Theor Appl Genet (1989) 78:365–368

Additive genetic variance within populations derived by single-seed descent and pod-bulk descent *

W. C. Wells [1] **and G. C. Weiser** [2]

[1] Hawaiian Sugar Planters' Association. P.O. Box 1057, Aiea, HI 96701-1057, USA
[2] Clemson University, Edisto Research and Education Center, Blackville, SC 29817, USA

Received September 15, 1988; Accepted March 17, 1989
Communicated by P. L. Pfahler

Summary. Breeders of self-pollinated legumes commonly use single-seed descent (SSD) or pod-bulk descent (PBD) to produce segregating populations of highly inbred individuals. We presented equations for the expected value of the additive genetic variance within populations derived by SSD $(E(V_A)_{SSD})$ and PBD $(E(V_A)_{PBD})$ in terms of the initial population size (N_0), the number of seed harvested per pod (M), the probability of survival of an individual (θ), and the generation at which the population is evaluated (S_t). Differences between $(E(V_A)_{SSD})$ and $(E(V_A)_{PBD})$ are due to differences in the expected amount of random drift which occurs with the two methods after the S_0 generation. With both methods, random drift occurs when progeny are sampled from heterozygous parents. An additional component of random drift occurs when sampled progeny fail to survive during SSD, or when sampling occurs amoung families during PBD. For values of N_0, M, θ, and S_t that are typical of soybean (*Glycine max* (L.) Merr.) breeding programs, $(E(V_A)_{SSD})$ will be greater than $(E(V_A)_{PBD})$. The ratio of $(E(V_A)_{SSD})$ to $(E(V_A)_{PBD})$ will: (1) increase as M and θ increase; (2) approach a value of 1.00 as N_0 increases; and (3) be a curvilinear function of S_t. Plant breeders should compare SSD and PBD based upon values of $(E(V_A)_{SSD})$ and $(E(V_A)_{PBD})$ and the expected cost of carrying out the two methods.

Key words: Soybeans – Breeding methods – Bulk populations – Random drift – Quantitative genetics

Introduction

Many breeders of self-pollinated crop species feel that the response to early generation selection for quantitative traits is too small and too erratic to justify the expense of carrying out this strategy. The alternative strategy of delaying selection until a high degree of homozygosity is attained is now popular. Single-seed descent (SSD), as described by Brim (1966), and pod-bulk descent (PBD) as described by Fehr (1987) are two methods commonly used by breeders of self-pollinated legumes to produce populations of nearly homozygous individuals. The relative efficiency of these two methods will depend upon the cost of each method and on the amount of additive genetic variance retained by each method. The objectives of this paper are: (1) to present equations for predicting the expected value of the additive genetic variance within populations derived by SSD and PBD; and (2) to discuss the conditions under which each breeding method retains the most genetic variation.

Methods

Assumptions

Equations for the expected value of the additive genetic variance (V_A) within populations derived by single-seed descent (SSD) and by pod-bulk descent (PBD) were derived based upon the following assumptions: (1) diploid inheritance; (2) two alleles per locus; (3) propagation solely by self-fertilization; (4) closed population; (5) no artificial or natural selection; (6) independent assortment; and (7) no mutation. For the sake of simplicity, equations for V_A will be presented for a single locus. The expected value of V_A for a character governed by multiple independent loci will be equal to the sum of the values for the individual loci. It will be assumed that the initial (S_0) generation of the population was produced from the random mating of individuals within the previous generation, or that it was produced by the selfing of F_1 generation plants which were produced from a single cross between two homozygous parents. In the latter case, the terms S_0 and F_2 refer to the same generation.

Breeding methods

For pod-bulk descent, we assume that a single pod containing M seeds is harvested from each of N parents to produce a popula-

---

* Contribution No. 2910 of the South Carolina Agricultural Experiment Station, Clemson University

tion of NM initial progeny. A sample of N' progeny is partitioned from the bulk, N of which survive to become productive progeny. The probability of survival (θ), refers to the probability that a sampled progeny (seed) will produce a pod bearing plant, and is equal to: $\theta = N/N'$. The value of N' is chosen to keep the value of N constant across generations ($N' = N/\theta$). It will be assumed that the value of M is constant across harvested pods.

