

Correlation between testcross performance of lines at early and late selfing generations *

R. Bernardo

Limagrain Genetics, R.R.1, Box 232A, 737 Co. Rd. 1200 N., Champaign, IL 61821, USA

Received August 3, 1990; Accepted December 5, 1990

Communicated by A. R. Hallauer

Summary. In hybrid breeding programs, testcross evaluation of lines can be done during the early stages of selfing (early testing) or delayed until the lines are near-homozygous. To evaluate the usefulness of early testing, the expected genetic and phenotypic correlations between testcross performance at different selfing generations were examined. The genetic correlation ($r_{G_n G_{n'}}$) between testcross performance of S_n and $S_{n'}$ ($n' > n$) individuals or lines is equal to the square root of the ratio of their testcross genetic variances, and it is a function of the inbreeding coefficients (F) at the two selfing generations, i.e., $r_{G_n G_{n'}} = [(1 + F_n)/(1 + F_{n'})]^{0.5}$. The genetic correlation between testcross performance of lines and their directly descended homozygous ($n' = \infty$) lines is 0.71 for S_1 , 0.87 for S_2 , 0.93 for S_3 , 0.97 for S_4 , 0.98 for S_5 , and 0.99 for S_6 lines. The effectiveness of early testing is limited mainly by nongenetic effects. The square root of testcross heritability at generation n sets the upper limit on the correlation between phenotypic value at generation n and genotypic value at homozygosity. The probabilities of correctly retaining S_n individuals or lines that have superior testcross performance at homozygosity ($n' = \infty$) indicate that early testing should be effective in identifying lines with above- and below-average combining ability. However, the risk of losing lines with superior combining ability is high if strong (best 10%) selection pressure is applied during early testing. If only a small proportion of lines is retained based on testcross performance and/or if the heritability of the trait is low, selfing for two or three generations prior to testcrossing may be desirable to increase the likelihood of retaining lines that perform well at homozygosity. The theoretical results in this study

support the testcross evaluation procedures for grain yield used by most maize (*Zea mays* L.) breeders.

Key words: Testcross – Early generation testing – Selfing – Genetic correlation – *Zea mays* L.

Introduction

In maize (*Zea mays* L.) hybrid breeding programs, inbred lines are developed from segregating base populations by self-pollination and testing for performance in hybrid combination (Hallauer 1990). During inbreeding, visual selection is done for plant and ear traits and disease resistance of the lines per se, while yield evaluation of the lines is based on their performance when crossed to elite inbred lines or single crosses (Bauman 1981).

One system of maize inbred line development involves visual selection among and within ear-to-row progenies for several selfing generations. Testcross evaluation is delayed until the number of selections is greatly reduced and the lines are near-homozygous (Hallauer 1990). A second system involves evaluation of testcrosses during the early generations of selfing (e.g., S_0 or S_1). Lines that do not perform well are discarded early to allow expenditure of resources on the more promising lines. This procedure, called “early testing,” relies on the assumption that the combining ability of a line is determined during the early stages of selfing and does not change substantially with continued inbreeding (Jenkins 1935; Sprague 1946). Most maize breeders use a compromise between early and late testing and evaluate testcrosses of lines intermediate generations of selfing (Bauman 1981).

* A contribution from Limagrain Genetics, a Groupe Limagrain company

The usefulness of early testing has been debated ever since it was first proposed (Hallauer 1990). Empirical data that support early generation testing in maize were presented by Jenkins (1935), Sprague (1946), Lonquist (1950), Hallauer and Lopez-Perez (1979), and Jensen et al. (1983). However, Richey (1945) and Payne and Hayes (1949) presented arguments against early testing. Whereas Jenkins concluded that “the inbred lines acquired their individuality as parents of top crosses very early in the inbreeding process and remained relatively stable thereafter,” Richey concluded that “the use of either selfed or crossed progeny performance as a basis for selecting among or within families in the initial stages of a breeding program is not warranted.” However, these conflicting conclusions have not been addressed on the basis of the expected genetic and phenotypic correlations between testcross performance at early and late selfing generations.

