

# CHANGES OF GENE-FREQUENCY IN *CEPAEA NEMORALIS* (L.); THE ESTIMATION OF SELECTIVE VALUES

BRYAN CLARKE

*Department of Zoology, West Mains Road, Edinburgh*

and

JAMES MURRAY

*Department of Zoology, University Museum, Oxford*

Received 15.iii.62

## 1. INTRODUCTION

IN the previous paper (Clarke and Murray, 1962), we have described the present distribution of phenotypes of the land snail *Cepaea nemoralis* (L.) on sand dunes at Berrow in Somerset. The careful records of Captain C. Diver and the late Professor A. E. Boycott allowed us to compare the phenotype-frequencies found in 1959-60 with those found in 1926. The results were striking. While the stability of steep morph-ratio clines has been maintained over the intervening period, at least two phenotypes show systematic changes of frequency. Thus although the highest proportions of dark brown shells are found in the same places as before, there has been a consistent overall decline in the frequency of this phenotype throughout the area sampled. Similarly there has been a general increase in the proportion of single-banded (00300) shells.

We have presented reasons for attributing these changes to the action of natural selection (Clarke and Murray, 1962). Because of the precision of the historical records, the evidence for genetic continuity provided by the morph-ratio clines, and the consistency of the changes in phenotype-frequencies, we consider this to be a unique opportunity for obtaining quantitative estimates of the forces of natural selection.

Both the genes ( $G^B$  for dark brown and  $M^3$  for 00300: Cain and Currey, in preparation) are inherited as Mendelian dominants. Their phenotypic expressions seem to be unaffected by direct environmental influences, and segregation ratios are apparently normal (Lamotte, 1951; Cain and Sheppard, 1957; Cain, King and Sheppard, 1960).

## 2. THE FITNESS OF THE DARK BROWN PHENOTYPE

### (i) *Changes in the frequency of dark brown*

Twenty areas on the Berrow dunes were sampled both in 1926 and in 1959-60. Six of the resulting pairs of samples contain no dark browns and yield no information on changes of gene-frequency. The remaining fourteen pairs are listed in table 1 with the percentage of dark brown shells in each sample, the total numbers of shells and

the frequencies of the  $G^B$  gene calculated from the Hardy-Weinberg ratio. Of the fourteen comparisons, twelve show a decrease in the frequency of the gene, and one shows no change. The increase in the fourteenth case is the result of a single brown individual in our sample of 49 snails from a locality where Boycott and Diver found no browns in a sample of 171.

In seven pairs of samples only one member of the pair contains

TABLE 1

*Selection coefficients of the dark brown phenotype at Berrow in Somerset*

% = phenotype frequencies

N = sample sizes

$p$  and  $q$  = gene frequencies

S = selection coefficients

$V_{(S)}$  = variances of S

Those columns referring to 1926 bear the subscript  $_0$ ; those referring to 1959-60, the subscript  $_n$ . For those samples in which no dark browns were found, column 3 lists a frequency based on the assumption that the sample contained one dark brown individual.

1	2	3	4	5	6	7	8	9	10	11	12
	% <sub>0</sub>	% <sub>assumed</sub>	% <sub>n</sub>	N <sub>0</sub>	N <sub>n</sub>	$p_0$	$p_n$	$q_0$	$q_n$	S	$V_{(S)}$
D 19	3.5	0.76	0	142	132	0.018	0.004	0.982	0.996	0.1390	0.00958
D 25	0.4	0.66	0	266	152	0.002	0.003	0.998	0.997	-0.0370	0.01696
D 29	2.1	0.29	0	48	342	0.011	0.001	0.989	0.999	0.2194	0.02014
D 30	0.9	0.45	0	230	220	0.005	0.002	0.995	0.998	0.0838	0.01308
D 34	0	0.58	2.0	171	49	0.003	0.010	0.997	0.990	-0.1109	0.01686
D 33	0.6	0.60	0	316	166	0.003	0.003	0.997	0.997	0	0.01277
D 36	1.1	1.96	0	187	51	0.006	0.010	0.994	0.990	-0.0472	0.01212
D 40	2.8	—	2.8	180	106	0.014	0.014	0.986	0.986	0	0.00462
D 46	4.2	—	1.7	168	59	0.021	0.009	0.979	0.991	0.0791	0.00920
D 48	13.4	—	10.1	469	89	0.069	0.052	0.931	0.948	0.0290	0.00120
D 50	16.5	—	13.8	133	145	0.086	0.072	0.914	0.928	0.0190	0.00096
D 52	22.2	—	11.5	427	357	0.118	0.059	0.882	0.941	0.0745	0.00035
D 54	7.2	—	1.9	249	580	0.037	0.010	0.963	0.990	0.1234	0.00122
D 56	12.2	—	5.9	115	304	0.063	0.030	0.937	0.970	0.0734	0.00117

