

SIMULATION OF GENETIC SYSTEMS BY AUTOMATIC DIGITAL COMPUTERS

III. SELECTION BETWEEN ALLELES AT AN AUTOSOMAL LOCUS

By J. S. F. BARKER*

[Manuscript received July 7, 1958]

Summary

A new approach to analysis of the effect of selection on gene frequencies is described. An electronic digital computer (the SILLIAC) is used to simulate the selection processes that operate in populations. The Monte Carlo method permits inclusion of stochastic processes so that results should simulate those in natural or experimental populations.

The programme simulates selection between two autosomal alleles and allows for selection at four stages of the life cycle, viz.:

- (i) Zygote selection: ability to survive from fertilization to sexual maturity.
- (ii) Reproductive selection: differential reproductive ability of different genotypes.
- (iii) Selection between gametes at meiosis in heterozygotes.
- (iv) Selection between gametes on their ability to take part in fertilization.

These selective values and population size can be varied to investigate different genetic situations.

To test the programme, two experiments reported in the literature have been simulated, viz.:

- (i) Selection between *ST* and *CH* chromosomal arrangements in the third chromosome of *Drosophila pseudoobscura* (Dobzhansky and Pavlovsky 1953).
- (ii) Selection between *glass* and wild type in *Drosophila melanogaster* (Merrell and Underhill 1956).

Close agreement was obtained with the results of Dobzhansky and Pavlovsky. The importance of an adequate estimate of generation length in making these comparisons is discussed. Agreement with the results of Merrell and Underhill was not so close. Possible reasons for this in terms of selective mating are discussed.

The results show that it is possible to simulate selection between two alleles at an autosomal locus by using automatic digital computers.

I. INTRODUCTION

In this paper, a new approach to the analysis of selection is described. Insofar as certain parameters must be specified at the beginning of the analysis, the results are restricted. However, as these parameters may be varied at will, the extent to which the analysis may be taken as general can be arrived at by repeated analysis with different basic parameters. In this way, the extent of the generality of the conclusions and the particular factors of most importance in determining the results are revealed.

* Department of Animal Husbandry, University of Sydney.

The method simulates the processes which go on in a population and calculates in an electronic digital computer (the SILLIAC) the effects which these processes will have on gene frequency. The method is basically that described by Fraser (1957*a*, 1957*b*) in which the Monte Carlo method has been adapted for genetic analysis of populations. This method does not involve the use of complex mathematics common to this field. In their place, long series of simple arithmetic steps are performed in an electronic computer. One feature of this approach is that the calculations follow sequences closely related to the life cycles of animals, and are therefore much easier to understand than more subtle methods. An important feature of this "arithmetic" method is the inclusion of "stochastic simulants", i.e. arithmetic processes which introduce random variation into the main sequence. For a specified set of population parameters, inclusion of a different "stochastic simulant" for each of a number of runs means that each series of calculations will produce a different answer. Thus a number of such answers can be combined to give both the mean and the variance of the processes under investigation.

In the analysis discussed here, one autosomal gene with two alleles A and a is segregating in a bisexual population in which there is no overlap of generations. A programme has been written which sets the SILLIAC to simulate this genetic model, allowing for selection between the two alleles at different stages of the life cycle. The various selection coefficients and the population size can be varied to investigate different genetic situations. In this paper, the operations of the programme are described and the validity of the simulation tested.

II. SELECTION PROCESSES

The numbers of fertilized zygotes are taken as the reference point for the beginning of each generation. The selection processes which go on in the population are as follows:

(i) The first selection is on the ability of genotypes to survive from fertilization to sexual maturity. This may be different in the two sexes which are therefore treated separately. It is assumed that each individual of a given genotype has a certain probability of surviving to sexual maturity. However, there is a chance element in this survival. Parents are selected, not in exact proportions according to the probability of survival, which is the mean result in an infinite population, but according to the operation of a random or chance effect on the probability. The number of parents so selected is calculated.

(ii) Once the parents are selected, they must generate gametes. Selection operates again in that certain females may be more fecund than others, or more readily mated, or more long-lived. In the male, some may be more aggressive at getting mates, or more long-lived. Again, the probability of a male or female producing gametes which take part in fertilization is specified and the gametes picked out by the operation of a chance effect on these probabilities, rather than as clear-cut proportions.

