Genet. Res., Camb. (1964), 5, pp. 341–353 With 6 text-figures Printed in Great Britain

The effect of continued selection of phenotypic intermediates on gene frequency

By R. N. CURNOW

Unit of Biometry, University of Reading

(Received 6 August 1963)

1. INTRODUCTION AND SUMMARY

The selection of animals or plants for high values of a certain character may favour not only genotypes associated with these high values but also genotypes associated with high variability. Any differences between genotypes in variability may therefore be of considerable importance in plant and livestock improvement programmes as well as in evolution. The effects of various selection procedures on variability have been studied in three recent experiments [Falconer & Robertson (1956), Falconer (1957) and Prout (1962)]. In these experiments one line was continued by selecting, in each generation, parents with values of a particular character near the population mean. Manning (1955, 1956) has described the effects of this kind of selection applied to cotton. Robertson (1956) derived and discussed the theory of such selection procedures when certain simplifying approximations can be made. We shall obtain some more general results and show that Robertson was incorrect in saying that the selection procedure would lead to gene fixation even if the heterozygotes are less variable than the homozygotes. The importance of the results is discussed in section 8.

The selection procedure to be considered is one in which individuals with values of a certain character sufficiently near the population mean are bound to be selected and all other individuals are bound not to be selected. The results can therefore be of direct interest only in artificial selection experiments. In natural selection, there will generally be a much weaker relationship between the probability of an individual surviving and the phenotypic value of any measurable character. Also, the selection pressure will probably be directed towards a fixed value rather than a population mean that changes from generation to generation. The complicated results obtained in this paper stem very largely from the fact that the population mean is the centre of the selection interval. This means that the selective values of the genotypes depend on the gene frequency and therefore vary from generation-togeneration.

2. A MODEL FOR THE GENETIC EFFECTS AND THE SELECTION PROCEDURE

We shall assume that there are only two alleles, A_1 and A_2 , at a particular autosomal locus and that they have an additive effect on the character being used for the selection. Further, we shall assume that the genetic effects at this locus are

unaffected by any changes in the gene frequencies that may be occurring at other loci. The frequencies of the A_1 and A_2 alleles will be written p and q respectively, where p+q=1. With an infinite closed population, no mutation and with random mating among the selected individuals, the situation can be represented as follows:—

Genotype	A_1A_1	A_1A_2	A_2A_2
Frequency	p^2	2pq	q^2
Average Value	0	a/2	a
Selective Value	$1 - s_1$	1	$1 - s_2$

We shall assume that the phenotypic values of the individuals of a given genotype are normally distributed about their average value with variance σ^2 for the two homozygous genotypes and $(\sigma/R)^2$ for the heterozygote genotype. Without any loss of generality, we shall take $\sigma^2 = 1$. This means that the value of a will be in units of one σ . The introduction of R, the ratio of the standard deviation of the homozygous genotypes to the standard deviation of the heterozygous genotype, is necessary because of the evidence that heterozygotes are sometimes less variable than homozygotes. We shall be particularly interested, therefore, in values of R > 1. R > 1 can be thought of as a crude model for a kind of homeostasis with respect to this particular character in which the heterozygotes are less sensitive than the homozygotes to environmental disturbances and to variations in the genotypic structure at the other loci. By assuming that the variances for the two homozygotes are equal, we are excluding loci at which there may be an additive effect of the genes on variability. The variance will contain components from at least three different sources of variability—environment, genetic effects at other loci and genotypeenvironment interactions. A reduced heterozygote variance may result from reductions in any or all of these components. We shall assume that the normality of the phenotypic distributions of each genotype persists through the period of selection and that the values of σ^2 and R remain unchanged.