dark browns. For the purpose of computation, the other member is assumed to include one dark brown individual. Column 3 of table 1 gives the assumed percentages of dark browns for those samples where the percentage was actually zero.

#### (ii) Coefficients of selection

Let  $G^B$  represent the gene for dark brown, and  $G^X$  all other members of the allelic series controlling the ground colour of the shell. According to the Hardy-Weinberg ratio the frequencies of the three genotypes should be:

Genotype	$G^B G^B$	$G^B G^X$	$G^X G^X$
Frequency	$p^2$	$2pq$	$q^2$

If  $S$  represents the selective disadvantage of the dark brown phenotype, then the fitness of the genotypes can be written as:

Genotype	$G^B G^B$	$G^B G^X$	$G^X G^X$
Fitness	1-S	1-S	1

The change of gene-frequency as a result of a single generation of selection is (Falconer, 1960):

$$\Delta q = \frac{Sq^2(1-q)}{1-S(1-q^2)}$$

By treating this expression as a differential equation and integrating, it is possible to express the selection coefficient  $S$  in terms of the initial and final gene-frequencies  $q_0$  and  $q_n$  and the number of generations  $t$  involved (Lush, Bulmer, personal communications):

$$S = \frac{\log_e \left[ \frac{q_n(1-q_0)}{q_0(1-q_n)} \right] - \frac{1}{q_n} + \frac{1}{q_0}}{t + \log_e \left( \frac{q_n}{q_0} \right) - \frac{1}{q_n} + \frac{1}{q_0}}$$

This expression allows us to calculate the variance of the coefficient of selection with respect to the two sample sizes  $N_0$  and  $N_n$ :

$$V_{(S)} = \frac{\left( \frac{1}{q_n} + \frac{1}{1-q_n} + \frac{1}{q_0^2} \right)^2 \frac{q_n(1-q_n)}{2N_n} + \left( \frac{1}{q_0} + \frac{1}{1-q_0} + \frac{1}{q_0^2} \right)^2 \frac{q_0(1-q_0)}{2N_0}}{\left[ t + \log_e \left( \frac{q_n}{q_0} \right) - \frac{1}{q_n} + \frac{1}{q_0} \right]^2}$$

Columns 11 and 12 on table 1 show the results of these calculations for the fourteen pairs of samples, assuming a generation time of three years (see below).

The various estimates of the selection coefficient can be combined into a weighted mean value:

$$\bar{S} = \frac{\sum \frac{S_i}{V_{(S_i)}}}{\sum \frac{1}{V_{(S_i)}}}$$

which has the variance:

$$V_{(\bar{S})} = \frac{1}{\sum \frac{1}{V_{(S_i)}}}$$

By combining the estimates in this way, the mean selective coefficient  $\bar{S}$  is found to be  $0.0619 \pm 0.0118$ . (S.E.)

## 3. THE FITNESS OF THE SINGLE-BANDED PHENOTYPE

## (i) Changes in the frequency of 00300

When considering the changes that have taken place in the frequency of the single-banded type at Berrow, all twenty comparable pairs of samples may be used. The percentages of 00300 shells, the sample sizes and the Hardy-Weinberg gene-frequencies are listed in table 2.

TABLE 2

*Selection coefficients for the single-banded (00300) phenotype at Berrow in Somerset*  
Columns and symbols as in table 1.

