Genet. Res., Camb. (1979), **33**, *pp.* 29–48
With 10 *text-figures*
Printed in Great Britain

Distribution of linkage disequilibrium with selection and finite population size

By P. J. AVERY* and W. G. HILL
Institute of Animal Genetics, West Mains Road, Edinburgh EH9 3JN†

(*Received* 4 *July 1978*)

SUMMARY

The effects of finite population size, occurring either as a bottleneck in a single generation followed by a large expansion or in all generations, are considered for models of two linked heterotic loci. Linkage is assumed to be tight because it is required if there is to be stable linkage disequilibrium, $D \neq 0$, in infinitely large populations. (D is the difference between gamete frequencies and the product of the gene frequencies.)

If a substantial perturbation of frequencies occurs as a result of a bottleneck but the population is subsequently very large, D may take hundreds of generations to return to its stable point. In finite populations, the distribution of D can be U-shaped, unimodal or bimodal. The correlation of D in successive generations is higher with tight linkage and is little affected by selection or the size of the population.

The utility of infinite population studies of linkage disequilibrium and its stable points is questioned, and considerable pessimism is expressed about the possibilities of distinguishing selection and sampling effects at linked loci.

1. INTRODUCTION

There has been considerable interest in recent years in theoretical and experimental studies on the population genetics of linked loci. In particular there is now an extensive literature on equilibrium properties of populations with two or more linked loci using deterministic (i.e. infinite population) models, on effects of genetic drift (in small populations) on linked loci without selection and on association of alleles at electrophoretic loci in laboratory or natural populations, usually of *Drosophila*. Lewontin (1974), Thomson (1977) and Hedrick, Jain & Holden (1978) have reviewed this area in some detail and Karlin (1975) gives a valuable summary of the two locus deterministic theory.

As with single loci, one of the main difficulties in multi-locus studies is to distinguish between selection and drift or sampling effects. Lewontin (1974) argues that if selection is the main cause of disequilibrium the same sign and magnitude of disequilibrium would be expected in different populations, whereas if random drift due to small population size is the main cause, values will differ randomly

* Present address: Department of Applied Statistics, University of Reading, Reading RG6 2AN.
† Address for reprints.

0016-6723/79/2828-7250 $01.00

between populations. For two loci, stable linkage disequilibrium can only be obtained if there is tight linkage, some non-additivity of effects between the loci, and also heterozygote superiority at individual loci such that they remain segregating. Apart from small population size, no such restraints have to be made for disequilibrium generated by chance sampling, with new variation generated by mutation. There has been some study of the joint effect of selection and sampling, for example to ask how resistant to perturbations due to finite population size are stable positions of linkage disequilibrium in an infinite population (Sved, 1968; Franklin & Lewontin, 1970; Felsenstein, 1974; Clegg, 1978; Avery, 1978).

In this paper we shall illustrate, in an essentially non-mathematical way, the effects of perturbations due to sampling on tightly linked loci. Firstly we consider the effect of a substantial perturbation, for example a population passed through a bottleneck of a single pair of individuals. The point of this is not to show that new disequilibrium is generated, which is obvious, but to illustrate how slowly populations would return to a stable point even if the subsequent population were infinitely large. In particular the linkage disequilibrium (gametic association) terms return much more slowly than gene frequencies. Then we turn to populations always of the same finite size and investigate the autocorrelation of linkage disequilibrium effects, and the distribution of the disequilibrium which can be bimodal in form.

2. MODEL

Consider a diploid population reproducing by random mating including random selfing, with no migration or mutation. The analysis will be of two loci each with two alleles, where

A, a are alleles at locus 1, and p is the frequency of A;
B, b are alles at locus 2, and q is the frequency of B;
x_1, x_2, x_3, x_4 are the frequencies of chromosomes AB, Ab, aB and ab respectively;
$D = x_1x_4 - x_2x_3 = x_1 - pq$ is the linkage disequilibrium;
$r = D/[p(1-p)\,q(1-q)]^{\frac{1}{2}}$ is the correlation of gene frequencies on the same gamete; and
c = recombination fraction between the loci.