In a single generation, selection will change the gene frequency from q to $q + \Delta q$ where

$$\Delta q = \frac{pq(s_1p - s_2q)}{1 - p^2 s_1 - q^2 s_2} \tag{1}$$

The mean value of the character in the whole population will be aq. The probabilities that individuals of the three genotypes are selected will be the probabilities that their phenotypes lie within the interval aq - X to aq + X, where 2X is the length of the selection interval measured in units of one σ . They are therefore,

$$P(A_{1}A_{1}) = F[aq + X] - [aq - X]$$

$$P(A_{1}A_{2}) = F\left[R\left(aq + X - \frac{a}{2}\right)\right] - F\left[R\left(aq - X - \frac{a}{2}\right)\right]$$

$$P(A_{2}A_{2}) = F[(aq + X - a)] - F[(aq - X - a)],$$
(2)

and

where

$$F(t) = \int_{-\infty}^{t} f(u) du$$

Effect of selection of phenotypic intermediates

with

$$f(u) = rac{1}{\sqrt{2\pi}} \exp{\{-rac{1}{2}u^2\}},$$

the standardized normal frequency function. Clearly

$$1 - s_1 = \frac{P(A_1A_1)}{P(A_1A_2)}$$
 and $1 - s_2 = \frac{P(A_2A_2)}{P(A_1A_2)}$, (3)

343

The proportion of individuals selected will be

$$P = p^{2} P(A_{1}A_{1}) + 2pqP(A_{1}A_{2}) + q^{2} P(A_{2}A_{2})$$
(4)

The gene frequency will tend towards an equilibrium value satisfying the equation

$$\Delta q = 0$$

We shall see later that there are sometimes equilibria additional to the obvious ones at $q = 0, \frac{1}{2}$ and 1.

There are two possible models for the selection procedure. In the first, the width of the selection interval, 2X, is kept constant and the size of the population is allowed to change or is controlled by random sampling the selected parents. In the second, the proportion selected is kept constant by allowing the width of the selection interval to vary with q. We shall return to this point later.

Robertson (1956) showed that when R = 1 (in his notation, $R = 1/\sqrt{k}$) and a and X/a are both small, $q = \frac{1}{2}$ is an unstable equilibrium and one or other allele will eventually become fixed. His statement that this is also true when R > 1 is not necessarily correct. We shall discuss the stability of the various equilibria for general values of a, X and R and show that even when a and X/a are small, $q = \frac{1}{2}$ will often be a stable equilibrium if the heterozygotes are less variable than the homozygotes.

3. THE BEHAVIOUR OF THE GENE FREQUENCY UNDER SELECTION

The graph relating Δq to q determines the future behaviour of the gene frequency. This graph can take one of four possible basic forms depending on the signs of $d\Delta q/dq$ at q = 0 and at $q = \frac{1}{2}$. The four basic forms are presented diagrammatically in Fig. 1. The labelling of the graphs S, I, FS and F refers to F for fixation possible, S for stability at $q = \frac{1}{2}$ and I for intermediate equilibrium other than $q = \frac{1}{2}$ stable. Because of the symmetry of the selection procedure about $q = \frac{1}{2}$, the sign of $d\Delta q/dq$ at q = 1 must always be the same as the sign of $d\Delta q/dq$ at q = 0 and the $(\Delta q, q)$ graph will always take one of the basic forms of Fig. 1 unless there are at least four more equilibria, $\Delta q = 0$. The existence of these extra equilibria does seem rather unlikely.

The equilibrium at $q = \frac{1}{2}$ is stable only if $d\Delta q/dq < 0$ at $q = \frac{1}{2}$ (graphs S and FS of Fig. 1). The fixation of one or other allele is only possible if $d\Delta q/dq < 0$ at q = 0 (graphs F and FS of Fig. 1). In graph S the gene frequency will tend eventually to a stable equilibrium at $q = \frac{1}{2}$. In graph I the gene frequency will tend to one or other of the stable equilibria lying between q = 0 and $\frac{1}{2}$ and between $q = \frac{1}{2}$ and 1. In graph



Fig. 1. The four types of behaviour of gene frequency under selection. The arrows indicate the direction in which the gene frequency will change.

FS, the gene frequency will either tend to a stable equilibrium at $q = \frac{1}{2}$ or one or other allele will become fixed depending on the initial gene frequency. In graph F, one or other allele will eventually become fixed.

