

A computer model of speciation by founder effects

B. CHARLESWORTH AND D. B. SMITH*

School of Biological Sciences, University of Sussex, Brighton BN1 9QG

(Received 11 August 1981 and in revised form 2 November 1981)

SUMMARY

A computer model of a two-locus genetic system with epistatic selection was used to investigate factors influencing the probability of the origin of reproductive isolation, due to a genetic revolution following a founder event (Mayr, 1954; Carson, 1975). Restricted population size can sometimes cause such a system to drift from one equilibrium to another, which can result in loss of fitness to hybrids with the ancestral population. The chance of such an event was found to be low unless the bottleneck in population size associated with the founder event was preceded by many generations of relaxed selection. It is highest when the bottleneck is not prolonged and when the population size during the bottleneck is not too small. It seems to be difficult to achieve a high degree of reproductive isolation in one step by this method, and it is concluded that it is unlikely to be a major cause of rapid speciation, although it could be a contributory factor.

1. INTRODUCTION

The classical view of allopatric speciation is that reproductive isolation between closely related species is acquired as a result of the gradual genetic divergence of spatially isolated populations (e.g. Muller, 1940; Mayr, 1942). Emphasis has usually been placed on the role of natural selection in creating distinct, harmonious combinations of genes in different populations, but which interact to produce inferior genotypes when hybrids are formed (cf. Dobzhansky, 1951, chap. 7). More recently, Mayr (1954, 1963, chap. 17) has argued that an isolated population of small size, such as may arise when a new habitat is colonised, might undergo a 'genetic revolution' in which genetic drift cause a shift in the state of the population from one set of coadapted genotypes to another. As a result, such a population might rapidly become partially reproductively isolated from its ancestral population. Carson (1968, 1970, 1975) has elaborated on this idea in his 'founder-flush' model of speciation, which stresses the importance of a period of relaxed selection associated with a population flush, preceding the reduction in population size involved in a founder event. Such a relaxation of selection allows an increase in the frequency of normally unfavourable recombinant genotypes, and so increases the chance that they will be established at high frequency by drift.

* Present address: Moltano Institute of Parasitology, University of Cambridge, Cambridge.

The concept of relatively rapid speciation by means of random events rather than selection has become a cornerstone of the theory of evolution by 'punctuated equilibria' (Gould & Eldredge, 1977; Stanley, 1979).

Little theoretical work seems to have been done to establish the conditions under which this model of speciation is likely to work, apart from the discussion of Templeton (1980*a, b*) which does not explicitly incorporate the effects of selection, and that of Lande (1980), which is concerned with morphological divergence rather than isolating mechanisms. As pointed out by Wright (1978), the genetic process involved is similar to that envisaged in his 'shifting balance' model of evolution (Wright, 1932), in which genetic drift causes populations to move from one selectively stable equilibrium point to another (a peak shift, in Wright's terminology). Templeton (1980*a, b*) has proposed the term 'genetic transilience' for this process, but we prefer to use the older term of Wright. A single locus with heterozygote disadvantage is the simplest genetic system capable of generating a peak shift, since a population fixed for either allele is stable to the introduction of its alternative. This is most likely to be of importance in the case of chromosomal re-arrangements which cause segregation disturbances when heterozygous. The probability of establishment of such rearrangements has been extensively analysed by Wright (1941), Bengtsson & Bodmer (1976), Lande (1979) and Slatkin (1981), and will not be considered further here. The models of Carson, Mayr and Templeton involve at least two loci with epistatic effects on fitness. Studies of the effects of drift on such systems have been made by Felsenstein (1974), Avery (1978) and Avery & Hill (1979), but their results cannot be used directly to investigate the probability of genetic revolutions. The purpose of this paper is to describe the results of some computer simulations that permit an investigation of the probability of peak shifts in relation to speciation. We shall not attempt to model peak shifts in relation to the evolution of the species as a whole, as conceived by Wright.

