

# An unbiased estimator for identifying lines useful for the improvement of elite single crosses, based on a combining ability model

T. HOHLS\*, G. P. Y. CLARKE†, P. E. SHANAHAN & H. O. GEVERS‡

*Department of Genetics, University of Natal, Private Bag X01, Scottsville, Pietermaritzburg, †Department of Statistics and Biometry, University of Natal, Pietermaritzburg and ‡Grain Crops Institute, A.R.C., Pietermaritzburg, South Africa*

Elite single crosses are frequently identified through combining ability analysis of the results obtained from a diallel cross of selected inbred lines. The parents of these elite single crosses can be improved further by identifying a source of favourable alleles that are not present in the single cross. An unbiased estimator of the potential of an inbred line to donate favourable alleles,  $\mu a(B+G)$ , that is easy to compute and free from the assumption of complete dominance, is proposed. The theory for the identification of donor inbred lines is based on a combining ability model. Five methods of identifying donor lines are compared theoretically with the proposed unbiased estimator. A worked example, based on a  $12 \times 12$  diallel cross across two environments, illustrates the relationships between the various statistics used to identify donors. The unbiased estimator,  $\mu a(B+G)$ , is shown to provide an appropriate means of identifying donor inbred lines.

**Keywords:** combining ability, diallel cross, donor inbred, favourable alleles, single crosses, unbiased estimator.

## Introduction

Elite single crosses ( $I_1 \times I_2$ ) do not necessarily contain all the favourable alleles controlling the quantitative trait of interest. To increase the performance of an elite single cross hybrid, a donor line ( $I_w$ ) that contains dominant favourable alleles not present in  $I_1$  or  $I_2$  must be identified. The donor is then crossed to one of the parents (e.g.  $I_1$ ), and selfed in the  $F_2$  (or backcrossed one or more generations to  $I_1$  and then selfed) to isolate a new line with improved hybrid performance when crossed to  $I_2$ . Dudley (1984, 1987, 1988) presented the theoretical framework for the identification of donor inbreds and the methodology for the incorporation of favourable alleles into breeding material. Although Dudley modified his analysis to allow for fewer restrictive assumptions, his analysis is still based on the assumptions of complete dominance, a constant mid-homozygote value for all loci, and no epistasis.

Gerloff & Smith (1988a,b) compared Dudley's estimator ( $\mu G'$ ) of the superiority measure of an inbred line (the superiority measure of  $I_w$  was defined as the relative number of loci where  $I_1$  and  $I_2$  were unfavourable and  $I_w$  was favourable) with two other measures that they proposed: the test cross to the single cross

(TC(SC)), which may be written as  $I_w \times (I_1 \times I_2)$ , and the upper bound statistic (UBND), which is defined as the minimum of  $((I_w \times I_1) - I_1, (I_w \times I_2) - I_2)$ . Gerloff & Smith (1988b) did not come to any definitive conclusions about which statistic provides the best indication of the superiority measure of a donor line.

Misevic (1989a,b) also compared these three statistics:  $\mu G'$ , UBND and TC(SC). Misevic (1989b) found that correlations among these estimators of the superiority measure of the donor lines were high, positive and significant.  $\mu G'$  and TC(SC) were highly correlated and, therefore, Misevic concluded that TC(SC) also provided reliable information. Zanoni & Dudley (1989) found that  $\mu G'$  estimates were highly correlated with the TC(SC) and UBND statistics, for all the traits that they considered.

Bernardo (1990) proposed an alternative statistic for determining the best donor line, termed the net improvement (NI) statistic. This statistic estimates the number of loci where favourable alleles can be gained, minus the number of loci where favourable alleles can be lost in the single cross. NI was compared with  $\mu G'$ , UBND and TC(SC). Correlations among the four estimators were high but the rank order of donor lines differed, depending on the statistic used (Bernardo, 1990).

The aims of this study were to: (i) relate the statistics, developed for the identification of donor inbreds, in

\*Correspondence.

terms of a combining ability model; (ii) indicate the bias introduced into the statistics through having to make the assumption of complete dominance; (iii) obtain an unbiased estimator of the potential of an inbred line to donate favourable alleles; and (iv) compare the estimators by means of a worked example based on a  $12 \times 12$  diallel cross experiment conducted among elite white modified opaque-2 maize inbred lines across two environments.

### Theory

The genotypic model used by Dudley was that developed by Comstock & Robinson (1948) where, for the genotypes  $++$ ,  $+ -$  and  $--$ , the genotypic values were designated as  $z + 2\mu$ ,  $z + \mu + a\mu$  and  $z$ , respectively;  $z$  is thus the base value of the  $--$  genotype,  $\mu$  is the midhomozygote value and  $a$  represents the degree of dominance.  $z$ ,  $\mu$ , and  $a$  are assumed to be constant for all loci. This model is used to enable direct comparisons with Dudley's (1988) method. Following Dudley (1984, 1987, 1988), for any three homozygous lines there are eight classes of loci (Table 1). If  $A, B, \dots, H$  are used to represent the numbers of loci in the eight classes then the genotypic means of  $I_1, I_2, I_w, I_1 \times I_2, I_1 \times I_w$ , and  $I_2 \times I_w$  can be determined according to the model of Comstock & Robinson (1948) (Table 2).

The theory developed by Dudley (1984, 1987, 1988) has been used to develop estimators based on the parameters of a combining ability model. The combining ability model used is that of Griffing (1956) method 4, model I.

