Theor Appl Genet (1996) 92:78–82

S. K. St. Martin · K. S. Lewers · R. G. Palmer
B. R. Hedges

A testcross procedure for selecting exotic strains to improve pure-line cultivars in predominantly self-fertilizing species

Received: 5 December 1994 / Accepted: 7 July 1995

Abstract Methods for identifying germplasm carrying alleles with the potential to improve a particular single-cross hybrid have been proposed and discussed in recent years. There is a need for similar methods to be used in breeding crops for which pure-line cultivars, rather than hybrids, are the goal. The objective of this research was to develop a method to identify germplasm lines with the potential to contribute favorable alleles not present in a specified pure line or set of pure lines. Given a set of adapted pure lines ($A_1, A_2, \ldots, A_m$) to be improved and a set of germplasm lines (P_1, $P_2, \ldots, P_f$), the procedure consists of producing all f × m possible hybrids and evaluating them along with the parents. The testcross statistic T_{ij} is defined by $T_{ij} = \gamma(F_{ij} - A_j) + (1 - \gamma)(F_{ij} - P_i)$, where A_j, P_i, and F_{ij} represent the performance of the j^{th} adapted line, the i^{th} germplasm line, and their hybrid, respectively. The statistic $\bar{T}_i = (1/m)\sum(T_{ij})$ is the mean value of T_{ij} over all adapted parents A_j. If $\gamma = (1/2)(1 + d')$, where d' = the mean degree of dominance, then T_{ij} measures the potential for alleles from P_i to improve A_j, and $\bar{T}_i$ measures the potential for alleles from P_i to improve the set A_1, $A_2, \ldots, A_m$. Use of data on soybean and peanut hybrids published by other researchers suggests that the value assumed for d' has little effect on the P_i chosen. The ability of the T_{ij} and $\bar{T}_i$ statistics to identify germplasm strains carrying rare favorable alleles should be assessed in empirical studies.

Key words Exotic germplasm · Selection of parents · Testcross

Introduction

Plant breeders have long been interested in using germplasm collections as resources for crop improvement. Such collections are considered reservoirs for favorable alleles not present in existing cultivars. Within the germplasm collection, the frequency of these favorable alleles may be low, which makes their identification and utilization difficult. Gerloff and Smith (1988 a) summarized the rationale underlying the development of methods to identify germplasm carrying these rare, favorable alleles.

Recently, several researchers have proposed and evaluated methods for identifying germplasm of this type in cross-fertilizing species. Dudley (1984 a,b, 1987 a,b) proposed the $lp_l\mu$, $lp_l\mu'$, and $\mu G'$ statistics for the identification of populations ($lp_l\mu$ and $lp_l\mu'$) or inbred lines ($\mu G'$) containing favorable dominant alleles not present in an elite single-cross hybrid cultivar. Gerloff and Smith (1988 a,b) proposed two additional statistics, UBND and TC(SC), for this purpose, and compared them with those proposed by Dudley. Pfarr and Lamkey (1992 a,b) evaluated these, along with two additional statistics, the testcross to an inbred line and performance per se. All these statistics are functions of the performance of: (1) the hybrid to be improved, (2) the parents of this hybrid, (3) the populations or inbreds to be evaluated as donors of favorable alleles, and (4) various hybrids involving these materials. Additional research

---

Joint contribution: OARDC (Journal Article No. 161-94), USDA-ARS, Iowa Agriculture and Home Economics Experiment Station (Journal Paper No. J-16109; Project 2985), and Agriculture and Agri-Food Canada. Salaries and research support for S. K. St. Martin provided by state and federal funds appropriated to the Ohio Agricultural Research and Development Center, Ohio State University

Communicated by A. L. Kahler

S. K. St. Martin (✉)
Department of Horticulture and Crop Science, Ohio Agric. Res. and Development Center (OARDC), Ohio State University, 2021 Coffey Rd., Columbus, OH 43210, USA

K. S. Lewers
Department of Agronomy, Iowa State Univ., Ames, IA 50011, USA

R. G. Palmer
USDA-ARS, Department of Agronomy, and Dep. of Zoology/Genetics, Iowa State University, Ames, IA 50011, USA

B. R. Hedges
Agriculture and Agri-Food Canada, Harrow Research Station., Harrow, Ontario NOR 1GO, Canada

has been carried out to compare these statistics (Dudley 1988; Zanoni and Dudley 1989; Hogan and Dudley 1991).