(iii) In the formation of gametes in heterozygotes, they may not be produced in a 1 : 1 ratio. Selection of this type may be specified and the number of A

and a gametes determined, again with the introduction of the chance effect. For example, "sex ratio" in *Drosophila pseudoobscura* (Sturtevant and Dobzhansky 1936) is of this type.

(iv) Once the gametes are formed, selection again may operate to determine what proportion of gametes of each kind are successful, for example, the S genes in plants (Bateman 1952).

It should be noted that if it is specified that there is no selection of types (ii), (iii), and (iv) above, then type (i) can be used as relative adaptive values, specifying overall selection.

TABLE 1
INFORMATION REQUIRED ABOUT ZYGOTES

Genotype	No. of Females	Zygote Selective Value	Genotypic Reproductive Coefficient	No. of Males	Zygote Selective Value	Genotypic Reproductive Coefficient
AA	N_{f_1}	w_f	x_{f_1}	N_{m_1}	w_m	x_{m_1}
Aa	N_{f_2}	$w_f(1-h_fs_f)$	x_{f_2}	N_{m_2}	$w_m(1-h_ms_m)$	x_{m_2}
aa	N_{f_3}	$w_f(1-s_f)$	x_{f_3}	N_{m_3}	$w_m(1-s_m)$	x_{m_3}
Totals	N_f		1	N_m		1

III. OPERATIONS OF THE PROGRAMME

At the start, the alleles A and a will be distributed with a certain frequency in the population and will be present in zygotes AA , Aa , and aa , the frequency of which must be specified. The starting point is taken as the moment of fertilization. The information required about zygotes can be tabulated as in Table 1. The intensity of selection is s ; introducing h makes the formulae general for all degrees of dominance (Lush 1948). With this information it is possible to take the population to the point where gametes are produced.

The first step is to determine the number of females which reach sexual maturity. To do this N_{f_1} , N_{f_2} , and N_{f_3} are multiplied by the appropriate zygote selective values to give the expected number of females surviving to sexual maturity, viz. N'_{f_1} , N'_{f_2} , and N'_{f_3} , whose sum is N'_f . The actual number will vary in a random way about the expected. The random choice is put in by generating a random number v_i , which lies between 0 and 1. This random number v_i is multiplied by N'_f and the product will lie somewhere between 0 and N'_f . If it lies between 0 and N'_{f_1} , it is taken to represent an AA individual, if it lies between N'_{f_1} and $N'_{f_1} + N'_{f_2}$ it is taken to be an Aa individual, and if it lies between $N'_{f_1} + N'_{f_2}$ and N'_f , it is taken to be an aa individual. This process is repeated with N'_f random numbers so that N'_f individuals are generated and classified as AA , Aa , and aa

in such a way that the mean proportions in an infinite population will be $N'_{f_1} : N'_{f_2} : N'_{f_3}$, but in a finite population will deviate from these values by chance. This is repeated for males and we now have the numbers of females and males at sexual maturity as follows:

Genotype	No. of Females	No. of Males
AA	N_{F_1}	N_{M_1}
Aa	N_{F_2}	N_{M_2}
aa	N_{F_3}	N_{M_3}
Totals	N'_f	N'_m

At this stage, a device is introduced to prevent the population increasing or decreasing indefinitely. The expected size (E) of the population is specified and the average number of offspring of each female is made to be E/N'_f as will be shown later. Thus the size of the population produced by females will tend towards E but will fluctuate around it according to chance. The relative reproductive abilities of different genotypes must also be introduced. Taking females first, let the average number of offspring of each genotype be P_1 for AA , P_2 for Aa , and P_3 for aa , so that

$$\frac{1}{3}(P_1 + P_2 + P_3) = E/N'_f.$$

As

$$x_{f_1} + x_{f_2} + x_{f_3} = 1,$$

$$3E/N'_f \times x_{f_1} = P_1,$$

$$3E/N'_f \times x_{f_2} = P_2,$$

$$3E/N'_f \times x_{f_3} = P_3,$$

P_1 , P_2 , and P_3 may thus be calculated.