For single-seed descent, we assume that a single seed will be harvested from each of N_n individuals of generation n to produce a population of N_n initial progeny, N_{n+1} of which will survive to become productive progeny. When the value of θ is less than 1.00, as in typical real populations, the values of N_n will decrease with each succeeding generation, and $N_t = N_0(\theta)^t$ when θ is constant across generations.

In the present paper, equations will be presented for the estimation of the additive genetic variance among the N productive progeny derived by PBD and the N_t progeny derived by SSD.

Expected value of the additive genetic variance

The additive genetic variance for a character measured in the S_t generation is equal to:

$$V_{AS_t} = V_{xS_t} \propto^2 \tag{1}$$

where V_{xS_t} is the variance of x among individuals of the t^{th} generation of the population; x is the number of A_1 alleles present at the A locus of an individual; and $\propto$ is the average effect of a gene substitution (Falconer 1985). For a population of infinite size, the value of V_{xS_t} is equal to:

$$V_{xS_t} = 2pq(1 + F_{S_t}) \tag{2}$$

where F_{S_t} is the coefficient of inbreeding for the S_t generation, p is the frequency of the A_1 allele in the S_0 generation of the population, and $q = 1 - p$ (Falconer 1981). For a finite population of the S_1 or later generation that has been derived by PBD, the expected value of V_{xS_t} is equal to:

$$E[V_{xS_t}] = 2pq[1 + F_{S_t} - (1/N)] - \sum_{n=1}^{t} E[V(\bar{x}_{S_n} - \bar{x}_{S_{n-1}})] \tag{3}$$

where $2pq/N$ is the expected value of the square of the random drift which occurs between the parents of the S_0 population and the $N\,S_0$ individuals; and $E[V(\bar{x}_{S_n} - \bar{x}_{S_{n-1}})]$ is the expected value of the square of the random drift which occurs between the S_n and the S_{n-1} generation. The expected value of the square of the random drift which occurs between the S_0 and the S_t generations is equal to the sum of the variance of the difference between the mean value of x in the progeny generation ($\bar{x}_{S_n}$) and the mean value in the parental generation ($\bar{x}_{S_{n-1}}$), across consecutive pairs of generations from the S_0 to the S_t generation. Summation of the variance of mean gene frequency changes across pairs of generations is valid when there is no correlation between changes in the mean gene frequency between generations. When generations are advanced in the absence of selection, as is assumed to be the case with both SSD and PBD, no correlation in gene frequency changes is expected to occur.

For PBD, the random drift variance between the parental and the progeny generations can be partitioned into two components:

$$V(\bar{x}_{S_n} - \bar{x}_{S_{n-1}}) = V(\bar{x}'_{S_{n-1}} - \bar{x}_{S_{n-1}}) + V(\overline{x_{ij} - x_i}) \tag{4}$$

where $V(\bar{x}'_{S_{n-1}} - \bar{x}_{S_{n-1}})$ is the variance between the simple ($\bar{x}_{S_{n-1}}$) and the weighted ($\bar{x}'_{S_{n-1}}$) mean value of x within the parental generation, where the weight of the i^{th} parent is equal to the number of productive progeny that it produces (k_i); and $V(\overline{x_{ij} - x_i})$ is the variance of the mean difference between the value of x for the i^{th} parent and the value of x for the j^{th} progeny of that parent. The first component ($V(\bar{x}'_{S_{n-1}} - \bar{x}_{S_{n-1}})$) is the result of uneven contribution of the parents to the population of productive progeny. The second component ($V(\overline{x_{ij} - x_i})$) is the result of sampling progeny genotypes derived from a heterozygous parental genotype. The difference between $\bar{x}'_{S_{n-1}}$ and $\bar{x}_{S_{n-1}}$ is equal to:

$$\bar{x}'_{S_{n-1}} - \bar{x}_{S_{n-1}} = \sum_{i=1}^{N} (k_i x_i - x_i)/N \tag{5}$$

and the expected value of the square of this difference is equal to:

$$E[V(\bar{x}'_{S_{n-1}} - \bar{x}_{S_{n-1}})] = E[V_k]\, E[V_{xS_{n-1}}]/N \tag{6}$$

where the expected value of V_k is equal to:

$$E[V_k] = (1 - 1/N)\,[1 - (N-1)/(MN-1)] \tag{7}$$

when the N productive progeny are a random sample of the NM initial progeny (Demple 1975), as will be the case with PBD. The term $V_{xS_{n-1}}$ is the variance of x among the parents. It is assumed that V_k and $V_{xS_{n-1}}$ are independent.