From (1), at
$$q = 0$$

$$\frac{d\Delta q}{dq} = \frac{s_1}{1 - s_1}$$

 $\frac{d\Delta q}{dq} = \frac{(ds_1/dq) - 2s_1}{2(2-s_1)}$

and at $q = \frac{1}{2}$,

Effect of selection of phenotypic intermediates 345

Therefore, from (2) and (3), the signs of $d\Delta q/dq$ at q = 0 and $\frac{1}{2}$ are the signs of

$$F\left[R\left(X+\frac{a}{2}\right)\right]+F\left[R\left(X-\frac{a}{2}\right)\right]-2F(X)$$
(5)

and

$$F\left(X+\frac{a}{2}\right)+F\left(X-\frac{a}{2}\right)+\frac{a}{2}\left[f\left(X-\frac{a}{2}\right)-f\left(X+\frac{a}{2}\right)\right]-2F(RX)$$
(6)

respectively. For given values of X, a and R, the signs of these two expressions can be determined and the possibilities of stable intermediate gene frequencies and of fixation investigated using the graphs of Fig. 1.



Fig. 2. Values of R and RX for which fixation of one or other allele can occur. The region of possible fixation is to the left and under the curve for each of the various values of a.

4. CONDITIONS UNDER WHICH AN ALLELE CAN BECOME FIXED

An allele can become fixed as a result of the selection procedure if and only if expression (5) is negative (graphs FS and F of Fig. 1). Expression (5) is the correct expression to use whether the selection keeps X fixed or P fixed. If P is kept fixed,

the value of X in (5) is that corresponding to q = 0 (or, equivalently, to q = 1) in (4),

$$P = 2F(X) - 1 \tag{7}$$

Figure 2 shows the curves defined by the change of sign of expression (5) for a range of values of a. The region to the left and under each curve contains, for that value of a, the values of X and R for which fixation is possible. Note that, for reasons of scale, the x-axis corresponds to values of RX not X. This means that the selection interval is measured in units of the heterozygote rather than the homozygote standard deviation. The slope of a line through the origin is 1/X and so the selection interval, 2X, or intensity, P, is constant along straight lines through the origin. A selection interval of length 3 homozygous standard deviations, for example, corresponds to a line on the graph passing through the origin and having a slope of 2/3. A detailed derivation and description of Fig. 2 is given in Appendix I.

Table 1. Probability that a particular homozygous individual is selected when $q = \frac{1}{2}$ [see equation (2), $P(A_1A_1) = P(A_2A_2)$]

a	Value of X					
	0.5	1.0	1.5	2.0	3.0	<u>4</u> ·0
0	0.383	0.683	0.866	0.954	0.997	1.000
$\frac{1}{2}$	0.372	0.668	0.854	0.948	0.996	1.000
1	0.341	0.625	0.819	0.927	0.994	1.000
<u>3</u> 2	0.296	0.559	0.761	0.891	0.988	0.999
2	0.242	0.477	0.685	0.840	0.977	0.999
4	0.061	0.157	0.308	0.200	0.841	0.977

Table 1. Probability that a particular heterozygous individual is selected when $q = \frac{1}{2}$ [see equation (2), $P(A_1A_2)$]

	Value of X					
R	0.5	1.0	1.5	2.0	3.0	<u>4</u> ·0
$\frac{1}{2}$	0.197	0.383	0.547	0.683	0.866	0.954
ī	0.383	0.683	0.866	0.954	0.997	1.000
32	0.547	0.866	0.976	0.997	1.000	1.000
$\overline{2}$	0.683	0.954	0.997	1.000	1.000	1.000

5. CONDITIONS FOR THE EQUILIBRIUM AT $q = \frac{1}{2}$ TO BE STABLE

From (4) when $q = \frac{1}{2}$, dX/dq = 0 if P is kept constant. Therefore expression (6) being negative (graphs S and FS of Fig. 1) is the condition for the equilibrium at $q = \frac{1}{2}$ to be stable whether it is the length of the selection interval, 2X, or the selection intersity, P, that is kept constant. If the selection intensity is kept constant, X in (6) is related to P by formula (4) with $q = \frac{1}{2}$, i.e.

$$P = \frac{1}{2} \left\{ F\left(\frac{a}{2} + X\right) - F\left(\frac{a}{2} - X\right) + F(RX) - F(-RX) \right\}$$

The probabilities of a particular homozygous individual and of a particular heterozygous individual being selected when $q = \frac{1}{2}$ are given in Table 1. *P* can be calculated as the average of these two probabilities.