A similar need for germplasm evaluation exists in self-fertilized crop species in which pure-line, rather than hybrid, cultivars are used. Examples of attempts to use F_1 hybrid data to predict the potential of a cross to produce superior progeny in advanced generations include the work of Smith and Lambert (1968), Cregan and Busch (1977), and Isleib and Wynne (1983 b). Kenworthy (1980) suggested the use of F_2 bulks derived from crosses to a common tester to assess the value of exotic strains. In other experiments, F_1 or later generations were used to study the combining ability of geographically or morphologically defined groups of germplasm strains (Paschal and Wilcox 1975; Isleib and Wynne 1983 a; Sweeney and St. Martin 1989). Despite the existence of these and other similar empirical studies, there has been no theoretical investigation on the design and analysis of testcross experiments conducted with the aim of identifying, in a collection of germplasms, strains carrying favorable dominant alleles not present in adapted cultivars. The objective of the present paper is to describe such a design and to discuss statistics derived from its analysis.

Description and analysis of experiment

Design

Let A_1, A_2, A_3, ..., A_m be a set of m homozygous, homogeneous lines selected to represent the adapted germplasm base. The A_js may be commercial cultivars, elite breeding lines, or ancestor lines that have contributed to the germplasm base. Let $P_1, P_2, P_3, ..., P_f$ be a set of f homozygous, homogeneous germplasm lines from which it is desired to identify lines with favorable genes that are rare or absent among the A_js. Ordinarily, the P_is would be selections from a larger germplasm collection, selected on the basis of performance per se, geographic origin, or other characteristics likely to maximize the potential for useful genetic diversity.

A factorial set consisting of all fm possible hybrids (F_{ij}) between the P_i parents and the A_j parents is produced. These hybrids, along with the parents, are then evaluated in agronomic trials. Because the most important type of comparison is that between a hybrid and its parents, a split-plot design, with parental pairs as main-plot factors and the three genotypes, P_i, A_j and F_{ij}, in sub-plots, may be beneficial.

The feasibility of the proposed experiment depends on the ability to produce quantities of hybrid seed sufficient for agronomic testing. Suitable methods for the production of hybrid seed are available for many autogamous species, e.g., wheat (*Triticum aestivum* L. em Thell.) (Allan 1980), barley (*Hordeum vulgare* L.) (Starling 1980), rice (*Oryza sativa* L.) (Coffman and Herrera 1980), tomato (*Lycopersicon esculentum* Mill.) (Rick 1980), and soybean [(*Glycine max* L. (Merr.)] (Burton and Carter 1983; Skorupska and Palmer 1989).

Analysis

The analysis of the testcross experiment is aimed at choosing, from among the P_i parents, one or more parents that carry favorable alleles not present in a particular A_j or in the collection of all A_js. Selection of these P_i parents is based on a statistic of the type

$$T_{ij} = \gamma(F_{ij} - A_j) + (1-\gamma)(F_{ij} - P_i) \qquad (1)$$

where F_{ij}, A_j, and P_i refer, respectively, to the mean phenotypes (e.g., yields) of hybrid F_{ij} and parents A_j and P_i, and γ is a weight. Additionally, the mean value ($\bar{T}_i$) of T_{ij} may be determined:

$$\bar{T}_i = \left(\sum_j T_{ij}\right)\Big/m.$$

The following theoretical development addresses the choice of an appropriate value for γ. The theoretical development shows that, given a properly chosen γ, a P_i with a large value of T_{ij} carries favorable alleles not present in A_j. Further, a P_i whose $\bar{T}_i$ is large carries favorable dominant alleles rare or absent in the set A_1, $A_2, \ldots, A_m$ of adapted parents.