Most of the analysis will deal with a special case of the symmetric model of fitness used by Lewontin & Kojima (1960) and Bodmer & Felsenstein (1967), defined as follows:

	BB	*Bb*	*bb*
AA	$1-2s+ks^2$	$1-s$	$1-2s+ks^2$
Aa	$1-s$	1	$1-s$
aa	$1-2s+ks^2$	$1-s$	$1-2s+ks^2$

Three examples will be used: additive, $k = 0$; multiplicative, $k = 1$; epistatic, $k = 1/s$. Although this is only one of an infinite set of epistatic models, it shows strong epistasis and has been discussed previously (Avery, 1978).

With these models and an infinitely large population there are at most two stable equilibria at which both loci are polymorphic. These are where both genes have frequency one-half ($p = q = 0{\cdot}5$), and the chromosome frequencies are $x_1 = 0{\cdot}5 - x_2 = 0{\cdot}5 - x_3 = x_4$, with $0 \leqslant x_1 \leqslant 0{\cdot}5$. Thus at a stable point $D = x_1 - 0{\cdot}25$ and $r = 4D$, with the values depending on the recombination fraction and fitness array. Lewontin & Kojima (1960), Bodmer & Felsenstein (1967) and Karlin (1975) show that for the additive model ($k = 0$), or other models where $k < 0$, there is only one stable polymorphism for all values of c, which is at $D = 0$. For other models, the equilibria are given by the roots of $16ks^2D^3 - (ks^2 - 4c)\,D = 0$, which are

$$D = 0 \text{ or } D = \pm\tfrac{1}{4}\sqrt{[1 - 4c/(ks^2)]}. \quad (1)$$

If $c < ks^2/4$, there are stable points at $D \neq 0$ and an unstable point at $D = 0$, while if $c > ks^2/4$ there is a stable point only at $D = 0$. Thus there are stable values with linkage disequilibrium if $c < s^2/4$ in the multiplicative model which would imply very tight linkage, and $c < s/4$ in the epistatic model. If $D \neq 0$ is stable, the zones of attraction in an infinitely large population are simply that if D is initially positive it moves to the positive stable value and vice versa if negative (Bodmer & Felsenstein, 1967).

3. POPULATION BOTTLENECKS

A major perturbation of chromosome frequencies away from their stable values can occur if the population passes through a very small bottleneck of numbers, the most extreme case being the foundation of a new colony by a single inseminated migrant female or by seed from one plant which has been ferilized by another single plant. This example provides a useful illustration of the rate at which stability is reached and the amount of population diversity possible, even though the subsequent population size is very large. The model taken is thus of a founder sample of four chromosomes from a population at the stable point, with the new populations then assumed to be infinitely large from the first generation so that deterministic formulae can be used.

The different groups of sampled populations are listed in Table 1, together with their probabilities in terms of the chromosome frequencies in the original population. For example, (1, 0, 0, 3) denotes a population with 1 *AB* and 3 *ab* chromosomes initially, comprising the diploid pair *AB*/*ab*, *ab*/*ab*; this population has probability $4x_1x_4^3$ and the initial correlation of frequencies is $r = 1$. In the symmetric fitness model (1, 0, 0, 3) is similar in terms of change in D or r to the population (3, 0, 0, 1) and, with the sign of D or r reversed, to (0, 1, 3, 0) and (0, 3, 1, 0). Founder samples in which one or more genes are fixed are excluded from Table 1 and subsequent discussion. Thus the sum of the probabilities of all the groups gives the probability of both loci remaining unfixed which in the simple case of $x_1 = x_2 = x_3 = x_4$ is given by $(1 - \frac{1}{8})^2 = \frac{49}{64}$. The subsequent deterministic changes in gene frequency and correlation of frequencies for each group are shown in Fig. 1 for the additive, multiplicative and epistatic models with $s = 0{\cdot}25$ and $c = 0{\cdot}01$

Table 1. *Classifications by arrangements of founder populations of four chromosomes as* (AB, Ab, aB, ab), *their corresponding correlation of gene frequencies* (r), *and their probabilities, in terms of gamete frequencies* (x_i) *in the base population*

