Theor Appl Genet (1998) 97 : 990–993

M. P. Koshy · G. Namkoong · J. H. Roberds

Genetic variance in the F_2 generation of divergently selected parents

Received: 5 April 1998 / Accepted: 22 April 1998

Abstract Either by selective breeding for population divergence or by using natural population differences, F_2 and advanced generation hybrids can be developed with high variances. We relate the size of the genetic variance to the population divergence based on a forward and backward mutation model at a locus with two alleles with additive gene action. The effects of population size and initial gene frequency are also explored. Larger parental population sizes increase the F_2 genetic variance if the initial probability distribution is uniform or U-shaped. However, population size has the opposite effect if the initial distribution of gene frequencies is skewed such as it would be with newly arriving alleles. These alleles contribute to the genetic variance sooner when the selection pressure is higher or when the effective population size is smaller.

Key words F_2 genetic variance · Divergent selection · Provenance hybrids

Introduction

When isolated populations have been subjected to long-term selection but with divergent selection goals, an inter-population genetic variance is created. The genetic variances in hybrid populations created by inter-crossing and in advanced generations of hybrids are functions of divergence in allele frequencies among the parent populations. In this paper, simple theoretical models for the genetic variance of the F_2 population are investigated and the effects of the magnitude of divergence in selection, of population size, and of the initial probability distribution of gene frequencies in the parental populations are explored.

Theoretical models for F_2 variance in crosses involving divergently selected lines

Slatkin and Lande (1994) analyzed models for the genetic variance in the F_2 generation based on the multivariate normal, house-of-cards, and pleiotropic models put forward earlier. These models consider isolated populations for which drift and mutation processes yield different combinations of genes. For these models, means and genetic variances within the parental populations are equal. Differences in allele frequencies and combinations contribute to the segregation variance in the F_2 generation. The magnitude of the segregation variance expressed in the F_2 generation depends on the type of genetic variation resulting from mutation. If variation is due to numerous alleles of moderate or small effect, the segregation variance increases linearly with the time of separation (Slatkin and Lande 1994). Because stabilizing selection imposes only a single constraint on the mean of the character, most alleles at the underlying loci are free to drift. When variation is maintained by low-frequency alleles, the increase in segregation variance is likely to be small under stabilizing selection.

These models are extended here to include divergent selection between populations. We first consider a simple case with two alleles at a single locus segregating in divergently selected lines.

Communicated by P. M. A. Tigerstedt

M. P. Koshy (✉) · G. Namkoong
Faculty of Forestry, University of British Columbia,
2424 Main Mall, Vancouver, BC, V6 T 1Z4, Canada

J. H. Roberds[1]
USDA Forest Service, Box 8002, Biltmore Hall,
Department of Forestry, North Carolina State University,
Raleigh, NC 27695-8002, USA

Present address:
[1]Southern Institute of Forest Genetics, USDA Forest Service,
Sucier, Mass., USA

Under the single locus, two-allele model, two scenarios are considered. In the first case, it is assumed that alleles in populations from different environments are under different natural selection pressures, where the selection coefficient for an allele changes between populations depending on the environment, and within each population a steady-state distribution of gene frequencies is reached based on a mutation-selection-drift balance. When fitness is not density dependent, Wright (1969) showed that the steady-state distribution of gene frequencies under selection, mutation, and drift is:

$$\mathrm{P}(q) = \mathrm{C}e^{-4\mathrm{N}sq}\, q^{4\mathrm{N}\mu - 1}(1-q)^{4\mathrm{N}\upsilon - 1} \tag{1}$$

$$\mathrm{C} = 1 \Big/ \int_0^1 e^{-4\mathrm{N}sq} q^{4\mathrm{N}\mu - 1}(1-q)^{4\mathrm{N}\upsilon - 1}\,\mathrm{d}q \tag{2}$$

where N is effective population size, s is the selection coefficient, and μ and υ are per locus forward and backward mutation rates.

