Theor Appl Genet (1984) 68:183–186

Genetic variance and drift in selfed and intermated populations derived from backcrossing *

T. S. Cox **

Research Geneticist, USDA-ARS, Agronomy Department, Kansas State University, Manhattan, KS 66506, USA

Received January 4, 1984; Accepted January 9, 1984
Communicated by A. L. Kahler

Summary. The genetic variance among random-mated lines derived from backcrossing (BC_gS_1 lines) depends upon the backcross generation (g) and the number (n) of BC_gF_1 plants crossed in generations 1 through g. There is little effect of n on genetic variance for $n > 6$. The genetic variance among BC_gF_2-derived lines is greater than that among BC_gS_1 lines for all g. If either BC_gF_2-derived or BC_gS_1 lines are used as a base population for recurrent selection, 8, 16, 32, and 64 BC_1F_1, BC_2F_1, BC_3F_1, and BC_4F_1 plants, respectively, should be used to avoid loss of donor alleles to drift.

Key words: Genetic variance – Genetic drift – Backcrossings – Population size – Selection

Introduction

In crosses involving exotic germplasm, backcrossing to the adapted parent proves effective in increasing the probability of obtaining progeny superior for quantitative traits (Kenworthy and Brim 1979; Lawrence and Frey 1975; Rodgers 1982). Cox (1984) showed that, assuming no dominance or epistasis, the genetic variance among backcross-F_2-derived lines is

$$\left[\prod_{k=0}^{g-1}\left(1-\frac{1}{2^k n}\right)\left(\frac{1}{2}\right)^g - \left(1-\frac{1}{n}\right)^g\left(\frac{1}{2}\right)^{2g} + \left(\frac{1}{2}\right)^{g+1} - \frac{1}{2mn}\right]\sigma_A^2,$$

where g is the backcross generation, n is the number of BC_gF_1 plants crossed and selfed per generation, m is the mean number of BC_gF_2-derived lines per BC_gF_1 family, σ_A^2 is the variance among F_∞-derived lines, and the adapted and exotic parents were both homozygous lines. Thus, genetic variance decreases linearly with percentage recurrent parent germplasm for $n = 1$; for large n, the genetic variances among BC_0F_2- and BC_1F_2-derived lines are approximately equal.

Dudley (1982) showed with computer simulation that, if the donor parent is inferior, backcrossing produces a base population in which recurrent selection may be practiced more efficiently; the optimum number of backcrosses depends on the relative contributions of positive alleles by the recurrent and donor parents. Dudley (1982) did not consider population size in his study.

The purposes of this study are (1) to determine the effect of finite population size during random backcrossing on the genetic variance in randomly mated populations derived from backcrossing, (2) to compare the genetic variances among selfed and randomly mated populations derived from backcrossing, and (3) to determine the effect of population size during backcrossing before selection on the limits to gain from selection.

Population structures

Cox (1984) described the development of BC_gF_2-derived lines. Recurrent and donor inbred parental lines are crossed, and the F_1 is backcrossed to the recurrent parent (producing the BC_1F_1) and selfed (producing the BC_0F_2). Backcrossing is continued, each BC_gF_1 being selfed to produce BC_gF_2 plants. The selfed progeny of a random individual BC_gF_2 plant is a BC_gF_2-derived line.

---

* Joint contribution of USDA-ARS and Journal Paper No. J-11224 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa. Project No. 2471

** Formerly Research Geneticist, USDA-ARS, Ames, Iowa, USA

In this paper, I also will consider BC_gS_0 plants, which are produced by randomly intermating, rather than selfing, BC_gF_1 plants. The selfed progeny of a random individual BC_gS_0 plant will be referred to as a BC_gS_1 line.