We now have N_{F_1} AA females tending to produce an average of P_1 offspring each, N_{F_2} Aa tending to produce P_2 , and N_{F_3} aa tending to produce P_3 . The actual numbers are formed by a further random process in which P_i is assumed to be the mean of a Poisson distribution. This is an ideal situation and, as Crow (1954) shows, will be most accurate where $P_i = 2$. By calculating e^{-P_1} and multiplying together a series of random numbers v_0, v_1, \dots, v_n until the product is less than e^{-P_1} , a random Poisson deviate n is formed which is taken to be the actual number of offspring produced by a particular AA female. This process is repeated N_{F_1} times for AA females. Using e^{-P_2} , it is repeated N_{F_2} times for Aa females, and using e^{-P_3} , N_{F_3} times for aa females. We now have the number of gametes from each female which take part in fertilization. They still have to be sorted into A and a . All those produced by AA females will be A ; call the total number of these F'_1 . All those from aa females will be a ; call the total number of these F'_3 . Call the total number from Aa females F'_2 . If there is selection between

gametes at meiosis in heterozygotes with selection coefficient of A equal to r_{f_1} , and of a equal to r_{f_2} , the expected proportions of A and a gametes will be

$$F'_2/2 \times r_{f_1} : F'_2/2 \times r_{f_2}.$$

Use of the same operation as was used before to choose the actual number of females of each genotype surviving to sexual maturity—in this case operating on the above expected proportions—makes the choice of A and a random. As F'_2 has been determined as the number of gametes from Aa females, this random transform must generate a total of F'_2 A and a gametes. By adding the numbers of A and a gametes we get the total of A and a gametes produced by females.

There is one further step to take: selection between gametes at fertilization. This is done by a random transform as before. The total number of gametes produced by females and that will take part in fertilization are made to equal $F'_1 + F'_2 + F'_3$, which is called F' .

These processes are repeated for males to get the numbers of male A and a gametes that will take part in fertilization, where the total is M' . The processes are not such that F' necessarily equals M' , while, for fertilization, there must be equal numbers of male and female gametes. However, they are not likely to be very different and the number of offspring generated is set to equal F' or M' , whichever is the smaller.

Fertilization is then performed by random combination of gametes. One male and one female gamete are taken and the combination tested to see if it is AA , Aa , or aa . Another pair is taken, and so on. As this is done, the numbers of male and female A and a gametes are counted as they are used and tested against the numbers available for fertilization, until all the gametes of each type from each sex are used. Each offspring individual, once its genotype is determined, is operated on by a random transform on an expected 1 : 1 ratio to determine if it is to be male or female.

The gene frequencies in males, in females, and overall are calculated from the total numbers of males and females of each genotype so generated in the offspring generation. The programme then prints out this information on the numbers of males and females of each genotype and the gene frequencies. It then uses these to produce the following generation, and so on. The programme can be set to run for any number of generations.

IV. RESULTS

The programme has been tested by trying to simulate experiments that are reported in the literature. Attempts have been made in some of these experiments to determine the nature and extent of the selection forces operating. These estimated selection coefficients can be used as specifications for the programme and the results of the simulated populations compared with those of the selection experiments. Two such experiments have been simulated, one with a large population, the other with a small.

(a) Large Population Size

Using population cages, Dobzhansky and Pavlovsky (1953) studied competition between the *ST* and *CH* chromosomal arrangements of *Drosophila pseudoobscura* in four populations. From their results, they calculated that the adaptive values of the genotypes were: *ST/ST* 0.895, *ST/CH* 1.000, and *CH/CH* 0.413.

