Theor Appl Genet (1990) 80: 349–352

Identifying populations useful for improving parents of a single cross based on net transfer of alleles *

R. Bernardo

Lifaco Genetics, RR1, Box 232A, Champaign, IL 61821, USA

Received February 2, 1990; Accepted May 3, 1990
Communicated by A. R. Hallauer

Summary. Theory and methods for identifying populations (P_y) with the highest frequency of favorable dominant alleles not present in an elite single cross ($I_1 \times I_2$) have been developed recently. During selection, new favorable alleles can be transferred from P_y to either I_1 or I_2 only at the risk of losing favorable alleles already present in the single cross. A "net improvement" (NI) statistic, which estimates the relative number of favorable alleles that can be gained from P_y minus the relative number of favorable alleles that can be lost from I_1 or I_2, is presented. NI is calculated as maximum $[(I_1 \times P_y - I_1 \times I_2)/2, (I_2 \times P_y - I_1 \times I_2)/2]$. Because $I_1 \times I_2$ is constant in an experiment, the method reduces to choosing P_y populations with the best mean performance in combination with either I_1 or I_2. For a set of maize (*Zea mays* L.) grain yield data, NI was highly correlated to three other statistics proposed for choosing populations, namely: (1) minimally biased estimate ($l\,\bar{p}_l\,\mu'$) of the relative number of favorable dominant alleles present in P_y but not in I_1 and I_2; (2) minimum upper bound on $l\,\bar{p}_l\,\mu$; and (3) predicted performance of the three-way cross $[P_y\,(I_1 \times I_2)]$. While $l\,\bar{p}_l\,\mu'$ estimates potential improvement likely to be achieved only through long-term recurrent selection, NI is probably a better predictor of short-term improvement in single-cross performance.

Key words: Favorable alleles – Populations – *Zea mays* L.

Introduction

The choice of parental germ plasm is crucial to the success of maize (*Zea mays* L.) pedigree breeding programs. Lines or populations (exotic, open-pollinated, synthetic, or improved populations) are potential sources of favorable alleles not present in an elite single cross ($I_1 \times I_2$). Dudley (1987a, b) developed theory and methods for ranking populations (P_y) or lines (I_w) based on the relative number of favorable dominant alleles found in the population or line, but not in the single cross. But transfer of alleles from P_y or I_w to either I_1 or I_2 occurs only at the risk of losing favorable alleles already present in the single cross. Hence, Bernardo (1990) proposed choosing I_w lines based on a "net improvement" (NI) statistic. NI estimates the number of loci where favorable alleles can be gained from I_w minus the number of loci where favorable alleles can be lost from I_1 or I_2. Based on computer simulation results and limited experimental data, NI may be more useful in applied pedigree breeding programs than other statistics suggested for identifying I_w lines, such as Dudley's (1987b) $\mu G'$, Gerloff and Smith's (1988) minimum upper bound, or the three-way cross.

The objective of this paper is to present a "net improvement" statistic for identifying populations (P_y) useful for improving the parents of an elite single cross ($I_1 \times I_2$).

Theory

Assume + and − are the favorable and less favorable alleles, respectively, affecting a quantitative trait. Four classes of loci exist for any single cross ($I_1 \times I_2$) (Table 1). Let i, j, k, and l be the numbers of loci in their respective classes. The average frequencies of + alleles in a population (P_y) are $\bar{p}_i, \bar{p}_j, \bar{p}_k$, and $\bar{p}_l$ at classes i, j, k, and l, respectively. The genotypic values of the three genotypes (+ +, + −, and − −) at a single locus are $z+2\mu$, $z+\mu+a\mu$, and z, respectively, where μ is half the difference between homozygote values and a is the degree of dominance. Assume complete dominance ($a=1$), negligible epistasis, and that z, μ, and a are constant for N total loci. The genotypic values

---

* A contribution from Lifaco Genetics, a subsidiary of Groupe Limagrain

Table 1. Frequencies of favorable alleles in the parents (I_1 and I_2) of a single cross and a population (P_y) for each of four possible classes of loci (from Dudley 1987a)

Locus class	Frequencies in:		
	I_1	I_2	P_y
i	1.0	1.0	$\bar{p}_i$
j	1.0	0.0	$\bar{p}_j$
k	0.0	1.0	$\bar{p}_k$
l	0.0	0.0	$\bar{p}_l$