Comparison of  $I_1$  and  $I_2$  is based on differences according to four classes of loci (Table 3). The difference between the general combining abilities of these two lines reflects the relative difference in frequency between the number of (C+D) and (E+F) loci, that is,  $g_2$  (general combining ability of line 2)  $- g_1 = \mu[(E+F) - (C+D)]$ . Similarly, comparison of the loci between  $I_1$  and  $I_w$  can be summarized according to differences in four classes of loci (Table 4). The difference between  $g_w$  and  $g_1 = \mu[(E+G) - (B+D)]$ . There are also four classes of loci that indicate differences in the general combining abilities of  $I_2$  and  $I_w$  (Table 5). The difference between  $g_w$  and  $g_2 = \mu[(C+G) - (B+F)]$ .

The specific combining ability ( $s$ ) of a cross is the deviation of the cross from the expected value based on the mean of the two parents. Thus, using the results shown in Table 2,

$$s_{12} = (I_1 \times I_2) - \frac{1}{2}(I_1 + I_2) = \mu(Ca + Da + Ea + Fa),$$

$$s_{1w} = (I_1 \times I_w) - \frac{1}{2}(I_1 + I_w) = \mu(Ba + Da + Ea + Ga), \text{ and}$$

$$s_{2w} = (I_2 \times I_w) - \frac{1}{2}(I_2 + I_w) = \mu(Ba + Ca + Fa + Ga).$$

**Table 1** Genetic status of classes of loci possible for three homozygous lines and the single crosses between them

Class of loci	Line			Cross		
	$I_1$	$I_2$	$I_w$	$I_1 \times I_2$	$I_1 \times I_w$	$I_2 \times I_w$
A	++	++	++	++	++	++
B	++	++	--	++	+ -	+ -
C	++	--	++	+ -	++	- +
D	++	--	--	+ -	+ -	--
E	--	++	++	- +	- +	++
F	--	++	--	- +	--	+ -
G	--	--	++	--	- +	- +
H	--	--	--	--	--	--

Using the notation of Dudley (1988).

$I_w$ : donor; +: favourable allele; -: unfavourable allele.

**Table 2** Genotypic values of  $I_1, I_2, I_w$  and crosses among them, under a general genetic model

$$I_1 = N(z + \mu) + \mu(A + B + C + D - E - F - G - H)$$

$$I_2 = N(z + \mu) + \mu(A + B - C - D + E + F - G - H)$$

$$I_w = N(z + \mu) + \mu(A - B + C - D + E - F + G - H)$$

$$I_1 \times I_2 = N(z + \mu) + \mu(A + B + Ca + Da + Ea + Fa - G - H)$$

$$I_1 \times I_w = N(z + \mu) + \mu(A + Ba + C + Da + Ea - F + Ga - H)$$

$$I_2 \times I_w = N(z + \mu) + \mu(A + Ba + Ca - D + E + Fa + Ga - H)$$

From Dudley (1988).

$N$ : total number of loci;  $z$ : value of the  $--$  genotype;  $\mu$ : mid-homozygote value;  $a$ : degree of dominance;  $A, B, \dots, H$ : number of loci in their respective classes (Table 1).

**Table 3** Four classes of loci discerning parental inbreds 1 and 2

Loci class	$I_1$	$I_2$
A+B	++	++
C+D	++	--
E+F	--	++
G+H	--	--

**Table 4** Four classes of loci discerning parental inbred 1 and donor inbred  $w$

Loci class	$I_1$	$I_w$
A+C	++	++
B+D	++	--
E+G	--	++
F+H	--	--

**Table 5** Four classes of loci discerning parental inbred 2 and donor inbred  $w$

Loci class	$I_2$	$I_w$
A+E	++	++
B+F	++	--
C+G	--	++
D+H	--	--

**Unbiased estimator,  $\mu a(B+G)$**

From the preceding discussion,  $s_{1w} + s_{2w} - s_{12} = \mu(Ba - Ca - Fa + Ga + Ba + Ca + Fa + Ga) = 2\mu a(B+G)$ , that is,  $\mu a(B+G) = (s_{1w} + s_{2w} - s_{12})/2$ . This statistic,  $\mu a(B+G)$ , provides a means of quantifying similarities between  $I_1$  and  $I_2$ , when they are dissimilar to  $I_w$ , as class B loci represent the loci where  $I_1$  and  $I_2$  have favourable alleles and  $I_w$  does not, whereas class G loci represent the loci where  $I_1$  and  $I_2$  do not have favourable alleles, whereas  $I_w$  does.

Similarly,  $\mu a(C+F) = (s_{2w} + s_{12} - s_{1w})/2$  and  $\mu a(D+E) = (s_{12} + s_{1w} - s_{2w})/2$ . The three statistics, namely  $\mu a(C+F)$ ,  $\mu a(D+E)$ ,  $\mu a(B+G)$ , provide information on the relationship between  $I_1$  and  $I_w$ ,  $I_2$  and  $I_w$ , and  $I_1$  and  $I_2$ , respectively.

It is proposed that the unbiased estimator of the donor potential of an inbred line,  $\mu a(B+G)$ , be used as the criterion for selection of the donor line. The elite hybrid is produced from two extremely divergent lines, generally producing the highest possible specific combining ability in the cross between the two parental lines when the number of class B loci is small (Table 1). Thus, in a diallel cross, as  $I_1$  and  $I_2$  usually come from the most divergent heterotic groups,  $\mu aG \geq \mu aB$  (Table 1). Simultaneous selection for a large number of B and G loci therefore results in indirect selection for heterosis in the two crosses:  $(I_1 \times I_w)$  and  $(I_1 \times I_w) \times I_2$  (if  $I_1$  is more closely related to  $I_w$  than  $I_2$ ) (Table 1). By crossing  $I_w$  to  $I_1$  the heterosis in the final cross will be maximized as the divergence between  $I_1$  and  $I_2$  will be maintained.

