

THE EFFECT OF GENE FREQUENCY DISTRIBUTION IN BASE POPULATIONS ON RESPONSE TO SELECTION IN SYNTHETIC POPULATIONS

B. GUNAWAN\* and J.W. JAMES

School of Wool and Pastoral Sciences  
The University of New South Wales  
Kensington, Australia, 2033

SUMMARY

Computer simulations using additive and dominance models were carried out to investigate the effect of different distributions of gene frequency in base populations on the rate of response to selection in synthetic lines derived from them.

From 5 base populations, synthetic lines were founded with selected individuals from the best two lines or from all five lines. These were compared with the best line.

Selection in a synthetic line is most efficient if one constituent population has a low frequency for a favourable gene common in other populations, is moderately efficient if gene frequencies of favoured alleles are generally low, and is least efficient if gene frequencies are generally intermediate.

INTRODUCTION

Choosing foundation animals from each of several populations to form a synthetic line is likely to be a better strategy than selecting from a single population because increased additive genetic variance in the synthetic line should result in faster response to selection (Gunawan and James, 1986a,b). However, while these studies showed an advantage for synthetic lines, they give no indication of what genetic properties underlie the advantage, since they depend on phenotypic measurements only. It is also of interest to discover what patterns of differences among base populations, at the genome level, are favourable for the foundation of synthetic lines.

At a locus whose frequency varies between base populations, the average variance within lines is, for an additive model  $2a^2[\bar{p}(1-\bar{p}) - V_p]$  where  $a$  is the effect of a gene substitution,  $\bar{p}$  is the average gene frequency, and  $V_p$  is the variance between lines in gene frequency. The variance between lines is  $4a^2V_p$ , while the variance within a synthetic formed from all lines is  $2a^2\bar{p}(1-\bar{p})$ . Thus the genetic variance within a synthetic line is greater by  $\frac{1}{2}V_G$  than the average within line variance, where  $V_G$  is the between line variance. If gene frequencies are distributed independently over loci, these variances may be summed over loci to give similar results at the multilocus level.

---

\* From June 1986: The Project for Animal Research and Development,  
B.P.T., P.O. Box 123, Ciawi, Bogor, Indonesia.

For a dominance model the solution is not so straightforward, and there is no clearcut single condition for an increase of additive genetic variance  $V_A$  in a synthetic line. With this model, additive genetic variance is high when the recessive allele is at high frequency. Averaged over all loci, a synthetic would probably have higher variance, but the exact condition depends on third and fourth moments of the gene frequency distribution. Thus for this model, simulation may be the easiest way of investigating the importance of various features of the genetic structure of the base populations.

In practice, of course, neither the details of genetic control of the trait or traits of interest, nor the frequencies of genes at the relevant loci will be known, so that direct guidelines cannot be obtained. Nevertheless, a deeper understanding of conditions favouring use of synthetic lines should eventually have practical applications. We have therefore conducted Monte Carlo simulations of selection in synthetic lines formed from base populations with various gene frequency distributions.

#### METHODS

The simulation model consisted of five populations of a diploid species, with a trait controlled by 20 unlinked loci of equal effect. All loci were assumed completely additive in some simulations, while all showed complete dominance of the favoured allele in others. No epistasis was present in any simulation.

Three different types of gene frequency distribution were considered.

- Case 1. Each locus has either  $p = 0.15$  or  $p = 0.35$ , the proportion of loci with each frequency varying between populations. Average gene frequency in base populations ranged from 0.19 to 0.27.
- Case 2. Each locus has either  $p = 0.4$  or  $p = 0.6$ , with average gene frequency in the base populations ranging from 0.48 to 0.56.
- Case 3. Each locus has  $p = 0.02, 0.05, 0.2, 0.3, 0.5$  or  $0.7$ . A gene rare in one population is common in the others. Average gene frequency ranged from 0.259 to 0.339.