Figure 3 shows the curves defined by the change of sign of expression (6). The regions above and to the left of the curves for each value of a contain the values of X and R for which the equilibrium at $q = \frac{1}{2}$ is stable. A detailed derivation and description of Fig. 3 is given in Appendix II.



Fig. 3. Values of R and RX for which the equilibrium at $q = \frac{1}{2}$ is stable. The region of stability is to the left and above the curve for each of the various values of a. The top curve defines the region of stability for all values of a.

To appreciate fully the effects on gene frequency of selecting phenotypic intermediates we need to consider simultaneously the conditions under which the equilibrium at $q = \frac{1}{2}$ is stable and the conditions under which fixation is possible. Figures 4, 5 and 6 show, in diagrammatic form, the effects of the selection on gene frequency for a < 1.21, 1.21 < a < 2.62 and a > 2.62 respectively. The scales of the three diagrams are not the same. To fix the approximate position of the curves we note that the curve defining the region of fixation, F, always has as an asymptote

when R becomes large the straight line $X = \frac{a}{2}$. Unfortunately it approaches this asymptote very slowly indeed. The existence of regions in which fixation and stability at $q = \frac{1}{2}$ are both possible emphasizes the importance of initial gene frequency.

Robertson's result that, when the selection pressure is intense and the genetic effects small, one or other gene will eventually become fixed if there are no differences in variability (i.e. R = 1) is correct. Figures 2, 3 and 4 show that his result



Fig. 4. Values of R and RX for which the equilibrium at $q = \frac{1}{2}$ is stable (S) and for which fixation of one or other allele can occur (F) when a < 1.21. In the region marked I, the gene frequency will tend to a stable intermediate equilibrium between either q = 0 and $\frac{1}{2}$ or $q = \frac{1}{2}$ and 1.

that fixation will also occur whenever R > 1 is incorrect. He omitted a factor of R (in his notation, $1/\sqrt{k}$) when evaluating the frequency of the heterozygote distribution at the population mean (Robertson (1956), p. 237, line 9). In fact with small genetic effects and reasonably intense selection, R > 1 will generally result in a stable equilibrium at $q = \frac{1}{2}$ and R < 1 in one or other allele becoming fixed.

6. THE EFFECT OF DOMINANCE

To show that the existence of a stable equilibrium at an intermediate gene frequency is not completely dependent on the assumption of additive gene effects, the situation when the A_1 allele is completely dominant to the A_2 allele will now be considered. The possibility of the dominant allele becoming fixed depends solely



Fig. 5. Values of R and RX for which the equilibrium at $q = \frac{1}{2}$ is stable (S) and for which fixation of one or other allele can occur (F) when $1 \cdot 21 < a < 2 \cdot 62$. In the region marked I, the gene frequency will tend to a stable intermediate equilibrium between either q = 0 and $\frac{1}{2}$ or $q = \frac{1}{2}$ and 1.

on the sign of Δq at q = 0. This depends on the sign of $s_1(q = 0)$, which in turn has the sign of

$$2[F(XR)-F(X)]$$

Therefore the dominant allele can become fixed if and only if the heterozygotes are as variable or more variable than the homozygotes.

The possibility of the recessive allele becoming fixed depends on the sign of $d\Delta q/dq$ at q = 1. This depends on the sign of $s_2(q = 1)$, which in turn depends on the sign of

$$F[R(X+a)] + F[R(X-a)] - 2F(X)$$

350

Comparing this with expression (5), the values of the selection interval, X, and the standard deviation ratio, R, for which a recessive allele can become fixed, are the same as those for the fixation of an additive allele that produces a homozygote difference of 2a instead of a. Referring to Fig. 2, fixation of a recessive allele is possible for a wider range of values of X and R than is the fixation of one or other allele when the genetic effects are additive.