1	2	3	4	5	6	7	8	9	10	11
	% <sub>o</sub>	% <sub>n</sub>	N <sub>o</sub>	N <sub>n</sub>	p <sub>o</sub>	p <sub>n</sub>	q <sub>o</sub>	q <sub>n</sub>	S	V(S)
D 19	9.6	11.4	135	132	0.049	0.059	0.951	0.941	0.0186	0.00132
D 23	6.1	16.5	99	109	0.031	0.086	0.969	0.914	0.0949	0.00172
D 24	2.7	12.3	150	106	0.014	0.064	0.986	0.936	0.1299	0.00216
D 26	3.5	6.4	371	220	0.018	0.033	0.982	0.967	0.0549	0.00115
D 25	2.6	5.9	265	152	0.013	0.030	0.987	0.970	0.0736	0.00193
D 29	14.9	7.6	47	342	0.078	0.039	0.922	0.961	-0.0755	0.00204
D 30	4.8	13.6	228	220	0.024	0.070	0.976	0.930	0.0969	0.00095
D 31	13.0	35.6	23	59	0.067	0.198	0.933	0.802	0.1167	0.00330
D 32	7.5	18.2	389	22	0.038	0.096	0.962	0.904	0.0885	0.00252
D 34	2.9	12.5	171	48	0.015	0.065	0.985	0.935	0.1262	0.00257
D 33	3.5	4.2	314	166	0.018	0.021	0.982	0.979	0.0144	0.00198
D 35	1.0	15.8	207	19	0.005	0.082	0.995	0.918	0.2147	0.00476
D 36	4.9	7.8	185	51	0.025	0.040	0.975	0.960	0.0437	0.00299
D 40	5.7	12.6	175	103	0.029	0.065	0.971	0.935	0.0749	0.00143
D 46	11.8	22.4	161	58	0.061	0.119	0.939	0.881	0.0687	0.00123
D 48	14.8	23.8	405	80	0.077	0.127	0.923	0.873	0.0538	0.00071
D 50	12.6	17.6	111	125	0.065	0.092	0.935	0.908	0.0360	0.00111
D 52	9.9	13.0	332	315	0.051	0.067	0.949	0.933	0.0273	0.00050
D 54	17.8	28.0	230	597	0.093	0.151	0.907	0.849	0.0545	0.00031
D 56	18.8	16.4	101	286	0.099	0.086	0.901	0.914	-0.0157	0.00081

Since the gene for the unbanded condition is epistatic to that for 00300 and since there is also an interaction of some kind between the genes for dark brown and for banding, the percentages of 00300 have been calculated only from the banded yellow, pink and pale brown shells in each sample (Clarke and Murray, 1962). Hence the sample sizes are slightly smaller than the corresponding ones in table 1.

Of the twenty comparisons, eighteen show an increase in the frequency of 00300, and two show a decrease.

## (ii) Coefficients of selection

If  $M^3$  represents the gene for 00300 and  $M^-$  the unmodified banded condition, then the frequencies of the genotypes and their fitness can be represented as:

Genotype	$M^3M^3$	$M^3M^-$	$M^-M^-$
Frequency	$p^2$	$2pq$	$q^2$
Fitness	1	1	1-S

In order to consider selection as acting always *against* the decreasing genotype, we can use a modification of the expressions given above for the dark brown condition. It would, of course, be possible to use the former expressions, but then selection would be expressed relative to the dominant gene and would be opposite in sign to that relative to the recessive. The relation of the two methods of expressing the selection coefficient is:

$$S_{(\text{dominant})} = \frac{-S_{(\text{recessive})}}{1 - S_{(\text{recessive})}}$$

Expressed as selection *against the recessive* the change in the frequency of  $M$  - becomes:

$$\Delta q = \frac{Sq^2(1-q)}{1-Sq^2}$$

and:

$$S = \frac{\log_e \left[ \frac{q_o(1-q_n)}{q_n(1-q_o)} \right] + \frac{1}{q_n} - \frac{1}{q_o}}{t + \log_e \left( \frac{1-q_n}{1-q_o} \right)}$$

and:

$$V_{(S)} = \frac{\left( \frac{1}{q_n} + \frac{1}{1-q_n} + \frac{1}{q_n^2} \right)^2 \frac{q_n(1-q_n)}{2N_n} + \left( \frac{1}{q_o} + \frac{1}{1-q_o} + \frac{1}{q_o^2} \right)^2 \frac{q_o(1-q_o)}{2N_o}}{\left[ t + \log_e \left( \frac{1-q_n}{1-q_o} \right) \right]^2}$$

