Theor Appl Genet (1994) 88: 653–661

M. A. Fabrizius · S. J. Openshaw

Methods to evaluate populations for alleles to improve an elite hybrid

Received: 3 August 1993 / Accepted: 26 October 1993

Abstract Elite hybrids can be improved by the introgression of favorable alleles not already present in the hybrid. Our first objective was to evaluate several estimators derived from quantitative genetic theory that attempt to quantify the relative number of useful alleles in potential donor populations. Secondly, we wanted to evaluate two proposed ways of determining relatedness of donor populations to the parents of the elite hybrid. Two experiments, each consisting of 21 maize populations of known pedigree, were grown at three and four environments in Minnesota in 1991. Yield and plant height means were used to provide estimates of each of the following statistics: (1) LPLU, a minimally biased statistic, (2) UBND, the minimum estimate of an upper bound, (3) NI, the net improvement, (4) PTC, the predicted three-way cross, and (5) TCSC, the testcross of the populations. These statistics are biased estimators of the relative number of unique favorable alleles contained within a population compared to a reference elite hybrid. Based on rank correlations, all statistics except NI ranked populations similarly. The percent novel germplasm relative to the single cross to be improved was positively correlated with the estimates of favorable alleles except when NI was used as the estimator. The relationship estimators agreed with the genetic constitution of the donor populations. Strong positive correlations existed between diversity, based on the relationship rankings, and all the estimator rankings, except NI. Potential donor populations were effectively identified by LPLU, UBND, PTC, and TCSC. NI was not a good estimator of unique favorable alleles.

Key words Testcross · Upper bound · Favorable alleles Populations

Introduction

Introgression of new germplasm is a vital part of plant breeding programs. While new germplasm can potentially contribute to a breeder's program, the identification of exotic populations with useful and unique alleles can be a difficult task. Dudley (1984) proposed a method for identifying populations that could contribute unique favorable alleles to a reference hybrid. Modifications were presented in a later paper (Dudley 1987). In these two papers, Dudley defined four classes of loci relative to a reference hybrid: *i*, which is homozygous favorable in the hybrid; *j* and *k*, which are heterozygous in the hybrid; and class ℓ, which is homozygous unfavorable. Expectations for various gen-

Table 1 Expected genotypic means of the inbreds and their crosses[a]. The genetic model consists of: N = total number of loci, $N = i + j + k + \ell$, z = the value of the unfavorable homozygote, u = half the difference between the favorable homozygote and the unfavorable homozygote, a = the general level of dominance, and p = the favorable gene frequency (Dudley, 1987)

$I_1 \quad = N(z+u) + (i+j-k-\ell)\,u$

$I_2 \quad = N\,(z+u) + (i-j+k-\ell)\,u$

$I_1{\times}I_2 \quad = N\,(z+u) + (i+j\mathrm{a}+k\,\mathrm{a}-\ell)\,u$

$I_1{\times}P_y \quad = N\,(z+u) + \{i\,(p_i+q_i\,a) + j\,(p_j+q_j\,a) + k\,[p_k(1+a)-1] + \ell\,[p_\ell(1+a)-1]\}\,u$

$I_2{\times}P_y \quad = N\,(z+u) + \{i\,(p_i+q_i\,a) + j\,[p_j\,(1+a)-1] + k\,(p_k+q_k\,a) + \ell\,[p_\ell(1+a)-1]\}\,u$

$P_y{\times}(I_1{\times}I_2) = N\,(z+u) + \{i\,(p_i+q_i\,a) + j\,[p_j+½\,(a-1)] + k\,[p_k+½\,(a-1)] + \ell\,(p_\ell\,a-q_\ell)\}\,u$

[a] I_1 and I_2 are the inbred parents of the reference single cross ($I_1{\times}I_2$), P_y is a donor population. I_1 has favorable alleles at *i*- and *j*-class loci and unfavorable at *k*- and ℓ-class loci. I_2 has favorable alleles at *i* and *k* and unfavorable at *j*- and ℓ-class loci

---

Communicated by A. R. Hallauer

M. A. Fabrizius (✉) · S. J. Openshaw
Department of Agronomy and Plant Genetics,
411 Borlaug Hall,
1991 Buford Circle,
University of Minnesota,
St. Paul, MN 55108,
USA

Journal Paper No. 20,683 of the Minnesota Agric. Exp. Stn., St. Paul, MN 55108. Project No. 1318. Part of a thesis submitted by the senior author in partial fulfillment of MS requirements

Table 2 Formulas and expectations of the superiority measure estimators

Case	LPLU
$1=q_{j0},\ q_{k1}$	$\tfrac{1}{4}\{(I_1\times P_y)+(I_2\times P_y)-(I_1\times I_2)-I_1]$ $\tfrac{1}{2}\ell[p_\ell(1+a)]\,u+\tfrac{1}{2}i\,(p_i+q_i a-1)\,u-\tfrac{1}{2}j\,q_j u+\tfrac{1}{2}k\,p_k\,u$
$2=q_{j0},\ q_{j1}$	$\tfrac{1}{2}(I_2\times P_y)-\tfrac{1}{4}[(I_1\times I_2)+I_2]$ $\tfrac{1}{2}\ell[p_\ell(1+a)]\,u+\tfrac{1}{2}i\,(p_i+q_i a-1)\,u+\tfrac{1}{2}j\,(1+a)\,(p_j-\tfrac{1}{2})\,u+\tfrac{1}{2}k[p_k+q_k a-\tfrac{1}{2}\,1+a)]\,u$
$3=q_{j1},\ q_{k0}$	$\tfrac{1}{4}[(I_1\times P_y)+(I_2\times P_y)-(I_1\times I_2)-I_2]$ $\tfrac{1}{2}\ell[p_\ell(1+a)]\,u+\tfrac{1}{2}i\,(p_i+q_i a-1)\,u+\tfrac{1}{2}j\,p_j u-\tfrac{1}{2}k\,q_k\,u$
$4=q_{k1},\ q_{k0}$	$\tfrac{1}{2}(I_1\times P_y)-\tfrac{1}{4}[(I_1\times I_2)+I_1]$ $\tfrac{1}{2}\ell[p_\ell(1+a)]\,u+\tfrac{1}{2}i\,(p_i+q_i a-1)\,u+\tfrac{1}{2}j\,[p_j+q_j a-\tfrac{1}{2}(1+a)]\,u+\tfrac{1}{2}k\,(1+a)\,(p_k-\tfrac{1}{2})\,u$