Theoretical development

First, consider a trait controlled by multiple loci, with two alleles (designated + and −) per locus and any degree of dominance, but no epistasis. Let the genotypic values for the k^{th} locus be $z + 2\mu_k$ for the + + genotype, $z + \mu_k(1 + d_k)$ for the + − genotype, and z for the − − genotype. Assume that larger genotypic values are desirable. Consider, for each pair (ij) of parents, the following classification of loci (after Dudley, 1984 a,b, 1987 a,b):

class e_{ij} P_i and A_j are both + +
class f_{ij} P_i is + + and A_j is − −
class g_{ij} P_i is − − and A_j is + +
class h_{ij} P_i and A_j are both − −.

Loci of class f_{ij} are of the greatest interest because they represent the potential contribution of P_i to the improvement of the adapted genotype. With this model, the genotypic values of the parents and F_1 are obtained by summing across loci:

$$P_i = Nz + 2\left(\sum_e \mu_k + \sum_f \mu_k\right); \qquad (2)$$

$$A_j = Nz + 2\left(\sum_e \mu_k + \sum_g \mu_k\right); \qquad (3)$$

and

$$F_{ij} = Nz + 2\sum_e \mu_k + \sum_f \mu_k(1 + d_k) + \sum_g \mu_k(1 + d_k). \qquad (4)$$

In these expressions, N = the number of loci, and subscripts on the Σs indicate classes of loci over which summation occurs. Now let $\gamma = (1/2)[1 +$

$(\Sigma_g \mu_k d_k)/(\Sigma_g \mu_k)]$. The expression $(\Sigma_g \mu_k d_k)/(\Sigma_g \mu_k)$ represents the weighted mean degree of dominance of class-g loci, with the weight equal to the magnitude of the effect of the locus. (The rationale for using this value of γ is that the resulting expression for T_{ij} depends only on parameters of class-f loci, provided that simplifying assumptions are made.) Substituting this value of γ and equations (2), (3), and (4) into equation (1) yields, upon algebraic simplification,

$$T_{ij} = \sum_f \mu_k \left[d_k + \left(\sum_g \mu_k d_k \right) \Big/ \left(\sum_g \mu_k \right) \right],$$

or

$$T_{ij} = \sum_f \mu_k (d_k + 2\gamma - 1).$$

If it is assumed, following Dudley (1984 a, b, 1987 a, b), that the degree of dominance, d_k, has the same value, d', for all loci, irrespective of class, then $T_{ij} = 2d' \Sigma_f \mu_k$, where $2\Sigma_f \mu_k$ represents the maximum potential genetic improvement from class-f loci, i.e., the difference between the performance of A_j and that of the best homozygous progeny obtainable from the cross $P_i \times A_j$. Under these assumptions, the statistic T_{ij} can be used to assess the potential of this cross in relation to the adapted parent A_j.

For the $\bar{T}_i$ statistic, let $\gamma = (1/2)[1 + (\Sigma_- n_k \mu_k d_k)/(\Sigma_- n_k \mu_k)]$, where Σ_- refers to summation over all loci k such that P_i carries the $- -$ genotype. The expression $(\Sigma_- n_k \mu_k d_k)/(\Sigma_- n_k \mu_k)$ is the weighted average degree of dominance for such loci. Now evaluation of $\bar{T}_i$ in terms of the model yields

$$\bar{T}_i = (1/m) \sum\nolimits_+ \mu_k (m - n_k) [d_k + (\sum\nolimits_- n_k \mu_k d_k)/(\sum\nolimits_- n_k \mu_k)], \tag{5}$$

where n_k is the number of A_j lines ($n_k = 0, 1, 2, \ldots,$ or m) that carry the $+ +$ genotype at the k^{th} locus and where Σ_+ refers to summation over loci k such that P_i carries the $+ +$ genotype. As before, if the degree of dominance is d' for all loci, then (5) simplifies to

$$\bar{T}_i = (2d'/m) \sum\nolimits_+ \mu_k (m - n_k). \tag{6}$$

Equation (6) suggests that a P_i with a high value of $\bar{T}_i$ carries important + alleles (indicated by the large value of μ_k) that are relatively rare among the A_j (indicated by the large value of $m - n_k$).