Because Dudley (1984, 1987, 1988) makes the assumption of complete dominance, the effect of this assumption was investigated by comparing Dudley's estimators for  $\mu(B+G)$ ,  $\mu(C+F)$  and  $\mu(D+E)$  with  $\mu a(B+G)$ ,  $\mu a(C+F)$  and  $\mu a(D+E)$ .

From Dudley (1988):

$$\mu(G+B) = [(I_2 \times I_w) + (I_1 \times I_w) - I_w - (I_1 \times I_2)]/2 = (s_{2w} + s_{1w} - s_{12} + g_w)/2, \text{ that is, assuming } a=1 \text{ biases the estimate by } +g_w/2, \text{ as } \mu a(B+G) = (s_{1w} + s_{2w} - s_{12})/2.$$

$$\mu(C+F) = [(I_2 \times I_w) + (I_1 \times I_2) - I_2 - (I_1 \times I_w)]/2 = (s_{2w} + s_{12} - s_{1w} + g_2)/2, \text{ that is, assuming } a=1 \text{ biases the estimate by } +g_2/2, \text{ as } \mu a(C+F) = (s_{2w} + s_{12} - s_{1w})/2.$$

$$\mu(D+E) = [(I_1 \times I_w) + (I_1 \times I_2) - I_1 - (I_2 \times I_w)]/2 = (s_{12} + s_{1w} - s_{2w} + g_1)/2, \text{ that is, assuming } a=1 \text{ biases the estimate by } +g_1/2, \text{ as } \mu a(D+E) = (s_{12} + s_{1w} - s_{2w})/2.$$

The proposed estimators of  $\mu a(C+F)$ ,  $\mu a(D+E)$  and  $\mu a(B+G)$  are free from the bias created by having to assume that  $a=1$ . However, as with all the other statistics proposed thus far, the theory on which  $\mu a(C+F)$ ,  $\mu a(D+E)$  and  $\mu a(B+G)$  are based relies on the assumptions of a constant midhomozygote value and no epistasis. Effects of failure of these assumptions are not known (Dudley, 1988).

If  $\mu a(B+G)$  is higher than  $\mu a(C+F)$  or  $\mu a(D+E)$ , then the incorrect cross has been selected as the elite cross as  $I_1$  and  $I_2$  should be the most divergent (Table 1). If  $\mu a(C+F) > \mu a(D+E)$  then  $I_w$  should be crossed to  $I_1$  to maximize the heterosis in the final cross as  $I_w$  would be more closely related to  $I_1$  than to  $I_2$ , whereas if  $\mu a(D+E) > \mu a(C+F)$  then  $I_w$  should be crossed to  $I_2$  as  $I_w$  would be more closely related to  $I_2$  than to  $I_1$  (Table 1).

The introduction of alleles not present in  $I_1$  and  $I_2$  by means of the crosses  $(I_1 \times I_w)$  or  $(I_2 \times I_w)$  may result in the loss of favourable alleles that  $I_1$  or  $I_2$ , respectively, contribute. An important decision is therefore whether to backcross the  $F_1$  to the inbred parent that is being improved or whether selfing should begin directly in the  $F_2$ . The decision of whether to backcross to the parental inbred should be based on the performance of the single cross. The criteria used by Dudley (1988) for deciding on the strategy in the  $F_2$  are biased by the assumption of complete dominance. One generation of backcrossing would probably be useful under most circumstances as this enables the retrieval of favourable alleles that could be lost when the chosen parental inbred is crossed to  $I_w$ .

**Theoretical comparison with other statistics**

*Net improvement statistic (NI)*

Bernardo (1990) proposed a net improvement (NI) statistic that accounts for the possibility of loss of favourable alleles from  $I_1 \times I_2$  when  $I_w$  is crossed to either  $I_1$  or  $I_2$ . Dudley (1988) considered this problem in deciding whether to self in the  $F_2$  or backcross to the parental line.

In the single cross  $I_1 \times I_w$ , + alleles are gained at class G loci, but + alleles can also be lost at class F loci (Table 1). If  $I_2$  is crossed to  $I_w$ , + alleles can be gained at class G loci, but + alleles can also be lost at class D loci (Table 1). Therefore, Bernardo proposed the use of the maximum of the two statistics:  $\mu(G-F)$  and  $\mu(G-D)$  for finding out the merit of the donor lines. The statistics are identical to those used by Dudley (1988)

for determining whether backcrossing would be required.

Bernardo (1990) assumed complete dominance, and therefore  $[(I_1 \times I_w) - (I_1 \times I_2)]/2$  and  $[(I_2 \times I_w) - (I_1 \times I_2)]/2$  were used as estimates of  $\mu(G-F)$  and  $\mu(G-D)$ , respectively.