The objectives of this paper are to: (1) present equations for the expected genetic and phenotypic correlations between testcross performance of lines at different selfing generations; (2) calculate these correlations for different levels of heritability; (3) calculate the probabilities of correctly retaining partially inbred lines that have superior testcross performance at homozygosity; and (4) examine the usefulness of early testing given the values of these correlations and probabilities.

Theory

Assume that lines are developed by n and n' ($n' > n$) generations of selfing from a large (conceptually infinite) population in Hardy-Weinberg and gametic equilibrium. Each S_n line is a descendant of an individual S_n line and of an individual S_0 plant, and no selection occurs during selfing (e.g., single-seed descent). S_n and $S_{n'}$ individuals or lines are crossed to a common tester T. The tester is either an inbred line or a population of conceptually infinite size. The reference population is the gene-orthogonal

population between T and the S_0 population (Melchinger 1988). Free recombination and negligible epistasis between loci are assumed.

At a single locus, coded values of the ++, +-, and -- genotypes are designated a , d , and $-a$, respectively. The frequency of the + allele is p in the population undergoing selfing and r in the tester. The corresponding frequencies of the - allele are $q = (1 - p)$ in the population and $s = (1 - r)$ in the tester.

Let F_n denote the coefficient of inbreeding of the S_n individual or line. If an individual plant in the S_n generation is crossed to T, $F_n = 1 - (1/2)^n$. If several plants are used to represent the S_n line in testcrosses, $F_n = 1 - (1/2)^{n-1}$. The genetic variance among S_n testcrosses is (Rawlings and Thompson 1962)

$$\text{Var}(\text{TC}_n) = (1 + F_n) \frac{1}{2} p q [a + (s - r) d]^2.$$

The genetic covariance between the testcross performance of S_n and $S_{n'}$ individuals or lines, can be obtained from the genetic structure of their testcrosses (Table 1). S_n plants are either ++, +-, or -- with frequencies of $p^2 + p q F_n$, $2 p q (1 - F_n)$, and $q^2 + p q F_n$, respectively. The $S_{n'}$ plants will be ++ or -- if they are descendants of ++ or -- S_n plants, respectively. If a +- S_n plant is selfed, the resulting $S_{n'}$ plants are either ++, +-, or -- with conditional frequencies of $\frac{1}{2} [(F_{n'} - F_n)/(1 - F_n)]$, $(1 - F_n)/(1 - F_n)$, and $\frac{1}{2} [(F_{n'} - F_n)/(1 - F_n)]$, respectively. In the absence of selection, the testcross means (μ) of S_n and $S_{n'}$ individuals or lines, are identical. Thus, the genetic covariance between testcrosses of S_n and $S_{n'}$ individuals or lines, is (from Table 1)

$$\begin{aligned} \text{Cov}(\text{TC}_n, \text{TC}_{n'}) &= (p^2 + p q F_n) (r a + s d)^2 \\ &\quad + 2 p q (1 - F_n) \left\{ \frac{1}{2} [a(r - s) + d] \right\} \left\{ \frac{1}{2} [(F_{n'} - F_n)/(1 - F_n)] \right. \\ &\quad \cdot (r a + s d + r d - s a) + [(1 - F_n)/(1 - F_n)] \frac{1}{2} [a(r - s) + d] \left. \right\} \\ &\quad + (q^2 + p q F_n) (r d - s a)^2 - \mu^2 \\ &= (p^2 + p q F_n) (r a + s d)^2 \\ &\quad + 2 p q (1 - F_n) \left\{ \frac{1}{2} [a(r - s) + d] \right\}^2 \\ &\quad + (q^2 + p q F_n) (r d - s a)^2 - \mu^2 \\ &= \text{Var}(\text{TC}_n). \end{aligned}$$

Therefore, the genetic covariance between testcrosses of S_n and $S_{n'}$ individuals or lines, is equal to the genetic variance among S_n testcrosses. This result is similar to that obtained by Melchinger (1988) who used a different approach.