2. THE MODEL AND ITS SIMULATION

We studied discrete-generation, random-mating populations in which selection was acting on two loci with alleles *A, a* and *B, b* respectively. The recombination fraction for this pair of loci is denoted by *R*. The state of the population in a given generation is described by the frequencies of alleles *A* and *B*, and the coefficient of linkage disequilibrium *D* (c.f. Felsenstein, 1974). It is easy to construct fitness matrices for two-locus systems such that two alternative stable equilibria can exist in an infinite population. In the simplest case, these differ only in the sign of *D* (Lewontin & Kojima, 1960). Models can also be constructed in which the alternative equilibria differ in gene frequency as well as *D*, and we chose to work mainly with such a model, described by Lewontin (1974, p. 296). The fitness matrix is shown in Table 1. With tight linkage, two stable equilibria exist, e.g. with *R* = 0.05, the alternative *D* values are +0.191 and -0.177, with corresponding alternative frequencies of *A* 0.580 and 0.539, and frequencies of *B* 0.560 and 0.518. The equilibrium frequencies of the gametes *AB, Ab, aB* and *ab* for the two

populations are 0.516, 0.044, 0.064, 0.376 and 0.102, 0.416, 0.437 and 0.045 respectively. The mean fitnesses of the two equilibrium populations are 0.674 and 0.637 respectively, whereas the mean fitness of an F_1 hybrid population formed by combining pairs of gametes at random from the alternative equilibrium populations is 0.520. This F_1 fitness loss corresponds to the degree of hybrid inviability or sterility that is brought about as a consequence of a peak shift at this pair of loci.

Table 1. *Fitness matrix used in most of the simulations*

	<i>AA</i>	<i>Aa</i>	<i>aa</i>
<i>BB</i>	0.5000	0.5000	0.3750
<i>Bb</i>	0.5625	1.000	0.3125
<i>bb</i>	0.3750	0.4375	0.3750

The initial conditions with respect to gamete frequencies determine which of these equilibria will be approached in an infinite population. In order to simulate the founder-flush model, the populations were started close to the $D > 0$ equilibrium. Selection was then relaxed for a specified number of generations, n , corresponding to the flush phase. The gene frequencies are unaltered by this, but D is reduced to $(1 - R)^n$ of its initial value. For large n , and large initial D , there is thus a substantial increase in the frequencies of the Ab and aB gametes.

A population bottleneck associated with a founder event was simulated by the following procedure. In generation $n + 1$, a finite population of size N was formed by sampling $2N$ gametes from the gamete pool of the previous generation, using pseudo-random numbers to decide the composition of each gamete. Diploid individuals were formed by pairing off gametes formed in this way. If either locus was found to have become fixed at the end of this process, the run was terminated and a new run started with the original gamete pool as the starting point. Otherwise, gamete frequencies were calculated, taking recombination into account, after modifying the array of genotypes by applying the chosen fitness matrix to their frequencies. This was repeated until m generations had elapsed, or the population became fixed at either locus. If the population remained polymorphic, infinite population trajectories were then computed for a number of generations, using the gamete frequencies of the m^{th} generation with finite population size as the starting point. It was then determined which of the two equilibria was being approached. This simulates the outcome of the return to large size of the founding population.

For a given fitness matrix, the parameters that can be varied are thus the recombination fraction (R), the number of generations in the flush period (n), the population size during the bottleneck (N), and the number of generations that the bottleneck persisted (m). One-hundred replicates were run for each parameter combination studied, including the runs in which at least one locus became fixed, preventing a transition from the $D > 0$ to the $D < 0$ equilibrium.