The resulting coefficients and their variances are shown in columns 10 and 11 of table 2. Combining the estimates of  $S$  we find that:

$$\bar{S} = 0.0521 \pm 0.0076 \text{ (S.E.)}$$

#### 4. DISCUSSION

In order to estimate the forces of natural selection, we have made a number of assumptions. In the first place, we have assumed that the Hardy-Weinberg law provides an adequate description of the distribution of gene-frequencies in our samples. This law, of course, applies only to large, randomly-mating populations in the absence of mutation, migration and selection. It is unlikely, however, that mutation and migration will have a serious effect on the gene-frequency distributions studied by us. Lamotte (1951) has presented evidence against assortive mating in *C. nemoralis*, and our sampling areas are small enough to be considered as panmictic units (Lamotte, 1951). However, since we are attempting to measure natural selection, we must determine its effect on our estimates of gene-frequency.

Barring unlikely contingencies such as meiotic drive, which seems to be ruled out by the good segregation ratios obtained from breeding

experiments (see above), the Hardy-Weinberg ratio will be re-established with the union of gametes in each generation. Let us consider the case in which the fitness of the three genotypes is represented by:

$$\begin{array}{ccc} p^2 & 2pq & q^2 \\ 1-S & 1-S & 1 \end{array}$$

The initial frequency of the recessive homozygote will be  $q^2$ . After selection is complete, which will be the extreme case for the adult population, the proportion of recessives in the population is:

$$\frac{q^2}{1-S(1-q^2)}$$

In order to determine how much difference to the estimation of  $S$  might be caused by equating this quantity with  $q^2$ , we have recalculated the selection coefficient of dark brown for a typical example (D 56) on the assumption that the Hardy-Weinberg ratio had been disturbed by a selective force of 6 per cent. Although there is a 2.4 per cent. difference in the estimated gene frequencies, the resulting effect on the value of  $S$  is negligible. Instead of 0.0734, it becomes 0.0735. Thus a moderate amount of natural selection will have little effect on the determination of the size of the selection coefficient, despite small errors in the estimation of gene frequency. It would be possible by iterative calculations to apply a correction to  $S$ , but the gain in accuracy is hardly worth while.

We have made two assumptions about the nature of the selection coefficients themselves. First, the fitness of the dominant homozygote and the heterozygote have been assumed to be equal. This is probably not true since the fitness of the heterozygote in a polymorphic system is likely to be greater than that of either homozygote. However, in the cases of both of the dominants which we have considered, the gene-frequencies, and therefore the percentages of dominant homozygotes, are low. This is particularly true of the dark browns, where homozygotes in excess of 1 per cent. are expected in only one sample out of twenty-eight. Hence the selection coefficients of the dominants can be considered to represent most nearly those of the heterozygotes.

Second, in order to arrive at an estimate of the mean selective coefficient in each case, we have assumed the value of  $S$  to be the same in each set of samples. It is possible to test the homogeneity of the estimates by calculating  $\chi^2$  with  $K-1$  degrees of freedom, where  $K$  is the number of individual estimates.

$$\chi^2_{(K-1)} = \sum \frac{(S_i - \bar{S})^2}{V_{(S_i)}}$$

In the case of dark brown,  $\chi^2_{(13)} = 12.87$ ; and the probability lies between 0.3 and 0.5. There is thus no evidence of heterogeneity

among the estimates of  $S$  for dark brown, despite the fact that in half of them an assumed value has been used for either the initial or final gene frequency. A word of caution is necessary, however, since this good agreement should be interpreted partly as a reflection of the rather large variances of the individual estimates. This is particularly true of those based on assumed frequencies. On the other hand if we consider only those seven pairs of samples where no assumptions were necessary, the mean value of  $S$  is only slightly altered. Instead of 0.0619 it becomes 0.0638. The change is accompanied by an increase in the standard error of  $\bar{S}$  from 0.0118 to 0.0122.

In the case of 00300, where individual variances are smaller,  $\chi^2_{(10)} = 33.03$ . Since this figure is significant at the 5 per cent. level there is evidence of real differences in the value of  $S$  among the twenty samples, not altogether a surprising conclusion considering the wide range of frequencies over the area sampled. We cannot therefore place much confidence in the standard error of  $S$  for 00300.

