Theor Appl Genet (2003) 107:1252–1256
DOI 10.1007/s00122-003-1375-0

R. Bernardo

Parental selection, number of breeding populations, and size of each population in inbred development

Received: 29 April 2003 / Accepted: 11 July 2003 /

Abstract Some breeders select inbreds from many F_2 or backcross breeding populations, each with relatively few progenies. Other breeders select inbreds from only a few breeding populations, each with many progenies. My objectives were to: (1) determine the relative importance of parental selection, number of breeding populations, and size of each population, and (2) find optimum combinations between number and size of breeding populations. I assumed that a breeder has resources to test a total of 2,000 recombinant inbreds for a quantitative trait that was controlled by 100 additive loci and had a heritability of 0.20, 0.60, or 1.0. The parental inbreds had an inherent pedigree structure due to advanced cycle breeding. The parental inbreds were ranked according to their mean performance, and breeding populations were made among all parents, the top 25% of parents, and the top 10% of parents. I found that the issue of number versus size of breeding populations was only secondary compared with the ability to identify, prior to making the crosses, the breeding populations with the highest mean performance. For a given level of effectiveness of parental selection, the selection response was largest when the maximum number of breeding populations was used. The effect of the number of breeding populations was minor, however, when selection was practiced among the parents or when heritability was less than 1.0. The results suggested that, in practice, large selection responses could be obtained with a wide range of combinations between number and size of breeding populations.

Keywords Parental selection · Breeding populations · Inbred development

Communicated by H.C. Becker

R. Bernardo (✉)
Department of Agronomy and Plant Genetics,
University of Minnesota,
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul,
MN 55108, USA
e-mail: berna022@umn.edu
Tel.: 1-612-6256282
Fax: +1-612-6251268

Introduction

New inbreds are most often developed from crosses among elite inbreds in cultivar development programs (Allard 1960, p 282). Specifically, two inbreds are first selected as parents of an F_2 or backcross breeding population. New inbreds are then developed by pedigree selection, single-seed descent, or the bulk method of breeding. A breeder typically creates, selfs, and selects in several breeding populations at a time. This scheme (which has become known as advanced cycle breeding) for developing new inbreds is widely used both in self-pollinated crops, such as soybean [*Glycine max* (L.) Merrill; Hartwig 1973] and wheat (*Triticum aestivum* L.; Heyne and Smith 1967), and in hybrid crops, such as maize (*Zea mays* L.; Hallauer 1990).

Due to a finite amount of resources available in a breeding program, trade-offs need to be made between the number of F_2 or backcross breeding populations and the size of each breeding population (Yonezawa and Yamagata 1978; Baker 1984). Suppose a breeder has the resources to evaluate a total of 2,000 experimental families or recombinant inbreds for a quantitative trait, such as yield. On one extreme, the breeder may decide to develop and evaluate 200 F_2 or backcross breeding populations, each with only 10 families or recombinant inbreds. On the other extreme, the breeder may decide to develop and evaluate 10 breeding populations, each with 200 families or recombinant inbreds. The breeder may decide to take the middle ground, e.g., 50 breeding populations, each with 40 families or recombinant inbreds. Or perhaps he or she may decide to develop and evaluate more progenies in some breeding populations and fewer progenies in others.

The issue of number versus size of breeding populations has been studied with two approaches (Wricke and Weber 1986): (1) minimizing the risk of excluding superior genotypes, or (2) maximizing the response to selection. Using the first approach, Yonezawa and Yamagata (1978) and Weber (1979) concluded that with a finite amount of resources, the number of breeding populations rather than the size of each population should be maximized, i.e., each breeding population should be represented by only one family or recombinant inbred. Using the second approach, Baker (1984), Wricke and Weber (1986), and Hühn (1996) considered the response to selection both among and within breeding populations. Baker found that when testing resources are available for a total of 2,000 families, the maximum predicted response is obtained by developing 50 to 100 breeding populations, each with 20 to 40 families. The approaches of minimizing risk versus maximizing selection response have therefore given conflicting results.