Table 1. Genetic structure of testcrosses at n and n' ($n' > n$) selfing generations^a

Generation n			Generation n'		
Genotype	Frequency	Testcross mean	Genotype	Conditional frequency ^b	Testcross mean
++	$p^2 + p q F_n$	$r a + s d$	++	1	$r a + s d$
+ -	$2 p q (1 - F_n)$	$\frac{1}{2} [a(r - s) + d]$	++	$(F_{n'} - F_n)/2(1 - F_n)$	$r a + s d$
			+ -	$(1 - F_n)/(1 - F_n)$	$\frac{1}{2} [a(r - s) + d]$
			--	$(F_{n'} - F_n)/2(1 - F_n)$	$r d - s a$
--	$q^2 + p q F_n$	$r d - s a$	--	1	$r d - s a$

^a p = frequency of the + allele in the population; r = frequency of the + allele in the tester; $q = 1 - p$; $s = 1 - r$; F_n = inbreeding coefficient at selfing generation n ; a , d , and $-a$ = coded values of ++, +-, and -- genotypes, respectively

^b Conditional frequency of a genotype in generation n' given the ancestral genotype in generation n

The genetic correlation between testcrosses of S_n and $S_{n'}$ individuals or lines is

$$\begin{aligned} r_{GnGn'} &= \text{Cov}(\text{TC}_n, \text{TC}_{n'}) [\text{Var}(\text{TC}_n) \text{Var}(\text{TC}_{n'})]^{-0.5} \\ &= (1 + F_n)^{1/2} p q [a + (s-r) d]^2 \\ &\quad \cdot \{ (1 + F_n)(1 + F_{n'}) [\frac{1}{2} p q [a + (s-r) d]^2]^2 \}^{-0.5} \\ &= [(1 + F_n)/(1 + F_{n'})]^{0.5}. \end{aligned}$$

Therefore, the genetic correlation between testcrosses of S_n and $S_{n'}$, individuals or lines, is equal to the square root of the ratio of their testcross genetic variances and is a function of the inbreeding coefficients at the two selfing generations.

The phenotypic correlation between testcrosses of S_n and $S_{n'}$, individuals or lines, is

$$r_{PnPn'} = r_{GnGn'} h_n h_{n'},$$

where h_n and $h_{n'}$ are the square roots of the heritabilities of S_n and $S_{n'}$ testcrosses, respectively. If the amount of nongenetic variance is constant for generations n and n' ,

$$r_{PnPn'} = r_{GnGn'} \{ h_n^2 (1 + F_n) / [(1 + F_n) + (1 + F_{n'}) (1/h_n^2 - 1)] \}^{0.5}.$$

The above equations for $r_{PnPn'}$ assume uncorrelated nongenetic effects on S_n and $S_{n'}$ testcross performance, and the phenotypic covariance is equal to $\text{Cov}(\text{TC}_n, \text{TC}_{n'})$. In experiments grown in a single environment or in several environments that are considered fixed, this assumption is met by using proper randomization procedures (Cockerham 1961). If the testcrosses are evaluated at multiple random environments, random error effects should be uncorrelated, but covariances due to genotype \times environment interaction effects may exist. The latter may arise if the reactions to different environmental conditions (e.g., locations) of an S_n individual or line and its directly descended $S_{n'}$ individual or line are similar. A nonzero covariance between genotype \times environment interaction effects causes bias in the estimates of phenotypic correlation obtained using the above equations.

Early testing will be effective if it allows identification of partially inbred lines that would eventually perform well in testcrosses at homozygosity. One measure of the effectiveness of early testing is the correlation between the phenotypic testcross value of an S_n individual or line and the true genetic value (i.e., value in the absence of nongenetic effects) of the testcross of a directly descended individual or line at an advanced generation (e.g., $n' = \infty$). This correlation, denoted $r_{PnGn'}$, is equal to

$$r_{PnGn'} = r_{GnGn'} h_n.$$

A second measure of the effectiveness of early testing is the probability of retaining S_n lines that are genetically superior in testcrosses at homozygosity ($n' = \infty$). Let tc_n represent the standardized testcross phenotypic value at generation n and tc_∞ represent the standardized testcross genetic value at $n' = \infty$. As-