Results and discussion

The genetic variance among BC_gS_0 plants (or BC_gS_1 lines) due to locus i is $2p_i(1-p_i)a_i^2$, where p_i is the frequency of the donor allele, a_i is half the difference between the two homozygous genotypic values at locus i, and dominance is absent. Furthermore, $p_i = \frac{x_g}{2n_g}$, where x_g out of n_g BC_gF_1 plants were heterozygous at locus i. Then,

$$\begin{aligned} V_{g(i)} &= (P(x_g))(V_{g(i)} \mid x_g) \\ &= \sum_{x_g=0}^{n_g} \frac{n_g!}{x_g!(n_g-x_g)!}\left(\frac{x_{g-1}}{2n_{g-1}}\right)^{x_g}\left(1-\frac{x_{g-1}}{2n_{g-1}}\right)^{n_g-x_g} \\ &\quad \times 2\left(\frac{x_g}{2n_g}\right)\left(1-\frac{x_g}{2n_g}\right)a_i^2, \end{aligned}$$

where x_{g-1} of n_{g-1} plants were heterozygous in the $BC_{g-1}F_1$. This expression reduces to

$$V_{g(i)} = \left[\left(1-\frac{1}{2n_g}\right)\left(\frac{x_{g-1}}{2n_{g-1}}\right)-\left(1-\frac{1}{n_g}\right)\frac{1}{2}\left(\frac{x_{g-1}}{2n_{g-1}}\right)^2\right]a_i^2.$$

Therefore,

$$\begin{aligned} V_{g+1(i)} &= \sum_{x_g=0}^{n_g} \frac{n_g!}{x_g!(n_g-x_g)!}\left(\frac{x_{g-1}}{2n_{g-1}}\right)^{x_g}\left(1-\frac{x_{g-1}}{2n_{g-1}}\right)^{n_g-x_g} \\ &\quad \times\left[\left(1-\frac{1}{2n_{g+1}}\right)\left(\frac{x_g}{2n_g}\right)-\left(1-\frac{1}{n_{g+1}}\right)\frac{1}{2}\left(\frac{x_g}{2n_g}\right)^2\right]a_i^2, \end{aligned}$$

which in turn reduces to

$$\begin{aligned} V_{g+1(i)} &= \left[\left(1-\frac{1}{2n_{g+1}}\right)-\frac{1}{4n_g}\left(1-\frac{1}{n_{g+1}}\right)\right] \\ &\quad \times\left(\frac{x_{g-1}}{4n_{g-1}}\right)a_i^2-\frac{1}{2}\left(1-\frac{1}{n_g}\right) \\ &\quad \left(1-\frac{1}{n_{g+1}}\right)\left(\frac{x_{g-1}}{4n_{g-1}}\right)^2a_i^2. \end{aligned}$$

Repeating the process for V_{g+2} and V_{g+3}, letting $g = 1$ $\left(\text{so that } \frac{x_{g-1}}{n_{g-1}} = 1\right)$, and summing over all loci i (assuming no linkage or epistasis), we obtain

$$V_1 = \left[\left(1-\frac{1}{2n_1}\right)-\left(1-\frac{1}{n_1}\right)\left(\frac{1}{2}\right)^2\right]\left(\frac{1}{2}\right)\sigma_A^2,$$

$$\begin{aligned} V_2 = \Bigg[&\left(1-\frac{1}{2n_2}\right)-\frac{1}{4n_1}\left(1-\frac{1}{n_2}\right) \\ &-\left(1-\frac{1}{n_1}\right)\left(1-\frac{1}{n_2}\right)\left(\frac{1}{2}\right)^3\Bigg]\left(\frac{1}{2}\right)^2\sigma_A^2, \end{aligned}$$

$$\begin{aligned} V_3 = \Bigg[&\left(1-\frac{1}{2n_3}\right)-\frac{1}{4n_2}\left(1-\frac{1}{n_3}\right) \\ &-\frac{1}{8n_1}\left(1-\frac{1}{n_2}\right)\left(1-\frac{1}{n_3}\right) \\ &-\left(1-\frac{1}{n_1}\right)\left(1-\frac{1}{n_2}\right)\left(1-\frac{1}{n_3}\right)\left(\frac{1}{2}\right)^4\Bigg]\left(\frac{1}{2}\right)^3\sigma_A^2, \end{aligned}$$

and

$$\begin{aligned} V_4 = \Bigg[&\left(1-\frac{1}{2n_4}\right)-\frac{1}{4n_3}\left(1-\frac{1}{n_4}\right) \\ &-\frac{1}{8n_2}\left(1-\frac{1}{n_3}\right)\left(1-\frac{1}{n_4}\right) \\ &-\frac{1}{16n_1}\left(1-\frac{1}{n_2}\right)\left(1-\frac{1}{n_3}\right)\left(1-\frac{1}{n_4}\right) \\ &-\left(1-\frac{1}{n_1}\right)\left(1-\frac{1}{n_2}\right)\left(1-\frac{1}{n_3}\right)\left(1-\frac{1}{n_4}\right)\left(\frac{1}{2}\right)^5\Bigg] \\ &\times\left(\frac{1}{2}\right)^4\sigma_A^2, \end{aligned}$$

where σ_A^2 is the variance among BC_0F_∞ plants (or derived lines).