Extension to multiple alleles

If a simplifying assumption is made concerning dominance, interpretation of T_{ij} and $\bar{T}_i$ may be extended to the case of multiple alleles. Assume that the genotypic value of any genotype $(B_{lk} B_{l'k})$ for any locus k is $z + a_{lk} + a_{l'k} + d' |a_{lk} - a_{l'k}|$. In this model, the degree of dominance d' is the same for all possible pairs of alleles at all loci. For each locus k, let parent P_i have genotype $B_{ik} B_{ik}$ and let the genotype of each parent A_j be denoted $B_{jk} B_{jk}$. For any given i, j, and k, let class I(j) contain those loci such that $a_{ik} > a_{jk}$ and class II(j) those loci such that $a_{ik} \leqslant a_{jk}$. The notation I(j) and II(j) indicates that classification of loci with respect to a given P_i depends on the adapted parent A_j. Obviously, class-I(j) loci are of greatest interest in connection with the use of P_i as a source of new alleles.

With $\gamma = (1/2)(1 + d')$, the model value of T_{ij}, as defined in (1), becomes

$$T_{ij} = 2d' \sum_{I(j)} (a_{ik} - a_{jk}).$$

This expression indicates that T_{ij} measures the potential of genes from P_i to improve A_j. The model value of $\bar{T}_i$ is

$$\bar{T}_i = (2d'/m) \sum_j \sum_{I(j)} (a_{ik} - a_{jk}),$$

indicating that $\bar{T}_i$ measures the potential of genes from P_i to improve the A_js as a group.

Examples illustrating effect of d′ on choice of P_i

Little is known about levels of dominance (d') for quantitative traits in autogamous species, and this fact may inhibit use of the T_{ij} and $\bar{T}_i$ statistics. Data obtained from three published experiments were employed to calculate T_{ij} statistics with a wide range of $\gamma = (1/2)(1 + d')$. The primary purpose of this calculation was to determine the effect of assumed values of d' on parental selection. A secondary purpose was to compare the results of selection using the T_{ij} statistic with those of selection for parental or hybrid yield per se.

The first example derived from the results of Paschal and Wilcox (1975), who reported seed yields of hybrids between two adapted soybean cultivars ('Beeson' and 'Wells') and three exotic lines. The three exotic lines (P_i) crossed with Beeson were not identical to the three crossed with Wells. The best P_i parent was chosen by using three methods: (1) the T_{ij} statistic with values of d' ranging from 0.3 to 1.2 (i.e., γ ranging from 0.65 to 1.10), (2) the best parental yield per se, and (3) the parent producing the best hybrid with A_j. The range of d' from 0.3 to 1.2 was chosen to represent a broad range of possibilities from incomplete dominance to overdominance.

The second example was based on the data of Isleib and Wynne (1983 a), who tested hybrids of 27 exotic lines of peanut (*Arachis hypogaea* L.) with a single adapted parent. The best exotic parent to improve seed and pod yield of the adapted parent was identified by the same three methods as in the first example.

The data of Nelson and Bernard (1984) on hybrid soybean were used as the third example. The three cultivars ('Beeson', 'Wells', and 'Williams') used as female parents were considered as A_j lines, and the male parents were considered as P_i lines for the sake of the example, although all parents in the experiment were, in fact, adapted. There were 3–8 male (P_i) parents for each A_j line in each of the 2 years (1977 and 1978) of their study. Seed yield was used as the trait of interest. The best P_i parent for each A_j was chosen using the same three methods as in the first example.