Table 2. Expectations of genotypic means of I_1, I_2, P_y, and the crosses among them with a general genetic model[a] (from Dudley 1987a)

$$
\begin{aligned}
I_1 &= N(z+\mu)+\mu(i+j-k-l)\\
I_2 &= N(z+\mu)+\mu(i-j+k-l)\\
P_y &= N(z+\mu)+\mu\{i[1+2\bar{q}_i(a-1)-2\bar{q}_i^2\, a]\\
&\qquad +j[1+2\bar{q}_j(a-1)-2\bar{q}_j^2\, a]\\
&\qquad +k[1+2\bar{q}_k(a-1)-2\bar{q}_k^2\, a]\\
&\qquad +l[1+2\bar{q}_l(a-1)-2\bar{q}_l^2\, a]\}\\
I_1\times I_2 &= N(z+\mu)+\mu(i+j\,a+k\,a-l)\\
I_1\times P_y &= N(z+\mu)+\mu\{i(\bar{p}_i+\bar{q}_i\, a)+j(\bar{p}_j+\bar{q}_j\, a)+k[\bar{p}_k(1+a)-1]\\
&\qquad +l[\bar{p}_l(1+a)-1]\}\\
I_2\times P_y &= N(z+\mu)+\mu\{i(\bar{p}_i+\bar{q}_i\, a)+j(\bar{p}_j(1+a)-1]+k(\bar{p}_k+\bar{q}_k\, a)\\
&\qquad +l[\bar{p}_l(1+a)-1]\}
\end{aligned}
$$

[a] N = total number of loci; z = value of the $--$ homozygote; μ = half the difference between homozygote values; i, j, k, and l = number of loci in their respective classes; a = degree of dominance; $\bar{p}$ = frequency of the + allele at a given locus class; $\bar{q}=1-\bar{p}$

of I_1, I_2, P_y, and the crosses among them can be expressed in terms of the above genetic parameters (Table 2).

I_1 and I_2 do not have the + alleles at class l loci. The product of the relative number of class l loci ($l\,\mu$) and the average frequency of the + alleles at class l loci ($\bar{p}_l$) estimates the number of + alleles that may be transferred from P_y to I_1 or I_2 to improve the single cross. Hence, Dudley (1987a) proposed the minimally biased statistic $l\,\bar{p}_l\,\mu'$ for identifying P_y populations with the highest concentration of new favorable alleles not found in I_1 or I_2.

Net improvement in single-cross performance results if a greater number of favorable alleles is gained than lost in I_1 or I_2 during selection. If the probability of fixing the + allele is 1.0, the potential loss of + alleles already present in the single cross is irrelevant. But even with strong selection pressure and large locus effects, the probabilities of fixing the + allele are probably closer to 0.5 (probability of fixation in the absence of selection) than to 1.0 (Bernardo 1990). Thus, the potential loss of favorable alleles from $I_1\times I_2$ also has to be considered when choosing P_y.

If P_y is crossed to I_2, + alleles can be lost from I_2 at class k loci. The potential loss of + alleles from I_2 is proportional to $k(1-\bar{p}_k)\,\mu=k\,\bar{q}_k\,\mu$, where $\bar{q}_k$ is the average frequency of the − allele at class k loci. If P_y is crossed to I_1, + alleles can be lost from I_1 at class j loci. The loss of + alleles from I_1 is proportional to $j(1-\bar{p}_j)\,\mu=j\,\bar{q}_j\,\mu$. Net gain of favorable alleles is maximized if (1) the chance of gaining + alleles at class l loci is largest, and (2) the chance of losing + alleles at class k or j locus is smallest. Therefore, a "net improvement" (NI) statistic for identifying populations is NI = maximum $[(l\,\bar{p}_l-k\,\bar{q}_k)\,\mu,\ (l\,\bar{p}_l-j\,\bar{q}_j)\,\mu]$ $=l\,\bar{p}_l\,\mu-$ minimum $(k\,\bar{q}_k\,\mu,\ j\,\bar{q}_j\,\mu)$. This NI statistic for identifying populations is analogous to that for identifying lines (Bernardo 1990). $(l\,\bar{p}_l-k\,\bar{q}_k)\,\mu$ and $(l\,\bar{p}_l-j\,\bar{q}_j)\,\mu$ are estimated by $(I_1\times P_y-I_1\times I_2)/2$ and $(I_2\times P_y-I_1\times I_2)/2$, respectively. Therefore, NI = maximum $[(I_1\times P_y-I_1\times I_2)/2,\ (I_2\times P_y-I_1\times I_2)/2]$. Because $I_1\times I_2$ is constant in an experiment, the method reduces to choosing P_y populations with the best mean performance in combination with either I_1 or I_2. If $I_1\times P_y > I_2\times P_y$, then $j\,\bar{q}_j > k\,\bar{q}_k$. The latter result implies that more + alleles can be lost from the single cross if P_y is crossed to I_1 than if P_y is crossed to I_2. Therefore, if $I_1\times P_y > I_2\times P_y$, P_y is crosssed to I_2 and I_1 is used as the tester. Otherwise, P_y crossed to I_1.