Group	Arrangements: Example	No.	r	General case	$x_1 = x_2 = x_3 = x_4$
1	(1, 0, 0, 3)	2*	1	$4x_1x_4\,(x_1^2+x_4^2)$	8/256
	(0, 1, 3, 0)	2	−1	$4x_2x_3\,(x_2^2+x_3^2)$	8/256
2	(2, 0, 0, 2)	1	1	$6x_1^2x_4^2$	6/256
	(0, 2, 2, 0)	1	−1	$6x_2^2x_3^2$	6/256
3	(2, 1, 0, 1)	4	$1/\sqrt{3}$	$12x_1x_4\,(x_1+x_4)\,(x_2+x_3)$	48/256
	(1, 2, 1, 0)	4	$-1/\sqrt{3}$	$12x_2x_3\,(x_1+x_4)\,(x_2+x_3)$	48/256
4	(1, 2, 0, 1)	2	1/3	$12x_1x_4\,(x_2^2+x_3^2)$	24/256
	(2, 1, 1, 0)	2	−1/3	$12x_2x_3\,(x_1^2+x_4^2)$	24/256
5	(1, 1, 1, 1)	1	0	$24x_1x_2x_3x_4$	24/256

* i.e. (1, 0, 0, 3) and (3, 0, 0, 1).

(additive and multiplicative) or 0·05 (epistatic). For these examples, $|r| = 0$, 0·6 and $1/\sqrt{5} = 0{\cdot}447$, respectively, at the stable point.

The return of the gene frequencies to their stable value is seen to be much more rapid than that of the correlation of frequencies. Even so, the trajectory of change in frequency may not have constant direction (group 3, e.g. founder *AB/AB*, *Ab/ab*). Despite the very strong selection coefficients on individual genotypes a few hundred generations are required for r to become essentially zero with the additive model, the approach to the appropriate non-zero stable point being somewhat quicker in the other models. In each case, there are one or more groups in which the change in the absolute value, $|r|$, is not monotonic, although r itself never changes sign.

For the multiplicative and epistatic models there is an unstable equilibrium at $p = q = 0{\cdot}5$, $D = 0$, thus, in an infinite population, there would be no departure from $D = 0$ in group 5. However, to be more realistic, a slight perturbation has been introduced to allow for sampling in the first generation after the bottleneck, to an initial frequency of *AB* of 0·26 and $p = q = 0{\cdot}5$, i.e. $D = 0{\cdot}01$, $r = 0{\cdot}04$ (graphs '5 per' in Fig. 1). The rate of departure from this point is slow, and indeed for the multiplicative model it takes about 200 generations to get from $r = 0{\cdot}01$ to $r = 0{\cdot}04$.

The main point illustrated by Fig. 1 is that, since the rate of return to equilibrium is so slow with tight linkage (and tight linkage is required for there to be stable points at $D \neq 0$), for very many generations no distinction between the models is possible. There are some other anomalies, however. Since groups 3 and 4 are the most common and in these $|r|$ initially increases, the mean value, $E(r^2)$, actually rises in early generations if the population sampled is in linkage equilibrium with $p = q = 0{\cdot}5$ (Table 2). As a comparison, values of $E(r^2)$ following a bottleneck for a deterministic model with no selection ($s = 0$) are also shown in

https://doi.org/10.1017/S0016672300018140 Published online by Cambridge University Press