Together, these three examples provided ten opportunities to choose the best P_i to improve an adapted parent (Tables 1, 2, and 3). In 8 of the 10 cases, the choice of P_i made by using the T_{ij} statistic did not depend on the value of d' used, within the range $0.3 \leqslant d' \leqslant 1.2$. The two exceptions were the choice of P_i to improve seed yield in the experiment of Isleib and Wynne (1983 a) (Table 2) and the choice of P_i to improve Williams soybeans from the 1977 data of Nelson and Bernard (1984) (Table 3). These data suggest that the results of using the T_{ij} statistic do not depend strongly on the value assumed for d'. Therefore, the T_{ij} statistic may be useful even in the absence of knowledge about the degree of dominance for quantitative traits.

Selection of P_i on the basis of parental yield per se agreed poorly with the other two methods, good agreement being obtained only in the 1977 data of Nelson and Bernard (1984) for the two parents Beeson and Wells (Table 3). Yield of the P_i parent per se, while easy to determine and undoubtedly useful for preliminary screening of large germplasm collections to identify potential parents, cannot indicate whether the favorable alleles in a relatively productive P_i are already present in the A_j.

If $d' = 1$, then selection based on hybrid yield gives the same result as selection on T_{ij}, because then $T_{ij} = F_{ij} - A_j$. This fact, and the relatively small effect of the assumed value of d' on results of selection using T_{ij}, implies good agreement between selection using T_{ij} and selection on hybrid yield. Agreement between these

Table 1 Results of using three methods (parental yield, hybrid yield, and T_{ij} statistic) to identify an exotic parental line carrying favorable alleles for yield not present in two adapted soybean cultivars (based on data reported by Paschal and Wilcox 1975)

Adapted cultivar	Number of parents evaluated	Best parent by method		
		Parental yield	Hybrid yield	T_{ij}[a]
Beeson	3	PI 68788	Korean	Korean for all d' ($0.3 \leqslant d' \leqslant 1.2$)
Wells	3	PI 60296-1	PI 60296-1	PI 60296-1 for all d' ($0.3 \leqslant d' \leqslant 1.2$)

[a] $T_{ij} = [(1+d')/2]$ (yield of hybrid − yield of adapted cultivar) + $[(1-d')/2]$ (yield of hybrid − yield of parent), where d' = assumed mean degree of dominance

Table 2 Results of using three methods (parental yield, hybrid yield, and T_{ij} statistic) to identify an exotic parental line carrying favorable alleles for seed and pod yield not present in an adapted peanut breeding line, NC Ac18000 (based on data reported by Isleib and Wynne 1983 a)

Trait	Number of parents evaluated	Best parent by method		
		Parental yield	Hybrid yield	T_{ij}[a]
Pod yield	27	PI 157543	PI 275745	PI 275745 for all d' ($0.3 \leqslant d' \leqslant 1.2$)
Seed yield	27	PI 157543	PI 275745	PI 275745 for $0.3 \leqslant d' \leqslant 1.07$; PI 262113 for $1.07 \leqslant d' \leqslant 1.2$

[a] $T_{ij} = [(1+d')/2]$ (yield of hybrid − yield of adapted cultivar) + $[(1-d')/2]$ (yield of hybrid − yield of parent), where d' = assumed mean degree of dominance

Table 3 Results of using three methods (parental yield, hybrid yield, and T_{ij} statistic) to identify a parental line carrying favorable alleles for yield not present in three adapted soybean cultivars (based on data reported by Nelson and Bernard 1984)

