Theor Appl Genet (1984) 68:35–41

Expectations of means and genetic variances in backcross populations*

T. S. Cox
Research Geneticist, USDA-ARS, Room 354, Throckmorton Hall, Kansas State University, Manhattan, KS 66506, USA

Received September 12, 1983
Communicated by A. L. Kahler

Summary. The genetic variance among F_2-derived lines of backcrosses (BC_gF_2-derived lines) depends on the backcross generation (g), the number of F_1 plants crossed and selfed in generations 1 through g, and the number of BC_gF_2-derived lines evaluated. Additive genetic variance decreases linearly with backcrossing when one BCF_1 plant per generation is crossed and selfed. The relationship is curvilinear if more than one BCF_1 plant is used; as the number of BCF_1 plants increases, additive genetic variance among BC_1F_2-derived lines approaches that among BC_0F_2-derived lines. The effect of population size on genetic variance is due both to fixation of alleles in previous generations and to sampling of genotypes in the population being evaluated. Dominance and repulsion linkage can cause small increases in genetic variance from BC_0 to BC_1.

Key words: Introgression – Population size – Sampling – Backcrossing – Linkage

Introduction

Utilization of exotic germplasm often involves backcrossing to an adapted parent before starting selection in order to restore agronomic suitability (MacKey 1963). Several recent theoretical (Bailey 1977; Baker 1976), computer-simulation (Dudley 1982; Reddy and Comstock 1976), and empirical (Lawrence and Frey 1975; Rodgers 1982) studies have investigated the distribution of progeny in populations derived from backcrosses and the implications for plant breeding for quantitative traits.

To improve a quantitative trait when one parent is inferior to the other, a plant breeder must make enough backcrosses to increase the mean sufficiently but not so many that genetic variance is reduced below the level required for successful selection. In this paper, the theoretical means and genetic variances of populations of random F_2-derived lines from successive backcrosses are expressed in terms of the number of backcrosses, the number of F_1 plants crossed and selfed in each backcross generation, and the number of F_2-derived lines evaluated per F_1 family.

Population structure

The types of populations considered in this paper have been utilized experimentally by Lawrence and Frey (1975), Rodgers (1982), and Cox and Frey (198).

Populations are developed as follows. A cross is made between recurrent and donor homozygous parents, the F_1 is backcrossed to produce BC_1F_1 and also selfed to produce the F_2 (herein called BC_0F_2) generation (Fig. 1). Individual BC_0F_2 plants are self-pollinated to produce BC_0F_2-derived lines. Pollen from a number (n_1) or BC_1F_1 plants is bulked and used to make the second backcross, producing the BC_2F_1; the same n_1 BC_1F_1 plants are selfed to produce the BC_1F_2. BC_1F_2 seed is bulked and planted, and BC_1F_2-derived lines produced as in the BC_0. The process is repeated through any number of backcrosses (Fig. 1).

BC_gF_2-derived lines may be propagated by self-pollination through any number of generations with bulking of seed of selfed plants within lines. Thus, for example, we refer to BC_2F_2-derived lines in the F_3, BC_2F_2-derived lines in the F_4, etc.

---

* Joint contribution of USDA-ARS and Journal Paper No. J-11095 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa. Project No. 2471