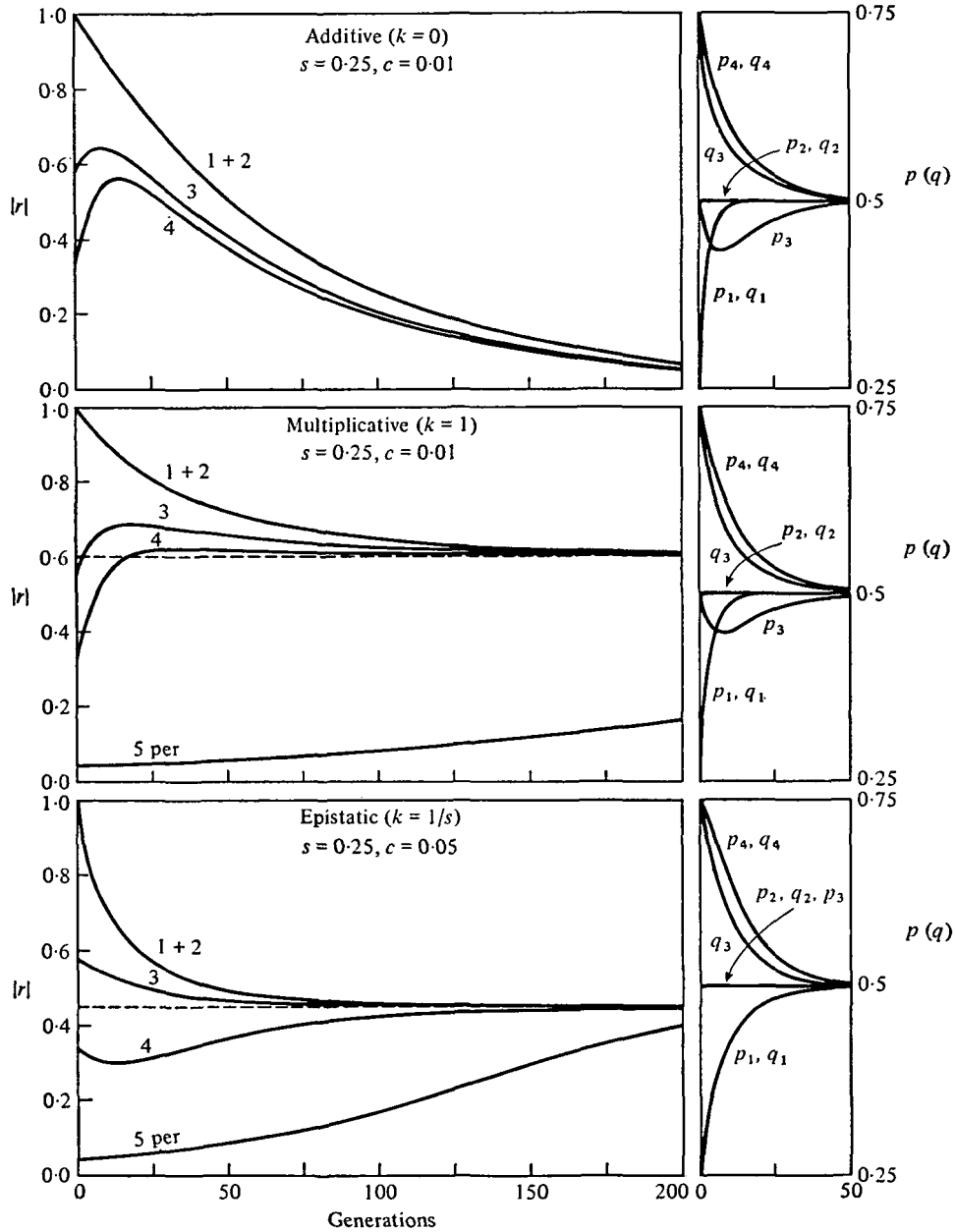

Fig. 1. Values of r, p and q are plotted against generations for various models, assuming an infinite population, after an initial bottleneck to four chromosomes. The numbers on the groups refer to the different groups obtained after bottlenecking as follows: group 1: (1003); 2: (2002); 3: (2101); 4: (1201); 5: (1111) (Table 1). The dotted lines refer to stable values. The graphs marked '5 per' refer to group 5 perturbed initially to $p = q = 0{\cdot}5$, $D = 0{\cdot}01$ (i.e. $r = 0{\cdot}04$).

https://doi.org/10.1017/S0016672300018140 Published online by Cambridge University Press

Table 2. *Mean values of r^2 in unfixed classes after a bottleneck of a population with $p = q = 0{\cdot}5$ and $D = 0$ to four chromosomes followed by an infinitely large population, for neutral ($s = 0$) and additive ($s = 0{\cdot}25$) models with $c = 0{\cdot}01$. Generation 0 refers to the four chromosomes of the bottleneck*

Generation	0	1	5	10	20	50	100	300
Neutral	0·3333	0·3267	0·3015	0·2726	0·2230	0·1220	0·0447	0·0008
Additive	0·3333	0·3452	0·3774	0·3840	0·3332	0·1547	0·0404	0·0002

Table 2, and it is seen that the additive model maintains more disequilibrium on average than the neutral model in early generations. Subsequently linkage disequilibrium is lost rather more rapidly with the additive model since selection increases the frequency of double heterozygotes, an effect noted by Clegg (1978).

4. FINITE POPULATIONS

The single bottleneck followed by a large population size exhibits some extreme effects of sampling on distribution of linkage disequilibrium; in general a population is never infinitely large so a stochastic model should always be used. Firstly let us assume that the population size and selective effects are both sufficiently large that the linkage disequilibrium is distributed around its stable point. This distribution was studied by Felsenstein (1974) for the case where the stable point is given by $D = 0$, and subsequently generalized by Avery (1978). Their results will be briefly reviewed and generalized to consider the correlation of D values in successive generations, which gives some impression of how long departures from the stable point persist.