However, without the assumption of complete dominance,

$$[(I_1 \times I_w) - (I_1 \times I_2)]/2 = \mu[Ba - B + C - Ca - F - Fa + Ga + G]/2$$

$$= \mu[(1-a)(C-B) + (1+a)(G-F)]/2, \text{ and}$$

$$[(I_2 \times I_w) - (I_1 \times I_2)]/2 = \mu[Ba - B - D - Da + E - Ea + Ga + G]/2$$

$$= \mu[(1-a)(E-B) + (1+a)(G-D)]/2.$$

Bernardo's statistic is therefore biased when  $a \neq 1$ . Bernardo's NI statistic can be expressed as the max.  $\{[(g_w - g_1 + s_{2w} - s_{12})/2, (g_w - g_2 + s_{1w} - s_{12})/2]\}$ .

*Three-way cross prediction (TWC)*

Jenkins's (1934) method B (also referred to as the TWC statistic) of three-way cross prediction is based on both GCAs and SCAs, and can be expressed as follows: the prediction of the three-way cross between the single cross  $(I_1 \times I_2)$  and  $I_3$  is:  $\hat{y}_{12.3} = \mu + g_3 + \frac{1}{2}(g_1 + g_2) + \frac{1}{2}(s_{13} + s_{23})$ . Introduction of favourable alleles not present in the elite single cross  $(I_1 \times I_2)$  would occur through one of the following two three-way crosses:  $(I_1 \times I_w) \times I_2$  or  $(I_2 \times I_w) \times I_1$ . Consider the situation where it is decided that  $I_1$  should be crossed to  $I_w$ .

For the cross  $(I_1 \times I_w) \times I_2$  the offspring with the highest value for the quantitative trait will be produced when the parents of the three-way cross have the maximum number of classes of favourable loci (Table 6). The genotypic value of the superior three-way cross will be:  $N(z + \mu) + \mu(A + B + Ca + Da + E + Fa + Ga - H)$ .

**Table 6** Progeny of  $(I_1 \times I_w) \times I_2$  with the highest selection potential (where parent 1 and parent 2 are homozygous)

Class of loci	Parent 1 ( $I_1 \times I_w$ )	Parent 2 $I_2$	Three-way cross $(I_1 \times I_w) \times I_2$
A	+	+	++
B	+	+	++
C	+	-	+-
D	+	-	+-
E	+	+	++
F	-	+	+-
G	+	-	+-
H	-	-	--

The superiority of this three-way cross over the elite single cross is calculated as:  $[N(z + \mu) + \mu(A + B + Ca + Da + E + Fa + Ga - H)] - [N(z + \mu) + \mu(A + B + Ca + Da + Ea + Fa - G - H)]$ . The superiority of this three-way cross over the elite single cross is  $= \mu((E - Ea) + (G + Ga))$ .

For the cross  $(I_1 \times I_w) \times I_2$ , the offspring with the lowest value for the quantitative trait will be produced when the parents of the three-way cross have the least number of classes of favourable loci (Table 7). The genotypic value of the inferior three-way cross will be:  $N(z + \mu) + \mu(A + Ba + Ca - D + Ea + Fa - G - H)$ . The superiority of this three-way cross over the elite single cross is  $= \mu((Ba - B) - (D + Da))$ .

For the three-way cross  $(I_1 \times I_w) \times I_2$ , following Table 6, the average of the extremes of the three-way cross over the elite single cross can be seen to be calculated as:  $\mu(E - Ea + G + Ga + Ba - B - D - Da)/2 = (s_{2w} - s_{12} + g_w - g_1)/2$ . This superiority measure of the three-way cross over the elite single cross was compared with the difference between Jenkins's (1934) method B prediction and the single cross, namely,  $g_1 + \frac{1}{2}(g_2 + g_w) + \frac{1}{2}(s_{12} + s_{1w}) - g_1 - g_2 - s_{12} = \frac{1}{2}(s_{2w} - s_{12} + g_w - g_1)$ . Jenkins's (1934) method B formula thus predicts three-way cross performance as the average between the best and worst offspring that would be expected from the three-way cross. This statistic is also identical to Bernardo's (1990) NI statistic for  $[(I_2 \times I_w) - (I_1 \times I_2)]/2$ .

Evidently Jenkins's (1934) method B (TWC), while providing a good approximation of the expected performance, will not reflect the actual magnitude of improvement to be obtained from the three-way cross after several cycles of selection. For this reason, use of the maximum of  $[\mu((E - Ea) + (G + Ga)), \mu((C - Ca) + (G + Ga))]$ , which refers to the best of the three-way crosses  $(I_1 \times I_w) \times I_2$  and  $(I_2 \times I_w) \times I_1$ , would be the best improvement statistic. However, these statistics cannot be solved in terms of the combining

**Table 7** Progeny of  $(I_1 \times I_w) \times I_2$  with the lowest selection potential

Class of loci	Parent 1 ( $I_1 \times I_w$ )	Parent 2 $I_2$	Three-way cross $(I_1 \times I_w) \times I_2$
A	+	+	++
B	-	+	+-
C	+	-	+-
D	-	-	--
E	-	+	+-
F	-	+	+-
G	-	-	--
H	-	-	--

ability information obtained from the single crosses of the diallel analysis, as the model is overparameterized and the incidence matrix is singular.

From the preceding discussion the optimum predictor of best three-way cross performance for the cross

$$(I_1 \times I_w) \times I_2 = \mu(A + B + Ca + Da + E + Fa + Ga - H). \quad (1)$$

The predictor, based on the average between the best and the worst progeny that are expected =  $\frac{1}{2}(g_1 + g_w) + g_2 + \frac{1}{2}(s_{12} + s_{2w})$

$$= \mu(A + \frac{1}{2}(Ba + B) + Ca - \frac{1}{2}(D - Da) + \frac{1}{2}(E + Ea) + Fa - \frac{1}{2}(G - Ga) - H) \quad (2)$$

The difference between eqns (1) and (2) =  $\frac{1}{2}\mu((B - Ba) + (D + Da) + (E - Ea) + (G + Ga))$ . This difference represents the extent to which Jenkins's (1934) method B (TWC) predictor and Bernardo's (1990) NI statistic are biased.