Moreover, the approaches used in the aforementioned studies had three important limitations. First, breeding populations were assumed to have been made from random crosses, rather than selected crosses, among the available inbreds. In practice, however, breeders have some prior knowledge regarding which breeding populations might be the most promising. Second, pedigree relationships that arise among inbreds due to advanced cycle breeding were not considered. These pedigree relationships would cause breeding populations to differ in their genetic variance, thereby rendering the approaches used in previous studies to be inapplicable. Third, linkage was ignored. Parental selection, pedigree relationships, and linkage are difficult to account for in an analytical approach, but they could be accounted for in a simulation experiment. My objectives in this simulation study were to: (1) determine the relative importance of parental selection, number of breeding populations, and size of each breeding population, and (2) find optimum combinations between number and size of breeding populations.

Materials and methods

General approach

I assumed that a breeder has resources to test a total of 2,000 recombinant inbreds for a quantitative trait. The number of recombinant inbreds in each breeding population was denoted by n_{RI}, and the number of breeding populations was $n_{\mathrm{BP}} = 2{,}000/n_{\mathrm{RI}}$. I considered the following combinations of n_{RI} and n_{BP} (in parenthesis): 1 (2,000), 5 (400), 10 (200), 20 (100), 40 (50), 80 (25), 125 (16), 200 (10), 250 (8), 500 (4), 1,000 (2), and 2,000 (1). I wrote a Fortran program to simulate the following steps: (1) develop 76 parental inbreds by advanced cycle breeding; (2) select parents of F_2 breeding populations; (3) develop a total of n_{BP} breeding populations, each with n_{RI} recombinant inbreds; and (4) determine the mean of the best 20 out of 2,000 (1%) recombinant inbreds.

Three methods were considered for selecting the parents of F_2 breeding populations. In the Random Parents method, pairs of inbreds were selected at random from the 76 parental inbreds. This first method therefore represented the lack of parental selection, as each inbred had an equal chance of being a parent of an F_2 breeding population. In the Top 25% of Parents method, the best 19 out of the 76 (25%) parental inbreds were first identified on the basis of their known genotypic values. Pairs of inbreds were then selected at random from the 19 best parental inbreds. In the Top 10% of Parents method, breeding populations were made by crossing pairs of inbreds from the best eight out of the 76 (approximately 10%) parental inbreds. The Top 25% of Parents method and Top 10% of Parents method therefore represented parental selection on the basis of their mean performance. Although the parents were chosen on the basis of their genotypic (rather than phenotypic) values, the three methods of parental selection can be viewed as equivalent to phenotypic selection with a fixed selection intensity but at three different levels of heritability (h^2).

The simulation experiment was repeated 5,000 times. The 5,000 repeats differed at random in the arrangement of loci into linkage groups, genotypes of inbreds, parents selected, and recombinant inbreds generated from the F_2 breeding populations.

Parental and recombinant inbreds

The number of parental inbreds simulated (76) was within the range of the number of elite inbreds (40 to 120) often found in different heterotic groups in maize (Bernardo 1996). Out of the 76 parental inbreds, four were founder inbreds, 18 were second-cycle inbreds, 27 were third-cycle inbreds, and 27 were fourth-cycle inbreds (Bernardo 2001). Selection was practiced during the development of these parental inbreds: each second-, third-, and fourth-cycle inbred had a genotypic value greater than the mean of the population from which the inbred was developed. The four founder inbreds were assumed unrelated. Three second-cycle inbreds were derived from the F_2 population of each of the six possible crosses among the founder inbreds. One third-cycle inbred was derived from the F_2 population of each of the 27 crosses between unrelated second-cycle inbreds. Finally, one fourth-cycle inbred was derived from each of the F_2 populations obtained by chain crossing the 27 third-cycle inbreds. For each of the n_{BP} breeding populations, a total of n_{RI} recombinant inbreds were then developed.