$l\,\bar{p}_l\,\mu'$ is a slightly biased statistic (Dudley 1987a), while NI is unbiased. $l\,\bar{p}_l\,\mu'$ estimates potential improvement in the single cross likely to be achieved only through long-term recurrent selection. NI is probably a better predictor of short-term gain.

If V_H and V_I are the variances of hybrid ($I_1\times P_y$, $I_2\times P_y$, or $I_1\times I_2$) and inbred lines means, respectively, the variances of the estimators are (1) $V(\mathrm{NI})=(1/2)\ V_H$ and (2) $V(l\,\bar{p}_l\,\mu')=[(3/16)\,V_H+(1/16)\,V_I]$ or $[(5/16)\,V_H+(1/16)\,V_I]$, depending on the case used to estimate $l\,\bar{p}_l\,\mu'$ (Dudley 1987a). Unless V_I is substantially greater than V_H, $V(l\,\bar{p}_l\,\mu')$ is expected to be smaller than $V(\mathrm{NI})$.

Empirical studies indicate partial to complete dominance for maize grain yield (Hallauer and Miranda 1981). If dominance is partial ($a<1$), maximum performance is achieved if both I_1 and I_2 carry the + allele. If P_y is crossed to I_2, potential gain of + alleles is proportional to $(l\,\bar{p}_l+j\,\bar{p}_j)\,\mu$, while loss of + alleles is proportional to $(k\,\bar{q}_k+i\,\bar{q}_i)\,\mu$. If P_y is crossed to I_1, potential gain of + alleles is proportional to $(l\,\bar{p}_l+k\,\bar{p}_k)\,\mu$, while loss of + alleles is proportional to $(j\,\bar{q}_j+i\,\bar{q}_i)\,\mu$. With arbitrary dominance, the expectation of NI is $\{[(l\,\bar{p}_l-k\,\bar{q}_k)(1+a)+(j\,\bar{p}_j-i\,\bar{q}_i)(1-a)]\,\mu/2,\ [(l\,\bar{p}_l-j\,\bar{q}_j)(1+a)+(k\,\bar{p}_k-i\,\bar{q}_i)(1-a)]\,\mu/2\}$. The terms that determine gain of + alleles ($l\,\bar{p}_l$, $j\,\bar{p}_j$, and $k\,\bar{p}_k$) have a positive contribution, while terms that determine loss of + alleles ($i\,\bar{q}_j$, $k\,\bar{q}_k$, and $j\,\bar{q}_k$) have a negative contribution to the expectation of NI. Because the expectation of NI reflects changes in the terms that determine net gain of favorable alleles as dominance decreases, NI is still useful for identifying P_y populations, even with partial dominance.

Illustration and application

The method is illustrated using testcross grain yields of 19 populations used to improve FRB73 (I_1) × FRMo17 (I_2). The data and a description of the populations used are in Dudley (1988). Crosses of the 19 P_y populations with FRB73 and FRMo17 were evaluated along with FRB73 × FRMo17, FRB73, and FRMo17 at four locations in Illinois in 1985 and 1986. In addition to NI and $l\,\bar{p}_l\,\mu'$, two other statistics suggested for choosing populations were considered. The predicted three-way cross (PTC) is calculated as $[(I_1\times P_y)+(I_2\times P_y)]/2$ and has the expectation $P_y(I_1\times I_2)=N(z+\mu)+\mu[i+j\,\bar{p}_j+k\,\bar{p}_k+l(2\,\bar{p}_l-1)]$. A minimum upper bound (UBND) on $l\,\bar{p}_l\,\mu$ (Gerloff and Smith 1988) is calculated as minimum $[(I_1\times P_y-I_1),\ (I_2\times P_y-I_2)]$. UBND has the expectation

Table 3. Estimate of NI, $l\,\bar{p}_l\,\mu'$, predicted three-way cross (PTC) means, and upper bound (UBND) for 19 populations used to improve FRB73 x FRMo17