The expected value of the random drift variance component that is the result of segregation at heterozygous loci is equal to:

$$E[V_{(\overline{x_{ij} - x_i})}] = 2pq(1 - F_{S_n})/N \tag{8}$$

where F_{S_n} is the inbreeding coefficient for the generation that is evaluated and $n \geq 1$. The sum of these components from the S_0 to the S_t generation is equal to:

$$E[V_{(\overline{x_{ij} - x_i})S_t}] = 2pq\, F_{S_t}/N. \tag{9}$$

Combining terms from Eqs. (3), (4), (5), (6) and (9), we can express the expected value of the genetic variance within the S_t generation of a population advanced by PBD as:

$$E[V_{xS_t}] = 2pq[(1 + F_{S_t})(N-1)/N] - \sum_{n=1}^{t} E[V_k]\, E[V_{xS_{n-1}}]/N. \tag{10}$$

With SSD, each of N_t S_0 individuals will contribute one descendant to the S_t generation. The $N_0 - N_t$ S_0 individuals which do not contribute descendants to the S_t generation may be ignored, since they will have no effect on the genetic constitution of the S_t generation. For SSD, therefore, the expected value of V_{xS_t} will be equal to:

$$E[V_{xS_t}] = 2pq(1 + F)(N_t - 1)/N_t \tag{11}$$

which is equivalent to the first term of Eq. (10) for PBD with N_t substituted for N.

Results and discussion

The expected value of the additive genetic variance within populations advanced by SSD and PBD will depend upon the probability of survival (θ), the initial S_0 population size (N_0), and the generation at which individuals are evaluated. The relative amount of genetic variation within populations derived by SSD and PBD will depend upon the amount of genetic variation lost due to the actual decrease in population size which occurs with SSD when $\theta < 1.00$, compared to the loss of genetic variation due to sampling among families within the bulk of initial progeny which occurs with PBD.

When $\theta = 1.00$, as will be the case with some methods of implementing SSD (Fehr 1987), the expected amount of additive genetic variation maintained by SSD will be greater than that maintained by PBD regardless of N_0 or

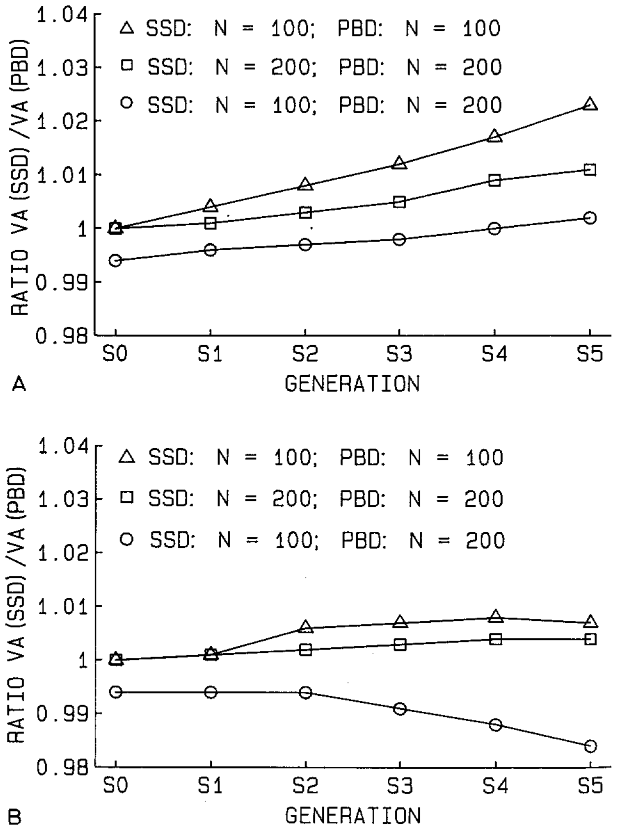

Fig. 1 A and B. Ratio of the expected value of the additive genetic variance (V_A) within a population derived by single-seed descent (SSD) to that within a population derived by pod-bulk descent (PBD) for $M = 3$ and initial S_0 population sizes of $N = 100$ and $N = 200$. **A** Probability of survival = 0.90. **B** Probability of survival = 0.80

the generation at which the population is evaluated. On the other hand, when $\theta < 1.00$, the relationship between the expected relative amount of genetic variation maintained by the two methods and the generation at which the population is evaluated is curvilinear (Fig. 1) and will depend upon the values of θ and N_0. In general, the ratio of genetic variance maintained by SSD to that maintained by PBD will increase in the early generations and decrease in later generations. The greater the value of θ, the greater the relative efficiency of SSD versus PBD. The greater the value of N_0, the less difference there is between the two methods.