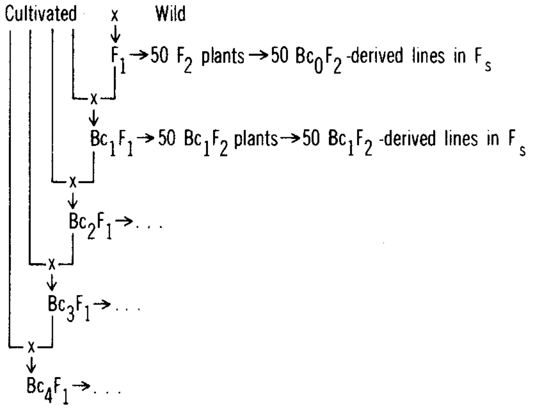

Fig. 1. Development of backcross populations

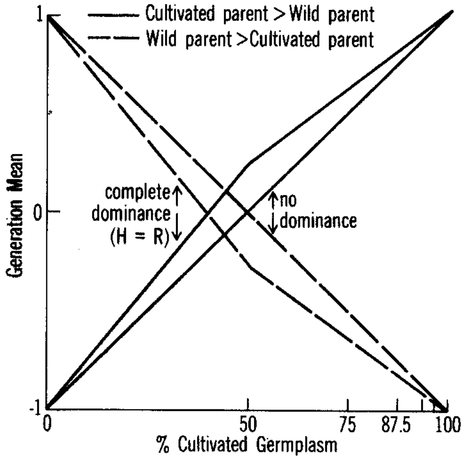

Fig. 2. Relation between generation means and percentages of cultivated germplasm with and without dominance (for BC_gF_2-derived lines in the F_3 from a wild × cultivated cross with backcrossing to the cultivated parent)

Means of backcross populations

The mean of the BC_gF_s generation in the absence of epistasis is

$$\mu_g = \{1 - (\tfrac{1}{2})^g\} R + (\tfrac{1}{2})^g (\tfrac{1}{2})^{s-1} H \tag{1}$$

where R is the mean of the recurrent parent and H the mean of the F_1. Depending on which parent has the higher value, the progeny mean will increase or decrease linearly with the percentage of cultivated germplasm in the population (Fig. 2).

Genetic variances of backcross populations

The genetic variance among BC_gF_2-derived lines was obtained for two cases: (1) an additive model with dominance and (2) a two-locus model with dominance, repulsion linkage, and no epistasis.

Additive model with dominance

The genetic variance among BC_gF_2-derived lines may be divided into two components: (1) variance among BC_gF_1 families and (2) mean variance among BC_gF_2-derived lines within BC_gF_1 families. Symbolically,

$$V_g = V_g(F_1) + \bar{V}_g(F_2/F_1). \tag{2}$$

The genetic variance among F_1 families is zero (i.e., $V_0(F_1) = 0$). For $g > 0$, $V_g(F_1)$ depends on the number of F_1 plants selfed to produce generation g and on the number of F_1 plants backcrossed in generations 1 through $(g-1)$ but is independent of the number of BC_gF_2-derived lines evaluated. $V_g(F_1)$ may be derived as follows.

Let n_g be the number of BC_gF_1 plants selfed to produce BC_gF_2-derived lines, and let k be the number of loci segregating.

Consider locus i at which x_g out of n_g BC_gF_1 plants are heterozygous. Let families derived from BC_gF_1 plants of the following genotypes have the indicated means:

BC_gF_1 genotype	Frequency	Family mean (BC_gF_2-derived lines in F_s)
B_iB_i	$1 - \frac{x_g}{n_g}$	a_i
B_ib_i	$\frac{x_g}{n_g}$	$(\frac{1}{2})^{s-1} d_i$

The variance among families, in the case of no dominance ($d_i = 0$), is

$$(V_g(F_1)_i | x_g) = \frac{x_g}{n_g}\left(1 - \frac{x_g}{n_g}\right) a_i^2. \tag{3}$$

The probability that x_g BC_gF_1 plants are heterozygous at locus i is given by the binomial formula:

$$P(x_g) = \frac{n_g!}{x_g!\,(n_g - x_g)!}\left(\frac{x_{g-1}}{2n_{g-1}}\right)^{x_g}\left(1 - \frac{x_{g-1}}{2n_{g-1}}\right)^{n_g - x_g}$$

where x_{g-1} plants were heterozygous in the $BC_{g-1}F_1$. The variance due to locus i is

$$\begin{aligned} V_g(F_1)_i &= \sum_{x_g=1}^{n_g} (P(x_g))(V_g(F_1)_i | x_g) \\ &= \sum_{x_g=1}^{n_g} \frac{n_g!}{x_g!\,(n_g - x_g)!}\left(\frac{x_{g-1}}{2n_{g-1}}\right)^{x_g} \\ &\quad \times \left(1 - \frac{x_{g-1}}{2n_{g-1}}\right)^{n_g - x_g}\left(\frac{x_g}{n_g} - \frac{x_g^2}{n_g^2}\right) a_i^2 \end{aligned} \tag{4}$$

which can be reduced to

$$V_g(F_1)_i = \left(1-\frac{1}{n_g}\right)\left(\frac{x_{g-1}}{2n_{g-1}}\right)\left(1-\frac{x_{g-1}}{2n_{g-1}}\right)a_i^2. \tag{5}$$