An examination of the individual contributions to the value of  $\chi^2$  for 00300 shows that one sample (D 29) contributes almost a quarter of the total. Since there is no independent reason for excluding this value, it is not legitimate to remove it from the calculations. However, it is interesting that without this extreme value,  $\chi^2$  is no longer significant, and  $\bar{S}$  is increased from 0.0521 to 0.0559.

The model that we have used assumes that selection takes place by steps in a series of discrete generations. In a long-lived animal such as *C. nemoralis* the process is continuous since there is no break between generations. This particular problem has been discussed by Haldane (1926). He has concluded that the model for selection based on overlapping generations is negligibly different from that based on discrete generations.

Finally, we must consider the generation time in *C. nemoralis*. For an animal with discrete generations, this is simply the period between two equivalent stages in successive generations. With overlapping generations, the important time is that at which each individual has made half its contribution to the next generation. Hence generation time should be measured from zygote formation to the mid-point in the reproductive life of the average individual. Even when maintained in optimum conditions where feeding and growth can continue throughout the winter, *C. nemoralis* rarely becomes adult within one year. It seems likely that the average individual begins to breed in its second autumn or third spring, when it is eighteen months to two years old. If we consider that adult snails are eliminated from the population at a rate of about 50 per cent. per year, then the generation time should be about three years.

Since these calculations are necessarily rather rough, we have indicated in fig. 1 the relation between generation time and the selection coefficient. Sample D 56 (dark brown) has again been used. The selection coefficient is plotted against generation time in



years. Over the range which we have investigated the relationship is very nearly linear. A reduction of the generation time from three to two years would reduce the value of the selection coefficient in this case from 7.3 to 4.9 per cent.

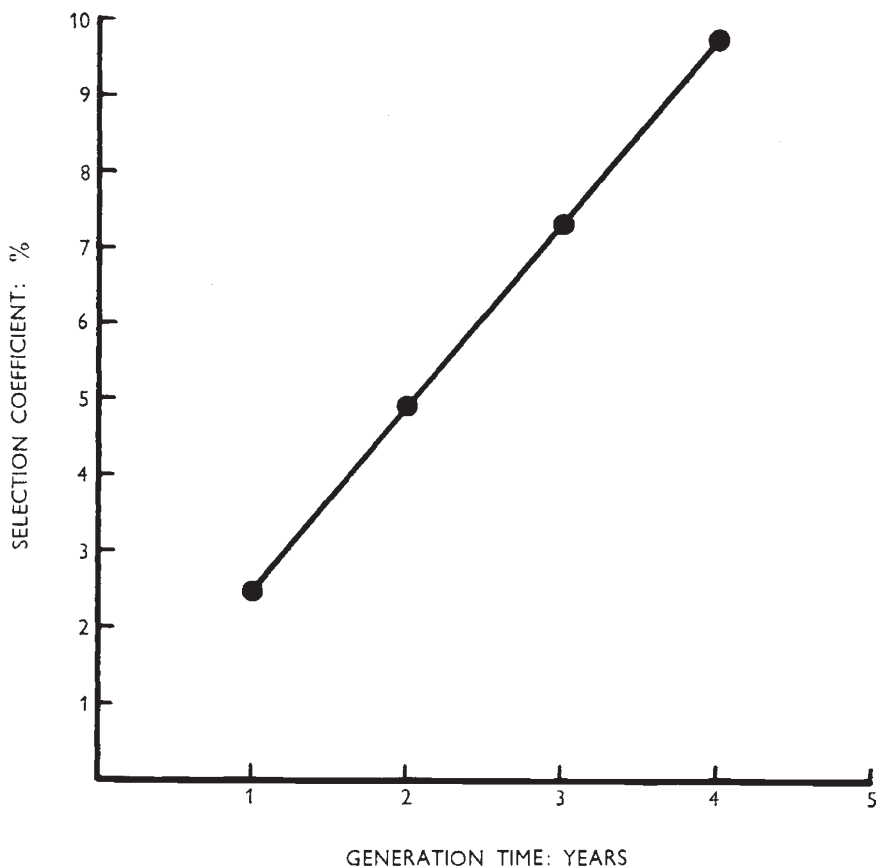


FIG. 1.—The relation of generation time to the magnitude of the selection coefficient in sample D 56 (dark brown).