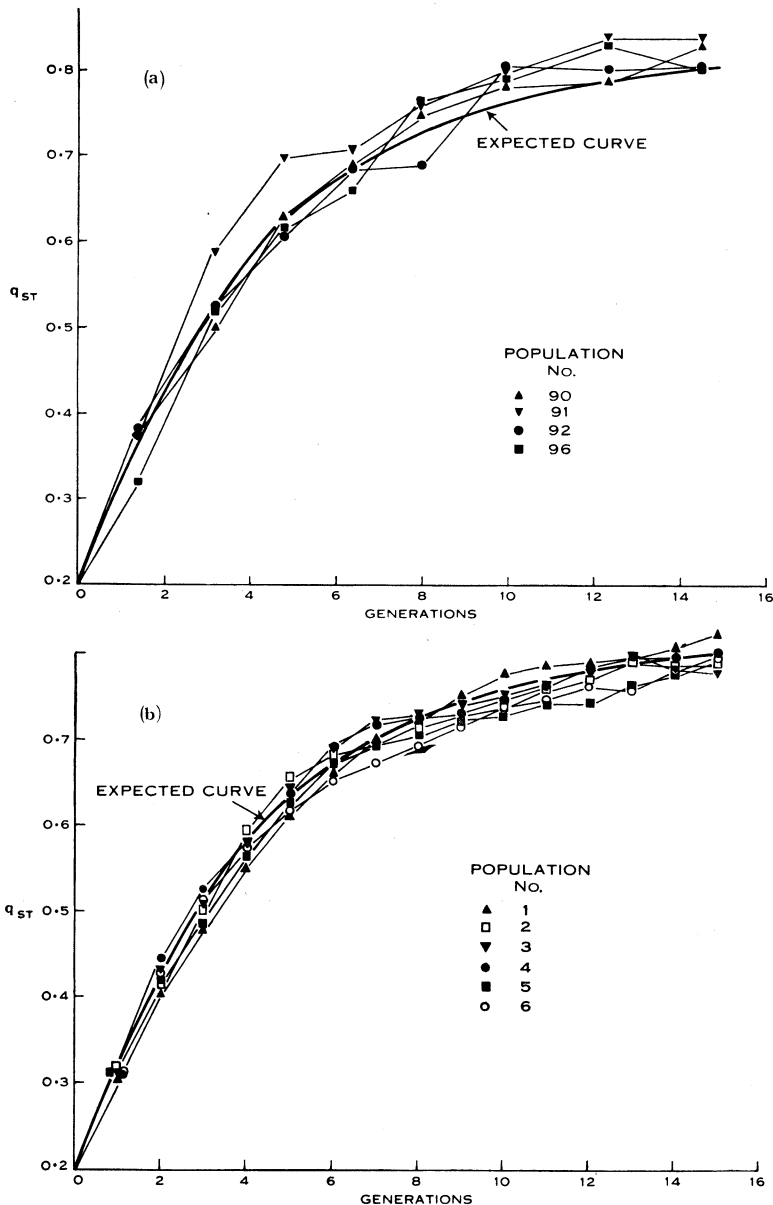


Fig. 1.—Frequencies of chromosomes with the *ST* gene arrangement. (a) Results of Dobzhansky and Pavlovsky's (1953) four experimental populations. (b) Results of six simulated populations. The smooth solid curve in both sets is the expected curve with adaptive values of *ST/ST* : *ST/CH* : *CH/CH* genotypes being 0.895 : 1 : 0.413 respectively.

These adaptive values have been used as relative zygote selective values for both males and females. It was specified that there was no genotypic reproductive selection and no gamete selection. However, chance effects still operate at these stages in the cycle. E was specified at 4000. Six replicate runs were made with these specifications. The results are shown in Figure 1. The generation length of the experimental population was assumed to be 25 days (Dobzhansky and Pavlovsky 1953).

(b) *Small Population Size*

Merrell and Underhill (1956) using population bottles, studied changes in gene frequency in various mutants of *D. melanogaster* when in competition with their wild-type alleles. The changes in frequency were considered to be mainly a function of selective mating.

TABLE 2
RESULTS OF FEMALE-CHOICE MATING TESTS
Results between *glass* (*gl*) and wild-type flies (quoted
from Merrell and Underhill 1956)

Female Genotype	Successful Male (%)	
	$+/+$	gl/gl
$+/gl$	96.3	3.7
gl/gl	100.0	0.0

The competition between *glass* (*gl*) and wild type has been simulated. Relative zygote selective values for both males and females have been taken as: $+/+$ 1.000, $+/gl$ 1.000, and gl/gl 0.908, as given by Merrell and Underhill. These are not calculated on the basis of given values of h and s but are overall values determined experimentally. They quote also the results of female-choice mating tests which are given in Table 2.