The computer programs will not be described in detail, but their general nature was as follows. The gene frequencies were specified for each base population, and a sample of 100 genotypes was generated with gene frequencies as probabilities of favourable alleles. Genotypic values were then calculated according to the model (additive or dominance). A random normal variate (environmental effect) with appropriate variance was then added to the genotypic value to give the phenotypic value. The environmental variance was chosen to give a heritability of about 0.4, though in individual populations in some cases the heritability ranged from about 0.3 to 0.4.

The base populations were ranked on phenotypic means over ten replicates (this always ranked populations correctly), and synthetic lines were founded with selected individuals from the best two lines (TB) and from all five lines (AL). For comparison, lines were founded from the best single line (OB). Selection within each line after foundation was simulated for ten generations, with ten replicates of each type.

Each selection line was founded with 20 individuals and, during the subsequent ten generations of selection, the best 20 out of 100 individuals were chosen as parents. Phenotypic means were calculated each generation and

averaged over replicates to measure selection responses. Rate of response to selection was estimated as the regression of phenotypic mean over replicates on generation number, for all gene frequency - gene action combinations. Additive genetic variance was calculated among the progeny of founder individuals as the variance of genotypic values for the additive model, and as the variance due to regression of genotypic value on number of favourable alleles for the dominance model.

#### RESULTS AND DISCUSSION

The initial means are shown in Table 1. The synthetic lines (TB, AL) had slightly lower means than the best line (OB).

TABLE 1. INITIAL PHENOTYPIC MEANS AVERAGED OVER 10 REPLICATES OF BASE POPULATIONS USED IN THE BEST (OB) AND SYNTHETIC LINES (TB, AL)

Line	Case 1		Case 2		Case 3	
	Additive	Dominance	Additive	Dominance	Additive	Dominance
OB	10.85±0.17	18.41±0.28	22.44±0.18	30.33±0.16	14.33±0.16	20.10±0.19
TB	10.34±0.15	17.77±0.21	21.95±0.20	30.01±0.13	13.86±0.15	19.58±0.19
AL	9.20±0.14	16.04±0.19	20.85±0.16	28.84±0.14	12.44±0.12	18.10±0.18

The phenotypic means after 10 generations of selection are presented in Table 2. Although initially lower, the means of the synthetic lines are now as high as, or higher than, the best line mean. Differences are significant for both types of gene action in Case 3, and for additive genes in Case 1. Thus greater response was achieved in synthetic lines. Regression estimates of rates of response are presented in Table 3, and confirm the greater response in synthetics, though only in Case 3 are the differences large.

TABLE 2 PHENOTYPIC MEANS AVERAGED OVER 10 REPLICATES AFTER 10 GENERATIONS OF SELECTION

Line	Case 1		Case 2		Case 3	
	Additive	Dominance	Additive	Dominance	Additive	Dominance
OB	32.7±0.5	36.3±0.3	38.5±0.3	38.8±0.2	29.7±0.6	32.1±0.6
TB	34.0±0.5	36.7±0.4	38.5±0.3	39.0±0.2	30.0±0.4	33.9±0.6
AL	33.6±0.3	36.0±0.4	38.2±0.1	39.0±0.2	31.7±0.5	36.3±0.3

The higher response rates in synthetic lines are due to higher additive variances. The variances ( $V_A$ ) estimated in the progeny of founders are presented in Table 4. As expected, the  $V_A$  values are higher in synthetic lines, and appreciably so in those cases where selection was markedly more effective (Case 3, Case 1 additive). For additive genes, genetic variance is proportional to heterozygosity while, for dominance genes,  $V_A$  is greatest when the recessive gene has a frequency of 0.75, and heterozygosity is 0.75 of its maximum value. Thus high average heterozygosity is expected to be associated with faster response, and Table 5 shows the expected heterozygosity in progeny of founders. This table was calculated using the expression of Nei (1963) for gene frequency change due to selection. Synthetic lines are generally more heterozygous, and this is especially so in Case 3.