Case decision (Dudley 1987):

The difference between two crosses can be written as follows;
$(I_1\times P_y)-(I_2\times P_y)=[(I_1\times I_2)-I_2]\,q_j-[(I_1\times I_2)-I_1]\,q_k$

Solve for lower limits: if $(I_1\times P_y)-(I_2\times P_y)$ is positive, set $q_k=0$, if negative set $q_j=0$. No solution for q exists when the equation is positive and greater than $(I_1\times I_2)-I_2$ or negative and greater than $(I_1\times I_2)-I_1$.

Solve for upper limits: if $(I_1\times P_y)-(I_2\times P_y)+(I_1\times I_2)-I_1$ is positive and less than $(I_1\times P_y)-I_2$, set $q_k=1$ if not set $q_j=1$.

	UBND
	$\min\{[(I_1\times P_y)-I_1],\ [(I_2\times P_y)-I_2]\}$
Minimum of:	$\ell\,[p_\ell(1+a)]\,u+i\,(p_i+q_i a-1)\,u+j\,(p_j+q_j a-1)\,u+k[p_k(1+a)]\,u$ $\ell\,[p_\ell(1+a)]\,u+i\,(p_i+q_i a-1)\,u+j\,[p_j(1+a)]\,u+k\,(p_k+q_k a-1)\,u$

	NI
	$\max\{\tfrac{1}{2}[(I_1\times P_y)-(I_1\times I_2)],\ \tfrac{1}{2}[(I_2\times P_y)-(I_1\times I_2)]\}$
Maximum of:	$\tfrac{1}{2}\ell\,[p_\ell(1+a)]\,u+\tfrac{1}{2}\{i\,(p_i+q_i a-1)\,u+jp_j(1-a)\,u-kq_k(1+a)\,u\}$ $\tfrac{1}{2}\ell\,[p_\ell(1+a)]\,u+\tfrac{1}{2}\{i\,(p_i+q_i a-1)\,u-jq_j(1+a)\,u+kp_k(1-a)\,u\}$

	PTC
	$\tfrac{1}{2}[(I_1\times P_y)+(I_2\times P_y)]$ $\ell\,[p_\ell(1+a)-1]\,u+N\,(z+u)+(p_i+q_i a)\,u+j\,[p_j-\tfrac{1}{2}(1-a)]\,u+k\,[p_k-\tfrac{1}{2}(1-a)]\,u$

	TCSC
	$P_y\times(I_1\times I_2)$ $\ell\,[p_\ell(1+a)-1]\,u+N\,(z+u)+(p_i+q_i a)\,u+j\,[p_j-\tfrac{1}{2}(1-a)]\,u+k\,[p_k-\tfrac{1}{2}(1-a)]\,u$

Therefore:

$$\begin{aligned}\mathrm{PTC}_i-\mathrm{PTC}_i'&=\mathrm{TCSC}_i-\mathrm{TCSC}_i'^{\,a}\\&=1\,[(p_\ell-p_\ell')\,(1+a)]\,u+i\,[(p_i-p_i')+(q_i-q_j')\,a]\,u+j\,(p_j-p_j')\,u+k\,(p_k-p_k')\,u\end{aligned}$$

[a] Empirical results failing to conform to this expectation could be due to sampling error, epistasis, or unequal genetic effects among loci

otypes were derived in terms of these classes and the genetic model outlined in Table 1. The relative number of favorable alleles in class ℓ, ($\ell p_\ell u$), will interest the breeder most since the hybrid to be improved has no favorable alleles at class-ℓ loci. This parameter, $\ell p_\ell u$, is what Gerloff and Smith (1988) dubbed the superiority measure of a population. It is the number of loci in the ℓ class, ℓ, times the favorable gene frequency of the ℓ class, p_ℓ, times the genetic effect of the loci, u. We will use $\ell p_\ell u$ to designate the true parameter value and LPLU to designate Dudley's (1987) estimator of this parameter.

Based on the same genetic model, a relatedness measure, also proposed by Dudley (1987), quantifies the degree of relationship between a population and the parents of the reference single cross. It is not, however, a measure of consanguinity (Pfarr and Lamkey 1992 b).

Dudley's (1987) modified statistic (LPLU), the predicted three-way cross (PTC) (Hallauer and Miranda 1988), the testcross of the population to the single-cross hybrid (TCSC) (Kramer and Ullstrup 1959; Stuber 1978), the minimum upper bound (UBND) (Gerloff and Smith 1988), and the net improvement statistic (NI) (Bernardo 1990 a), are all biased estimates of the relative number of favorable alleles at the class-ℓ loci, i.e., $\ell p_\ell u$. A bias exists because there is a difference between the expectations of these statistics and the true parameter value, e.g.,

$\mathrm{E(UBND)}-2\ell \mathrm{p}_{\ell}\mathrm{u}=\ell \mathrm{p}_{\ell}\mathrm{u(a-1)}+i(\mathrm{p}_i+q_i\mathrm{a}-1)\mathrm{u}+j(\mathrm{p}_j+\mathrm{q}_j\mathrm{a}-1)\mathrm{u}+k[\mathrm{p}_k(1+\mathrm{a})]\mathrm{u}$. Therefore the bias is caused by partial or overdominance at the ℓ, i and j loci and effects at the k-class loci. Expectations for the statistics are given in Table 2.