Genetic model and phenotypic values

The quantitative trait was controlled by $l = 100$ loci. Each locus had four alleles, denoted by +, +′, −, and −′. The frequency of each allele among the founder inbreds was 1/4 at each locus. The effects of the loci followed a geometric series, i.e., the quantitative trait was jointly controlled by few loci with large effects and by many loci with small effects. The genotypic values of the four homozygotes at locus k were arbitrarily set as $(0.98)^k$ for $(+/+)_k$, $1/2(0.98)^k$ for $(+'/+')_k$, $-1/2(0.98)^k$ for $(-/-)_k$, and $-(0.98)^k$ for $(-'/-')_k$ (Bernardo 2001). Dominance and epistasis were absent. The genotypic value of each parental inbred and recombinant inbred was therefore obtained by summing the genotypic values across all loci. Linkage among the loci was generated by randomly locating the l loci on 10 chromosomes. The sizes of the chromosomes (ranging from 128 to 241 centimorgans) and of the entire genome (1,749 centimorgans) corresponded to those in a published maize linkage map (Senior et al. 1996).

The mean of the base population (i.e., crosses among the founder inbreds) was $\mu = 0$. The additive genetic variance (V_A) at locus k, for a population in linkage equilibrium, was $(5/16)(0.98)^{2k}$. The variance among recombinant inbreds, across all 100 loci, was $2V_A = 14.891$. The base-population heritability for recombinant inbreds, which was equal to $h^2 = 2V_A/(2V_A + V_E)$, was either low (0.20) or moderately high (0.60). Selection based on the mean performance at several environments was assumed, and V_E was assumed to comprise both within-environment error variance and genotype-environment interaction variance. For comparison, a perfect h^2 of 1.0 was also considered. The phenotypic value of a recombinant inbred was obtained by adding a random nongenetic effect, drawn from a normal distribution with a mean of zero and a variance of V_E, to the genotypic value.

Comparison of different combinations of nBP and nRI

The best 20 recombinant inbreds out of 2,000 were identified regardless of the breeding population from which they originated. The best 20 inbreds were chosen on the basis of their phenotypic values. The mean of the 20 best inbreds, denoted by $\bar{X}_{0.01}$, was then calculated on the basis of their known genotypic values. The $\bar{X}_{0.01}$ was expressed as the number of additive genetic standard deviations among recombinant inbreds, i.e., $(2V_A)^{1/2}$. The mean $\bar{X}_{0.01}$ and standard deviation of $\bar{X}_{0.01}$ were calculated across the 5,000 repeats of the simulation experiment.

The total number of pair-wise combinations among the selected parents was 171 in the Top 25% of Parents method and 28 in the Top 10% of Parents method. I did not consider n_{BP} values less than 171 in the Top 25% of Parents method and less than 28 in the Top 10% of Parents method. For the Random Parents method, Top 25% of Parents method, and Top 10% of Parents method, two or more F_2 populations may have had the same pair of parental inbreds. I considered this as acceptable, given that breeding populations are sometimes duplicated among breeders in large, multi-station breeding programs.

Other models

Three variants of the general model were studied. First, a model with two alleles (instead of four) at each of the 100 loci was studied. The genotypic values of the homozygotes at locus *k* were $(0.98)^k$ for $(+/+)_k$ and $-(0.98)^k$ for $(-/-)_k$. The frequency of each allele among the founder inbreds was 1/2 at each locus. Second, a model with $l = 20$ additive loci was studied. The genotypic effects under this model were calculated in the same way as for the model with 100 loci. Third, resources were assumed available for testing a total of 1,000 or 4,000 (instead of 2,000) recombinant inbreds. The number of breeding populations was therefore $n_{BP} = 1{,}000/n_{RI}$ or $4{,}000/n_{RI}$.

Results and discussion

Non-additive gene action (i.e., dominance or epistasis) is often negligible or assumed negligible in self-pollinated crops (Moll and Stuber 1974). Likewise, testcross means of families in a cross-pollinated crop behave in an additive manner (Hallauer and Miranda 1981, p 28; Bernardo 2002, p 79) even if dominance is present. The assumption of additive gene action in this study implied that the results are applicable both to selection in self-pollinated crops and testcross selection in cross-pollinated crops. In this study, the size of a breeding population refers to the number of progenies that are first evaluated for a quantitative trait. If 1,000 F_2 plants are initially grown, but only 40 families derived by single-seed descent are evaluated for a quantitative trait such as yield, then the size of the breeding population is 40 rather than 1,000. For convenience, recombinant inbreds are considered in this study, although the results should be generally applicable to selection at earlier generations of selfing.