For generation g + 1,

$$V_{g+1}(F_1)_i = \sum_{x_g=1}^{n_g}\frac{n_g!}{x_g!\,(n_g-x_g)!}\left(\frac{x_{g-1}}{2n_{g-1}}\right)^{x_g} \times\left(1-\frac{x_{g-1}}{2n_{g-1}}\right)^{n_g-x_g}\left(1-\frac{1}{n_{g+1}}\right)\left(\frac{x_g}{2n_g}\right)\left(1-\frac{x_g}{2n_g}\right)a_i^2 \tag{6}$$

which can be reduced, in a similar way, to

$$V_{g+1}(F_1)_i = \left[\left(1-\frac{1}{2n_g}\right)\left(1-\frac{1}{n_{g+1}}\right)\left(\frac{x_{g-1}}{4n_{g-1}}\right) - \left(1-\frac{1}{n_g}\right)\left(1-\frac{1}{n_{g+1}}\right)\left(\frac{x_{g-1}}{4n_{g-1}}\right)^2\right]a_i^2. \tag{7}$$

Repeating the process for generation g + 2, summing over loci (assuming absence of linkage and epistasis), and then letting $g=1$ $\left(\text{so that } \frac{x_{g-1}}{n_{g-1}} = \frac{x_0}{n_0} = 1\right)$, we have

$$V_1(F_1) = \left[\left(1-\frac{1}{n_1}\right)\left(\frac{1}{2}\right) - \left(1-\frac{1}{n_1}\right)\left(\frac{1}{4}\right)\right]\sigma_A^2,$$

$$V_2(F_1) = \left[\left(1-\frac{1}{2n_1}\right)\left(1-\frac{1}{n_2}\right)\left(\frac{1}{4}\right) - \left(1-\frac{1}{n_1}\right)\left(1-\frac{1}{n_2}\right)\left(\frac{1}{16}\right)\right]\sigma_A^2,$$

$$V_3(F_1) = \left[\left(1-\frac{1}{4n_1}\right)\left(1-\frac{1}{2n_2}\right)\left(1-\frac{1}{n_3}\right)\left(\frac{1}{8}\right) - \left(1-\frac{1}{n_1}\right)\left(1-\frac{1}{n_2}\right)\left(1-\frac{1}{n_3}\right)\left(\frac{1}{64}\right)\right]\sigma_A^2, \quad \text{etc.} \tag{8}$$

where σ_A^2 is the variance among BC_0F_∞-derived lines (total potential additive variance).

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

$$V_g(F_1) = \left[\prod_{k=0}^{g-1}\left(1-\frac{1}{2^k n}\right) - \left(1-\frac{1}{n}\right)^g\left(\frac{1}{2}\right)^g\right]\left(\frac{1}{2}\right)^g\sigma_A^2, \quad (g>0).$$

$$V_0(F_1) = 0. \tag{9}$$

In the case of $d_i \neq 0$, the coefficients of d_i^2 and $a_i d_i$ may be calculated in the same way, so that

$$V_g(F_1) = \left[\prod_{k=0}^{g-1}\left(1-\frac{1}{2^k n}\right) - \left(1-\frac{1}{n}\right)^g\right]\left[\left(\frac{1}{2}\right)^g\right] \times\left[\sigma_A^2 + \left(\frac{1}{2}\right)^{2(s-1)}\sigma_D^2 - \left(\frac{1}{2}\right)^{s-2}\sigma_{AD}\right]. \tag{10}$$

where $\sigma_D^2 = \sum_i d_i^2$ and $\sigma_{AD} = \sum_i a_i d_i$.