Recent evaluations of LPLU, UBND, and TCSC (Hogan and Dudley 1991; Pfarr and Lamkey 1992 a,b) have used donor populations with a defined percentage of new germplasm relative to the reference single cross to be improved. This genetic structure is used to study the effect of novel germplasm on the superiority-measure estimates provided by the statistics. In these studies, the estimates were highly correlated with the genetic structure of the populations.

The ability of the estimates to detect differences between populations has been questioned due to bias effects and standard errors of the estimators. Misevic (1989) claimed that the estimators could detect differences between populations, while Pfarr and Lamkey (1992 b) observed that LPLU could only distinguish between the most-diverse populations.

Several studies have also evaluated Dudley's (1987) relatedness statistic. Results indicate that its performance is consistent with pedigree information (Dudley 1988; Zanoni and Dudley 1989; Hogan and Dudley 1991; Pfarr and Lamkey 1992 b).

The first objective of the present study was to evaluate statistics for their ability to effectively identify populations that contain unique favorable alleles relative to a reference maize hybrid. The second objective was to evaluate the effectiveness of Dudley's (1987) relatedness measure. To do this, two series of donor populations were created with 0, 25, 50, 75, or 100% novel germplasm relative to the reference single cross to be improved; one series of donor populations was evaluated for potential to improve the single cross A679×A682, and the other to improve FR902×LH82.

Methods

Two sets of genetic material were used for evaluating the superiority measure estimators. Donor populations in Experiment 1 were evaluated for their ability to contribute favorable alleles to the single cross, A679xA682. In Experiment 2, FR902xLH82 was used as the reference single-cross. The donor populations are actually single cross and backcross populations as indicated by the pedigrees in Table 3. Each experiment consists of 21 donor populations. These populations have a certain percentage of novel germplasm relative to the single cross, 0.0, 25, 50, 75, or 100%. These may not be exact percentages however, since some relationship may exist between inbreds, for example between A679 and FR902. A backcross (BC) series will be defined as the single cross and each parental backcross, for a total of seven BC series. BC series 1 contains populations with 0.0% novel germplasm, BC series 2–5 contain populations with 25, 50, or 75% novel germplasm, and BC series 6 and 7 contain populations that have 100% novel germplasm relative to the reference single cross. This structure allows comparisons of the estimators with the relative expectations based on pedigree.

Table 3 Donor population pedigrees and the percent novel germplasm they contain relative to the reference single cross. Reference single crosses are: A679 × A682 for Experiment 1 and FR902 × LH82 for Experiment 2

BC series	Novel[a] (%)	Donor	Exp 1	Exp 2
1	0	P_1	(A679xA682)	(FR902xLH82)
1	0	P_2	(A679xA682) A679	(FR902xLH82) FR902
1	0	P_3	(A679xA682) A682	(FR902xLH82) LH82
2	50	P_4	(A679xFR902)	(FR902xD474)
2	25	P_5	(A679xFR902) A679	(FR902xD474) FR902
2	75	P_6	(A679xFR902) FR902	(FR902xD474) D474
3	50	P_7	(A679xP3)	(FR902xA679)
3	25	P_8	(A679xP3) A679	(FR902xA679) FR902
3	75	P_9	(A679xP3) P3	(FR902xA679) A679
4	50	P_{10}	(A682xP3)	(LH82xP3)
4	25	P_{11}	(A682xP3) A682	(LH82xP3) LH82
4	75	P_{12}	(A682xP3)P3	(LH82xP3) P3
5	50	P_{13}	(A682xLH82)	(LH82xA682)
5	25	P_{14}	(A682xLH82) A682	(LH82xA682) LH82
5	75	P_{15}	(A682xLH82) LH82	(LH82xA682) A682
6	100	P_{16}	(FR902xP3)	(D474xA679)
6	100	P_{17}	(FR902xP3) FR902	(D474xA679) D474
6	100	P_{18}	(FR902xP3) P3	(D474xA679) A679
7	100	P_{19}	(P3xLH82)	(P3xA682)
7	100	P_{20}	(P3xLH82) P3	(P3xA682) P3
7	100	P_{21}	(P3xLH82) LH82	(P3xA682) A682

[a] Relative to reference single cross

Except for P3 and D474, the genetic background of the inbreds in these experiments is known. Based on testcross evaluations at Minnesota, D474 appears to be of stiff stalk (BSSS) origin, while P3 crosses well with lines of BSSS and non-BSSS origin. Inbreds A679 and FR902 belong to the BSSS heterotic group. A679 is an early B73 and FR902 is a B14 type. The other two inbreds, A682 and LH82, are not in the BSSS heterotic group. A682 is of C103 origin while LH82 is from the W153R family.

The experiments were conducted during the summer of 1991. Experiment 1 was grown at four locations and Experiment 2 at three locations. The three locations in common were the Rosemount Experiment Station at Rosemount, Minnesota, the Southwest Experiment Station at Lamberton, Minnesota, and the Southern Experiment Station at Waseca, Minnesota. The fourth location used in Experiment 1 was Olivia, Minnesota in cooperation with DeKalb Plant Genetics. A randomized complete block design was used with three-to-five replications per location. P_y refers to a donor population and I_1 and I_2 are the parent inbreds of the reference hybrid to be improved. Each block included $I_1 \times P_y$, $I_2 \times P_y$ (the inbred donor population test crosses), duplicate plots of the single cross to be improved ($I_1 \times I_2$), and duplicate plots of the inbred parents per se (I_1 and I_2). The TCSC, ($P_y \times (I_1 \times I_2)$), was only evaluated for Experiment 1 populations. The TCSC progeny were grown in a separate trial at Rosemount, Lamberton, and Waseca with either four or five replications.