The genotypic mean of the best 20 out of 2,000 recombinant inbreds ($\bar{X}_{0.01}$) was strongly influenced by the method of parental selection. Developing recombinant inbreds from crosses among the best 10% of the parental inbreds (Top 10% of Parents method) led to the highest $\bar{X}_{0.01}$; developing recombinant inbreds from crosses among the best 25% of the parental inbreds (Top 25% of Parents method) led to intermediate $\bar{X}_{0.01}$; and developing recombinant inbreds from random crosses among all available parental inbreds (Random Parents method) led to the lowest $\bar{X}_{0.01}$ (Fig. 1). One method remained superior to another regardless of the size of each breeding population (n_{RI}), the number of breeding populations (n_{BP}), or the heritability (h^2) of the trait.

The Random Parents method not only led to the lowest $\bar{X}_{0.01}$, but it also led to the largest standard deviation of $\bar{X}_{0.01}$, particularly when only n_{BP} = 1, 2, or 4 breeding populations were used (Fig. 1). In other words, the failure to select among the parents of breeding populations not only led to the lowest selection response, but it also caused the most erratic selection response. The standard deviations of $\bar{X}_{0.01}$ with the Top 25% of Parents method and Top 10% of Parents method were low and were comparable to each other. Parental selection reduces the variation among the means of the inbreds used as parents. This leads to reduced variation among the means of breeding populations and, consequently, among the $\bar{X}_{0.01}$ values.

For each method of parental selection, $\bar{X}_{0.01}$ was highest when the maximum number of breeding populations was used (Fig. 1). For example, when h^2 was 0.60 and the parental inbreds were chosen at random, the $\bar{X}_{0.01}$ decreased from 2.89 when recombinant inbreds were selected from n_{BP} = 2,000 breeding populations (each with n_{RI} = 1 recombinant inbred), to 2.88 when inbreds were selected from n_{BP} = 100 breeding populations (each with n_{RI} = 20 recombinant inbreds), and to 2.36 when inbreds were selected from n_{BP} = 1 breeding population (with n_{RI} = 2,000 recombinant inbreds). The results for the Random Parents method were therefore consistent with the findings of Yonezawa and Yamagata (1978) and Weber (1979): each breeding population should be represented by only one family or recombinant inbred, and the number of breeding populations should be maximized.

The number versus size of breeding populations had a smaller effect on $\bar{X}_{0.01}$, however, when selection was performed among the parents of the breeding populations. When h^2 was 0.60 and the top 25% of parents was selected, the $\bar{X}_{0.01}$ ranged from 3.38 to 3.41 when the number of breeding populations ranged from n_{BP} = 10 to 100 (Fig. 1). When h^2 was 0.60 and the top 10% of parents was selected, the $\bar{X}_{0.01}$ ranged from 3.60 to 3.63 when the number of breeding populations ranged from n_{BP} = 4 to 16. A low h^2 caused a further decrease in the effect of the number versus size of breeding populations. When h^2 was 0.20 and the top 10% of parents was selected, the $\bar{X}_{0.01}$ ranged from 3.09 to 3.10 when the number of breeding populations ranged from n_{BP} = 4 to 16. These results suggest that, in practice, large selection responses can be obtained with a wide range of combinations between number and size of breeding populations and that, again, the issue of parental selection is more important than the issue of number versus size of breeding populations. This conclusion also applied to the other models studied (i.e., 100 loci with two alleles per locus, 20 loci with four alleles per locus, and resources for testing a total of 1,000 or 4,000 recombinant inbreds; results not shown).