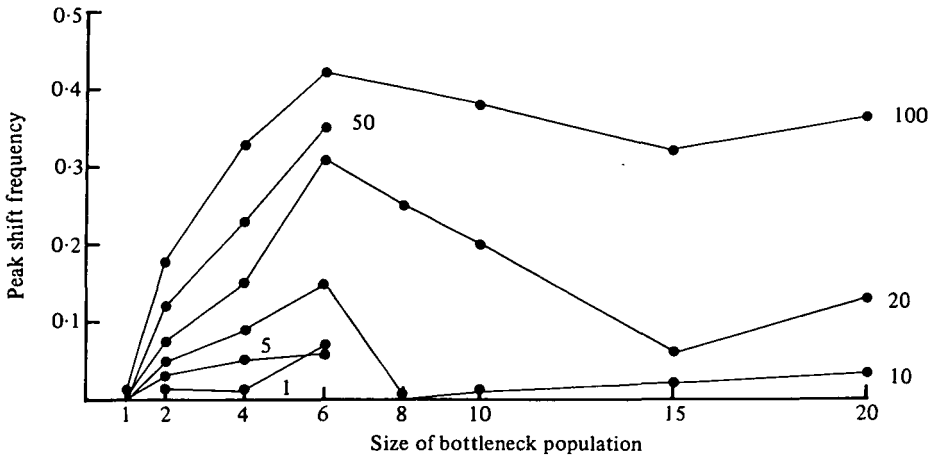


Fig. 1. The frequency of a peak shift is plotted as a function of the population size during the bottleneck, for various values of the number of flush generations, indicated against each line. The bottleneck lasted 4 generations in each case.

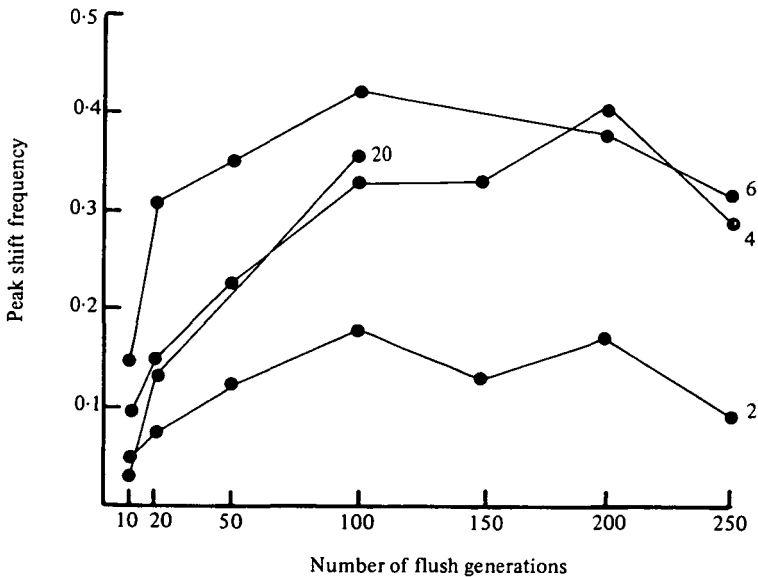


Fig. 2. The frequency of a peak shift is plotted as a function of the number of flush generations, for various values of the population size, indicated against each line. The bottleneck lasted 4 generations.

3. SIMULATION RESULTS

The results of the simulations may most conveniently be summarized by means of graphs displaying the effects of variation in the different parameters (Figs. 1–5). All the simulations displayed in these graphs were run using the fitness matrix of Table 1, and R was 0.05 except in Fig. 5.

(i) *Effects of population size and number of flush generations*

Figs. 1 and 2 shown the effects of N and n on the frequency of a peak shift for the case when there are four generations of restricted population size. There is a clear tendency for this frequency to increase with the number of flush generations, as expected intuitively (Carson, 1975). This tendency levels off when n exceeds

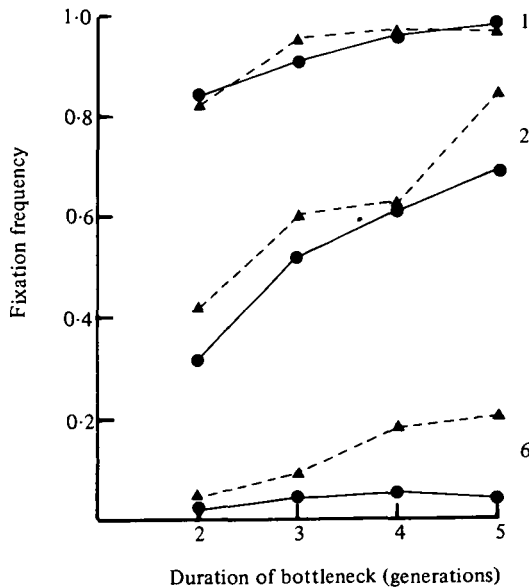


Fig. 3. The frequency of fixation at either locus is plotted against m , the number of generations for which the bottleneck persisted, for various value of n and N , the number of flush generations and size of the bottleneck population. The dashed lines are for $n = 100$ and the full lines for $n = 1$. The numbers against the lines are the values of N .