Population (P_y)[a]	$l\,\bar{p}_l\,\mu'$ [b]	NI	PTC	UBND	Cross to: [c]
BS11(FR)C7	1.02	0.28	10.10	3.54	FRMo17
BSCB1(R)C10	0.76	−0.11	9.98	3.27	FRB73
RBS10(R)C2	0.59	−0.28	9.23	2.90	FRMo17
JF10 x BS13(S1)C2	0.50	−0.37	9.44	2.72	FRB73
Caribbean Flint	0.49	−0.08	9.04	2.15	FRMo17
BSTL(S2)C4	0.46	−0.42	9.02	2.71	FRMo17
IF10	0.44	−0.46	9.26	2.65	FRMo17
RSSSC	0.40	−0.41	9.25	2.42	FRB73
JF10	0.38	−0.51	9.14	2.54	FRMo17
SA PICII(M)C1	0.35	−0.54	8.76	2.47	FRMo17
BS16(S2)C3	0.32	−0.57	9.09	2.42	either line
IF10 x BS13(S1)C2	0.27	−0.57	8.99	2.22	FRB73
BS18	0.27	−0.58	8.98	2.24	FRB73
Mexican Dent	0.26	−0.53	8.58	2.10	FRMo17
Antigua(M)C6	0.23	−0.66	8.70	2.24	FRMo17
BS12(HI)C6	0.20	−0.46	8.46	1.72	FRMo17
BS16 x RSSSC	0.12	−0.74	8.68	1.94	FRB73
NHG(M)I14(E6+7)(S1)C1	−0.02	−1.06	8.10	1.43	either line
Cateto	−0.06	−0.95	8.06	1.66	FRMo17

[a] See Dudley (1988) for a description of the populations. Populations are documented in the following references: BS11(FR)C7 and BSCB(R)C10, Hallauer et al. (1974); RBS10(R)C2, Lambert (1985); Caribbean Flint, Mexican Dent, and Cateto, Gerrish (1983); BSTL(S2)C4, Russell et al. (1971); IF10 and JF10, Hanson and Moll (1986); RSSSC, Kauffmann and Dudley (1979); SA PICII(M)C1, Alexander and Spencer (1982); BS16(S2)C3, Hallauer and Smith (1979); BS18 and BS12(HI)C6, Hallauer and Russell (1986); and NHG(M)I14(E6+7)(S1)C1, Gardner (1977).

[b] SE's of NI, PTC, and UBND are 0.26, 0.26, and 0.46, respectively. Average (over cases used to estimate $l\,\bar{p}_l\,\mu'$ (Dudley 1987a)) SE of $l\,\bar{p}_l\,\mu'$ is 0.20

[c] P_y is crossed to FRMo17 if (FRB73 x P_y) > (FRMo17 x P_y); otherwise, P_y is crossed to FRB73

Table 4. Correlations[a] among estimates of NI, $l\,\bar{p}_l\,\mu'$, predicted three-way cross (PTC) mean, and upper bound (UBND) for 19 populations used to improve FRB73 x FRMo17

	$l\,\bar{p}_l\,\mu'$	PTC	UBND
NI	0.95	0.86	0.82
$l\,\bar{p}_l\,\mu'$		0.94	0.94
PTC			0.94

[a] All correlation coefficients are significantly greater than zero ($P = 0.05$)

$2[l\,\bar{p}_l\,\mu + \text{minimum}\ (k\,\bar{p}_k\,\mu, j\,\bar{p}_j\,\mu)]$. Thus, both PTC and UBND/2 are biased estimators of $l\,\bar{p}_l\,\mu$ and could seriously overestimate the value of P_y for improving a single cross.

BS11 (FR) C7 had the largest estimates of NI, $l\,\bar{p}_l\,\mu'$, PTC, and UBND (Table 3). Grain yields (t ha^{-1}) of BS11 (FR) C7 when crossed to FRB73 and FRMo17 were 10.78 and 9.42, respectively (Dudley 1988). The average yield of FRB73 x FRMo17 was 10.23. For BS11 (FR) C7, NI was calculated as maximum[(10.78 − 10.23)/2, (9.42 − 10.23)/2] = 0.28. Because $I_1 \times P_y\,(10.78) > I_2 \times P_y\,(9.42)$, BS11 (FR) C7 should be crossed to FRMo17 (I_2). Other populations with large NI values were Caribbean Flint (−0.08) and BSCB1 (R) C10 (−0.11). BS11 (FR) C7 and BSCB1 (R) C10, but not Caribbean Flint, have undergone several cycles of selection for combining ability for grain yield (Hallauer et al. 1974; Gerrish 1983).

