$) and differences in β ($\beta_{W1} - \beta_{W2}$). These then reflect differences in relative numbers of loci at classes other than G. For example, the difference between TC(SC)s for I_{W1} and I_{W2} is $A_{W1} - A_{W2} - H_{W1} + H_{W2} + B_{W1} - B_{W2} + C_{W1} - C_{W2} + E_{W1} - E_{W2} + G_{W1} - G_{W2}$. The variance of $-(A-H)/2 - \sum_N (u_N + z_N)/2$, β_{SD_I}, is $(1/4)[var(A) + var(H)]$ if A and H are uncorrelated. This is less than the variance of $\beta_{TC(SC)}$, which is $var(A) + var(H) + var(B) + var(C) + var(E)$. Half the UBND (and the UBND itself) will be more highly correlated with G than TC(SC) because var[minimum(E, C)] is at the most maximum[var(E), var(C)], which is less than var[$\beta_{TC(SC)}$]. The variance of the minimum of E and C has as its lower bound 0.68[var(C)] when both E and C have the same variance in a group of inbreds, are uncorrelated, and are assumed to have a normal distribution. If A through H have the same variance, the variance of β_{SD_I} will be (1/2)[var(A)]. This is lower than the lower bound of the variance of minimum(E, C). Of these three estimators of G, SD_I is expected to have the greatest genetic correlation with G.

The measurements needed to calculate SD_I and UBND can also be used to estimate C+F and D+E (Dudley 1984a). These functions are estimates of the effects of loci where I_W has alleles like I1 but unlike I2, and those where I_W is like I2 but unlike I1, respectively. They provide an estimate of the relationship of I_W to I1 and I2, respectively. This information might be useful in determining with which inbred to cross I_W in a pedigree breeding program to maintain the heterotic pattern, and is unknown when only $TC(SC) = I_W \times (I1 \times I2)$ is grown. The advantage for the TC(SC) is that only one cross is needed to evaluate each inbred, and the inbreds per se need not be grown. This reduces testing resources approximately one-third compared with SD_I and eliminates the need to evaluate inbreds. The UBND also has the advantage over SD_I of not requiring estimation of I_W. Twice as many resources are needed for evaluation using UBND than TC(SC).

The effect of environmental variance is unpredictable without knowing something about the variances of A through H in a sample of inbred lines to be tested. Because the number of measurements used in calculating the estimating statistics varies, this may greatly affect the relative correlations between these estimators and G. The potential problems in estimating unadapted and adapted inbred performance and the differences in testing resources required indicates that further research is needed to evaluate the possible estimators of G in terms of their expected phenotypic correlations with G and efficiency in use of resources.

References

Beyer WH (1968) Handbook of tables for probability and statistics, 2nd edn. CRC Press, Boca Raton, Florida, 642 pp

Brown, WL (1965) Performance and quality in cereals as influenced by genotype and environment. In: Irving GW Jr, Hoover SR (eds) Food quality: effects of production, practices and processing. Am Assoc Adv Sci Publ 77. Washington/DC, pp 38–98

Brown WL (1983) Genetic diversity and genetic vulnerability – an appraisal. Econ Bot 37: 4–12

Burton GW, Davies WE (1984) Handling germplasm of cross-pollinated forage crops. In: Holden JHW, Williams JT (eds) Crop genetic resources: conservation and evaluation. Allen and Unwin, London, pp 180–190

Busch RH, Janke JC, Frohberg RC (1974) Evaluation of crosses among high and low yielding parents of spring wheat (*Triticum aestivum* L.) and bulk prediction of line performance. Crop Sci 14: 47–50

Comstock RE, Robinson HF (1948) The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. Biometrics 4: 254–266

Dudley JW (1984a) A method of identifying lines for use in improving parents of a single cross. Crop Sci 24: 355–357

Dudley JW (1984b) A method for identifying populations containing favorable alleles not present in elite germplasm. Crop Sci 24: 1053–1054

Duvick DN (1981) Genetic diversity in corn improvement. Proc Annu Corn Sorghum Industry Res Conf 36: 48–60

Eberhart SA (1971) Regional maize diallels with US and semi-exotic varieties. Crop Sci 11: 911–914

Frey KJ (1983) Genes from wild relatives for improving plants. In: Yap TC, Graham KM, Sukaimi J (eds) Crop improvement research. Proc 4th Int Congr Soc Adv Breed Res Asia Oceania (SABRAO). Bangi, Selangor, Malaysia, pp 1–20

Goodman MM (1985) Exotic maize germplasm: status, prospects, and remedies. Iowa State J Res 59: 497–528

Green JM (1948) Inheritance of combining ability in maize hybrids. J Am Soc Agron 40: 58–63

Hallauer AR, Miranda JB Fo (1981) Quantitative genetics in maize breeding. Iowa State University Press, Ames, Iowa, 468 pp

Harlan JR (1976) Genetic resources in wild relatives of crops. Crop Sci 16: 329–333

Harlan JR (1984) Evaluation of wild relatives of crop plants. In: Holden JHW, Williams JT (eds) Crop genetic resources: conservation and evaluation. Allen and Unwin, London, pp 212–222

Hayes HK, Johnson IJ (1939) The breeding of improved selfed lines of corn. J Am Soc Agron 31: 710–724

Josephson LM (1982) Development and utilization of populations in maize breeding. Proc 5th South African Maize Breed Symp, Publ 5. Department of Agriculture and Fisheries, Pietermaritzburg, South Africa, pp 22–24

Kramer HH, Ullstrup AJ (1959) Preliminary evaluations of exotic maize germ plasm. Agron J 51: 687–689

Lonnquist JH, Gardner CO (1961) Heterosis in intervarietal crosses in maize and its implication in breeding procedures. Crop Sci 1: 179–183

Plucknett DL, Smith NJH, Williams JT, Anishetty NM (1983) Crop germplasm conservation and developing countries. Science 220: 163–169

Spencer J (1980) Evaluation of tropical germ plasm. Proc 4th South African Maize Breed Symp, Tech Commun 172. Department of Agriculture and Fisheries, Pietermaritzburg, South Africa, pp 20–35

Stuber CW (1978) Exotic sources for broadening genetic diversity in corn breeding programs. Proc Annu Corn Sorghum Res Conf 33: 34–47