$\bar{V}_g(F_2/F_1)$, i.e., the mean variance within families, is independent of the number of F_1 plants selfed in any backcross generation, but does depend on the total number of F_2 plants selfed in generation g. Consider the class of segregating F_1 families comprising a total of v F_2 plants having the following genotypic frequencies and means:

F_2 plant genotype	Frequency	Genotypic mean of F_2-derived line in F_s
$B_i B_i$	w/v	a_i
$B_i b_i$	y/v	$(\frac{1}{2})^{s-2} d_i$
$b_i b_i$	z/v	$-a_i$

Then, if $d_i = 0$, the genetic variance within segregating families is

$$\left(\frac{w}{v}\right)a_i^2 + \left(\frac{z}{v}\right)(-a_i)^2 - \left(\frac{w-z}{v}\right)^2 a_i^2 \tag{11}$$

where $w + y + z = v$.

The variables w, y, and z follow the trinomial distribution:

$$P(x, y, z) = \frac{v!}{w!\,y!\,z!}\left(\frac{1}{4}\right)^w\left(\frac{1}{2}\right)^y\left(\frac{1}{4}\right)^z,$$

and the expected variance within classes of F_1 families is

$$\bar{V}_g(F_2/F_1)_i = P(BC_g F_1 = Bb) \times\sum\left[P(w, y, z)\,|\,BC_g F_1 = Bb\right]\left[\frac{w+z}{v} - \left(\frac{w-z}{v}\right)^2\right]a_i^2$$

$$= \left(\frac{1}{2}\right)^g \sum_{w=1}^{v}\sum_{y=1}^{v-w}\frac{v!}{w!\,y!\,z!}\left(\frac{1}{4}\right)^w\left(\frac{1}{2}\right)^y\left(\frac{1}{4}\right)^z \times\left[\frac{w+z}{v} - \left(\frac{w-z}{v}\right)^2\right]a_i^2 \tag{12}$$

$$= \left(1-\frac{1}{v}\right)\left(\frac{1}{2}\right)^{g+1}a_i^2.$$

If an average of m F_2 plants per F_1 family are selfed, then $v = (\frac{1}{2})^g m n_g$ and

$$\bar{V}_g(F_2/F_1)_i = \left[\left(\frac{1}{2}\right)^{g+1} - \frac{1}{2\,m n_g}\right]a_i^2, \quad m n_g > 2^g. \tag{13}$$

The coefficient of d_i^2 may be calculated in an analogous fashion, so that

$$\bar{V}(F_2/F_1) = \left[\left(\frac{1}{2}\right)^{g+1} - \frac{1}{2\,m n_g}\right]\left[\sigma_A^2 + \left(\frac{1}{2}\right)^{2s-3}\sigma_D^2\right]. \tag{14}$$

Since $V_g = V_g(F_1) + \bar{V}_g(F_2/F_1)$,

$$V_1 = \left[\left(1-\frac{1}{n_1}\right)\left(\frac{1}{4}\right)\right]\left[\sigma_A^2 + \left(\frac{1}{2}\right)^{2s-2}\sigma_D^2 - \left(\frac{1}{2}\right)^{s-2}\sigma_{AD}\right] + \left[\frac{1}{4} - \frac{1}{2m_1 n_1}\right]\left[\sigma_A^2 + \left(\frac{1}{2}\right)^{2s-3}\sigma_D^2\right],$$

$$V_2 = \left[\left(1-\frac{1}{2n_1}\right)\left(1-\frac{1}{n_2}\right)\left(\frac{1}{4}\right) - \left(1-\frac{1}{n_1}\right)\left(1-\frac{1}{n_2}\right)\left(\frac{1}{16}\right)\right] \times \left[\sigma_A^2 + \left(\frac{1}{2}\right)^{2s-2}\sigma_D^2 - \left(\frac{1}{2}\right)^{s-2}\sigma_{AD}\right] + \left[\frac{1}{8} - \frac{1}{2m_2 n_2}\right]\left[\sigma_A^2 + \left(\frac{1}{2}\right)^{2s-3}\sigma_D^2\right],$$

and

$$V_3 = \left[\left(1-\frac{1}{4n_1}\right)\left(1-\frac{1}{2n_2}\right)\left(1-\frac{1}{n_3}\right)\left(\frac{1}{8}\right) - \left(1-\frac{1}{n_1}\right)\left(1-\frac{1}{n_2}\right)\left(1-\frac{1}{n_3}\right)\left(\frac{1}{64}\right)\right] \times \left[\sigma_A^2 + \left(\frac{1}{2}\right)^{2s-2}\sigma_D^2 - \left(\frac{1}{2}\right)^{s-2}\sigma_{AD}\right] + \left[\frac{1}{16} - \frac{1}{2m_3 n_3}\right]\left[\sigma_A^2 + \left(\frac{1}{2}\right)^{2s-3}\sigma_D^2\right], \quad m_g n_g > 2^g. \tag{15}$$