100 or so (Fig. 2); this is due to the fact that D is vanishingly small after 100 generations without selection, so that no further increase in the Ab and aB gametes can occur. In all the cases studied by us, the frequency of a peak shift never exceeded 10% unless the flush phase lasted at least 10 generations; 20 generations seem to be required for the frequency to exceed 20%. The highest frequency was around 40%, with 50 or 100 unselected generations.

There is a tendency for the frequency of a peak shift to rise rapidly with N when N is small. This is due to the high probability of fixation at one or other locus with small N , as displayed in Fig. 3. For values of N above 6, increases in N tend to reduce the frequency of a peak shift, at least when n is low. For n of 100, the frequency of a peak shift is rather insensitive to the population size when N exceeds 6.

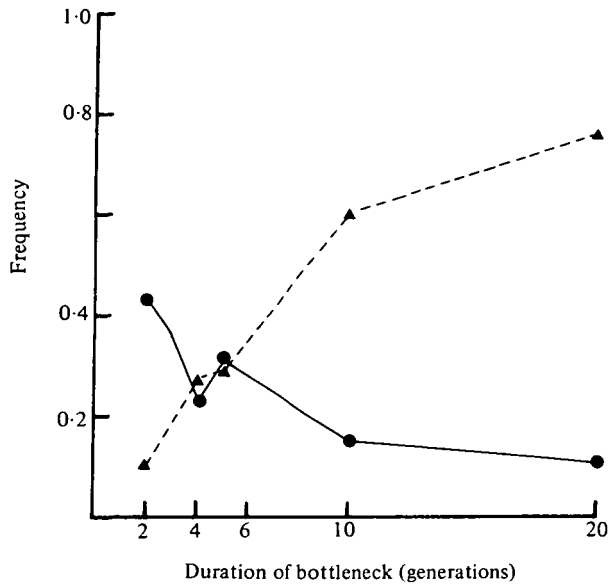


Fig. 4. The frequency of fixation at either locus (dashed line) and the frequency of a peak shift (full line) are plotted against the number of bottleneck generations, with a bottleneck size of 4, and 50 flush generations.

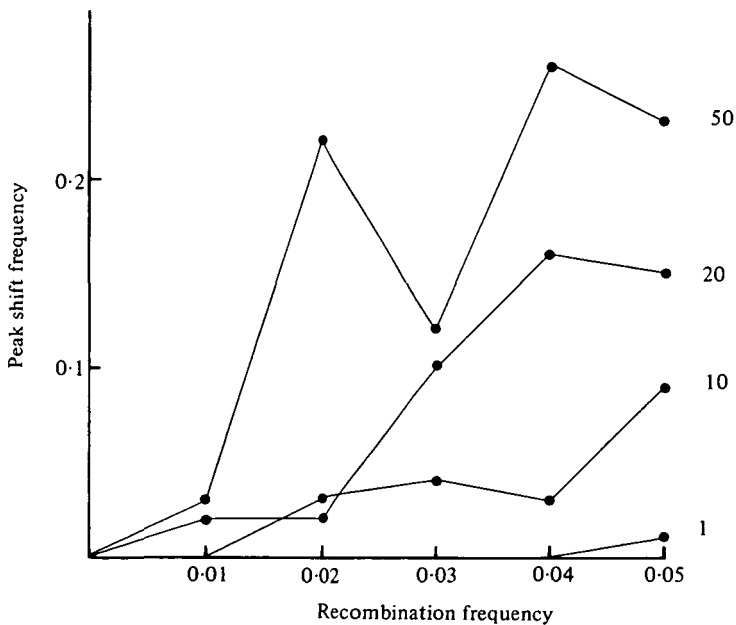


Fig. 5. The frequency of a peak shift is plotted against R , the recombination fraction, for different numbers of flush generations, indicated against the lines. The bottleneck population lasted 4 generations, with a size of 4.