For the second case, the distribution of gene frequencies in intermediate generations before steady-state is reached is considered. Here, a mutation occurs once at a locus in the population, and divergent selection takes place by selecting in opposing directions between the two different populations. The distribution of gene frequencies in intermediate generations is obtained by numerical methods using transition probability matrices. Let $\mathbf{P}(\mathrm{N}, \mathrm{N}s)$ denote the square matrix with elements P_{jk} defining the Wright-Fisher stochastic process.

$$\mathrm{P}_{\mathrm{jk}} = \binom{2N}{k} (q + \Delta q)^{\mathrm{k}}(1 - q - \Delta q)^{2\mathrm{N}-\mathrm{k}} \tag{3}$$

$(0 \le j,\ k \le 2\mathrm{N})$ (Ewens 1979)

where $q = j/2\mathrm{N}$, and $\Delta q = sq(1-q)$ is the expected change in gene frequency per generation, $s = ia/\sigma$. (i = selection intensity, σ = phenotypic standard deviation) The expected frequencies of genes at generation t with initial frequency $j/(2\mathrm{N})$ can be represented by the vector $\mathbf{v}(t, \mathrm{N}, \mathrm{N}s)$. Then $v_{j(0)} = j/(2\mathrm{N})$, and $\mathbf{v}(t, \mathrm{N}, \mathrm{N}s) = \mathbf{P}(\mathrm{N}, \mathrm{N}s)\, \mathbf{v}(t-1, \mathrm{N}, \mathrm{N}s)$ (Hill and Rashbash 1986) The basic assumptions of the model are as follows:

1) Additive gene action (no dominance)
2) No linkage disequillibrium in F_0, F_1, and F_2
3) Loci are independent
4) No mutation is considered after the F_1 crosses
5) Panmixia

Under constant environmental conditions the gene frequencies of an infinitely large population will maintain a steady-state determined by counteracting but systematic pressures of selection, migration and mutation. This stationary distribution of gene frequencies can be considered in three different ways. First, it may be taken as the relative frequencies of values of q at a particular locus over a large number of generations considering all conditions to be constant. Second, it may be considered as the probability distribution of the frequencies of all loci having the same systematic pressures of the same magnitude in a population. Finally, it can represent the probability distribution of gene frequencies at a particular locus in a number of random isolated populations of the same size (Li 1976).

In the following analyses, we first consider a single locus in a hybrid population derived from a sample of parental populations. We then extend the model to multiple, unlinked, independently acting loci and use the stationary distribution as the probability distribution of the frequencies of all loci having same systematic pressures of the same magnitude in a population.

Based on the steady-state distribution of the gene frequencies, we can obtain the mean and expected genetic variance at a single locus for a population using the formulae:

$$\text{Mean} = a[p - q] \tag{4}$$

$$\text{Genetic variance} = 2pq\mathrm{a}^2 \tag{5}$$

where p and q are allele frequencies, and a is the average effect of gene substitution. From the preceding, the mean per locus can be expressed as

$$E(Y_{\mathrm{i}}) = \int_0^1 (P(q;\, Ns,\, N\mu,\, Nv))\; a[(1 - 2q_{\mathrm{i}})]\,\mathrm{d}q_{\mathrm{i}} \tag{6}$$

Based on Eqs. 1 and 5, the expected per locus genetic variance is,

$$E(V_{\mathrm{i}}) = \int_0^1 (P(q;\, Ns,\, N\mu,\, Nv))\; 2(1 - q_{\mathrm{i}})(q_1)a^2\,\mathrm{d}q_{\mathrm{i}} \tag{7}$$

The mean and variance per locus in the F_2 population generated by inter-crossing the F_1 progeny obtained from the cross between the divergent parental populations is based on the gene frequencies in the parental population. In the F_1, a gene frequency which is equal to the average of the two parental populations can be expected. Then, based on the gene frequencies of the parental populations, the mean and genetic variance at a locus in the F_2 can be seen as

$$\text{Mean } (\mathrm{F}_2) = a[(1 - (q_1 + q_2)/2) - (q_1 + q_2)/2] \tag{8}$$

$$\text{Genetic variance } (\mathrm{F}_2) = (1 - (q_1 + q_2)/2)(q_1 + q_2)a^2 \tag{9}$$