All trials consisted of two row plots 6.7 m long with 0.76 m between rows. A timely planting in early May was achieved for all locations. Plots were overplanted and thinned to 59 000 plants ha^{-1} at the 6–9 leaf stage. Standard management levels and cultural practices were used for herbicides, cultivation, and fertilization.

Entries were evaluated for plant height and grain yield. Plant height was recorded as the visual mean from the soil surface to the flag leaf of all the plants in the plot. Grain moisture at harvest was recorded and grain yield was adjusted to 155 g kg^{-1} moisture. All trials were machine harvested.

Separate analyses of variance for yield and plant height were computed for each experiment in each location and for the combined analysis across locations. Estimates of LPLU, and the j and k class relative number of allele estimates, $jp_j\mathrm{u}$, $jq_j\mathrm{u}$, $kp_k\mathrm{u}$, and $kq_k\mathrm{u}$, were calculated according to the method described by Dudley (1987). The UBND estimates were calculated according to Gerloff and Smith (1988), NI according to Bernardo (1990 b), and PTC according to Hallauer and Miranda (1988).

Table 4 Estimates of the relative number of alleles for yield at the *j*- and *k*-class loci

Donor	Experiment 1[a]					Experiment 2[b]				
	jp_ju	jq_ju	kp_ku	kq_ku	Case[c]	jp_ju	jq_ju	kp_ku	kq_ku	Case
P_1	1.16	1.84	1.11	1.11	4	1.39	1.39	1.99	1.48	2
P_2	2.22	0.78	0.78	1.43	1	1.93	0.86	0.86	2.62	1
P_3	0.54	2.46	1.68	0.54	3	0.63	2.16	2.84	0.63	3
P_4	2.68	0.32	0.32	1.89	1	2.27	0.52	0.52	2.95	1
P_5	2.70	0.30	0.30	1.91	1	2.50	0.29	0.29	3.19	1
P_6	2.32	0.68	0.68	1.53	1	1.93	0.86	0.86	2.62	1
P_7	2.47	0.53	0.53	1.69	1	1.97	0.82	0.82	2.65	1
P_8	2.55	0.45	0.45	1.76	1	2.49	0.29	0.29	3.18	1
P_9	2.00	1.00	1.00	1.22	1	1.51	1.28	1.28	2.20	1
P_{10}	0.60	2.40	1.61	0.60	3	0.46	2.33	3.02	0.46	3
P_{11}	0.30	2.70	1.92	0.30	3	0.45	2.33	3.02	0.45	3
P_{12}	0.88	2.12	1.34	0.88	3	1.03	1.76	2.44	1.03	3
P_{13}	0.43	2.57	1.79	0.43	3	0.57	2.21	2.90	0.57	3
P_{14}	0.15	2.85	2.06	0.15	3	0.22	2.56	3.25	0.22	3
P_{15}	0.96	2.05	1.26	0.96	3	0.82	1.97	2.65	0.82	3
P_{16}	–	–	–	–	–	1.64	1.15	1.15	2.33	1
P_{17}	–	–	–	–	–	1.61	1.18	1.18	2.30	1
P_{18}	1.91	1.09	1.09	1.12	1	1.47	1.32	1.32	2.16	1
P_{19}	1.51	1.49	1.11	1.11	4	1.21	1.58	2.27	1.21	3
P_{20}	1.65	1.35	1.11	1.11	4	1.24	1.55	2.24	1.24	3
P_{21}	1.23	1.77	1.11	1.11	4	1.18	1.60	2.29	1.18	3
SE	0.125	0.101	0.101	0.101	1	0.132	0.108	0.108	0.108	1
	–	–	–	–	2	0.054	0.054	0.210	0.216	2
	0.101	0.202	0.125	0.101	3	0.108	0.216	0.132	0.108	3
	0.196	0.202	0.051	0.051	4	0.210	0.216	0.054	0.054	4

[a] Based on combined means from Lamberton, Waseca, Rosemount, and Olivia

[b] Based on combined means from Lamberton, and Rosemount

[c] Case used to calculate the estimate: 1=qj0, qk1, 2=qj0, qj1, 3=qj1, qk0, 4=qk1, qk0 (Table 2)

Each of the following statistics, LPLU, jp_ju, jq_ju, kp_ku, and kq_ku, can be calculated from four different equations. The choice of which equation to use is dependent upon the observed data. Different equations are necessary such that an allele frequency value between zero and one can be estimated.

The statistics were compared to each other using Spearman's rank correlation coefficient (Conover 1971). The statistics were evaluated by their ability to correctly rank populations with respect to percent novel germplasm in the donor populations. BC series 2–5 each have populations with 25, 50, and 75% novel germplasm; thus, in effect, the BC series in conjunction with the percent novel germplasm creates a 4×3 factorial. Analysis of this factorial provided an F statistic to test the significance of the linear effect of new germplasm on the superiority measure estimates within a BC series. The 21 populations were also divided into 0, 25, 50, 75, and 100% novel germplasm groups for an examination of the linear effect of new germplasm across a wider range of genetic diversity.

Dudley's (1987) method for relatedness was calculated and the values were checked against pedigree background. Bernardo's (1990 b) method for determining the relationship of the population to I_1 or I_2 was evaluated, and discrepancies from Dudley's (1987) measure were noted.Diversity, according to Dudley's (1987) estimate, and estimates of the relative number of unique favorable alleles, were also compared.

Results and discussion

Combining data

Experiment 2 and TCSC data at Waseca were extremely variable and were consequently not used in the final analysis. Bartlett's test for heterogeneity of error indicated that errors for yield ($P< 0.001$) and plant height ($P<.005$) were non-homogeneous across locations for these two trials. Heterogeneity of error was less significant for Experiment 1.

Significant genotype-by-location interaction (G×L) existed for yield and plant height in Experiment 1 when all locations were used. When only data from Lamberton and Rosemount were used, no plant height G×L was indicated. G×L also was not significant for Experiment 2 plant height. The G×L mean square was used for calculating the standard errors of all statistics. When the G×L mean square was non-significant ($p>0.20$) it was pooled with the error variance and the pooled variance was used for calculating the standard errors.

