SIMULATION OF GENETIC SYSTEMS BY AUTOMATIC DIGITAL COMPUTERS

II. EFFECTS OF LINKAGE ON RATES OF ADVANCE UNDER SELECTION

By A. S. Fraser*

[*Manuscript received July* 4, 1957]

Summary

Rates of progress of single populations under selection pressure have been simulated by an automatic electronic computer. Varying intensities of selection and tightness of linkage are compared, showing that linkage produces no qualitative effect on the rates of advance at values greater than 0·005, i.e. 0·5 per cent. recombination.

I. Introduction

Several of the sequences discussed in the introductory paper of this series (Fraser 1957) have been used as the basis of a programme which simulates the effect of selection on a single population of cross-breeding organisms, in which there is no overlap of generations. The programme is based on a genetic system of six loci, and provision has been made to allow variation of the recombination between loci from $r = 0{\cdot}5$ to $r = 0$, i.e. from independent assortment to complete linkage. It is therefore possible using this programme to produce results which simulate the rates of advance under selection for different degrees of linkage.

II. Structure of the Programme

The general sequence of the programme is as follows:

(i) *Given* the recombination values, form the vector of frequencies of recombinants and non-recombinants; this is termed $\{f_i\}$.

(ii) *Given* the vector of frequencies of types of haploid genotypes (i.e. types of gametes) which is termed $\{g_i\}$, then this vector is set to sum to unity, i.e.

$$\Sigma g = 1.$$

(iii) Transform $\{g_i\}$ by sequential summation, i.e. transform $\{g_i\}$ to $\{\hat{g}_i\}$, on the operation

$$\hat{g}_i = \sum_0^i g.$$

(iv) Generate a random number, S_j, and test this against $\{\hat{g}_i\}$ by counting the number of times that $S_j > \hat{g}_i$ until $\hat{g}_{i+1} \geqslant S_j > \hat{g}_i$.

Take the binary digital formula of i as the genetic structure of a haploid gamete; term this a_0. Repeat to produce a second gamete; term this a_1.

Then $\{a_0, a_1\}$ is the genetic structure of a diploid genotype, i.e. of a single progeny.

*Animal Genetics Section, C.S.I.R.O., University of Sydney.

(v) Repeat sequence (iv) N times, producing a set of N diploid genotypes, i.e. produce N progeny. This is termed $[\{a_0 \cdot a_1\}]$.

(vi) Determine the logical sum of a_0 and a_1. Sum the digits of the logical sum to give the phenotype. This is repeated for the N separate diploid genotypes, giving a set of N phenotypes corresponding to the set of N diploid genotypes; term this $\{P_i\}$.

(vii) Re-order $\{P_i\}$ in terms of descending value, re-sorting $[\{a_0 \cdot a_1\}]$ to correspond.

(viii) Take the first N' values from $[\{a_0 \cdot a_1\}]$ as the parents of the next generation.

(ix) Take $\{a_0 \cdot a_1\}_i$, and form the vector $[a_0/a_1]_{2^n}$ as detailed in the recombination section of the introductory paper to this series (Fraser 1957). This vector is of the types of gametes produced by a_0/a_1, and the relative frequencies of the types of gametes is given by $\{f_i\}$, i.e. the frequency of the ith term in the vector of types of gametes is f_i.

(x) Each individual is assumed to form the same number of gametes; therefore, if the number of parents is N' and the vector of frequencies of types of gametes, $\{f_i\}$, is divided by N', and each term is added to $\{g_i\}$, the vector of frequencies of types of gametes produced by N' parents is in the position specified by $[a_0/a_1]$. This is illustrated for a genetic system of two loci:

Given

$$a_0 = 0\ 1,$$

and

$$a_1 = 1\ 0,$$

and

$$\{f_i\} = \begin{Bmatrix} 0{\cdot}3 \\ 0{\cdot}2 \\ 0{\cdot}2 \\ 0{\cdot}3 \end{Bmatrix},$$

i.e. $r = 0{\cdot}2$, then

$$[a_0/a_1] = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 1 \\ 1 & 0 \end{bmatrix},$$

and if $N' = 10$, then

$$\{g_i\} = \begin{Bmatrix} 0{\cdot}02 \\ 0{\cdot}03 \\ 0{\cdot}03 \\ 0{\cdot}02 \end{Bmatrix}.$$