Fig. 6. Values of R and RX for which the equilibrium at $q = \frac{1}{2}$ is stable (S) and for which fixation of one or other allele can occur (F) when $2 \cdot 62 < a$. In the region marked I, the gene frequency will tend to a stable intermediate equilibrium between either q = 0 and $\frac{1}{2}$ or $q = \frac{1}{2}$ and 1.

The effect of complete dominance rather than additive gene action is therefore to restrict the conditions under which the dominant allele can become fixed and enlarge the conditions under which the recessive allele can become fixed. Whenever neither allele can become fixed, there must be a position of stable equilibrium at some intermediate gene frequency. Indeed there is probably a stable intermediate equilibrium whenever the heterozygotes are less variable than the homozygotes. Clearly, the general picture with complete dominance is very similar to that already described under the assumption of additive gene action. When R = 1, the equilibrium at $q = 1/\sqrt{2}$ is, as stated by Robertson (1956), unstable.

7. SELECTION OF PHENOTYPIC EXTREMES

Since the gene frequency in the total population must remain constant, the change in gene frequency from selecting individuals with phenotypes outside the interval about the mean must be of equal magnitude but of opposite sign to the change in gene frequency from selecting individuals with phenotypes inside the interval. The effect on gene frequency of random mating individuals selected from outside the interval can therefore be inferred from the effect on gene frequency of random mating individuals selected from inside the interval. For example, where fixation was possible it is now impossible and where $q = \frac{1}{2}$ was stable for an additive gene it is now unstable.

The effect on gene frequency of disassortative mating (i.e. mating individuals selected from opposite extremes of the phenotypic distribution) cannot, as was suggested by Robertson (1956), be inferred in this way. The mating of the selected individuals is not random and therefore the assumption of a Hardy–Weinberg equilibrium before each selection invalid.

8. DISCUSSION

We can formulate an approximate general rule that holds for sufficiently intense selection and for loci having genetic effects on the mean value of the character that are small compared with its overall phenotypic variability. This rule is that the selection of individuals in any region near the population mean will favour the least variable genotype irrespective of its position on the scale of mean values. Therefore, one or other allele will eventually become fixed unless there is overdominance on the scale of variability with the heterozygote less variable than either homozygote. This is true whether the selection interval remains fixed or varies with the population mean. Therefore, if environmental variability is under genetic control, selection of phenotypic intermediates will result in a reduced amount of variation, both genetic and environmental, unless the heterozygote is less variable than either homozygote. In this latter case genetic variance in both mean values and in the amount of environmental variation will persist. For an interesting discussion of the effects of stabilizing selection on variability, see Prout (1962).

The main purpose of this present paper is to show that when selection, whether artificial or natural, operates on a metric character, slight differences between genotypes in variability can be of crucial importance in deciding whether or not an allele will be eliminated from the population. Part of the effect of selection on the population mean may therefore be indirect acting through relationships between the means and variances of the various genotypes.

In a single generation, Δq will often be very small indeed. In the time-scale of laboratory experiments, the results obtained in this paper are probably only of interest in suggesting the likely direction of changes in gene frequency.

I would like to thank the referees for helpful suggestions about the presentation of these results.

REFERENCES

- FALCONER, D. S. (1957). Selection for phenotypic intermediates in Drosophila. J. Genet. 55, 551-561.
- FALCONER, D. S. & ROBERTSON, A. (1956). Selection for environmental variability of body size in mice. Z. indukt. Abstamm.- u. VererbLehre, 87, 385-391.
- MANNING, H. L. (1955). Response to selection for yield in cotton. Cold. Spr. Harb. Symp. quant. Biol. 20, 103-110.
- MANNING, H. L. (1956). Yield improvement from a selection index technique with cotton. *Heredity*, **10**, 303-322.
- **PROUT**, T. (1962). The effects of stabilizing selection on the time of development in *Drosophila* melanogaster. Genet. Res. **3**, 364–382.

ROBERTSON, A. (1956). The effect of selection against extreme deviants based on deviation or on homozygosis. J. Genet. 54, 236-248.