Autocorrelation of linkage disequilibrium

Let the stable point be denoted $\tilde{p}$, $\tilde{q}$ and $\tilde{D}$ and let $\mathbf{d}_t' = (p_t - \tilde{p},\ q_t - \tilde{q},\ D_t - \tilde{D})$ denote the transpose of the vector of departures from this stable point at generation t. Assuming these departures to be small,

$$\mathbf{d}_{t+1} = \mathbf{A}\mathbf{d}_t + \mathbf{e}_t \tag{2}$$

(Felsenstein, 1974) where $\mathbf{A}$ is the matrix of first order terms from the Taylor's expansion of p_{t+1}, q_{t+1} and D_{t+1} in terms of p_t, q_t and D_t evaluated at $\tilde{p}$, $\tilde{q}$ and $\tilde{D}$, and $\mathbf{e}_t$ is a random vector with mean zero, independent of $\mathbf{d}_t$.
From (2),

$$E(\mathbf{d}_{t+1}\mathbf{d}_t') = \mathbf{A}E(\mathbf{d}_t\mathbf{d}_t') + E(\mathbf{e}_t\mathbf{d}_t'). \tag{3}$$

Since $\mathbf{d}_t$ has mean $\mathbf{0}$, $\mathbf{d}_t$ and $\mathbf{e}_t$ are independent and for the symmetric model,

$$\text{Cov}\ (p_t - \tilde{p},\ D_t - \tilde{D}) = \text{Cov}\ (q_t - \tilde{q},\ D_t - \tilde{D}) = 0$$

(Avery, 1978), it follows from (3) that

$$\text{Cov}\ (D_{t+1},\ D_t) = a_{33}\ \text{Var}\ (D_t), \tag{4}$$

where a_{33} is an element of $\mathbf{A}$. When the distribution of D has stabilized, Var (D_t) =

https://doi.org/10.1017/S0016672300018140 Published online by Cambridge University Press

Var (D_{t+1}) so the correlation, ρ, of disequilibrium in successive generations is, from (4),

$$\rho = a_{33}. \tag{5}$$

For the symmetric fitness models, from Avery (1978),

$$\rho = \frac{(1-s-c+ks^2/2)\,(1-s+ks^2/4-4ks^2\tilde{D}^2)}{(1-s+ks^2/4+4ks^2\tilde{D}^2)^2}, \tag{6}$$

with $k = 0$ (additive model), 1 (multiplicative) or $1/s$ (epistatic). If $\tilde{D} = 0$, (6) reduces to

$$\rho = (1-s-c+ks^2/2)/(1-s+ks^2/4);$$

or if $\tilde{D} \neq 0$, substituting in (1),

$$\rho = (1-s+c)/(1-s+ks^2/2-c).$$

Equation (5) shows that, to the degree of approximation used, the autocorrelation of D values is a function solely of the deterministic changes in D and does not include terms in population size. The correlation is therefore just a first order approximation to the proportion of the deviation of D from its stable point which is retained in the next generation in an infinite population, i.e. $(D_{t+1}-\tilde{D})/(\tilde{D}_t-\tilde{D})$.

Table 3. *The correlation (ρ) of linkage disequilibrium in successive generations*

	Model of fitness and value of s						
		Additive		Multiplicative		Epistatic	
Recombination fraction (c)	Neutral 0	0·25	0·1	0·25	0·1	0·25	0·1
0·01	0·9900	0·9867	0·9889	0·9854[1]*	0·9917	0·8786[2]*	0·9681[5]*
0·025	0·9750	0·9667	0·9722	0·9878	0·9751	0·9118[3]*	1·0000
0·05	0·9500	0·9333	0·9444	0·9551	0·9474	0·9697[4]*	0·9730
0·1	0·9000	0·8667	0·8889	0·8898	0·8920	0·9538	0·9189
0·2	0·8000	0·7333	0·7778	0·7592	0·7812	0·8308	0·8108
0·5	0·5000	0·3333	0·4444	0·3673	0·4488	0·4615	0·4865

* Stable point at 1, $\tilde{D} = \pm 0{\cdot}15$; 2, $\tilde{D} = \pm 0{\cdot}2291$; 3, $\tilde{D} = \pm 0{\cdot}1936$; 4, $\tilde{D} = \pm 0{\cdot}1118$; 5, $\tilde{D} = \pm 0{\cdot}1936$. Otherwise $\tilde{D} = 0$.