Combined location means for the reference hybrid, its parents, and all population-by-inbred crosses were used in all calculations. Experiment 1 TCSC data can only be compared to other Experiment 1 data at two locations (Rosemount and Lamberton). Data for populations 16 and 17 were only available from these two locations.

J- and *k*-class allele estimates

The *j*-class loci are defined by loci that contain favorable alleles in I_1 and unfavorable alleles in I_2. The *k*-class loci are those loci that contain favorable alleles in I_2 and unfa-

Table 5 Estimator values for Experiment 1 yield (mg/ha) and plant height (cm). Combined data from Lamberton and Rosemount

Novel	Donor	LPLU		UBND		NI		PTC		TCSC	
(%)		Yield	Ht	Yield	Ht	Yield	Ht	Yield	Ht	Yield	Ht
0	P_1	–0.21	4.0	2.07	28.5	–1.45	–6.3	7.97	201	7.88	194
0	P_2	–0.34	2.8	1.16	27.0	–1.26	–8.0	8.36	203	8.14	200
0	P_3	–0.10	3.4	1.24	19.1	–0.81	–2.7	8.19	200	7.63	200
25	P_5	–0.15	3.8	0.90	22.6	–0.75	–3.8	8.74	205	9.25	210^{b}
25	P_8	–0.07	2.4	1.01	15.9	–0.65	–3.2	8.90	203	9.30	203
25	P_{11}	–0.03	3.3	0.57	16.3	-0.35^{b}	-1.6^{b}	8.32	200	8.32	196
25	P_{14}	0.04	2.2	0.53	11.3	-0.18^{b}	-1.3^{b}	8.47	198	7.80	191
50	P_4	0.09	8.1	1.01	34.8	-0.32^{b}	-1.3^{b}	9.23	214	9.56	211^{b}
50	P_7	0.32	5.0	1.99	27.6	-0.35^{b}	–3.8	9.69	208	9.44	203
50	P_{10}	0.46	5.6	2.32	24.1	-0.24^{b}	-0.7^{b}	9.30	205	8.83	200
50	P_{13}	0.42	3.6	1.84	20.7	-0.08^{b}	–3.2	9.22	201	9.21	199
75	P_6	0.55	7.8	2.59	38.7^{b}	-0.20^{b}	–3.8	10.14	213	9.70	209^{b}
75	P_9	0.49	6.1	2.91	32.6	–0.47	–4.1	10.03	210	9.72	209^{b}
75	P_{12}	0.75	6.5	3.39	33.0	-0.20^{b}	–3.5	9.87	206	9.41	206
75	P_{15}	0.87^{b}	3.6	3.92	29.1	-0.21^{b}	–7.4	10.13	201	9.59	196
100	P_{16}	0.92^{b}	10.1^{b}	4.04^{b}	40.9^{b}	-0.17^{b}	-0.2^{b}	10.89^{b}	218^{b}	10.49^{b}	212^{b}
100	P_{17}	0.90^{ab}	8.3^{a}	4.29^{b}	40.4^{b}	-0.35^{b}	–3.5	10.53	211	10.64^{b}	207^{b}
100	P_{18}	0.93^{ab}	9.2^{ab}	4.36^{b}	42.0^{b}	-0.32^{b}	–2.7	10.78^{b}	214	10.67^{b}	207^{b}
100	P_{19}	1.12^{ab}	4.4^{a}	4.73^{b}	32.6	-0.13^{b}	–7.4	11.07^{b}	204	10.66^{b}	201
100	P_{20}	1.12^{ab}	8.6^{b}	4.74^{b}	40.7^{b}	-0.13^{b}	–3.2	10.89^{b}	211	10.62^{b}	207^{b}
100	P_{21}	1.10^{ab}	3.9^{a}	4.70^{b}	31.5	-0.14^{b}	–8.0	10.80^{b}	202	10.70^{b}	199
		Standard errors									
		0.124	0.87	0.352	2.47	0.176	1.24	0.215	1.5	0.302	2.8
		0.164^{a}	1.16^{a}								

[a] These estimations have the standard error with the same note
[b] Within two standard errors of the best population

vorable alleles in I_1. The *j*- and *k*-class relative number of allele estimates follow expectations based on pedigree. An interesting comparison can be made between the relative allele estimates in the *j*-class loci and those in the *k* class (Table 4). Based on yield and plant height data (only yield is shown) in Experiment 1, $jp_j u$ estimates for P_4–P_9 were greater than $kp_k u$ estimates for P_{10}–P_{15}. These two estimates can also be compared to each other in P_{16}–P_{21} where once again $jp_j u > kp_k u$. This could be due to class-*j* loci outnumbering class-*k* loci, favorable gene frequency at the *j*-class loci being greater than at the *k* class, or the equal genetic effects (u) assumption not being correct, i.e., class-*j* loci have a greater effect on yield than class-*k* loci.

For the germplasm in Experiment 2, $kp_k u > jp_j u$ for both yield and plant height data. This is the reverse of Experiment 1 results. Therefore in this germplasm, *k*-class loci outnumber *j*-class loci, $p_k > p_j$, or loci in the *k* class have more effect on the traits than the loci in the *j*-class. A comparison of estimated allele frequencies (data not shown) indicates that other factors are involved besides a difference between p_k and p_j.

These results suggest that, in Experiment 1, A679 is a better inbred than A682 and, in Experiment 2, LH82 is a better inbred than FR902. These results conform to breeders' experiences.

Superiority measures

Breeders will be most interested in the relative number of favorable alleles in class ℓ, the class for which the reference hybrid has no favorable alleles. Class *j* and *k* are interesting as a check for the theory or for determining to which parent, I_1 or I_2, the population is more related. Each statistic presented in Tables 5 and 6 was evaluated as an estimator of the relative number of ℓ-class favorable alleles ($\ell p_\ell u$), i.e. the superiority measure of a population.