APPENDIX I

Detailed description of conditions for fixation to be possible (Fig. 2)

Fixation is always possible, if the homozygotes are as variable or less variable than the heterozygotes (i.e. $R \leq 1$). If the genetic effects are large, $[a > 2F^{-1}(\frac{3}{4}) =$ 1·349], fixation is always possible if the length of the selection interval, 2X, is less than the genetic range, a, i.e. for lines through the origin of slope greater than 2/a. If the selection interval is greater than the genetic range, fixation is possible only if the ratio of the homozygote to heterozygote standard deviation, R, is sufficiently near R = 1. For sufficiently large genetic effects any values for the selection interval and the standard deviation ratio will allow fixation to occur. The region in which fixation cannot occur becomes progressively less as the genetic effects increase.

If the genetic effects lie between $a = 2/\sqrt{e} = 1.213$ and a = 1.349, fixation is always possible when the standardized selection interval, 2X, is sufficiently small but there is a range of intermediate values of the standard deviation ratio for which fixation is impossible when the standardized selection interval is above a certain minimum value but still less than the genetic range. When the selection interval exceeds the genetic range, fixation is only possible for values of the standard deviation ratio sufficiently near R = 1.

If the genetic effects are small (a < 1.213), the *a* curves cut the *R*-axis in the two points given by $\log R^2/R^2 = a^2/4$. When a = 1.213, this equation has only one root in *R*, namely $R = \sqrt{e}$, and the curve touches the *R*-axis. If the selection interval is less than the genetic range, fixation is possible either if the standard deviation ratio is large or if it is near R = 1. For $a = \frac{1}{2}$ and 1 in Fig. 2 the values of *R* above which fixation is possible when the selection interval is less than the genetic range all exceed 2.9 and so lie outside the range of the graph. If the selection interval is larger than the genetic range, fixation is possible only if *R* is sufficiently near R = 1.

The only additional point that may help in constructing further curves is that when a < 1.349 and the selection interval equals the genetic range, fixation is possible only if R is less than the solution of

$$F(aR) = 2F(a/2) - \frac{1}{2}$$

APPENDIX II

Detailed description of conditions for stability at $q = \frac{1}{2}$ (Fig. 3)

When the genetic effects are small, the equilibrium is stable if and only if the homozygotes are slightly more variable than the heterozygotes and the standardized selection interval, 2X, is not too large. As the genetic effects increase, the region of stability decreases until a is approximately a = 2, when it starts to increase again. For very large genetic effects the equilibrium will be stable unless the standardized selection interval, 2X, is large or the heterozygotes are much more variable than the homozygotes ($R \ll 1$). The uppermost curve in Fig. 3 defines the region within which the equilibrium is stable whatever the size of the genetic effects. The corresponding relationship between X and R is found by choosing a to maximize expression (6) and then equating expression (6) to zero for this value of a. The maximizing value of a is a function of the selection interval, 2X, being the solution of the equation

$$\frac{1}{2}aX\tanh\frac{aX}{2} = X^2$$

For any particular pair of values for the standard deviation ratio and the genetic effects, the equilibrium will be unstable if the standardized selection interval is sufficiently large. If the selection interval equals the genetic range, the equilibrium will be stable if the standard deviation ratio is sufficiently large provided only that the genetic effect is less than a = 2.62 [the solution of the equation

$$a[f(0)-f(a)]+2F(a) = 3]$$

The smallest value of R for which the equilibrium is stable when the selection interval equals the genetic range is found by equating the left-hand side of this equation to

$$4F(Ra)-1$$

When the standardized selection interval is very small (X = 0), the equilibrium is stable if and only if

$$R > [\frac{1}{4}a^2 + 1] \exp\left\{-\frac{a^2}{8}\right\}$$

This lower bound to the value of the standard deviation ratio increases from R = 1 to $R = 2/\sqrt{e}$ (the value on the boundary of the region where stability exists for all values of the genetic effects) as a increases from 0 to 2. It then decreases from $R = 2/\sqrt{e}$ to R = 1 as a increases from 2 to 3.18 and from R = 1 to R = 0 as a increases from 3.18 to $+\infty$.