#### *Predictors of Gerloff & Smith (1988) (UBND and TC(SC))*

Gerloff & Smith (1988a,b) proposed two other estimators: UBND and TC(SC). The TC(SC) statistic is based on the prediction of the performance of the three-way cross:  $(I_1 \times I_2) \times I_w$ . The TC(SC) statistic is computed as:  $[(I_1 \times I_w) + (I_2 \times I_w)]/2 = \frac{1}{2}(g_1 + g_2) + g_3 + \frac{1}{2}(s_{13} + s_{23})$ . The TC(SC) statistic is not believed to be capable of indicating the potential of a donor line as three-way crosses are planned to maximize heterosis in the final cross. The donor line should always be crossed with one of the parental lines first, and therefore prediction of the performance of  $(I_1 \times I_2) \times I_w$  is unjustified. The second statistic that Gerloff & Smith (1988a,b) proposed would provide a better indication of the merit of a donor line if it were not based on the assumption of  $a = 1$ . The UBND statistic is computed as the minimum of  $[(I_w \times I_1) - I_1, (I_w \times I_2) - I_2]$ . The reasoning behind selecting the minimum value of  $(I_1 \times I_w) - I_1$  and  $(I_2 \times I_w) - I_2$  is that for  $I_1$  crossed to  $I_w$ , class E loci must be minimized, whereas for  $I_2$  crossed to  $I_w$ , the number of class C loci must be kept to a minimum (Table 1). Class E loci must be minimized for  $I_1$  as class E loci represent the loci where only  $I_2$  and  $I_w$  have favourable alleles (Table 1). Similarly, class C loci represent loci where only  $I_1$  and  $I_w$  have favourable alleles (Table 1).

$$(I_w \times I_1) - I_1 = (Ba - B) + (Da - D) + (Ea + E) + (Ga + G),$$

and

$$(I_w \times I_2) - I_2 = (Ba - B) + (Ca + C) + (Fa - F) + (Ga + G).$$

Gerloff & Smith (1988a,b) therefore assume that  $a = 1$ , so that the minimum of  $2(G + E)$  and  $2(G + C)$  is selected for. The UBND statistic only provides a true indication of the superiority measure of a line when  $a = 1$ . Gerloff & Smith's (1988a,b) UBND statistic = minimum of  $[g_w - g_1 + s_{1w}, g_w - g_2 + s_{2w}]$ .

#### **Worked example**

The six methods of identifying potential donor inbred lines that have been discussed, namely  $\mu G'$ , NI,  $\mu a(B + G)$ , TWC, TC(SC) and UBND, have been applied to the yield data obtained from a  $12 \times 12$  diallel cross of maize conducted at two sites: Ukulinga and Cedara, which will be referred to as experiments 91U and 91C, respectively. The 12 maize inbred lines come from three heterotically divergent backgrounds, termed the P, M and F heterotic groups (Gevers & Whyte, 1987). The lines are: FO215W(P), SO713W(P), RO460W(M), SO507W(M), RO465W(M), RO452W(M), SO181W(M), RO550W(F), KO54W(F), RO558W(F), RO594W(F) and RO504W(F).

The highest yielding hybrid at Ukulinga (experiment 91U) was SO713W(P)  $\times$  RO452W(M). Therefore,  $I_1 = \text{SO713W(P)}$  and  $I_2 = \text{RO452W(M)}$  for 91U. The other 10 inbreds from the  $12 \times 12$  diallel cross were used as potential donors ( $I_w$ ) of favourable alleles not present in the elite single cross. Following the diallel analysis of the yield data of 91C, the single cross SO507W(M)  $\times$  RO558W(F) was identified as the elite single cross to be improved upon. Therefore,  $I_1 = \text{SO507W(M)}$  and  $I_2 = \text{RO558W(F)}$  for 91C. The other 10 inbreds from the  $12 \times 12$  diallel cross were used as potential donors.

The estimators obtained from each of these methods have been presented for each potential donor inbred line (Table 8). The rankings of the 10 inbreds changed according to the statistic being used (Table 8). Because the NI and TWC statistics have been shown to be theoretically identical, it is not surprising that these statistics produced an identical ranking of the donor lines (Table 8). The rank correlations (Snedecor & Cochran, 1989) among the six methods confirm that rankings of the donors were more similar for the different estimators in experiment 91U than for 91C (Table 9). The  $\mu a(B + G)$  estimator is strongly correlated with the  $\mu G'$  and TC(SC) statistics (Table 9). As found by Misevic (1989b) and Zanoni & Dudley (1989), the  $\mu G'$  and TC(SC) statistics are highly correlated (Table 9).