Based on yield data, the estimators identified nearly the same populations within two standard errors of the best population. The net improvement statistic, however, was unable to separate a small group of populations as the best due to its insensitivity, i.e., its relatively-large standard error. The NI statistic was unable to provide estimates significantly greater than zero due to lack of precision and its negative bias. Plant height results were more variable, NI estimates were very different from the other statistic estimates for Experiment 1 but were similar for Experiment 2.

Populations with 25% novel germplasm were not always identified as having a significant superiority measure by LPLU and UBND. However, it is not the absolute values of the estimates that are of use but rather the value of a population relative to other populations being evaluated in the same experiment. The inability to detect small estimates significantly different from zero is not a major

Table 6 Estimator values for Experiment 2 yield (mg/ha) and plant height (cm). Combined data from Lamberton and Rosemount

Novel (%)	Donor	LPLU		UBND		NI		PTC	
		Yield	Ht	Yield	Ht	Yield	Ht	Yield	Ht
0	P_1	–0.24[a]	4.0[a]	2.32	23.5	–1.63	–3.8	8.20	174
0	P_2	–0.04	4.3[a]	1.64	24.0	–0.90	–3.5	8.00	184
0	P_3	–0.18	4.8	0.91	21.0	–0.81	–1.0	8.41	172
25	P5	0.04	2.9	0.66	12.0	–0.25	–0.3	8.16	182
25	P_8	0.37	6.6	1.33	23.0	0.08[b]	1.8	8.82	189
25	P_{11}	–0.07	6.4	0.77	21.5	–0.53	2.0	8.62	175
25	P_{14}	0.14	6.6	0.72	21.0	–0.09[b]	2.8	9.04	176
50	P_4	0.19	5.0	1.42	21.0	–0.33	–0.5	8.46	186
50	P_7	0.69	7.3	3.03	29.0	–0.13[b]	0.0	9.46	191
50	P_{10}	0.14	9.8	1.19	31.5	–0.32	3.8	9.05	182
50	P_{13}	0.42	11.8	2.00	36.5	–0.15	5.3	9.61	186
75	P_6	0.63	5.8	2.97	23.0	–0.23	0.0	9.33	188
75	P_9	0.98	12.8[a]	4.52	41.0	–0.30	5.0	10.04	199[b]
75	P_{12}	0.43[a]	11.3[a]	2.92	38.0	–0.60	3.5	9.62	187
75	P_{15}	0.72[a]	15.0[ab]	3.07	45.5[b]	–0.11[b]	7.3[b]	10.20	193
100	P_{16}	1.32[b]	11.3[a]	4.93[b]	38.0	0.17[b]	3.5	10.71[b]	198[b]
100	P_{17}	0.95	9.3[a]	4.25	34.0	–0.23	1.5	9.98	193
100	P_{18}	1.36[b]	12.3[a]	5.37[b]	40.0	0.04[b]	4.5	10.81[b]	198[b]
100	P_{19}	0.88[a]	13.8[a]	4.17	43.0	–0.33	6.0	10.52[b]	196[b]
100	P_{20}	0.79[a]	14.0[a]	4.06	43.5[b]	–0.44	6.3[b]	10.36	198[b]
100	P_{21}	0.91[a]	16.5[ab]	4.20	48.5[b]	–0.27	8.8[b]	10.59[b]	198[b]
		Standard errors							
		0.108 0.143[a]	0.94 1.24[a]	0.306	2.66	0.153	1.33	0.188	1.6

[a] These have the standard error that is so marked

[b] Within two standard errors of the best population

concern when the goal is identification of the best populations in the experiment. The problems that cause this are of concern, however. The primary problem appears to be insensitivity of the statistics. Because of this, exotic populations that have little donor value will be difficult to distinguish from one another if the populations are all similar. Either the bias is masking small differences or the standard error is too large. The statistics are similar in their sensitivity based on standard errors except for NI which has a larger standard error and is less able to detect differences among populations. The masking due to the bias will change with the choice of estimator and the genetic populations being evaluated. LPLU is minimally biased based on expectations and should be less susceptible to these changes; however, NI seems to be especially susceptible to change of genetic background. The correlation between estimator rankings of the populations are compared in Tables 7 and 8. Note the difference in correlations of NI with the other estimators for plant height in Experiment 2 compared to Experiment 1. In addition, NI is highly variable in how it ranks populations with respect to the other statistics in different environments. For example, its rank correlation with the other statistics changes depending on the number of locations being analyzed. The other estimators are more consistent relative to each other. PTC, LPLU, UBND, and TCSC all rank the populations similarly.

Two approaches were used for comparing expected statistic ratios with pedigrees. The four different series of backcross populations (series 2–5) at three levels of percent novel germplasm (25, 50, 75%) were analyzed as a 4×3 factorial. The interaction term was used as the error for testing significance of the individual linear contrasts within a BC series. PTC, LPLU, and UBND increased linearly with percent novel germplasm within a backcross series. TCSC showed this relationship for only one BC series and NI did not show this relationship at all for yield. Results based on plant-height analysis were similar but less consistent; some BC series showed no linear effect of percent novel germplasm on the superiority measure.

The estimates were also evaluated by comparing percent novel germplasm in the donor population, 0, 25, 50, 75, 100%, with the mean estimates of each class (Table 9). A strong percent germplasm linear effect on the estimates results in high correlation coefficient values. LPLU, PTC, and TCSC values were significant, and UBND approached significance (*P*-value approximately 0.09). The higher *P*-value was due primarily to the 0% novel germplasm group (P_1-P_3). These populations are expected to have a large bias for UBND estimates. UBND for population 1 had an especially large bias due to p_j and p_k, both at 0.5. To have less bias, UBND needs one of these to be closer to zero. With the exception of NI effects based on plant height in Experiment 2, only NI had non-significant linear effects.