Mean (M_{F_2}) and genetic variance (V_{F_2}) of the population can be defined as follows:

$$\begin{aligned} E(M_{\mathrm{F}_2}) = \int_0^1\!\!\int_0^1 & (P(q_1;\, Ns_1,\, N\mu,\, \mathrm{N}v)) \\ & \times (P(q_2;\, Ns_2,\, N\mu,\, Nv))a[(1 - (q_1 + q_2)/2) \\ & - [(q_1 + q_2)/2]]\,\mathrm{d}q_1\,\mathrm{d}q_2 \end{aligned} \tag{10}$$

$$\begin{aligned} E(V_{\mathrm{F}_2}) = \int_0^1\!\!\int_0^1 & (P(q_1;\, Ns_1,\, N\mu,\, Nv))(P(q_2;\, Ns_2,\, N\mu,\, Nv)) \\ & \times (q_1 + q_2)(1 - (q_1 + q_2)/2)a^2\,\mathrm{d}q_1\mathrm{d}q_2 \end{aligned} \tag{11}$$

In the second scenario, consider that artificial selection is carried out with selection intensity i on a trait with phenotypic standard deviation σ, selection differential $i\sigma$, and effective population size N. Assume that at all loci gene action is additive and that there are only two alleles at a locus. Natural selection is not considered here. Then if the frequency of the allele giving the higher value is q, the number of loci affecting a trait is n, and the probability distribution of gene frequencies is obtained for populations selected divergently for t generations. The expected genetic variance in F_0 can then be written as

$$E(V_{F0}) = n \sum_{j=1}^{m} v_j q_j (1 - q_j) a^2 \tag{12}$$

where $m = 2N - 1$, and v_j is an element in the vector of frequencies for transition matrix.

Similarly the expected variance in the F_2 generation is given by

$$E(V_{F_2}) = n \sum_{j=1}^{m} v_{1j} v_{2j} (q_{1j} + q_{2j})\ (1 - (q_{1j} + q_{2j})/2) a^2 \tag{13}$$

Assuming independence between loci, the expected genetic variance for multiple loci can be obtained by summation.

Numerical examples

Values for the F_2 genetic variance resulting from crosses involving parental populations that have reached a steady-state distribution are presented in Fig. 1. Here N, μ, and υ with values of 500, 0.0001, and 0.00001, respectively, are assumed across populations, while s values differ between parental populations and a is assumed equal to 1 at all loci.

The genetic variance of the F_2 progeny can be seen to increase as parental populations diverge because of selection. At a single locus, selection in each of the parental populations either favoring the same allele or a different allele with consequent reduction or increase in the F_2 additive genetic variance form the neutral (0, 0) case. The symmetry is forced by the equivalence of allele effects in the F_2 regardless of the origin of the alleles between the parental populations. Once the fixation or steady-state point is reached in the two populations, further increase is not possible.

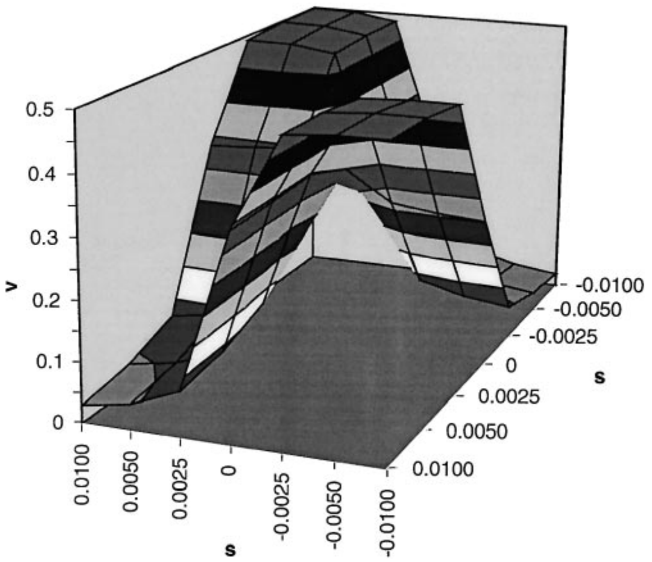

Fig 1. Additive genetic variance in the F_2 generation subsequent to crosses of divergent parental populations plotted against selection levels (s). Segregating locus is comon in the parental populations

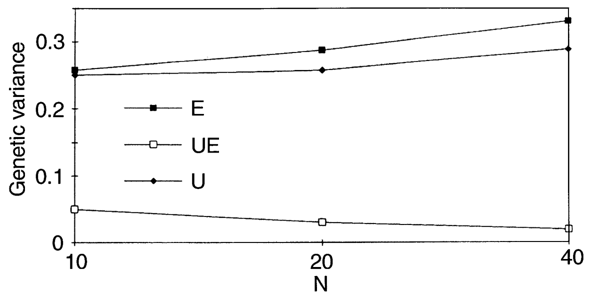

Fig 2. Additive genetic variance in the F_2 generation when divergently selected parental populations with different selection coefficients ($s = 0.01$and -0.01) and initial probability distribution of gene frequencies (E, UE, and U) were crossed. Effective population size 10, 20, and 40