All formulas apply to BC_gF_2 plants when $s = 2$.

In general, if n BCF_1 plants are crossed and selfed in each generation 1 to g and a total of mn BC_gF_2 plants per generation are selfed,

$$V_g = V_g(F_1) + \bar{V}_g(F_2/F_1) = \left[\prod_{k=0}^{g-1}\left(1-\frac{1}{2^k n}\right) - \left(1-\frac{1}{n}\right)^g\left(\frac{1}{2}\right)^g\right]\left[\left(\frac{1}{2}\right)^g\right] \times \left[\sigma_A^2 + \left(\frac{1}{2}\right)^{2s-2}\sigma_D^2 - \left(\frac{1}{2}\right)^{s-2}\sigma_{AD}\right] + \left[\left(\frac{1}{2}\right)^{g+1} - \frac{1}{2mn}\right]\left[\sigma_A^2 + \left(\frac{1}{2}\right)^{2s-3}\sigma_D^2\right], \quad mn > 2^g \tag{16}$$

and

$$\lim_{\substack{n\to\infty \\ m\to\infty}} V_g = \left[\frac{3}{2} - \left(\frac{1}{2}\right)^g\right]\left(\frac{1}{2}\right)^g\sigma_A^2 + \left[2 - \left(\frac{1}{2}\right)^g\right] \times \left(\frac{1}{2}\right)^{g+2s-2}\sigma_D^2 - \left[1 - \left(\frac{1}{2}\right)^g\right]\left(\frac{1}{2}\right)^{g+s-2}\sigma_{AD}. \tag{17}$$

V_g closely approaches this limit for $mn \geqq n \geqq 12$.

The sign of σ_{AD} depends on whether the products $a_i d_i$ for the separate loci are predominantly positive or negative. The sign of $a_i d_i$ will be according to the following chart:

Effect of allele from cultivated parent	Dominance of cultivated allele		
	Dominant	No dominance	Recessive
Positive	+	0	–
Negative	–	0	+

Thus, for example, if most alleles from the recurrent parent are positive and dominant, σ_{AD} is positive and V_g is reduced. To the extent that alleles behave additively or that dominant and positive alleles are dispersed between the parents, the term involving σ_{AD} is reduced.

For BC_gF_2-derived lines in the F_4 or later generation, σ_D^2 and σ_{AD} are relatively small, so that if mn is large,

$$V_g \cong \left[\prod_{k=0}^{g-1}\left(1-\frac{1}{2^k n}\right) - \left(1-\frac{1}{n}\right)^g\left(\frac{1}{2}\right)^g + \frac{1}{2}\right]\left[\left(\frac{1}{2}\right)^g\right]\sigma_A^2 \tag{18}$$

and

$$\lim_{\substack{n\to\infty \\ m\to\infty}} V_g \cong \left[\frac{3}{2} - \left(\frac{1}{2}\right)^g\right]\left(\frac{1}{2}\right)^g\sigma_A^2.$$

The relationship between V_g and g for different combinations of n and m is shown in Fig. 3.