(ii) *Effects of number of generations in the bottleneck.*

Fig. 4 shows the effect of m on the frequency of a peak shift and the frequency of fixation at either locus, for the case of $N = 4$ and $n = 50$, which is favourable for the occurrence of a peak shift. It can be seen that the probability of loss of variability increases rapidly with m , so that the chance of a peak shift declines.

(iii) *Effects of recombination fraction*

Fig. 5 shows the effects of R on the frequency of a peak shift, for the case of $m = 4$ and $N = 4$. There is a strong tendency for the frequency to increase with increasing R . This is due to two factors. First, the equilibrium value of D in the initial population decreases with increasing R (Lewontin, 1974, p. 296). Second the rate at which D decays with relaxed selection also increases with R . Values of R above 0.05 were not studied, as only one stable equilibrium exists in such cases.

4. DISCUSSION

The results described above show that the following conditions are most favourable for a peak shift following a bottleneck in population size associated with a founder event.

(a) A large number of flush generations with relaxed selection. Peak shifts are rare when this number is less than ten.

(b) The population spends only a few generations in the bottleneck, so that the chance of fixation at the selected loci is small.

(c) The population size during the bottleneck is neither very small nor very large; small population size is more unfavourable than large when the number of flush generations is large.

(d) The recombination fraction is as high as is compatible with the existence of two alternative stable equilibria.

All these conclusions are, of course, subject to the artificialities of the model, and are based on the study of a single fitness matrix. The conclusions agree qualitatively with what one would expect intuitively about the effects of the various parameters (cf. Carson, 1975; Templeton, 1980*a, b*), but obviously the actual values of the peak shift probabilities will be affected by the nature of the model (see below). Our results do establish, however, that a peak shift may occur with non-negligible probability under suitable conditions, particularly when there is a large number of generations without selection before the bottleneck. It may be questioned, however, whether selection is likely to be totally relaxed during phases of rapid population growth, in view of the fact that strong selection under normal conditions is required.

The models of Carson (1975) and Templeton (1980*a, b*) assume that selection is relaxed during the bottleneck phase as well as during the flush phase. This does not seem to be assumed in Mayr's model of genetic revolutions (Mayr, 1954, 1963);

we have assumed the operation of selection at its normal intensity during the bottleneck. It may be seen from Fig. 4, and by comparing Fig. 2 with Fig. 3, that the chance of loss of variability during the bottleneck has a major influence on the net probability of a peak shift. For example, with $m = 4$ and $n = 100$, this chance of loss of variability changes from 0.2 with $N = 6$ to 0.6 with $N = 2$, while the chance of a peak shift changes from 0.4 to 0.15. The corresponding values of the chance of a peak shift conditional on no loss of variability are 0.5 and 0.38,

Table 2. *Fitness matrix for model with strong selection*

	<i>AA</i>	<i>Aa</i>	<i>aa</i>
<i>BB</i>	0.20	0.05	0.15
<i>Bb</i>	0.05	1.00	0.05
<i>bb</i>	0.15	0.05	0.20

which do not differ significantly. Relaxation of the very strong selection in favour of heterozygotes assumed here is thus likely to reduce the overall chance of a peak shift, due to an increased loss of variability in the absence of such selection.

Another restriction of our model is that we only studied a fitness matrix in which the two alternative equilibria had rather similar gene frequencies. This is likely to facilitate a peak shift, since the system needs to change in effectively only one dimension, that of linkage disequilibrium.