As expected, the standard error of $l\,\bar{p}_l\,\mu'$ (0.20) was smaller than that of NI (0.26) (Table 3). Based on NI, $l\,\bar{p}_l\,\mu'$, PTC, and UBND, the population BS11 (FR) C7 was identified as the best population for improving the FRB73 x FRMo17, although the rankings of the 18 other populations depended on which of the four estimators was used. Correlations among the four estimators were high (Table 4). The correlation with NI was highest for $l\,\bar{p}_l\,\mu'$ and lowest for UBND.

To maintain the heterotic pattern of the single cross, Dudley (1987a) suggested crossing P_y to I_1 or to I_2 if the term $[(I_2 \times P_y) - (I_1 \times P_y) + (I_1 - I_2)/2]$ is positive or negative, respectively. An alternative method is to cross P_y with I_2 if $(I_1 \times P_y) > (I_2 \times P_y)$; otherwise P_y is crossed to I_1. These two methods for determining the parent to cross with P_y showed agreement except for the unrelated (to FRB73 or FRMo17) populations IF10 and JF10 (Table 3). IF10 and JF10 are cycle 10 populations from full-sib selection in Indian Chief and Jarvis, respectively (Hanson and Moll 1986). Whereas IF10 and JF10 should both be crossed to FRB73 based on Dudley's method, these two populations should be crossed to FRMo17 based on the method suggested herein.

Estimation of NI requires similar testing resources as $l\,\bar{p}_l\,\mu'$, PTC, and UBND. With $n\,P_y$ populations, estimation of all four statistics requires evaluating $n\,I_1 \times P_y$ and $n\,I_2 \times P_y$ crosses in a sufficient number of environments. In addition, $I_1 \times I_2$ is evaluated to estimate NI and $l\,\bar{p}_l\,\mu'$, while I_1 and I_2 are evaluated to estimate $l\,p_l\,\mu'$ and UBND. Estimation of the four statistics does not require evaluation of the population per se, which is difficult if the populations are exotic and/or unadapted. If three-way cross, i.e., $P_y(I_1 \times I_2)$ (Gerloff and Smith 1988), data are used instead of PTC, half the amount of testing resources is needed compared to NI.

Acknowledgements. The author thanks Drs. J.W. Dudley, A.R. Hallauer, S.J. Openshaw, and an anonymous reviewer for their comments and suggestions.

References

Alexander DE, Spencer J (1982) Registration of South African photoperiod-insensitive maize composites I, II, and III. Crop Sci 22:158

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

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

Dudley JW (1987b) Modification of methods for identifying inbred lines useful for improving parent of elite single crosses. Crop Sci 27:944–947

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

Gardner CO (1977) Population improvement in maize. In: Walden DB (ed) Maize breeding and genetics. Proc Int Maize Symp Genet Breed, Urbana/IL, Sept. 1978. Wiley and Sons, New York, pp 207–228

Gerloff JE, Smith OS (1988) Choice of method for identifying germ plasm with superior alleles. 1. Theoretical results. Theor Appl Genet 76:209–216

Gerrish EE (1983) Indications from a diallel study for interracial hybridization in the corn belt. Crop Sci 23:1082–1087

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

Hallauer AR, Russell WA (1986) Registration of BS18 maize germ plasm. Crop Sci 26:838

Hallauer AR, Smith OS (1979) Registration of BS13(S2) C1 and BS16 maize germ plasm. Crop Sci 19:755

Hallauer AR, Eberhart SA, Russell WA (1974) Registration of maize germ plasm. Crop Sci 14:340–341

Hanson WD, Moll RH (1986) An analysis of changes in dominance-associated gene effects under intrapopulation and interpopulation selection in maize. Crop Sci 26:268–273

Kauffmann KD, Dudley JW (1979) Selection indices for corn grain yield, percent protein, and kernel weight. Crop Sci 19:583–588

Lambert RJ (1985) Reciprocal recurrent selection of maize in a high-yield environment. Maydica 29:419–430

Russell WA, Penny LH, Hallauer AR, Eberhart SA, Scott GE, Guthrie WD, Dicke FF (1971) Registration of maize germ plasm synthetics. Crop Sci 11:140–141