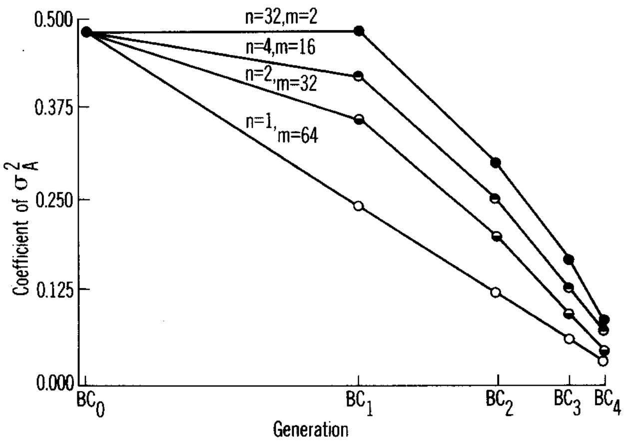

Fig. 3. Change in genetic variance with backcrossing (in absence of dominance) for four combinations of number of F_1 families (*n*) and number of lines per family (*m*). The total number of lines is constant ($mn = 64$)

Repulsion linkage

The breakage of possible repulsion linkages lessens the rate of decrease in additive variance, compared with the example of independent assortment. Consider two loci linked in repulsion with recombination frequency r. Given that a large number of BC_1F_1 and BC_2F_1 plants are crossed and selfed ($n \rightarrow \infty$) and that large populations of BC_gF_2-derived lines are evaluated ($m \rightarrow \infty$), the 16 possible genotypes of BC_gF_2-derived lines will occur in the frequencies given in Table 1. Genetic variances computed from Table 1 are:

Generation	V_g
BC_0	$r a^2 + \left(\frac{1}{4}\right)^{s-2} \left(r^2 - r + \frac{1}{2}\right) d^2$
BC_1	$\left(-\frac{1}{2} r^2 + \frac{5}{4} r\right) a^2$
	$+ \left(\frac{1}{4}\right)^{s-2} \left(-\frac{1}{2} r^3 + r^2 - \frac{3}{4} r + \frac{3}{8}\right) d^2$
BC_2	$\left(\frac{1}{4} r^3 - \frac{7}{8} r^2 + r\right) a^2$
	$+ \left(\frac{1}{4}\right)^{s-2} \left(\frac{1}{4} r^4 - \frac{3}{4} r^3 + \frac{7}{8} r^2 - \frac{1}{2} r + \frac{7}{32}\right) d^2$

For $r = \frac{1}{2}$, these expressions give the same results as formula 17, because σ_{AD} is 0.

The relationship of V_g to r for $d = 0$ is depicted in Fig. 4. $V_1 > V_0 > V_2$ for any value of r except $\frac{1}{2}$ and 0.

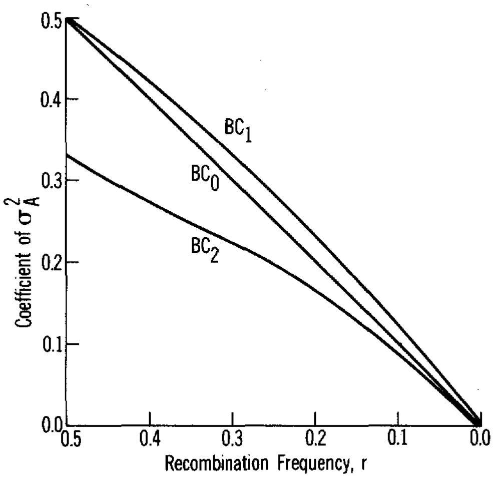

Fig. 4. Relationship between genetic variance and recombination frequency for two loci linked in repulsion

Genetic variance is increased with backcrossing because of the additional opportunity for recombination at meiosis in the BC_1F_1 plants. However, for this recombination to cause a large relative increase in variance from the BC_0F_2 to the BC_1F_2, r must be very small, causing drastic reductions in the absolute variance (Fig. 4). The dominance component of variance would be reduced by recombination, causing a decrease in genetic variance from BC_0F_2 to BC_1F_2 for $d > 0$.

Discussion

Wright (1931), by considering the binomial sampling of 2N gametes, showed that a fraction, 1/2N, of all segregating loci will become fixed with every generation of random mating (with no selection, mutation, or migration) within a population of N individuals. With no dominance, the average allelic frequencies, and thus the expected genotypic mean, will not change, but the variance of the allelic frequency will. Accordingly, the genetic variance after t generations will be

$$V_t = \left(1 - \frac{1}{2N}\right)^t V_0 \tag{19}$$

where V_0 is the initial genetic variance (Bulmer 1980).