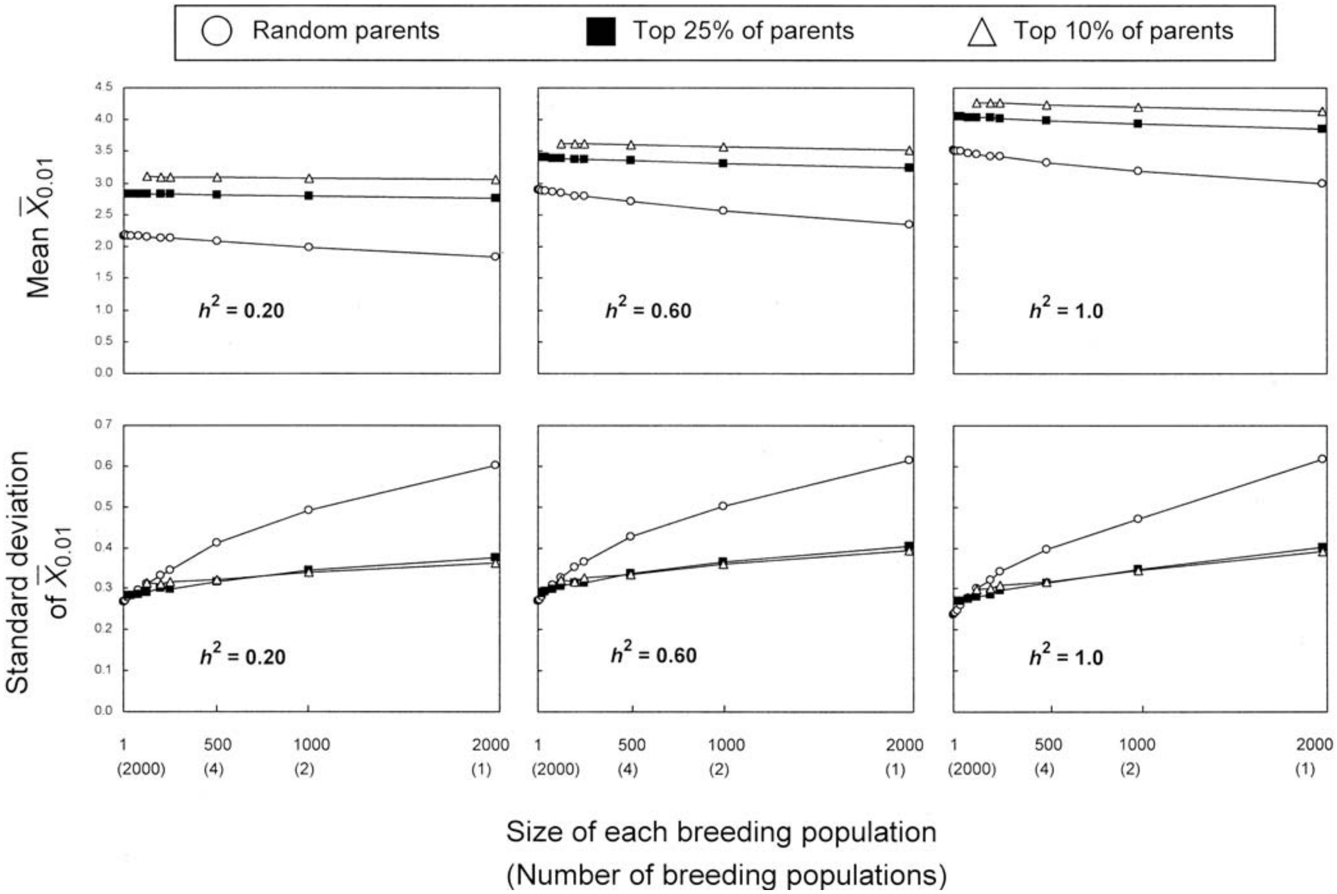

Fig. 1 Mean and standard deviation of the genotypic value of the best 20 out of 2,000 recombinant inbreds ($\bar{X}_{0.01}$) for different numbers of F_2 breeding populations, sizes of each breeding population, methods of selecting the parents of each breeding population, and heritability (h^2) of the trait