sume tc_n and tc_∞ have a bivariate normal distribution with a correlation equal to $r_{PnG\infty}$. With truncation selection at generation n , lines with testcross performance greater than z_{α_i} are selected, where $P(tc_n > z_{\alpha_i}) = \alpha_i$. In this paper, proportions selected (α_i) equal to 50% and 10% are considered. The conditional probability that the genetic testcross value of an S_∞ line is in the upper α_i of the distribution, given that the phenotypic testcross value of its ancestral S_n individual or line is in the upper α_i of the distribution, is equal to

$$P(tc_\infty > z_{\alpha_i} | tc_n > z_{\alpha_i}) = P(tc_\infty > z_{\alpha_i}, tc_n > z_{\alpha_i}, r_{PnG\infty}) / \alpha_i.$$

Values of $P(tc_\infty > z_{\alpha_i}, tc_n > z_{\alpha_i}, r_{PnG\infty})$ were obtained from bivariate normal distribution tables prepared by the U.S. National Bureau of Standards (1959).

Results and discussion

Usefulness of early testing

Expected genetic correlations between testcrosses of lines at different selfing generations range from 0.71 to 1.00 (Table 2). The genetic correlation between testcrosses at generations n and n' ($r_{GnGn'}$) increases as the difference between their inbreeding coefficients ($F_{n'} - F_n$) decreases. The genetic correlation between testcrosses of individual S_0 plants or S_1 lines and their directly descended homozygous lines (S_∞) is $(1/2)^{0.5} = 0.71$. This result supports Jenkins' (1935) conclusion that "the individuality as parents" of inbred lines in testcrosses is established early during inbreeding. The amount of increase in $r_{GnG\infty}$ diminishes with each additional selfing generation. $r_{GnG\infty}$ is equal to 0.87 for S_2 , 0.93 for S_3 , 0.97 for S_4 , 0.98 for S_5 , and 0.99 for S_6 lines.

Nongenetic effects decrease the effectiveness of early generation testing substantially. Specifically, the square root of testcross heritability at generation n (h_n) sets the upper limit on the correlation between phenotypic value at generation n and genotypic value at homozygosity ($r_{PnG\infty}$) (Table 3). Because $r_{GnG\infty}$ is expected to be at least 0.71, the effectiveness of early generation testing (as indicated by $r_{PnG\infty}$) is limited primarily by nongenetic effects. If S_1 lines are testcrossed, $r_{P1G\infty}$ values are 0.35 for $h^2 = 0.25$, 0.50 for $h^2 = 0.50$, and 0.61 for $h^2 = 0.75$. Increases in $r_{PnG\infty}$ are small beyond the S_3 generation (or S_2 generation if individual plants rather than lines are testcrossed).

Table 2. Expected genetic correlations between testcrosses at n and n' selfing generations

Generation n		n' line							
Individual	Line	F_n	S_2	S_3	S_4	S_5	S_6	S_7	S_∞
S_0	S_1	0	0.82	0.76	0.73	0.72	0.71	0.71	0.71
S_1	S_2	0.5		0.92	0.89	0.88	0.87	0.87	0.87
S_2	S_3	0.75			0.97	0.95	0.94	0.94	0.93
S_3	S_4	0.875				0.98	0.98	0.97	0.97
S_4	S_5	0.9375					0.99	0.99	0.98
S_5	S_6	0.96875						1.00	0.99

Table 3. Expected correlation between testcross phenotypic value at generation n and true genetic value at homozygosity ($r_{PnG\infty}$) for different ratios of nongenetic variance [$\text{Var}(E)$] to testcross genetic variance among S_0 individuals [$\text{Var}(TC_0)$]

Generation n		Var(E):Var(TC ₀)					
Indi- vidual	Line	3:1		1:1		1:3	
		$r_{PnG\infty}$	h_n^2	$r_{PnG\infty}$	h_n^2	$r_{PnG\infty}$	h_n^2
S_0	S_1	0.35	0.25	0.50	0.50	0.61	0.75
S_1	S_2	0.50	0.33	0.67	0.60	0.78	0.82
S_2	S_3	0.57	0.37	0.75	0.64	0.86	0.84
S_3	S_4	0.60	0.38	0.78	0.65	0.89	0.85
S_4	S_5	0.62	0.39	0.80	0.66	0.91	0.85
S_5	S_6	0.62	0.40	0.81	0.66	0.92	0.86