(xi) Repeating these sequences gives a new set of values of $\{g_i\}$. If we specify the whole of the above sequence as a transform $\overline{S}$, then

$$\overline{S} : \{g_i\}_j \rightarrow \{g_i\}_{j+1},$$

and the general sequence is

$$\overline{S} : \{g_i\}_0 \rightarrow \{g_i\}_1 \rightarrow \{g_i\}_2 \ . \ . \ . \ . \rightarrow \{g_i\}_j,$$

expressing the continued transformation of the genetic structure of the original population.

III. Information Generated by the Programme

The programme was set to print certain information for each generation, namely $\{g_i\}_j$, $\overline{P}_j$, and $\sigma_P{}^2$. Each run of the programme gave values simulating the rate of advance under selection over a period of 20–25 generations. Runs were made covering twenty variations of the genetic constants. Two runs were made at each set of values of the genetic constants, starting at different points in the random number sequence, thus giving two replicates.

IV. Values of the Genetic Constants

The genetic structure of the original population, i.e. $\{g_i\}_0$, was set at two values which are termed the P and Q values. They were set as shown below.

(a) P *Value of* $\{g_i\}_0$

Only those types of gametes listed below were represented in the original population.

Type of Gamete	Frequency
0 1 1 0 1 1	0·1111
0 1 1 1 0 1	0·1111
0 1 1 1 1 0	0·1111
1 0 1 0 1 1	0·1111
1 0 1 1 0 1	0·1111
1 0 1 1 1 0	0·1111
1 1 0 0 1 1	0·1111
1 1 0 1 0 1	0·1111
1 1 0 1 1 0	0·1111

It can be seen that if the six loci are separated into two linkage groups of three loci, then this specification of the genetic structure of the original population is such that two cross-overs must occur before individuals homozygous for the 1 type of allele can occur.

This programme does not allow any variation of the dominance relations, being fixed on complete dominance of the 1 alleles. The P value of $\{g_i\}_0$ therefore specified, if the phenotypic component of the 1 allele at each locus is 1·0, that the mean phenotype of the original population was 5·2. Since the maximum phenotype is 6·0, the runs with the value of $\{g_i\}_0$ are essentially on a quantal basis.

(*b*) Q *Value of* $\{g_i\}_0$

Only those types of gametes listed below were represented in the original population.

Type of Gamete	Frequency
0 0 1 0 0 1	0·1111
0 0 1 0 1 0	0·1111
0 0 1 1 0 0	0·1111
0 1 0 0 0 1	0·1111
0 1 0 0 1 0	0·1111
0 1 0 1 0 0	0·1111
1 0 0 0 0 1	0·1111
1 0 0 0 1 0	0·1111
1 0 0 1 0 0	0·1111

The second value of $\{g_i\}_0$ is such that four cross-overs must occur before individuals homozygous for the 1 type of allele can occur. The mean of the original phenotype was 3·2.

The 20 runs of the programme, each replicated, are therefore divided into two sets of 10 runs: one with the P value of $\{g_i\}_0$, the other with the Q value of $\{g_i\}_0$.

The runs were made at two sizes of population. The first value was set to correspond to a medium-sized population with a low intensity of selection (of the order of selection experiments with larger animals). The number of progeny, N, was set at 100, the number of parents, N', at 50. The second value was set to correspond to a small population with a high intensity of selection (of the order of selection experiments with *Drosophila*). The number of progeny was set at 40, the number of parents at 4. These two values of the size of population are termed the "large" and "small" values respectively. Runs of the programme were made at five different values of the linkage relations of the six loci. These are listed below:

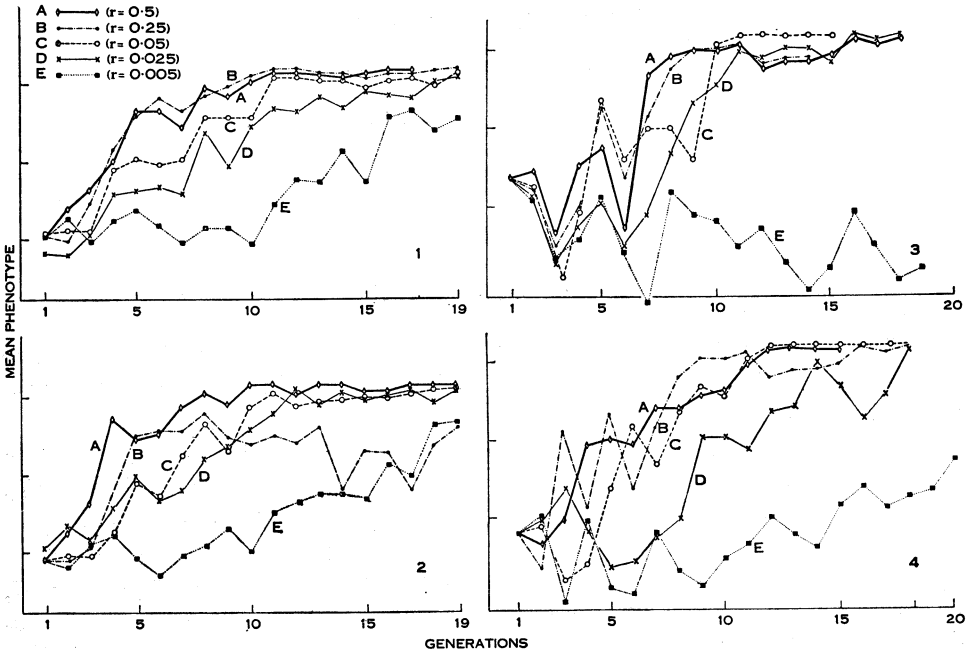
Linkage	Locus					
	a	*b*	*c*	*d*	*e*	*f*
A	0·5	0·5	0·5	0·5	0·5	
B	0·25	0·25	0·5	0·25	0·25	
C	0·05	0·05	0·5	0·05	0·05	
D	0·025	0·025	0·5	0·025	0·025	
E	0·005	0·005	0·5	0·005	0·005	

The A value of the linkage relations specifies genetic independence of the loci, and the other values specify various degrees of linkage of the loci within two "chromosomes" of three loci, i.e. *abc* and *def*.

V. Results

(a) *The* P *Set of Runs*

The results of the runs at the P value of $\{g_i\}_0$ are shown in Figures 1, 2, 3, and 4, in which the values of P for each generation are plotted against generation. The



Figs. 1–4.—The simulated effects of selection are shown as mean phenotype of progeny plotted against generation. Figure 1 shows the P set of runs for a population of size 100 progeny/50 parents. Figure 2 shows the replicates of these runs at different starting points in the set of random numbers. Figures 3 and 4 are as for Figures 1 and 2, but for a population of size 40/4. The five selection lines show the effect of different degrees of linkage from A (independence) to E (tight linkage).

first feature of these results is the similarity with experimental data. The programme is, as far as inspection of these "selection lines" can tell, simulating natural systems.

Considering the large populations, i.e. Figures 1 and 2, there is a correlation of rate of advance with tightness of linkage. The effect of linkage is negligible for $r = 0{\cdot}25$, slight for $r = 0{\cdot}05$ and $0{\cdot}025$, and marked for $r = 0{\cdot}005$. At the latter value there are distinct breaks in the selection lines.

The small populations, i.e. Figures 3 and 4, show, even though there is a much greater intensity of selection, a distinctly slower rate of advance and far greater variability from generation to generation. The effect of linkage is considerably

exaggerated, particularly in the E run of Figure 4, in which no advance occurred even after 20 generations of selection.

Comparisons between the replicate runs (Figs. 5 and 6) show fairly close agreement in the large populations, but in one run, the B run of Figure 2, the selection line drifted downwards for several generations before regaining its earlier advance. In the small populations, differences between replicates are, as is expected, much greater.