Values of ρ are given in Table 3 for a range of models, and compared with the value $1-c$, which equals the ratio D_{t+1}/D_t when there is no selection. The correlation is seen to depend mainly on the recombination fraction rather than fitness values, although it must be emphasized that the population sizes have been assumed to be sufficiently large that D never departs far from its stable point.

Monte Carlo simulation

Under these assumptions of reasonably large population size, D has a normal distribution and values for its variance have been given by Felsenstein (1974) and Avery (1978). It is apparent from some of Avery's results, however, that the distributions of D about $\tilde{D}$ are somewhat skewed when $\tilde{D} \neq 0$. To obtain a fuller

description of the distribution some Monte Carlo simulations have been made using a procedure described by Avery (1978). These results can also be used to show how the distribution changes with time, say for populations drawn from the same base population at which there is segregation around the stable point.

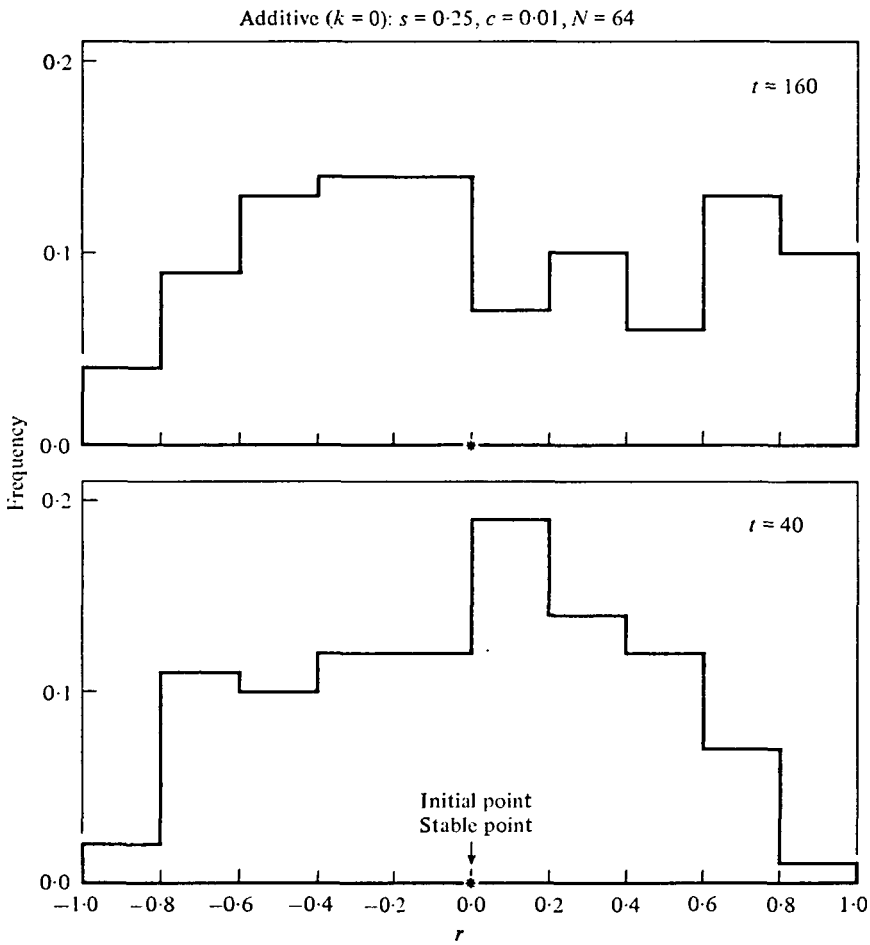

Fig. 2. Frequency distribution of r obtained by simulation (100 replicates) for an additive model with $s = 0{\cdot}25$, $c = 0{\cdot}01$ and $N = 64$ after $t = 40$ and 160 generations. Initially $r = 0$ (stable point).