Table 4. Conditional probabilities that the testcross genetic value of a homozygous line ($n' = \infty$) is in the upper α tail of the distribution, given that the testcross phenotypic value of the ancestral individual or line at generation n is in the upper α tail of the distribution. Probabilities are for different heritability values of S_0 testcrosses (h_0^2)

Generation n		$h_0^2=0.25$		$h_0^2=0.50$		$h_0^2=0.75$	
Indi- vidual	Line	$\alpha=0.5$	$\alpha=0.1$	$\alpha=0.5$	$\alpha=0.1$	$\alpha=0.5$	$\alpha=0.1$
		S_0	S_1	0.61	0.24	0.67	0.32
S_1	S_2	0.67	0.32	0.73	0.44	0.79	0.54
S_2	S_3	0.69	0.37	0.77	0.51	0.83	0.63
S_3	S_4	0.70	0.39	0.79	0.54	0.85	0.67
S_4	S_5	0.71	0.40	0.80	0.56	0.86	0.70
S_5	S_6	0.71	0.40	0.80	0.57	0.87	0.72

The probabilities of correctly retaining S_n lines that are genetically superior in testcrosses at homozygosity ($n' = \infty$) are higher at later than at earlier selfing generations (Table 4). As expected, $P(tc_\infty > z_{\alpha i} | tc_n > z_{\alpha i})$ increases as heritability increases. For the levels of heritability considered, the probability of retaining genetically superior individuals is at least 0.61, even at early selfing generations (S_0 individuals or S_1 lines), if half of the lines are selected ($\alpha = 50\%$). If the breeder selects 50 out of 100 S_n lines based on testcross performance, at least 30 of the 50 lines selected are expected to be superior (i.e., in the top 50%) at homozygosity. Thus, early testing should be effective in discriminating between lines with above- and below-average combining ability. $P(tc_\infty > z_{\alpha i} | tc_n > z_{\alpha i})$ values at early generations are low if strong selection pressure is applied ($\alpha = 10\%$). If the breeder selects 10 out of 100 S_1 lines based on testcross performance, only 3 out of the 10 lines selected are expected to be superior (i.e., in the top 10%) at homozygosity if S_0 testcross heritability is equal to 0.50. The probabilities of correctly retaining the best 50 and 10% of lines are not directly comparable,

because the definition of "superior" is not consistent in the two cases (i.e., above the population mean versus in the top 10% of the population). However, regardless of the proportion of lines selected based on testcross performance, increments in $P(tc_\infty > z_{\alpha i} | tc_n > z_{\alpha i})$ are small beyond the S_3 generation (or S_2 generation if individuals instead of lines are testcrossed).

The heritability of maize grain yield is usually less than 0.30 (Hallauer and Miranda 1981). For such traits with low heritability, selfing for two or three generations prior to testcrossing may be desirable to increase the likelihood of retaining lines that perform well at complete homozygosity. Based on both $r_{PnG\infty}$ and $P(tc_\infty > z_{\alpha i} | tc_n > z_{\alpha i})$ values, there is little advantage in delaying testcross evaluation beyond the S_3 generation (or S_2 generation if individual plants rather than lines are testcrossed). Bauman (1981) indicated that 22% of maize breeders surveyed delay testcrossing of lines beyond the S_3 stage. Among the breeders surveyed, 18% evaluated testcrosses of lines at the S_1 stage, 33% at the S_2 stage, and 27% at the S_3 stage. The theoretical results in this study support the testcross evaluation procedures used by most maize breeders.

Comparison with other results

Jensen et al. (1983) studied the phenotypic correlation between testcross performance at different selfing generations in maize. For grain yield, the average phenotypic correlation between S_1 and S_4 performance when crossed to different testers was 0.67, with a heritability of 0.80. The calculated phenotypic correlation between S_1 and S_4 testcross performance is 0.61, which is close to the value obtained by Jensen et al. (1983).