Year of data	Adapted cultivar	Number of parents evaluated	Best parent by method		
			Parental yield	Hybrid yield	T_{ij}[a]
1977	Beeson	3	Cutler 71	Cutler 71	Cutler 71 for all d'($0.3 \leqslant d' \leqslant 1.2$)
1977	Wells	5	Bonus	Bonus	Bonus for all d'($0.3 \leqslant d' \leqslant 1.2$)
1977	Williams	5	Bonus	Calland	Kanrich for $0.3 \leqslant d' < 0.45$; Calland for $0.45 \leqslant d' \leqslant 1.2$
1978	Beeson	7	Bonus	Corsoy	Corsoy for all d'($0.3 \leqslant d' \leqslant 1.2$)
1978	Wells	5	Bonus	Cutler 71	Cutler 71 for all d'($0.3 \leqslant d' \leqslant 1.2$)
1978	Williams	8	Beeson	Cutler 71	Cutler 71 for all d'($0.3 \leqslant d' \leqslant 1.2$

[a] $T_{ij} = [(1+d')/2]$ (yield of hybrid − yield of adapted cultivar) + $[(1-d')/2]$ (yield of hybrid − yield of parent), where d' = assumed mean degree of dominance

methods is indicated in Tables 1, 2, and 3. Selection based on F_{ij}, if it gives results similar to that based on T_{ij}, would not require evaluation of either P_i or A_j, only the hybrid.

Empirical evaluation of the proposed testcross experiment and the T_{ij} statistic should be carried out to determine their worth. For example, the relationship of the T_{ij} and $\bar{T}_i$ statistics to the mean and variance of lines derived from adapted × exotic crosses could be assessed.

References

Allan RE (1980) Wheat. In: Fehr WR, Hadley HH (eds) Hybridization of crop plants. American Society of Agronomy and Crop Science Society of America, Madison, Wisconsin, USA, pp 709–720

Burton JW, Carter TE Jr (1983) A method for production of experimental quantities of hybrid soybean seed. Crop Sci 23:388–390

Coffman WR, Herrera RM (1980) Rice. In: Fehr WR, Hadley HH (eds) Hybridization of Crop Plants. American Society of Agronomy and Crop Science Society of America, Madison, Wisconsin, USA, pp 511–522

Cregan PB, Busch RH (1977) Early generation bulk hybrid testing of adapted hard red spring wheat crosses. Crop Sci 17:887–891

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

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

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

Dudley JW (1987 b) Modifications of methods for identifying inbred lines useful for improving parents 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

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

Gerloff JE, Smith OS (1988 b) Choice of method for identifying germplasm with superior alleles. II. Computer simulation results. Theor Appl Genet 76:217–227

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

Isleib TG, Wynne JC (1983 a) Heterosis in testcrosses of 27 exotic peanut cultivars. Crop Sci 23:832–841

Isleib TG, Wynne JC (1983 b) F_4 bulk testing in testcrosses of 27 exotic peanut cultivars. Crop Sci 23:841–846

Kenworthy WJ (1980) Strategies for introgressing exotic germplasm in breeding programs. In: Corbin FT (ed) World soybean research conference II: proceedings. Westview Press, Boulder, Colorado, USA, pp 217–223

Nelson RL, Bernard RL (1984) Production and performance of hybrid soybeans. Crop Sci 24:549–553

Paschal EH, Wilcox JR (1975) Heterosis and combining ability in exotic soybean germplasm. Crop Sci 15:344–349

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

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

Rick CM (1980) Tomato. In: Fehr WR, Hadley HH (eds) Hybridization of crop plants. American Society of Agronomy and Crop Science Society of America, Madison, Wisconsin, USA, pp 669–680

Skorupska H, Palmer RG (1989) Genetics and cytology of the *ms6* male-sterile soybean. J Hered 80:304–310

Smith EL, Lambert JW (1968) Early generation testing in spring barley. Crop Sci 8:490–493

Starling TM (1980) Barley. In: Fehr WR, Hadley HH (eds.) Hybridization of crop plants. American Society of Agronomy and Crop Science Society of America, Madison, Wisconsin, USA, pp 189–202

Sweeney PM, St. Martin SK (1989) Testcross evaluation of exotic soybean germplasm of different origins. Crop Sci 29:289–293

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