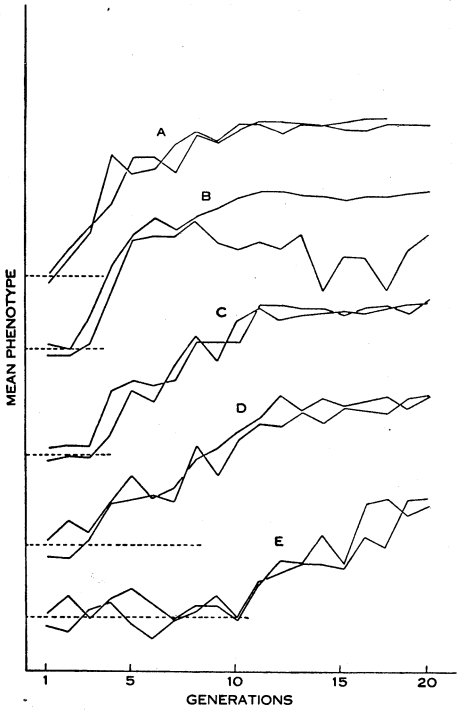


Fig. 5

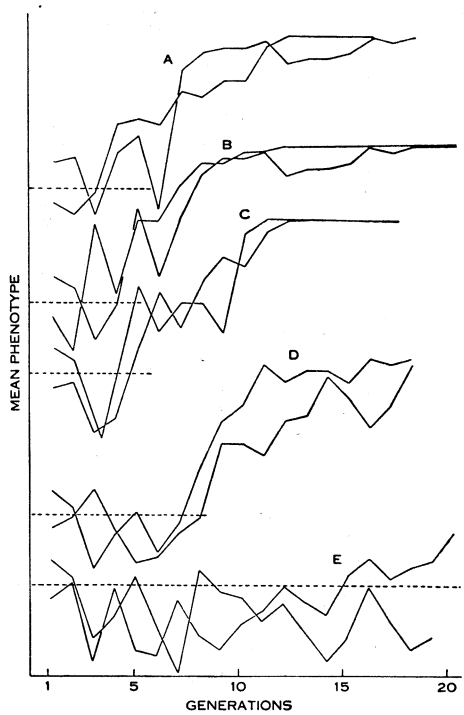


Fig. 6

Figs. 5 and 6.—The replicate runs shown separately in Figures 1–4 are shown plotted together.

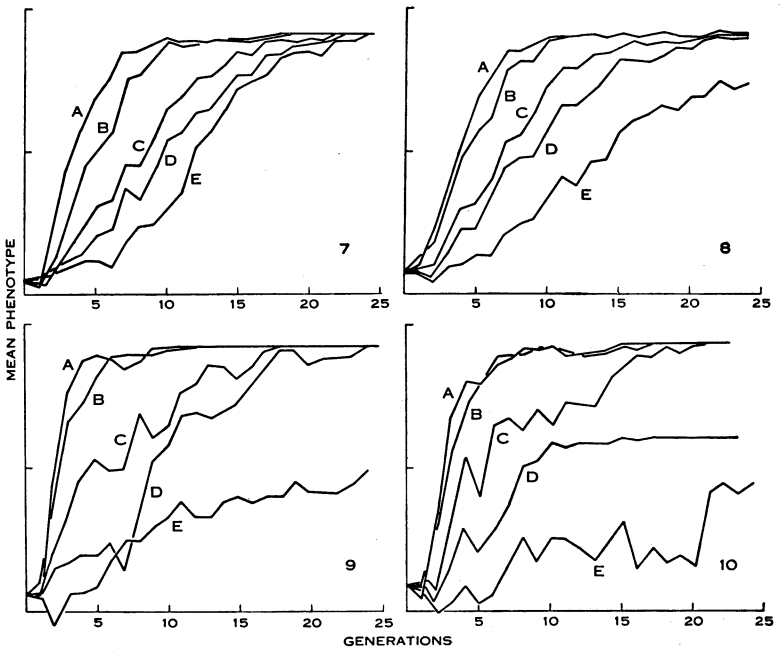
(*b*) *The* Q *Set of Runs*

The results of the runs at the Q value of $\{g_i\}_0$ are shown in Figures 7, 8, 9, and 10.

The large populations show a clearer relation of the rate of advance under selection to the amount of recombination. Even the B runs ($r = 0{\cdot}25$) show a slower rate of advance, which was not apparent in the P set of runs.