In the BSSS maize population, Hallauer and Lopez-Perez (1979) reported an average (for five different testers) genetic correlation equal to 0.34 between grain yield testcross performance of S_1 and S_8 lines. This value is lower than the expected correlation of 0.71. The reasons for this discrepancy are unclear. Genetic correlations estimated in experiments are subject to large standard errors, and it is unclear whether the empirical estimate of 0.34 is significantly different from the expected value of 0.71. Aside from sampling error, other factors that may contribute to discrepancies between estimated and expected genetic correlations include selection and/or drift during inbreeding, linkage, and epistasis.

¹ In this paper, S_0 refers to the noninbred, segregating generation, e.g., the F_2 following a cross between two homozygous lines. Responses to the questions in Bauman's (1981) study indicated that the breeders surveyed used S_0 to refer to the F_1 of a biparental cross and S_1 to refer to the F_2 progeny produced by selfing the F_1 . This same notation was used by Jensen et al. (1983). Thus, the S_1, S_2, \dots, S_5 generations in Bauman's and Jensen et al.'s papers are referred to as S_0, S_1, \dots, S_4 in this paper.

Hallauer and Lopez-Perez (1979) reported significant changes in the average testcross performance of the S_1 and S_8 lines studied. These differences may be partly due to natural selection and/or genetic drift. Because linkage and epistasis affect the variance among testcrosses (Melchinger 1988), the genetic correlation is also affected by these two factors. However, epistasis is relatively unimportant compared to additive and dominance effects in maize grain yield (Hallauer and Miranda 1981). In the absence of free recombination, the covariance between average effects of a gene substitution at two linked loci contributes to the variance among testcrosses. This covariance is positive with coupling phase and negative with repulsion phase linkages. The coefficient of this covariance is larger in later than in earlier selfing generations (Melchinger 1988). Compared to the expected $r_{GnGn'}$ values with unlinked loci, $r_{GnGn'}$ values will be larger with coupling phase linkages and smaller with repulsion phase linkages. However, the BSSS population studied by Hallauer and Lopez-Perez (1979) was formed by intermating 16 inbred lines and was maintained for many generations by random mating. Thus, it is difficult to envision that linkage had substantial effects on the estimates of genetic correlation obtained.

References

- Bauman LF (1981) Review of methods used by breeders to develop superior inbreds. Proc Corn Sorghum Industry Res Conf 36:199–208
- Cockerham CC (1961) Implications of genetic variances in a hybrid breeding program. Crop Sci 1:47–52
- Hallauer AR (1990) Methods used in developing maize inbreds. Maydica 35:1–16
- Hallauer AR, Lopez-Perez E (1979) Comparisons among testers for evaluating lines of corn. Proc. Corn Sorghum Industry Res Conf 34:57–75
- Hallauer AR, Miranda JO (1981) Quantitative genetics in maize breeding. Iowa State University Press, Ames/IA
- Jenkins MT (1935) The effect of inbreeding and of selection within inbred lines of maize upon the hybrids made after successive generations of selfing. Iowa State Coll J Sci 3:429–450
- Jensen SD, Kuhn WE, McConnell RL (1983) Combining ability studies in elite U.S. maize germ plasm. Proc Corn Sorghum Industry Res Conf 38:87–96
- Lonnquist JH (1950) The effect of selecting for combining ability within segregating lines of corn. Argon J 42:503–508
- Melchinger AE (1988) Means, variances, and covariances between relatives in hybrid populations with disequilibrium in the parent populations. In: Weir BS, Eisen EJ, Goodman MM, Namkoong G (ed) Proc 2nd Int Conf Quantit Genet. Sinauer, Sunderland/Ma, pp 400–415
- Payne KT, Hayes HK (1949) A comparison of combining ability in F_3 and F_2 lines of corn. Agron J 41:383–388
- Rawlings JO, Thompson DL (1962) Performance level as criterion for the choice of maize testers. Crop Sci 2:217–220
- Richey FD (1945) Isolating better foundation inbreds for use in corn hybrids. Genetics 30:455–471
- Sprague GF (1946) Early testing of inbred lines of corn. J Am Soc Argon 38:108–117
- U.S. National Bureau of Standards (1959) Tables of the bivariate normal distribution function and related functions. US Govt Printing Office, Washington, DC