Furthermore, the degree of reproductive isolation from the ancestral population induced by a peak shift is not very impressive, at any rate with the fitness matrix we have studied. With $R = 0.05$, the mean fitness of the F_1 hybrid between the two alternative equilibrium populations is about 79% of the mean of fitnesses of the parental populations. Smaller values of R produce greater effects; with $R = 0$, the F_1 would have a mean fitness of 66% of the mean of the parents, and the corresponding value for the F_2 is 83%. But, as can be seen from Fig. 5, the chance of a peak shift is drastically reduced with very low recombination. An alternative way to achieve greater hybrid breakdown is by stronger epistatic selection. The fitness matrix of Table 2 generates an F_1 fitness 20% of that of the parents, with $R = 0.05$. The chance of a peak shift with this model is similar to that with the earlier one; with $N = 2$, $n = 50$ and $m = 4$, the frequency was 16% (cf. Fig. 1). Such extreme selection seems biologically implausible, however.

It therefore seems unlikely from our results that significant reproductive isolation could often be achieved by means of a peak shift in a two-locus system due to a single bottleneck in population size. It is important to note in this context that the main evidence for the role of founder effects in speciation comes from groups like the Hawaiian *Drosophila*, where migration to a new locality is very frequently associated with speciation (Carson, 1970). For example, the numerous picture-winged *Drosophila* species of Hawaii, the newest island in the archipelago, are all descended from species present on the neighbouring island of Maui, none of which is represented on Hawaii. Our results suggest that the most that a two-locus peak shift could contribute to this high chance of speciation associated with

colonisation is to provide a stimulus to *subsequent* evolutionary change, in response to the altered genetic environment caused by a peak shift. It could, of course, be argued that quite different results might be obtained in multi-locus systems. This seems unlikely, for the following reason. Two extreme forms of multi-locus system with epistatic selection can be envisaged. One consists of a set of more or less independent two-locus systems. Independent peak shifts in each sub-system could combine to create a high degree of isolation, but the chance of their occurring simultaneously in a single founder event is very low. The alternative type of system is when the entire set of loci behaves as a block, as simulated by Franklin & Lewontin (1970) for example. Here, the equilibrium frequencies of the repulsion gametes, which would have to come to predominate after a peak shift, are kept very low in frequency, so that a peak shift must be improbable. Such systems also maintain a very high genetic variance in fitness, which does not seem consistent with observations on species like *D. melanogaster*. Even the two-locus systems of Table 1 maintains an unrealistically high level of genetic variance at equilibrium; with $R = 0.05$, the component of genetic variance in fitness contributed by the marginal fitnesses of each gamete is $0.0175 \times$ the squared mean fitness. This may be compared with a value of 0.009 estimated by Mukai *et al.* (1974), for viability effects of the entire second chromosome of *D. melanogaster*.

Several successive bottlenecks in population size after a founder event might help to raise the probability of achieving a high degree of hybrid breakdown. The difficulty with this is that it leads to a depletion of variability, due to chance fixation of alleles. As we have seen, this is unfavourable to the occurrence of a peak shift. However, the resulting changes in allele frequencies could, as suggested by Mayr (1954, 1963), stimulate evolutionary change at other loci. This possibility runs up against the difficulty that there is no evidence for reduced heterozygosity at enzyme loci in newly arisen species of *Drosophila* on Hawaii (Craddock & Johnson, 1979), nor in the rapidly speciating North American minnows (Avisé, 1977). This lack of evidence for effects of founder events on heterozygosity also creates difficulties for models of the rapid origin of ethological isolation by founder effects (Kaneshiro, 1976; Wills, 1977; Lande, 1981). If loci affecting courtship behaviour can become fixed due to genetic drift caused by founder effects, why should enzyme loci have remained apparently immune? We conclude that an important role of bottlenecks in population size as a cause of rapid speciation remains to be established.

We thank H. L. Carson, D. Charlesworth, J. Maynard Smith and A. R. Templeton for their comments on the manuscript.

REFERENCES

- AVERY, P. J. (1978). The effects of finite population size on models of linked overdominant loci. *Genetical Research* **31**, 239–254.
- AVERY, P. J. & HILL, W. G. (1979). Distribution of linkage disequilibrium with selection and finite population size. *Genetical Research* **33**, 29–48.