In general, if $n_1 = n_2 = n_3 = \ldots = n$,

$$\begin{aligned} V_g = \Bigg[&\left(1-\frac{1}{2n}\right)-\sum_{k=2}^{g}\frac{1}{2^k n}\left(1-\frac{1}{n}\right)^{k-1} \\ &-\left(1-\frac{1}{n}\right)^g\left(\frac{1}{2}\right)^{g+1}\Bigg]\left(\frac{1}{2}\right)^g\sigma_A^2, \end{aligned}$$

and

$$\lim_{n\to\infty} V_g = \left(\frac{1}{2}\right)^g\left(1-\left(\frac{1}{2}\right)^{g+1}\right)\sigma_A^2.$$

As n increases from 1 to 4, there is a significant increase in V_g; however, increasing n beyond approximately 6 has little effect on V_g (Fig. 1).

The genetic variance among BC_gS_0 plants or BC_gS_1 lines is less than that among BC_gF_2 plants or derived lines (Cox 1984) for any backcross generation and a given *n* > 1 (Fig. 2). This is true because BC_gF_2 plants are not in Hardy-Weinberg equilibrium. In using a backcross generation as a base population for selection, therefore, initial response may be greater if BC_gF_2-derived lines are the unit of selection in the first cycle. Selected lines must then be intermated if recurrent selection is to be practiced as in the Dudley (1982) simulation study.

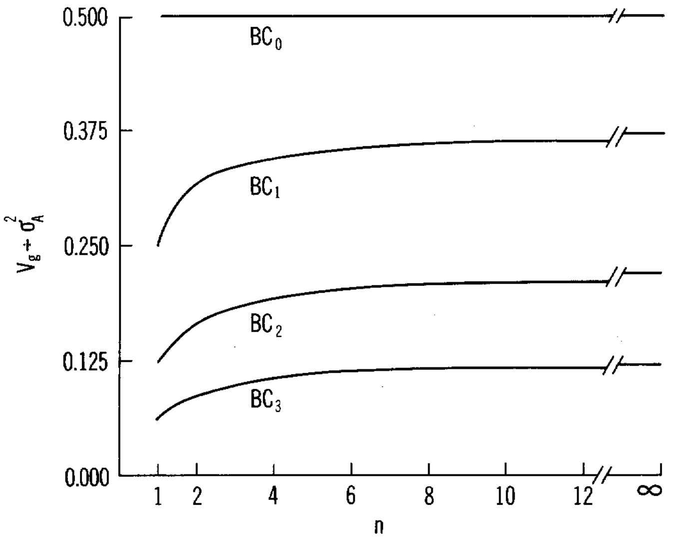

Fig. 1. Genetic variance among BC_gS_1 lines (V_g) for $1 < n < \infty$

Whatever the initial unit of selection, the ultimate gain from selection in a backcross population is limited by the number of loci still segregating for the donor allele. If n_g BC_gF_1 plants are selfed or intermated and no alleles have been lost due to drift in generations 1 through $g-1$, the probability for a given locus that no BC_gF_1 plant carries the donor allele is $(1-(\frac{1}{2})^g)^{n_g}$. The probability that at least one BC_gF_1 plant carries the donor allele is $1-(1-(\frac{1}{2})^g)^{n_g}$; this also represents the fraction of loci segregating in the population.