A principal components analysis (PCA) was conducted to define further the relationship among the six estimators and to obtain a better indication of the

**Table 8** Ranking of the 10 potential donor inbred lines (using the yields of the genotypes from experiments 91U and 91C) according to the six estimators

Donor	$\mu G'$	NI	$\mu a(B+G)$	$\mu a(C+F)$	$\mu a(D+E)$	TWC	TC(SC)	UBND
(a) Experiment 91U								
FO215W(P)	0.12(2)	-0.42 <sup>2</sup> (1)	-0.52(2)	1.17	0.69	1.62 <sup>1</sup> (1)	1.09(1)	0.52 <sup>1</sup> (1)
RO460W(M)	-0.47(9)	-0.88 <sup>1</sup> (7)	-1.45(9)	0.18	1.67	1.17 <sup>2</sup> (7)	-0.59(9)	-1.39 <sup>2</sup> (7)
SO507W(M)	-0.79(10)	-1.48 <sup>1</sup> (10)	-1.80(10)	0.68	1.18	0.56 <sup>2</sup> (10)	-1.31(10)	-1.61 <sup>2</sup> (10)
RO465W(M)	-0.29(7)	-0.46 <sup>1</sup> (3)	-1.17(6)	-0.29	2.14	1.58 <sup>2</sup> (3)	-0.23(7)	-1.50 <sup>2</sup> (9)
SO181W(M)	-0.29(8)	-0.80 <sup>1</sup> (6)	-1.34(8)	0.40	1.46	1.24 <sup>2</sup> (6)	-0.23(6)	-0.81 <sup>2</sup> (6)
RO550W(F)	-0.09(4)	-0.92 <sup>2</sup> (8)	-1.26(7)	1.68	0.17	1.12 <sup>1</sup> (8)	-0.42(8)	-1.49 <sup>1</sup> (8)
KO54W(F)	-0.10(5)	-0.79 <sup>2</sup> (5)	-0.58(3)	1.16	0.70	1.25 <sup>1</sup> (5)	0.36(4)	-0.19 <sup>1</sup> (4)
RO558W(F)	-0.16(6)	-0.96 <sup>2</sup> (9)	-0.85(5)	1.20	0.66	1.08 <sup>1</sup> (9)	-0.02(5)	-0.60 <sup>1</sup> (5)
RO594W(F)	0.37(1)	-0.44 <sup>2</sup> (2)	-0.49(1)	1.51	0.35	1.60 <sup>1</sup> (2)	0.72(2)	-0.18 <sup>1</sup> (3)
RO504W(F)	0.03(3)	-0.49 <sup>1</sup> (4)	-0.63(4)	0.55	1.31	1.56 <sup>2</sup> (4)	0.56(3)	0.13 <sup>2</sup> (2)
(b) Experiment 91C								
FO215W(P)	0.88(1)	-0.39 <sup>2</sup> (2)	-0.92(3)	0.27	1.63	1.93 <sup>2</sup> (2)	0.81(1)	-0.24 <sup>2</sup> (1)
SO713W(P)	0.37(8)	-1.02 <sup>2</sup> (9)	-1.35(9)	0.49	1.41	1.29 <sup>2</sup> (9)	-0.24(8)	-1.07 <sup>2</sup> (3)
RO460W(M)	0.20(9)	-0.58 <sup>1</sup> (6)	-1.19(6)	2.63	-0.74	1.74 <sup>1</sup> (6)	-0.48(9)	-2.64 <sup>1</sup> (10)
RO465W(M)	0.42(5)	-0.32 <sup>1</sup> (1)	-0.96(4)	2.69	-0.80	2.00 <sup>1</sup> (1)	-0.02(4)	-2.24 <sup>1</sup> (9)
RO452W(M)	0.41(6)	-0.70 <sup>1</sup> (8)	-0.91(2)	1.99	-0.10	1.61 <sup>1</sup> (8)	-0.08(5)	-1.60 <sup>1</sup> (6)
SO181W(M)	0.16(10)	-1.07 <sup>1</sup> (10)	-1.30(8)	1.74	0.15	1.24 <sup>1</sup> (10)	-0.57(10)	-1.84 <sup>1</sup> (8)
RO550W(F)	0.44(4)	-0.69 <sup>2</sup> (7)	-1.15(5)	-0.07	1.96	1.63 <sup>2</sup> (7)	-0.12(7)	-1.51 <sup>2</sup> (5)
KO54W(F)	0.41(7)	-0.51 <sup>2</sup> (3)	-1.26(7)	-0.39	2.28	1.80 <sup>2</sup> (3)	-0.09(6)	-1.80 <sup>2</sup> (7)
RO594W(F)	0.67(2)	-0.52 <sup>2</sup> (4)	-0.84(1)	0.17	1.72	1.79 <sup>2</sup> (4)	0.44(2)	-0.70 <sup>2</sup> (2)
RO504W(F)	0.54(3)	-0.56 <sup>2</sup> (5)	-1.48(10)	-0.06	1.95	1.76 <sup>2</sup> (5)	0.15(3)	-1.23 <sup>2</sup> (4)

$\mu a(C+F)$  and  $\mu a(D+E)$  statistics express the relationship between the donor inbred and the parental lines.

$\mu$ : midhomozygote value;  $a$ : degree of dominance;  $\mu G'$ : estimator of Dudley (1988).

$B$ : number of loci where the parents have favourable alleles but the donor does not.

$\mu a(C+F)$ : measure of the relationship between parent 1 and the donor inbred;  $\mu a(D+E)$ : measure of the relationship between parent 2 and the donor inbred.

NI: net improvement statistic (Bernardo, 1990), where  $^1 = (g_w - g_2 + s_{1w} - s_{12})/2$  and  $^2 = (g_w - g_1 + s_{2w} - s_{12})/2$ .

TWC: three-way cross prediction, where  $^1 = (I_1 \times I_w) \times I_2$  and  $^2 = (I_2 \times I_w) \times I_1$ .