From these, relative genotypic reproductive coefficients for males have been taken as: $+/+$ 0.48, $+/gl$ 0.48, and gl/gl 0.04.

As there are no data on the occurrence or extent of selective mating by males when given a choice of females, the coefficients for females must be assumed to be identical for each genotype. The female coefficients are therefore: $+/+$ 0.33, $+/gl$ 0.33, and gl/gl 0.33. No other selection was involved in the simulated populations and E was specified as 180. The results are shown in Figure 2. The curve shown for the experimental populations is the average of nine populations as given by Merrell and Underhill. Generation length in these populations has been taken as 24 days (Merrell 1953).

V. DISCUSSION

(a) Large Population Size

It will be noted from Figure 1 that four of the six simulated populations (Nos. 3, 4, 5, and 6) have identical frequencies of *ST* in generation 1. This results from an inherent difficulty in the programme. Dobzhansky and Pavlovsky's populations were started with 600 *ST/CH* individuals and 900 *CH/CH*. With the given adaptive values and assuming no random processes, all the heterozygotes and 372 of the *CH/CH* individuals will contribute gametes to form generation 1. In the population cage, these flies will give rise to a population of at least 4000 in generation 1. However, in the simulated populations with a Poisson distribution of progeny numbers,

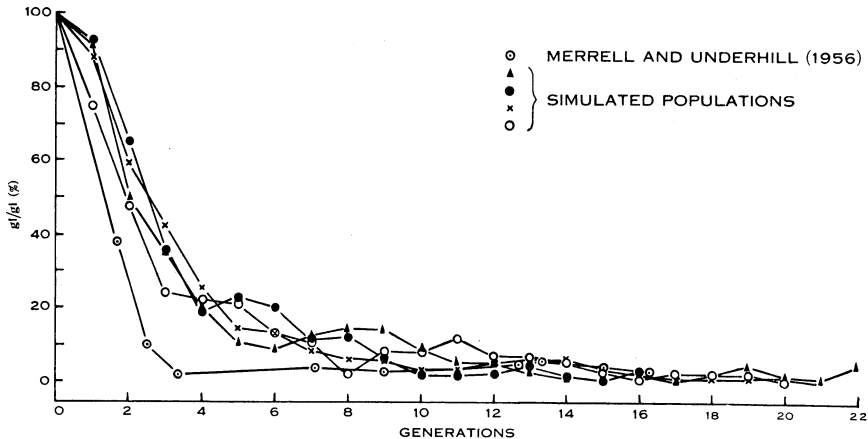


Fig. 2.—Results of competition between wild type and *glass*. ○ Average of the nine experimental populations of Merrell and Underhill (1956). ▲, ●, ×, ○ Simulated populations.

it takes the population two or three generations to reach equilibrium size. Therefore, the expected numbers of each genotype in generation 1 with a population size of 4000 have been calculated and these four simulated populations started from this point. The first two runs (simulated populations Nos. 1 and 2) were done using the numbers of each genotype as used by Dobzhansky and Pavlovsky as the starting point in generation 0.

Figure 1 shows that the results of the experimental and simulated populations agree fairly closely. One difference is the relative lack of generation to generation variation in the simulated compared to the experimental populations. This may be due to:

- (i) In the simulated populations, generations are discrete, i.e., there is no generation overlap.
- (ii) The variation in the experimental populations will include sampling variations, as each point is based on a sample of 300 chromosomes.
- (iii) The adaptive values in the simulated populations are constant. In experimental populations, they would probably show some variation from generation to generation. This would increase generation to generation variation in the experimental populations.

It appears that overlapping generations have little effect on the results (except for the variation effects noted above). The question of actual generation length is important in this connection. Dobzhansky and Pavlovsky's adaptive value estimates depend on the assumption of a generation length of 25 days. If this is incorrect, the adaptive values are also incorrect. There is a suggestion from their results that they may be slightly incorrect. In their Figure 2, all except one of the results lie within the 95 per cent. confidence limits for the expected values of ST frequency. However, the estimates at 80, 120, and 160 days lie mainly below the expected curve but with the mean biased upwards by population 91, while those at 250, 310, and 365 days lie mainly above the expected curve. Therefore, assuming a slightly different generation length may give a better fit to the data. This question of generation length in population models has been discussed briefly by Moree (1955).