When backcrossing is studied, binomial and trinomial sampling of genotypes, and not gametes, is more appropriately used because mating is not random and mean allelic frequencies change from generation to generation. Thus, the present study utilized sampling of F_1 or F_2 plants. The asymmetric nature of the backcross populations considered herein resulted in formulas for genetic variance that are more complicated than formula 19. The formulas may be compared by rearranging, as an example, formula 15 for the BC_3:

$$V_3 = \left(1 - \frac{1}{n_3}\right) \left[\left(1 - \frac{1}{4n_1}\right)\left(1 - \frac{1}{2n_2}\right)\left(\frac{1}{2}\right)^3 - \left(1 - \frac{1}{n_1}\right)\left(1 - \frac{1}{n_2}\right)\left(\frac{1}{2}\right)^6\right] \sigma_A^2 + \left(1 - \frac{1}{(\frac{1}{2})^3 m_3 n_3}\right)\left(\frac{1}{2}\right)^4 \sigma_A^2 .$$

Terms involving n_1 and n_2 signify reduction in variance due to fixation of alleles in previous generations. They are analogous to the term $(1 - 1/2N)$ in formula 19, except that the effect of F_1 plant number in earlier generations (grandparental, great-grandparental,

Table 1. F_2 plant genotype, value of the line in F_s derived from that genotype, and genotypic frequencies in the BC_0, BC_1, and BC_2 in the case of repulsion linkage between loci A and B with recombination frequency r

F_2 plant genotype	F_2-derived line value	Genotypic frequency		
		BC_0	BC_1	BC_2
AABB	a	$\frac{1}{4}r^2$	$-\frac{1}{8}r^3+\frac{1}{8}r^2+\frac{1}{8}r$	$\frac{1}{16}r^4-\frac{1}{8}r^3+\frac{1}{8}r$
AABb	$\frac{1}{2}a+\left(\frac{1}{2}\right)^{s-1}d$	$-\frac{1}{2}r^2+\frac{1}{2}r$	$\frac{1}{4}r^3-\frac{1}{2}r^2+\frac{1}{2}r$	$-\frac{1}{8}r^4+\frac{3}{8}r^3-\frac{3}{8}r^2+\frac{3}{8}r$
AAbb[a]	0	$\frac{1}{4}r^2-\frac{1}{2}r+\frac{1}{4}$	$-\frac{1}{8}r^3+\frac{3}{8}r^2-\frac{5}{8}r+\frac{5}{8}$	$\frac{1}{16}r^4-\frac{1}{4}r^3+\frac{1}{8}r^2+\frac{3}{4}r+\frac{7}{16}$
AaBB	$\frac{1}{2}a+\left(\frac{1}{2}\right)^{s-1}d$	$-\frac{1}{2}r^2+\frac{1}{2}r$	$\frac{1}{4}r^3-\frac{1}{2}r^2+\frac{1}{4}r$	$-\frac{1}{8}r^4-\frac{3}{8}r^3-\frac{3}{8}r^2+\frac{1}{8}r$
AaBb	$\left(\frac{1}{2}\right)^{s-2}d$	$r^2-r+\frac{1}{2}$	$-\frac{1}{2}r^3+r^2-\frac{3}{4}r+\frac{1}{4}$	$\frac{1}{4}r^4-\frac{3}{4}r^3+\frac{7}{8}r^2-\frac{1}{2}r+\frac{1}{8}$
Aabb	$-\frac{1}{2}a+\left(\frac{1}{2}\right)^{s-1}d$	$-\frac{1}{2}r^2+\frac{1}{2}r$	$\frac{1}{4}r^3-\frac{1}{2}r^2+\frac{1}{2}r$	$-\frac{1}{8}r^4+\frac{3}{8}r^3-\frac{3}{8}r^2+\frac{3}{8}r$
aaBB[a]	0	$\frac{1}{4}r^2-\frac{1}{2}r+\frac{1}{4}$	$-\frac{1}{8}r^3+\frac{3}{4}r^2-\frac{1}{4}r+\frac{1}{8}$	$\frac{1}{16}r^4-\frac{1}{4}r^3+\frac{3}{8}r^2-\frac{1}{4}r+\frac{1}{16}$
aaBb	$-\frac{1}{2}a+\left(\frac{1}{2}\right)^{s-1}d$	$-\frac{1}{2}r^2+\frac{1}{2}r$	$\frac{1}{4}r^3-\frac{1}{2}r^2+\frac{1}{4}r$	$-\frac{1}{8}r^4+\frac{3}{8}r^3-\frac{3}{8}r^2+\frac{1}{8}r$
aabb	$-a$	$\frac{1}{4}r^2$	$-\frac{1}{8}r^3+\frac{1}{8}r^2+\frac{1}{8}r$	$\frac{1}{16}r^4-\frac{1}{8}r^3+\frac{1}{8}r$