TC(SC) and UBND: estimators proposed by Gerloff & Smith (1988a,b), where for UBND,  $^1 = (g_w - g_1 + s_{1w})$  and  $^2 = (g_w - g_2 + s_{2w})$ , and the numbers in brackets indicate the rank of the donors according to the various estimators.

most favourable donor inbred for experiments 91U and 91C. The similarity matrix was calculated as the correlation matrix for the six statistics obtained on each of the 10 potential donors. The first two principal components ( $Z_1$  and  $Z_2$ ) accounted for most of the total variation for both 91U and 91C (Table 10). For the PCAs of both 91U and 91C,  $Z_1$  reflects the general donor potential of an inbred line, in that the inbred with the most negative  $Z_1$  is the most desirable, as all the parameters have negative vector loadings and the highest values of the parameters are most favourable (Table 10). For the PCA of 91U,  $Z_2$  represents a contrast between the two estimators NI and TWC on the one hand, and the remaining estimators on the other (Table 10(a)). This suggests that the NI and TWC

estimators provide a different perspective of the donor potential of the lines, compared with the other four estimators. For the PCA of the estimators of 91C,  $Z_2$  represents a contrast between NI, TWC and  $\mu a(B+G)$ , and the remainder. The third significant principal component for the PCA of 91C,  $Z_3$ , represents a comparison between the  $\mu a(B+G)$  statistic and the others. The  $\mu a(B+G)$  statistic has been set apart from the other statistics, suggesting that this statistic, in being free from the assumption of  $a = 1$ , provides a different perspective of the potential of a donor line.

For both experiments, 91U and 91C, the two inbreds, namely FO215W(P) and RO594W(F), have the best general donor potential (Table 8; Fig. 1). When either of these inbreds is used as the donor for the

**Table 9** Rank correlations between the six estimators of the donor potential of an inbred line (yields from experiments 91U and 91C)

Estimator	$\mu G'$	NI	$\mu a(B+G)$	TWC	TC(SC)
(a) Experiment 91U					
NI	0.82**				
$\mu a(B+G)$	0.94**	0.86*			
TWC	0.82**	1.00**	0.86**		
TC(SC)	0.92**	0.87**	0.97**	0.87**	
UBND	0.87**	0.80**	0.90**	0.80**	0.97**
(b) Experiment 91C					
NI	0.78**				
$\mu a(B+G)$	0.70*	0.64*			
TWC	0.78**	1.00**	0.64*		
TC(SC)	0.96**	0.84**	0.71**	0.84**	
UBND	0.83**	0.48	0.53	0.48	0.78**

Rank correlations were calculated according to the method of Spearman (Snedecor & Cochran, 1989).

\*, \*\*: significantly different from zero at the 0.05 and 0.01 probability levels, respectively.

**Table 10** Latent vectors of the most important principal components for the PCA of the six estimators used to identify donor inbreds in experiments 91U and 91C (yield)

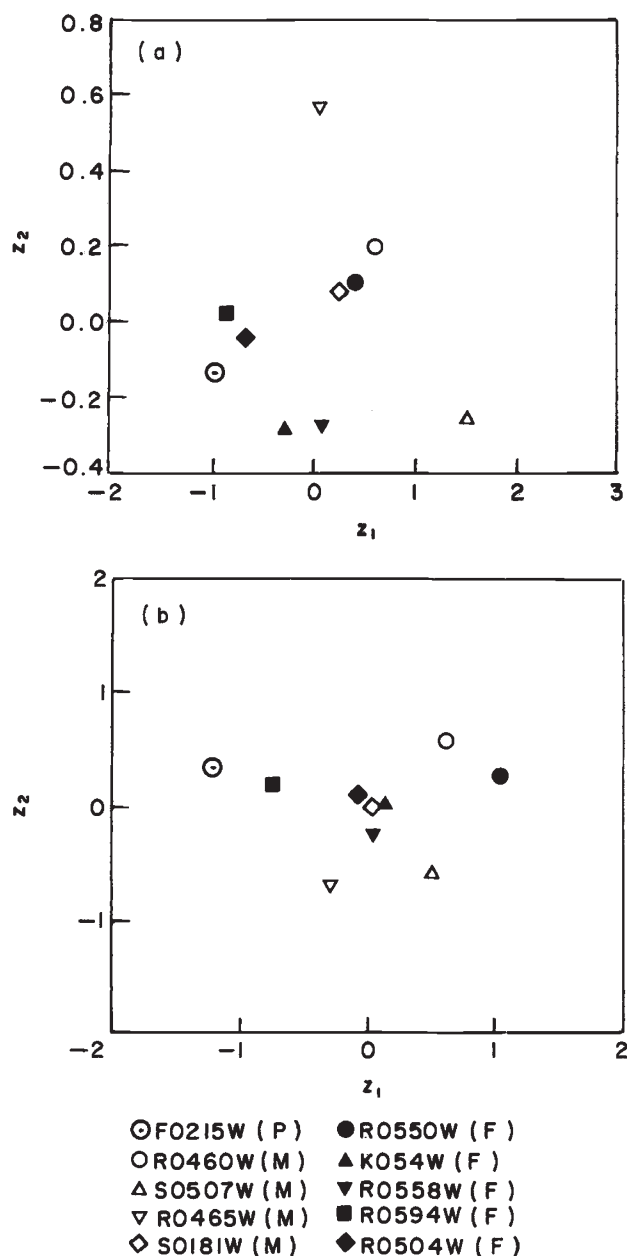
Inbred	(a) Experiment 91U		(b) Experiment 91C		
	$Z_1$	$Z_2$	$Z_1$	$Z_2$	$Z_3$
$\mu G'$	-0.41	-0.04	-0.48	0.26	0.13
NI	-0.40	0.53	-0.40	-0.48	0.23
$\mu a(B+G)$	-0.42	-0.30	-0.33	-0.15	-0.93
TWC	-0.40	0.54	-0.40	-0.48	0.23
TC(SC)	-0.44	-0.16	-0.49	0.21	0.10
UBND	-0.38	-0.56	-0.32	0.64	0.04
% variation	85.20	10.42	64.68	24.79	10.44