- AVISE, J. C. (1977). Genic heterozygosity and the rate of speciation. *Paleobiology* **3**, 422–432.
- BENGTSSON, B. O. & BODMER, W. F. (1976). On the increase of chromosome mutations under random mating. *Theoretical Population Biology* **9**, 260–281.
- CARSON, H. L. (1968). The population flush and its consequences. In *Population Biology and Evolution* (ed. R. C. Lewontin), pp. 123–137. Syracuse, N.Y.: Syracuse University Press.
- CARSON, H. L. (1970). Chromosome tracers of the origin of species. *Science* **168**, 1414–1418.
- CARSON, H. L. (1975). The genetics of speciation at the diploid level. *American Naturalist* **109**, 83–92.
- CRADDOCK, E. M. & JOHNSON, W. E. (1979). Genetic variation in Hawaiian *Drosophila*. V. Chromosomal and allozymic diversity in *Drosophila silvestris* and its homosequential species. *Evolution* **33**, 137–155.
- DOBZHANSKY, T. (1951). *Genetics and the Origin of Species*. New York: Columbia University Press.
- FELSENSTEIN, J. (1974). Uncorrelated genetic drift of gene frequencies and linkage disequilibrium in some models of linked overdominant polymorphisms. *Genetical Research* **24**, 281–294.
- FRANKLIN, I. & LEWONTIN, R. C. (1970). Is the gene the unit of selection? *Genetics* **65**, 707–734.
- GOULD, S. J. & ELDREDGE, N. (1977). Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* **3**, 115–151.
- KANESHIRO, K. (1976). Ethological isolation and phylogeny in the *planitibia* subgroup of Hawaiian *Drosophila*. *Evolution* **30**, 740–745.
- LANDE, R. (1979). Effective deme sizes during long-term evolution estimated from rates of chromosomal rearrangement. *Evolution* **33**, 234–251.
- LANDE, R. (1980). Genetic variation and phenotypic evolution during allopatric speciation. *American Naturalist* **116**, 463–479.
- LANDE, R. (1981). Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences, U.S.A.* **78**, 3721–3725.
- LEWONTIN, R. C. (1974). *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.
- LEWONTIN, R. C. & KOJIMA, K-I. (1960). The evolutionary dynamics of complex polymorphisms. *Evolution* **14**, 458–472.
- MAYR, E. (1942). *Systematics and the Origin of Species*. New York: Columbia University Press.
- MAYR, E. (1954). Change of genetic environment and evolution. In *Evolution as a process* (eds. J. S. Huxley, A. C. Hardy & E. B. Ford), pp. 157–180. London: Allen & Unwin.
- MAYR, E. (1963). *Animal Species and Evolution*. Cambridge, Mass: Harvard University Press.
- MUKAI, T., CARDELLINO, R. A., WATABABE, T. K. & CROW, J. F. (1974). The genetic variance for viability and its components in a local population of *Drosophila melanogaster*. *Genetics* **78**, 1195–1208.
- MULLER, H. J. (1940). Bearings of the *Drosophila* work on systematics. In *The New Systematics* (ed. J. S. Huxley), pp. 185–268. London: Systematics Association.
- SLATKIN, M. (1981). Fixation probabilities and fixation times in a subdivided population. *Evolution* **35**, 477–488.
- STANLEY, S. M. (1979). *Macroevolution: Pattern and Process*. San Francisco: W. H. Freeman.
- TEMPLETON, A. R. (1980a). The theory of speciation via the founder principle. *Genetics* **94**, 1011–1038.
- TEMPLETON, A. R. (1980b). Modes of speciation and inferences based on genetic distances. *Evolution* **34**, 719–729.
- WILLS, C. J. (1977). A mechanism for rapid allopatric speciation. *American Naturalist* **111**, 603–605.
- WRIGHT, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the 6th International Congress of Genetics*, **1**, 356–366.
- WRIGHT, S. (1941). On the probability of fixation of reciprocal translocations. *American Naturalist* **75**, 513–522.
- WRIGHT, S. (1978). Review of Modes of Speciation (M. J. D. White). *Paleobiology* **4**, 373–379.