The small populations show a markedly different pattern from the large populations. Firstly, there are no differences between the A and B runs, i.e. reducing recombination from 0·5 to 0·25 has no effect. Secondly, comparing the A and B runs in the large and small populations, there is a clear difference in the rate of advance; the higher intensity of selection in the small populations is effective in the Q set of runs, whereas it was not effective in the P set of runs. Thirdly, the effect of the

more intense selection in the small populations is not very marked in the run at $r = 0{\cdot}05$ (the C runs) and there are possible discontinuities in the rate of advance. Fourthly, discontinuities are more marked in the D runs and are very clear in the E runs. The E run of Figure 9 shows two advances separated by a period of 10 generations when no advance occurred. Finally, in the D run of Figure 10 fixation of a recessive allele, i.e. an 0 allele, occurred.



Figs. 7–10.—As for Figures 1–4, but for the Q set of runs at population sizes of 100/50 (Figs. 7 and 8) and 40/4 (Figs. 9 and 10) respectively.

VI. Discussion

There has been considerable discussion on the importance of linkage in quantitative variation, led by Mather in his construction of the "polygene concept" (cf. Mather 1943). The essence of this concept is that the individual genes determining the variation of a quantitative character occur in linked units, which he terms polygenes. The initial evidence on which he based the concept came from the results of selection experiments. "Selection experiments have shown that such combinations exist in *Drosophila melanogaster* (Mather 1941; Wigan 1942; Sismanidis 1942) and they can be inferred from Winter's (1929) results with maize. The great storage capacity is well brought out by Payne's (1918) selection for increase in the number of scutellar bristles in *Drosophila*. . . . . Sismanidis (1942) has confirmed Payne's results and has been able to assign the sudden responses to selection to recombinations in particular chromosomes" (Mather 1943). Sismanidis's results are illustrated in Figure 11, which is after Figure 5 of Mather (1943). The E run from Figure 10 is shown in Figure 12 for comparison with Sismanidis's results.

It is clear from the runs of this programme that the detection of the occurrence of linkage in a genetic system from heterogeneity of the rate of advance under selection is unlikely to be possible in large populations. Further where linkage is detectable it is likely to be fairly close, i.e. $r \simeq 0{\cdot}005$.

The programme used to produce the simulation of the effects of selection suffers from two major defects. These are (1) limitation of the number of loci to less than seven, and (2) lack of provision for other dominance relations than complete dominance. It has, therefore, not been prepared in a form for general use, and a second programme has been written which does not contain these defects. This programme will be discussed in a later paper, and prepared in a form for general use on any ILLIAC-type computer.

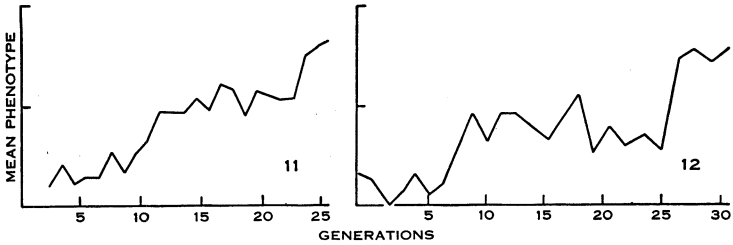


Fig. 11.—Selection for number of chaetae in *Drosophila* (after Mather 1943).
Fig. 12.—The *E* run from Figure 10 shown for comparison with Figure 11.

VII. Acknowledgments

I am indebted to the staff of the Adolph Basser Computing Laboratory, University of Sydney, particularly Dr. J. O. Bennett, Dr. B. Chartres, and Mr. J. Butcher. Miss J. Ogilvie prepared and checked the many tapes required in this study.

VIII. References

Fraser, A. S. (1957).—*Aust. J. Biol. Sci.* **10**: 484–91.

Mather, K. (1941).—*J. Genet.* **41**: 159–74.

Mather, K. (1943).—*Biol. Rev.* **18**: 32–64.

Payne, F. (1918).—*Ind. Univ. Stud.* **5**: 1.

Sismanidis, A. (1942).—*J. Genet.* **44**: 204–15.

Wigan, L. G. (1942).—*Heredity* **3**: 53–65.

Winter, I. (1929).—*J. Agric. Res.* **39**: 451–70.