Under scenario two, the magnitude of the F_2 additive genetic variance expressed varies depending on the initial probability distribution of gene frequencies in the parental populations. Uniform (E), skewed (UE), and U-shaped distributions are considered. When the allele is of recent mutation origin, a skewed distribution with higher probability for lower gene frequency values is expected. When a selectively neutral allele prevails in the population for a long period of time, a U shaped distribution can be expected. Assuming that an allele is formed due to mutation in a population, the probability distribution in the next generation can be obtained using binomial probability distribution function. In Fig. 2, the F_2 genetic variance following 60 generations of selection in the parental populations (with $s = 0.01$ and -0.01) before crossing and effective population sizes of 10, 20, and 40 for different distributions of initial gene frequencies is given.

While the uniform and U-shaped initial distribution cases showed increases in the F_2 genetic variance with increases in the effective population size, the skewed distribution case showed a decrease in the F_2 genetic variance with increase in N. This reflects the fact that once a mutation event has occurred, the newly arriving alleles contribute to the genetic variance sooner when either selection pressure is higher or when effective population size is smaller. Compared with larger populations, the initial gene frequency of the newly arriving allele remains higher and leads to a higher F_2 genetic variance.

Conclusion

It is possible to generate greater genetic variances in F_2 populations by crossing divergently selected parents than by crossing parents from replicate populations. Considering parental populations with no selection as the control, progeny from crosses of divergently selected parents gave 46.14%, 48.07%, and 48.34% increases in genetic variance, while crosses within selected lines showed a reduction of 77.57%, 87.27%, and 91.29% for s values 0.0025, 0.005 and 0.01, respectively. Though the increase in genetic variance described is for a trait controlled by a single locus and is a special case, it does provide an indication of what can happen when an increase in multivariate generalized variance is possible, such as for a broader range of trait combinations.

Wright (1922) demonstrated the advantage of subdivided populations with a regular crossing and selection scheme for rapid improvement. Various other studies have shown that when gene action is additive, there is no advantage in selection response for a subdivided population structure (Madalena and Hill 1972; Rathie and Nicholas 1980; Enfield et al. 1986). However, there is considerable advantage for lower effective population size when the allele being favored is of new mutation origin. Therefore, in the later stages of a breeding program when much of the response to selection is derived from new mutations, divergent selection and the formation of hybrid F_2 populations can be advantageous. Additional investigations on the variance in F_2 populations to include non-additive gene actions, gametic migration among populations, and repeated cycles of selection with crossing are required to fully explore the advantages provided by the formation of these populations in selection programs. If natural populations have diverged in different selection demes, the increase in genetic variance in a F_2 population can be substantial.

Acknowledgements This work was supported by a strategic grant from Natural Sciences and Engineering Research Council of Canada.

References

Enfield FD, Ankelsaria F (1986) An evaluation of multiple peak epistasis and population structure in directional selection program. In: Third World Cong Genet Appl Livestock Production. University of Nebraska Institute of Agriculture and Natural Resources, Lincoln, Neb., pp 283–294

Ewens WJ (1979) Mathematical population genetics. Longman, London

Hill WG, Rashbash R (1986) Models of long term artificial selection in finite population. Genet Res 48 : 41–50

Li CC (1976) First course in population genetics. Boxwood, Pacific Grove, Calif.

Madalena FE, Hill WG (1972) Population structure in artificial selection programmes: simulation studies. Genet Res 20 : 75–99

Rathie KA, Nicholas FW (1980) Artificial selection with differing population structures. Genet Res 36 : 117–131

Slatkin M, Lande R (1994) Segregation variance after hybridization of isolated populations. Genet Res 64 : 51–56

Wright S (1922) The effects of inbreeding and crossbreeding on guinea pigs. III. Crosses between highly inbred families. US Dep Agric Bull 1121 : 1–60

Wright S (1969) Evolution and the genetics of populations, vol 2: the theory of gene frequencies. University of Chicago, Chicago