TABLE 3. REGRESSIONS (b) OF PHENOTYPIC MEANS ON GENERATION NUMBERS

Line	Case 1		Case 2		Case 3	
	Additive	Dominance	Additive	Dominance	Additive	Dominance
OB	2.00±0.05	1.53±0.11	1.43±0.10	0.71±0.08	1.39±0.05	1.08±0.05
TB	2.11±0.06	1.60±0.12	1.43±0.12	0.71±0.09	1.45±0.05	1.26±0.07
AL	2.16±0.06	1.67±0.15	1.50±0.10	0.76±0.11	1.68±0.05	1.59±0.10

These results indicate that synthetics are likely to give markedly more rapid response when they have much higher heterozygosity than individual base populations. This condition is perhaps most readily achieved by having one or more favourable alleles rare in each base population, while the same alleles are common in other base populations. This is the situation in Case 3.

TABLE 4. ADDITIVE GENETIC VARIANCE ( $V_A$ ) AVERAGED OVER 10 REPLICATES IN PROGENY OF FOUNDERS

Line	Case 1		Case 2		Case 3	
	Additive	Dominance	Additive	Dominance	Additive	Dominance
OB	5.98±0.25	10.22±0.57	6.74±0.42	4.99±0.35	3.78±0.16	4.90±0.53
TB	6.58±0.29	11.34±0.89	6.93±0.37	4.93±0.50	4.39±0.31	5.23±0.32
AL	7.0 ±0.23	10.70±0.62	7.06±0.54	5.01±0.40	4.87±0.32	7.16±0.77

However, when gene frequencies are generally intermediate, even though they may vary as much on average as in Case 3, the increase in heterozygosity in the synthetic is not very great, because heterozygosity in all constituent populations is already close to maximum, and only small increases can occur. All results are consistent with this interpretation, and indicate that the particular pattern of gene frequencies in base populations may be important in determining whether synthetics are desirable.

In the simple theory as used by Howe and James (1973), the effect of forming a synthetic line on additive variance is to increase  $V_A$  by  $\frac{1}{2}V_G$ , or as a proportion by  $\frac{1}{2}\phi^2$ , where  $\phi$  is the ratio of between to within population genetic standard deviations. In these simulations,  $\phi$  was about 0.49 in Case 1, 0.41 in Case 2 and 0.57 in Case 3. The increases in variance expected in the 3 cases on this model would be 0.12, 0.08 and 0.16 approximately in the AL line. While the increases did rank as expected from the  $\phi$  values,  $\phi$  was not a reliable predictor, especially for dominance models. The model used by Howe and James does assume complete additivity. Thus the present results indicate that such simple models, though useful, may be rather inaccurate in detailed predictions.

TABLE 5. EXPECTED HETEROZYGOSITY AVERAGED OVER 20 LOCI IN PROGENY OF FOUNDERS

Line	Case 1		Case 2		Case 3	
	Additive	Dominance	Additive	Dominance	Additive	Dominance
OB	0.46	0.41	0.36	0.35	0.22	0.16
TB	0.48	0.43	0.36	0.35	0.26	0.19
AL	0.49	0.42	0.37	0.37	0.32	0.28

#### REFERENCES

- GUNAWAN, B. and JAMES, J.W., 1986a. Efficiency of response to selection in synthetic populations of *Drosophila*. I. Estimates of genetic parameters in the base populations and evaluation of theory. *Génétique, Sélection, Evolution*. (submitted for publication).
- GUNAWAN, B. and JAMES, J.W. 1986b. Efficiency of response to selection in synthetic populations of *Drosophila*. II. Response to two-trait selection. *Génétique, Sélection, Evolution*. (submitted for publication).
- HOWE, R.R. and JAMES, J.W. 1973. Response to selection in synthetic lines of *Drosophila melanogaster*, *Aust. J. Biol. Sci.* 26, 613-23.
- NEI, M. 1963. Effect of selection on the components of genetic variance. IN Hanson, W.D. and Robinson, H.F. (eds.) *Statistical genetics and plant breeding*. NAS-NRC. Publ. 982, 501-515.