[a] Parental genotype

etc.) is succesively reduced, since each BCF_1 generation contributes only $\frac{1}{2}$ of the genes of the following generation. The term $(1-1/n_3)$ signifies reduction of variance due to a finite number of BCF_1 plants in the present generation. It has nothing to do with fixation of alleles but is due to random deviations from the expected ratio of heterozygous to homozygous F_1 plants. The term involving $m_3 n_3$ likewise is due to deviations from a 1AA:2Aa:1aa ratio among F_2 plants from heterozygous F_1s.

Taking into account sampling of genotypes in generation t would change formula 19 to

$$V_t=\left(1-\frac{1}{2N}\right)^t\left(1-\frac{1}{N}\right)V_0 .$$

In evolutionary situations which involve many generations the effect of population size in generation t on V_t is small, compared with the cumulative effect of small population size in generations 1 through $t-1$. Also, the effect on V_t of the number of progeny evaluated in generation t is small in most quantitative genetic studies, since a large population generally is evaluated (though it may be descended from a small number of individuals). With backcrosses, however, a small number of BC_gF_1 plants may be selfed to produce the BC_gF_2, with a considerable reduction in variance. Variances among BC_0F_2- and BC_1F_2-derived lines are equal for infinite n. Figure 3 indicates that this upper limit of V_g is closely approached when $n=32$ and $m=2$. Indeed, increasing n beyond 12 produces little change in V_g. For any value of mn, the maximum V_g is obtained when $m=1$.

Hill (1966) gave the expected additive variance among BC_1F_1 families as $\frac{1}{4}D$ (equal to $\frac{1}{4}\sigma_A^2$ in the present terminology) and that among F_2 plants within BC_1F_1's as $\frac{1}{4}D$. This gives a total variance of the BC_1F_2 of $\frac{1}{2}D$ ($\frac{1}{2}\sigma_A^2$), equal to V_1 in formula 15 as m_1 and n_1 become very large.

Acknowledgement. Dr. Dan M. Rodgers, in discussions with the author, first suggested considering the effect of BCF_1 plant numbers on genetic variance. The author wishes to thank the World Food Institute at Iowa State University for the support it provided through a graduate assistantship.

References

Bailey TB (1977) Selection limits in self-fertilizing populations following the cross of homozygous lines. In: Pollack E, Kempthorne O, Bailey TB (eds) Proc Int Conf Quant Genet. The Iowa State University Press, Ames, Iowa, pp 399–412

Baker RJ (1976) Some thoughts about backcrossing and quantitative traits. Wheat Newslett 22:43–44

Bulmer MG (1980) The mathematical theory of quantitative genetics. Clarendon Press, Oxford

Cox TS, Frey KJ (1984) Genetic variation for grain yield and related traits in sorghum introgression populations. Theor Appl Genet 68:145–153

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

Hill J (1966) Recurrent backcrossing in the study of quantitative inheritance. Heredity 21:85–120

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

MacKey J (1963) Autogamous plant breeding based on already high-bred material. In: Akerberg E, Hagberg A (eds) Recent plant breeding research, Svalof 1946–1961. John Wiley, New York, pp 73–88

Reddy BVS, Comstock RE (1976) Simulation of the backcross breeding method. 1. The effect of heritability and gene number on fixation of desired alleles. Crop Sci 16:825–830

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

Wright S (1931) Evolution in Mendelian populations. Genetics 16:97–159