## 5. CONCLUSIONS

We have estimated the magnitude of the selective forces necessary to bring about observed changes in the proportions of two phenotypes in populations of *Cepaea nemoralis* at Berrow in Somerset. Over the period 1926 to 1959-60, the dark brown phenotype has been at an average disadvantage of about 6.2 per cent., while the 00300 phenotype has been at an average advantage of about 5.2 per cent. per generation. These figures are based on an average generation time of three years. In the unlikely event that the generation time in *C. nemoralis* is as low as two years, then these estimates would be reduced to 3.8 and 2.8 per cent. respectively.



There is no evidence of variation from place to place in the value of the selection coefficient of dark brown. On the other hand, it is likely that there are real differences in the magnitude of the selection coefficient of 00300 at different localities. The heterogeneity among the estimates of the selection coefficient of 00300 is of importance for two reasons. First, it may explain the differences in the frequency of the single-banded type within the area. Secondly, it shows that the selective value is not associated with a single obvious feature of the habitat. It is likely that each value is a result of the combined action of many different forces.

We have no information on fluctuations in the magnitude of the selection coefficients during the intervening period between the collections of the two sets of samples. However, any fluctuations in the direction and intensity of the selection would necessarily increase the intensity of selection for a part of the time. Hence our estimates are minimum values for the magnitude of selection.

Finally, with natural selection of this intensity, there can be little scope for the action of dispersive processes in the determination of gene-frequencies.

## 6. SUMMARY

1. Assuming a generation time of three years, the dark brown phenotype in populations of *Cepaea nemoralis* at Berrow in Somerset has been at an overall disadvantage of 6.2 per cent. per generation over the period 1926 to 1959-60.

2. During the same period, the single-banded (00300) phenotype has enjoyed an overall selective advantage (expressed as the disadvantage of the normally banded phenotype) of 5.2 per cent. per generation.

3. With this level of selection, negligible errors are introduced by calculating gene-frequencies from the Hardy-Weinberg ratio.

4. No heterogeneity is demonstrable in the individual estimates of the selection coefficient of dark brown. Heterogeneity among the estimates of the selection coefficient of 00300 suggests real differences in the value of  $S$  in different parts of the area studied.

5. Over the range investigated, a direct linear relationship obtains between length of generation time and magnitude of the selection coefficient.

*Acknowledgments.*—We would like to express our appreciation to Captain C. Diver, C.B., C.B.E., for his generous offer of the data collected by Professor Boycott and himself. We are grateful to Dr A. J. Cain and Dr P. M. Sheppard for their help and encouragement at all times. Professor J. L. Lush and Dr M. G. Bulmer have given valuable statistical advice, and Elizabeth Murray has very kindly helped in the preparation of the manuscript. The work was carried out partly in the Department of Zoology, Oxford, where one of us (J. M.) held a National Science Foundation Pre-doctoral Fellowship, and partly in the Department of Zoology, Edinburgh, where the other (B. C.) received support from the University of Edinburgh. We would like to thank Professor Sir Alister Hardy, F.R.S., and Professor M. M. Swann, F.R.S. for the help and facilities they have provided.

## 7. REFERENCES

- CAIN, A. J., AND CURREY, J. D. Area effects in *Cepaea*. In preparation.
- CAIN, A. J., KING, J. M. B., AND SHEPPARD, P. M. 1960. New data on the genetics of polymorphism in the snail *Cepaea nemoralis* L. *Genetics*, 45, 393-411.
- CAIN, A. J., AND SHEPPARD, P. M. 1957. Some breeding experiments with *Cepaea nemoralis* (L.). *J. Genet.*, 55, 195-199.
- CLARKE, B., AND MURRAY, J. 1962. Changes of gene-frequency in *Cepaea nemoralis* (L.). *Heredity*, 17, 445-465.
- FALCONER, D. S. 1960. Introduction to quantitative genetics. Edinburgh and London : Oliver and Boyd.
- HALDANE, J. B. S. 1926. A mathematical theory of natural and artificial selection, IV. *Proc. Camb. phil. Soc.*, 23, 607-615.
- LAMOTTE, M. 1951. Recherches sur la structure génétique des populations naturelles de *Cepaea nemoralis* (L.). *Bull. biol., Suppl.* 35, 1-239.