(b) *Small Population Size*

The rate of change in percentage of gl/gl from the first to the third generation is similar in the simulated and experimental populations, but slightly lower in the former. It may be that *glass* females are not quite as efficient as wild type in mating, whereas they may have been assumed the same in the simulated populations.

In the first generation, however, and from the third to the twelfth, the curves show different trends. The lag (i.e. slower change) in the simulated populations in the first generation is difficult to explain. The experimental populations were started by placing one wild-type male into an established *glass* population. The immediate decrease in gl/gl frequency is surprising, as one would not expect this one male to be so successful in mating when so many gl/gl males are present. However, the first estimate in the experimental populations is at 30 days, and would include both first and second generation flies. Thus, the lag in the first generation may be real, but followed by a larger decrease in gl/gl frequency in the second generation than is shown by the simulated populations. This could operate through selective mating against gl/gl females. Some factor such as this, increasing the selection against *glass*, must be operating.

This is shown by the fact that seven runs were made in the SILLIAC but three of these went to 100 per cent. *glass* in the first generation. Merrell and Underhill (1956) state that, of their 10 experimental populations, only one went to 100 per cent. *glass*.

The slower change in the simulated populations from the third to the twelfth generation could be explained by the fact that, in these populations, the genotypic reproductive coefficients are constant throughout. It seems likely that as the frequency of *glass* males decreases, their mating success relative to wild type will also decrease. The number of females available to the *glass* males will be the important factor. *glass* males have a lower mating ability than wild type so that with a higher frequency of wild-type males than *glass* males in the population, the former would cover most of the females. This would result in an increase in the rate of change in gl/gl frequency.

The results of the two comparisons made above show that it is possible to simulate in an automatic digital computer the operations of selection between two alleles at an autosomal locus.

This programme will be used to generate families of curves relating changes of gene frequency of two autosomal alleles in competition to the parameters of population size and nature and intensity of selection.

VI. ACKNOWLEDGMENTS

I am indebted to the staff of the Adolph Basser Computing Laboratory, particularly to Dr. J. M. Bennett, Dr. B. A. Chartres, and Mr. J. C. Butcher; to Dr. A. S. Fraser and Dr. P. J. Claringbold for their tuition in various aspects of programming; and to Dr. J. M. Rendel and Dr. A. S. Fraser for their critical examination of the manuscript.

VII. REFERENCES

- BATEMAN, A. J. (1952).—Self-incompatibility systems in angiosperms. I. Theory. *Heredity* **6**: 285–310.
- CROW, J. F. (1954).—Breeding structure of populations. II. Effective population number. In "Statistics and Mathematics in Biology." Ch. 43. p. 543–56.
- DOBZHANSKY, TH., and PAVLOVSKY, OLGA (1953).—Indeterminate outcome of certain experiments on *Drosophila* populations. *Evolution* **7**: 198–210.
- FRASER, A. S. (1957a).—Simulation of genetic systems by automatic digital computers. I. Introduction. *Aust. J. Biol. Sci.* **10**: 484–91.
- FRASER, A. S. (1957b).—Simulation of genetic systems by automatic digital computers. II. Effects of linkage on rates of advance under selection. *Aust. J. Biol. Sci.* **10**: 492–9.
- LUSH, J. (1948).—"The Genetics of Populations." (Mimeographed Edition.) (Iowa State College: Ames.)
- MERRELL, D. J. (1953).—Selective mating as a cause of gene frequency changes in laboratory populations of *Drosophila melanogaster*. *Evolution* **7**: 287–96.
- MERRELL, D. J., and UNDERHILL, J. C. (1956).—Competition between mutants in experimental populations of *Drosophila melanogaster*. *Genetics* **41**: 469–85.
- MOREE, R. (1955).—The question of generation length in population models of *Drosophila melanogaster*. *Drosophila Inform. Serv.* **29**: 142–3.
- STURTEVANT, A. H., and DOBZHANSKY, TH. (1936).—Geographical distribution and cytology of "sex-ratio" in *Drosophila pseudoobscura* and related species. *Genetics* **21**: 473–90.