single cross: SO713W(P) × RO452W(M), as  $\mu a(C+F) > \mu a(D+E)$ , SO713W(P) should be crossed to the donor inbred to maximize the heterosis in the cross with RO452W(M). For 91C, as  $\mu a(D+E) > \mu a(C+F)$  for either of these inbred lines, RO558W(F) should be crossed to the donor inbred, to maximize the heterosis in the cross with SO507W(M). The  $\mu a(C+F)$  and  $\mu a(D+E)$  statistics show that the inbred lines from the P heterotic group are more closely related to the lines from the F group than to the inbreds from the M group (Table 8).

## Conclusions

Both FO215W(P) and RO594W(F) were identified as superior donors of favourable alleles for the different elite single crosses of experiments 91U and 91C. Therefore FO215W(P) and RO594W(F) have favourable gene combinations that should be utilized.

The  $\mu a(B+G)$  statistic produces similar rankings of the inbred lines to the other estimators. The  $\mu a(B+G)$  statistic is easier to compute than the  $\mu G'$  statistic of Dudley (1988), and is free from the assumption of



**Fig. 1** Plot of the two leading principal components for the principal components analysis of the potential donor estimators used on the yield data of (a) experiment 91U and (b) experiment 91C.

complete dominance. In contrast to the other estimators, which only require combining ability information (obtained from the single crosses), calculation of the  $\mu G'$  estimator of Dudley (1988) also requires that the parental inbreds and donor inbred be grown. The  $\mu G'$ , NI, TWC and UBND estimators are biased by the assumption of complete dominance, namely  $a = 1$ . The

PCA scores determined from a PCA on the estimators obtained using all six methods of identifying potential donor inbreds would not produce accurate rankings when used to identify potential inbred donors, as several of the estimators are biased. The PCA of the six estimators indicated that  $\mu a(B+G)$  provides a different perspective of the potential of a donor line. The results obtained from the  $\mu a(C+F)$  and  $\mu a(D+E)$  statistics agreed well with pedigree information. The unbiased estimator,  $\mu a(B+G)$ , provides an appropriate means of identifying inbred donors of favourable alleles.

### Acknowledgements

The authors gratefully acknowledge the assistance of the Agricultural Research Council (South Africa) in financing the field trials.

### References

- BERNARDO, R. 1990. An alternative statistic for identifying lines useful for improving parents of an elite single cross. *Theor. Appl. Genet.*, **80**, 105–109.
- COMSTOCK, R. E. AND ROBINSON, H. F. 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, J. W. 1984. A method of identifying lines for use in improving parents of a single cross. *Crop Sci.*, **24**, 355–357.
- DUDLEY, J. W. 1987. Modification of methods for identifying inbred lines useful for improving parents of elite single crosses. *Crop Sci.*, **27**, 944–947.
- DUDLEY, J. W. 1988. Theory for identification of lines or populations useful for improvement of elite single crosses. In: Weir, B. S., Eisen, E. J., Goodman, M. M. and Namkoong, G. (eds) *Proceedings of the 2nd International Conference on Quantitative Genetics*, pp. 451–461. Sinauer, Sunderland, MA.
- GERLOFF, J. E. AND SMITH, O. S. 1988a. Choice of method for identifying germplasm with superior alleles. I. Theoretical results. *Theor. Appl. Genet.*, **76**, 209–216.
- GERLOFF, J. E. AND SMITH, O. S. 1988b. Choice of method for identifying germplasm with superior alleles. II. Computer simulation results. *Theor. Appl. Genet.*, **76**, 217–227.
- GEVERS, H. O. AND WHYTHE, I. V. 1987. Patterns of heterosis in South African maize breeding material. In: *Proceedings of the Seventh S. African Maize Breeding Symposium*, Potchefstroom, 1986. Technical Communication of the Department of Agriculture and Water Supply, R.S.A., no. **212**, pp. 21–26.
- GRIFFING, B. 1956. Concept of general and specific combining ability in relation to diallel crossing systems. *Aust. J. Biol. Sci.*, **9**, 463–493.
- JENKINS, M. T. 1934. Methods of estimating the performance of double crosses in corn. *J. Am. Soc. Agron.*, **26**, 199–204.



- MISEVIC, D. 1989a. Identification of inbred lines as a source of new alleles for improvement of elite maize single crosses. *Crop Sci.*, **29**, 1120–1125.
- MISEVIC, D. 1989b. Evaluation of three test statistics used to identify maize inbred lines with new favourable alleles not present in elite single crosses. *Theor. Appl. Genet.*, **77**, 402–408.
- SNEDECOR, G. W. AND COCHRAN, W. G. 1989. *Statistical Methods*, 8th edn. Iowa State University Press, Ames, IA.
- ZANONI, U. AND DUDLEY, J. W. 1989. Testcross evaluation of F<sub>2</sub> populations from maize inbreds selected for unique favourable alleles. *Crop Sci.*, **29**, 589–595.