Table 7 Superiority measure rank comparisons determined from yield data. Experiment 1 (17 *df*[a]), [Exp. 1, 2 locations (19 *df*)][b], and Exp. 2 (19 *df*)

Exp	NI		PTC		LPLU		TCSC
	1	2	1	2	1	2	1
UBND	0.26 (0.47*)[b]	0.29	0.85** (0.86**)	0.85**	0.84** (0.89**)	0.90**	– (0.81**)
NI			0.43 (0.65**)	0.47*	0.65** (0.74**)	0.54*	– (0.47*)
PTC					0.91** (0.95**)	0.92**	– (0.93**)
LPLU							– (0.87**)

*,** Significant at P=0.05 and P=0.01 levels

[a] Degrees of freedom

[b] Determined from data at only Rosemount and Lamberton, Minnesota

Table 8 Superiority measure rank comparisons determined from plant height data. Experiment 1 (17 df_a), [Exp. 1, 2 locations (19 *df*)][b], and Exp. 2 (19 *df*)

Exp	NI		PTC		LPLU		TCSC
	1	2	1	2	1	2	1
UBND	–0.34 (–0.05)[b]	0.82**	0.80** (0.84**)	0.80**	0.87** (0.92**)	0.92**	– (0.63**)
NI			–0.10 (0.15)	0.66**	0.02 (0.20)	0.96**	– (0.19)
PTC					0.81** (0.90**)	0.79**	– (0.89**)
LPLU							– (0.71**)

*,**Significant at P=0.05 and P=0.01

[a] Degrees of freedom

[b] Determined from data at only Rosemount and Lamberton, Minnesota

Table 9 Correlations determined by comparing percent novel germplasm with estimates of the superiority measure (SM), three degrees of freedom

SM	Yield			Plant height		
	Exp 1		Exp 2	Exp 1		Exp 2
	All locs[a]	Two locs[b]		All locs[a]	Two locs[b]	
LPLU	0.99**	0.99**	1.00**	0.90*	0.94**	0.99**
UBND	0.89*	0.90*	0.89*	0.81*	0.83*	0.93*
NI	0.73	0.84*	0.70	0.28	0.08	0.95**
PTC	1.00**	1.00**	1.00**	0.88*	0.95**	1.00**
TCSC		0.99**			0.97**	

*,**Significant at P=0.05 and P=0.01

[a] Includes Lamberton, Waseca, Rosemount, and Olivia (14 *df*)

[b] Includes Lamberton, and Rosemount (16 *df*)

Relationship

Dudley (1987) also proposed a statistic that indicates to which parent of the reference hybrid the population is more related. The *j* and *k*-classes are related to the reference hybrid parents I_1 and I_2. If a population has more *j*-class favorable alleles than *k*, we should expect the population to be more related to I_1, and vice versa. The number of unfavorable alleles at each class should also be considered. According to Dudley (1987):

$$\begin{aligned} \text{relationship} &= (jp_j u + kq_k u) - (jq_j u + kp_k u) \\ &= (I_2 \times P_y) - (I_1 \times P_y) + 0.5(I_1 - I_2) \qquad (1) \end{aligned}$$

This equation indicates a relatedness to I_1 when positive and to I_2 when negative. Intermediate values, closer to zero, represent populations related to neither inbred more strongly than the other. The signs and relative values of Dudley's (1987) relationship estimates conform to expectations based on pedigrees, for both yield and plant height in the two experiments (Figs. 1 and 2).

Bernardo's method (1990 b) to determine to which parent the population is most related differs from Dudley's (1987) relationship measure by a constant. Dudley (1987) includes half the difference of the measured parental traits in his algorithm (equation 1). Bernardo (1990 b) simply states that if $I_1 \times P_y > I_2 \times P_y$ the population is more related to I_2. In our experiments the two methods were in agreement for all but a few populations. It appears that the two methods can differ for populations that are not related to either inbred more strongly than the other. Dudley's (1987) method agreed more with pedigree information, especially when the trait value differed greatly between the two parents. For example, plant height differs greatly between FR902 (161 cm) and LH82 (133 cm), the parents in Experiment 2. Consequently Bernardo's (1990b) method agreed less with Dudley's (1987) method in Experiment 2. Dudley's (1987) relatedness measure showed better agreement with pedigree information. P_2 is an example of this when Experiment 2 plant height data was used. P_2 contains mostly FR902 germplasm yet Bernardo's (1990 b) method indicates that this population is more related to LH82 than FR902.

Populations 4–21 were ranked, based on the absolute value of their relatedness measure. Relationship to either parent gives large values. We expected greater superiority-measure estimates for populations less related to the parents of the hybrid to be improved. LPLU, UBND, PTC, and TCSC conformed with this expectation. NI estimates had no correlation with relatedness based on yield, a negative correlation based on Experiment 1 plant height, and a positive correlation based on Experiment 2 plant height. Also, TCSC based on plant height was not correlated with relatedness in Experiment 1.