Baker (1984) studied the issue of number versus size of breeding populations in a different context and, consequently, arrived at conclusions different from those in this study. Baker assumed that F_2 breeding populations were made at random and that selection proceeded in two steps. First, the best five breeding populations were identified on the basis of the mean performance of the n_{RI} families in each population. Second, the best five families within each of the best five breeding populations were selected. In this context, Baker concluded that the optimum combination involves either 50 breeding populations with 40 families in each population, or 100 breeding populations with 20 families in each population. In the current study, however, no distinction was made between selection among and selection within breeding populations. In the Random Parents method, the best 20 out of 2,000 recombinant inbreds were selected regardless of the breeding population from which they were developed. The assumption behind the Top 25% of Parents method and the Top 10% of Parents method was that prior performance data are available for selecting breeding populations, and that none of the testing resources for recombinant inbreds was devoted to identifying the top 25% or top 10% of the breeding populations. In practice, this can be achieved by best linear unbiased prediction, in which the breeding value of an inbred is estimated from field trial data that are routinely generated in a breeding program (Panter and Allen 1995; Bernardo 2002, p 227).

Current breeding programs differ widely in the number of breeding populations that are developed and evaluated each year. In the wheat breeding program at the University of Minnesota, for example, about 300 new breeding populations are created each year (J.A. Anderson, personal communication, 2001). But in the barley (*Hordeum vulgare* L.) breeding program, about 75 breeding populations are created each year (K.P. Smith, personal communication, 2002). From personal discussions, I have found that some maize breeders create only 10 breeding populations each year, whereas other maize breeders create up to 50 breeding populations each year. The conclusion from this study is that the issue of number versus size of breeding populations is only secondary compared with the ability to identify, prior to making the crosses, the breeding populations with the highest mean performance. Breeding programs that differ in the number and size of breeding populations can therefore be equally successful in developing improved inbreds.

References

Allard RW (1960) Principles of plant breeding. Wiley, New York

Baker RJ (1984) Quantitative genetic principles in plant breeding. In: Gustafson JP (ed) Gene manipulation in plant improvement. Plenum Press, New York, pp 147–176

Bernardo R (1996) Best linear unbiased prediction of maize single-cross performance. Crop Sci 36:50–56

Bernardo R (2001) What if we knew all the genes for a quantitative trait in hybrid crops? Crop Sci 41:1–4

Bernardo R (2002) Breeding for quantitative traits in plants. Stemma Press, Woodbury, Minn.

Hallauer AR (1990) Methods used in developing maize inbreds. Maydica 35:1–16

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

Hartwig EE (1973) Varietal development. In: BE Caldwell (ed) Soybeans: improvement, production, and uses. Am Soc Agron, Madison, Wis., pp 187–210

Heyne EG, Smith GS (1967) Wheat breeding. In: KS Quisenberry (ed) Wheat and wheat improvement. Am Soc Agron, Madison, Wis., pp 269–306

Hühn M (1996) Optimum number of crosses and progeny per cross in breeding self-fertilizing crops. I. General approach and first numerical results. Euphytica 91:365–374

Moll R, Stuber CW (1974) Quantitative genetics-Empirical results relevant to plant breeding. Adv Agron 26:277–313

Panter DM, Allen FL (1995) Using best linear unbiased predictions to enhance breeding for yield in soybean. I. Choosing parents. Crop Sci 35:397–405

Senior ML, Chin ECL, Lee M, Smith JSC, Stuber CW (1996) Simple sequence repeat markers developed from maize sequences found in the GENBANK database: map construction. Crop Sci 36:1676–1683

Weber WE (1979) Number and size of cross progenies from a constant total number of plants manageable in a breeding program. Euphytica 28:453–456

Wricke G, Weber WE (1986) Quantitative genetics and selection in plant breeding. de Gruyter, Berlin

Yonezawa K, Yamagata H (1978) On the number and size of cross combinations in a breeding programme of self-fertilizing crops. Euphytica 27:113–116