When values of M, N_0, and θ that are typical for a soybean (*Glycine max* (L.) Merr.) breeding program are substituted into Eqs. (7), (10) and (11), the expected amount of genetic variation maintained by SSD will be greater than that maintained by PBD in the generations from which lines are typically derived for evaluation (S_2, S_3, S_4, or S_5); in later generations SSD will lose its advantage (Fig. 1). In the early generations, the loss of genetic variance due to the failure of initial progeny to survive is less than that due to the loss of families caused by sampling within the bulk of initial progeny. When the value of θ is at least 0.80, the advantage of SSD actually increases from the S_1 to the S_4 generation. However, at some later generation, the loss of genetic variation caused by the premature death of individuals will be greater than that caused by random sampling among families present in the bulk of initial progeny, and relative amounts of genetic variation within populations maintained by the two methods will converge and actually reverse in ranking. The lower the probability of survival and the smaller the initial population size, the sooner this will occur.

Most breeders choose to select among lines derived from populations produced by SSD or PBD at the S_2, S_3, or S_4 generation. At these generations, SSD has less than a 1% advantage over PBD when $\theta = 0.80$ and $N_0 = 200$ (Fig. 1 B); when $\theta = 0.90$ and $N_0 = 100$, SSD has an advantage of between 1% and 2% over PBD (Fig. 1 A). Surprisingly, when the probability of survival is high, the expected amount of genetic variation within a population derived by SSD from an S_0 population of 100 individuals will be very similar to that within a population derived by PBD from an S_0 population of 200 individuals (Fig. 1 A). On the other hand, when the probability of survival is relatively low, the expected amount of genetic variation within a population derived by PBD from an S_0 population of 200 individuals will be appreciably greater than that within a population derived by SSD from an S_0 population of 100 individuals (Fig. 1 B).

In the present paper, SSD and PBD were compared in the context of advancing populations of soybeans which had three seeds per pod. Pod-bulk descent may also be used to advance populations of species which produce more than three seeds per pod. Another commonly used method for creating populations of highly inbred individuals is bulk-population descent (BPD), in which the entire plant is harvested and all of the resulting seed is placed in the bulk of initial progeny. Equations presented in the present paper may be used for predicting the amount of genetic variation within populations derived by BPD when there is no selection pressure and an equal number of seeds is harvested from each plant. As the number of initial progeny contributed by each parent to the bulk of initial progeny increases, the value of the variance of parental contributions to the population of productive progeny (V_k) will asymptotically increase to a limit of 1.0, and consequently the amount of genetic variance lost by sampling among families will increase to a limit of $V_{xS_{n-1}}/N$ [Eq. (6)]. The relative efficiency of SSD over PBD or BPD for the maintenance of genetic variation will, therefore, increase as the number of seeds per pod or plant increases.

The actual relative efficiencies of SSD and PBD depend upon the relative amount of genetic variation within populations produced by the two methods and the relative cost of carrying out these two methods. In our paper, we have presented equations for the estimation of the amount of additive genetic variance within populations derived by the two methods. Plant breeders can use estimates from these equations with estimates of the relative cost of carrying out these two methods to predict which method would be most efficient for their particular needs.

Acknowledgements. The editorial and statistical assistance of M. L. Roose, University of California, Riverside, and P. M. Burrows, Clemson University, was appreciated by the authors.

References

Brim CA (1966) A modified pedigree method of selection in soybeans. Crop Sci 16:220

Demple L (1975) A note on increasing the limit of selection through selection within families. Genet Res 24:127–135

Falconer DS (1981) Introduction to quantitative genetics. Longman, New York

Falconer DS (1985) A note on Fisher's 'average effect' and 'average excess'. Genet Res 46:337–347

Fehr WR (1987) Breeding methods for cultivar development. In: Wilcox JR (ed) Soybeans: Improvement, production, and uses. 2nd edn. American Society of Agronomy, Madison; pp. 249–293