Implications

Estimates based on plant-height data were highly variable and differed more from expectations than did estimates based on yield. Probably fewer loci control plant height

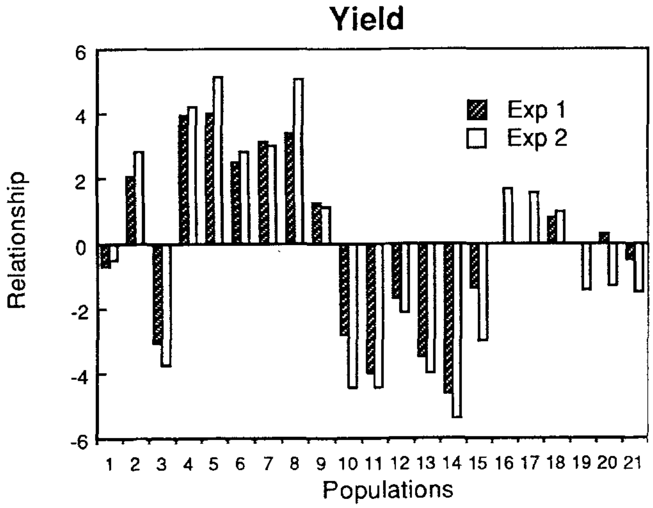

Fig. 1 Dudley's relationship estimates based on yield data. Positive values indicate the population is more related to I_1 and negative values indicate more relationship to I_2

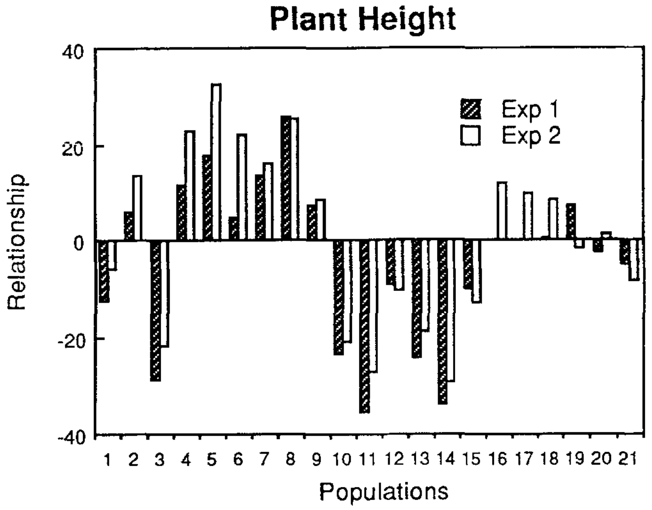

Fig. 2 Dudley's relationship estimates based on plant height data. Positive values indicate the population is more related to I_1 and negative values indicate more relationship to I_2

than yield. Because of this, several alleles may have a large genetic effect (u) on plant height. Because yield may be controlled by more loci the assumption of equal genetic effects among loci is most likely closer to being true than is the case for plant height.

Of the estimators we evaluated, NI is the only one that can be readily discarded as an estimator of $\ell p_\ell u$. It gave different results from the other estimators for yield in both experiments and for plant height in Experiment 1. Secondly, it seemed to be more influenced by environments and genetic background. The net improvement statistic (NI) was proposed to identify populations that could provide an immediate contribution to a reference hybrid. This is based on the idea that favorable alleles gained at class ℓ must contribute more than alleles lost at the *j*- or *k*-class loci in order for a net gain to occur. Examination of the expectations (Table 2) can provide a probable reason for such different results for NI estimates.

NI is the only estimator we evaluated that had a negative bias due to *j*- and *k*-class loci effects. All the other estimators are biased positively by these effects. In addition, bias due to partially-dominant alleles in the *j* and *k*-class is a positive effect for NI and a negative one for the other estimators. NI is truly estimating the net good of a population and not necessarily the unique alleles available at class-ℓ loci. If not for its inability to detect differences between populations, NI could be a useful statistic for evaluating populations for immediate improvement of a hybrid considering all loci classes, not just class-ℓ loci.

For estimating the superiority measure, the statistics LPLU, UBND, PTC, and TCSC all appear to give similar results, with similar ability to detect differences among populations. The estimator to use depends upon the resources available and the goal of the donor evaluation, e.g., identifying populations to use in parent building, or for direct improvement of the hybrid. Less expensive mass testing of populations could be best done with TCSC since it requires approximately half the resources of the other estimators.

The effectiveness of LPLU was as great as, or greater than, that of the other estimators. If the most accurate measure of a population is needed, it should be used. It could effectively identify the best and worst donor populations, and it is minimally effected by the other loci classes. Also, since the relative number of *j*- and *k*-class allele estimates are provided from the same data, Dudley's theory (1987) provides more information than the other methods. This information is important in deciding relationship to the inbreds, and potential gain from using a population.

Acknowledgements We thank Dekalb Plant Genetics and Dr. Marv Boerboom at Olivia, Minnesota for their cooperation in this study.

References

Bernardo R (1990 a) An alternative statistic for identifying lines useful for improving parents of an elite single cross. Theor Appl Genet 80:105–109

Bernardo R (1990 b) Identifying populations useful for improving parents of a single cross based on net transfer of alleles. Theor Appl Genet 80:349–352

Conover WJ (1971) Practical nonparametric statistics. John Wiley and Sons, New York

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

Dudley JW (1987) Modification of methods for identifying populations to be used for improving parents of elite single crosses. Crop Sci 27:940–943

Dudley JW (1988) Evaluation of maize populations as sources of favorable alleles. Crop Sci 28:486–491

Gerloff JE, Smith OS (1988) Choice of method for identifying germplasm with superior alleles. I. Theoretical results. Theor Appl Genet 76:209–216

Hallauer AR, Miranda JB Fo. (1988) Quantitative genetics in maize breeding. 2nd edn. Iowa State University Press, Ames, Iowa, pp 361–362

Hogan RM, Dudley JW (1991) Evaluation of a method for identifying sources of favorable alleles to improve an elite single cross. Crop Sci 31:700–704

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

Misevic D (1989) Evaluation of three test statistics used to identify maize inbred lines with new favorable alleles not present in an elite single cross. Theor Appl Genet 77:402–408

Pfarr DG, Lamkey KR (1992 a) Comparison of methods for identifying populations for genetic improvement of maize hybrids. Crop Sci 32:670–677

Pfarr DG, Lamkey KR (1992 b) Evaluation of theory for identifying populations for genetic improvement of maize hybrids. Crop Sci 32:663–669

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

Zanoni U, Dudley JW (1989) Comparison of different methods of identifying inbreds useful for improving elite maize hybrids. Crop Sci 29:577–582