The left half of Table 1 shows that approximately 8, 16, 32, and 64 BC_gF_1 plants must be used in the BC_1, BC_2, BC_3, and BC_4, respectively, to prevent fixation at more than 1 to 2% of the loci. If small plant numbers are used in previous generations, fixation of alleles is more severe. For example, if only one BC_gF_1 plant is backcrossed in generations 1 through $g-1$, the fraction of loci segregating in generation g is $(\frac{1}{2})^{g-1}(1-(\frac{1}{2})^{n_g})$, where $n_g BC_gF_1$ plants are selfed or intermated

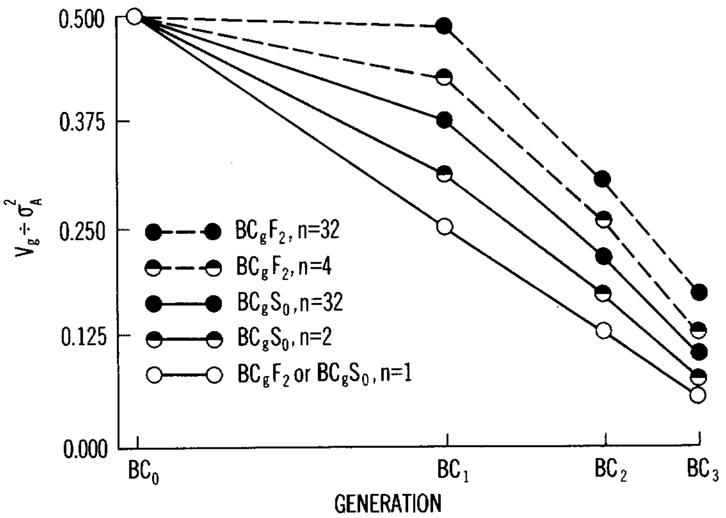

Fig. 2. Genetic variance (V_g) among BC_gF_2 plants or BC_gS_0 plants with n = 1, 2, 4, and 32. The curve for BC_gF_2 plants with n = 2 is coincident with the curve for BC_gS_0 plants with n = 32. Spacing of generations on the x-axis is proportional to percent recurrent parent germplasm

Table 1. The proportion of loci segregating in the BC_g population, given that either a large number of BCF_1 plants are used in the BC_1 through BC_{g-1} (no previous drift) or one BC_gF_1 plant per generation is used in the BC_g through BC_{g-1} (previous drift)

No of plants in BC_gF_1	No previous drift				No of plants in BC_gF_1	Previous drift			
	BC_1	BC_2	BC_3	BC_4		BC_1	BC_2	BC_3	BC_4
1	0.50	0.25	0.13	0.06	1	0.50	0.25	0.13	0.06
2	0.75	0.44	0.23	0.12	2	0.75	0.38	0.19	0.09
4	0.94	0.68	0.41	0.23	4	0.94	0.47	0.23	0.12
8	0.99	0.90	0.66	0.40	8	0.99	0.50	0.25	0.13
16	1.00	0.99	0.88	0.64	16	1.00	0.50	0.25	0.13
32	1.00	1.00	0.98	0.87	32	1.00	0.50	0.25	0.13
64	1.00	1.00	1.00	0.98	64	1.00	0.50	0.25	0.13

(Table 1, right half). Ultimate gain from selection is very reduced in this situation.

The fraction of loci segregating in the BC_g when n_1 through n_{g-1} are variable is difficult to formulate; it equals

$$\sum_{x=0}^{n_{g-1}} \frac{n_{g-1}!}{x!\,(n_{g-1}-x)!}\, p^x (1-p)^{n_{g-1}-x} \left[1-\left(1-\frac{x}{n_{g-1}}\right)^{n_g}\right].$$

where a fraction $2p$ of $BC_{g-2}F_1$ plants were heterozygous and n_{g-1} and n_g plants were used in the $BC_{g-1}F_1$ and BC_gF_1, respectively. To have available most of the genetic variability when commencing selection, the breeder should cross at least eight BC_1F_1 plants and double the number with each succeeding generation (Table 1, left half). During subsequent selection, large population sizes must be maintained because many donor alleles will be at low frequency.

References

Cox TS (1984) Expectations of means and genetic variances in backcross populations. Theor Appl Genet 68:35–41

Dudley JW (1982) Theory for transfer of alleles. Crop Sci 22:631–637

Kenworthy WJ, Brim CA (1979) Recurrent selection in soybeans. 1. Seed yield. Crop Sci 19:315–318

Lawrence PK, Frey KJ (1975) Backcross variability for grain yield in oat species crosses (*Avena sativa* L. x *A. sterilis* L.). Euphytica 24:77–85

Rodgers DM (1982) Improvement of cultivated barley (*Hordeum vulgare*) with germplasm introgressed from *H. spontaneum*. PhD Dissertation. Iowa State University, Ames Iowa. University Microfilm No 82-21223