The additive model, with $s = 0{\cdot}25$ and $c = 0{\cdot}01$ is illustrated in Fig. 2 for a population initially at $r = D = 0$ (the stable point). A histogram of r is plotted for $N = 64$ and $t = 40$ and 160 generations, the distribution having stabilized by about 80 generations. No fixation occurred in these replicates and the gene frequencies p and q depart little from 0·5. Notice that the distribution is almost uniform over the range -1 to $+1$, and the fact that there is stability at $r = 0$ in an infinite population is not very informative. After the distribution has stabilized, the standard deviation of r is SD(r) = 0·55, very close to the value of 0·54 pre-

https://doi.org/10.1017/S0016672300018140 Published online by Cambridge University Press

dicted from Felsenstein's theory (Avery, 1978). For larger values of N, SD(r) declines in proportion to $1/\sqrt{N}$, as predicted from both theoretical and simulated results (Avery, 1978).

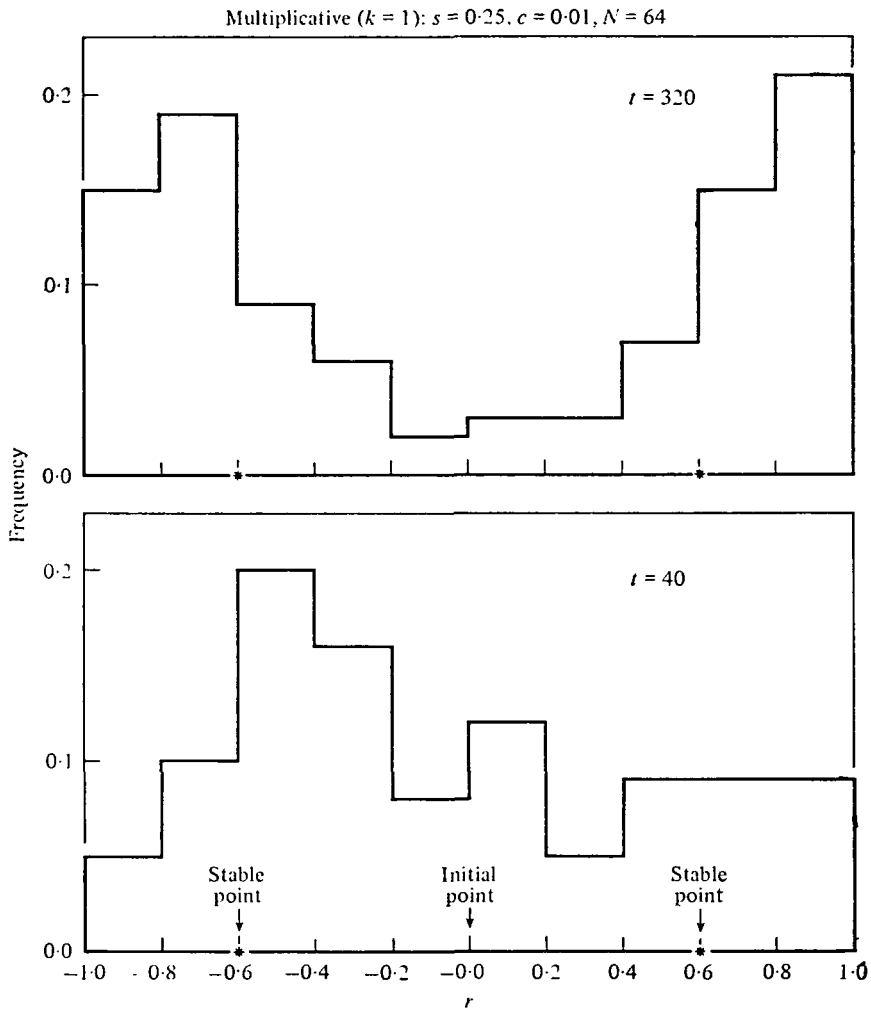

Fig. 3. As Fig. 2 but for 100 replicates of a multiplicative model with s = 0·25, c = 0·01 and N = 64 after t = 40 and 320 generations. Initially r = 0 (point of unstable equilibrium).

When there is not linkage equilibrium at the stable point the situation is more complicated. Figure 3 shows the multiplicative model, with otherwise similar parameters to the additive model in Fig. 2 and started at r = 0, when a symmetric bimodal distribution is generated but the modes are outside the stable points at $r = \pm 0{\cdot}6$, indeed quite close to unity. Similar results are shown in Fig. 4 for the epistatic model and a range of population sizes, although in all cases the starting point is taken at the positive stable point, $r = 1/\sqrt{5}$. Unless N is very large, however, selection is not strong enough to prevent values of r or D changing in sign.

https://doi.org/10.1017/S0016672300018140 Published online by Cambridge University Press

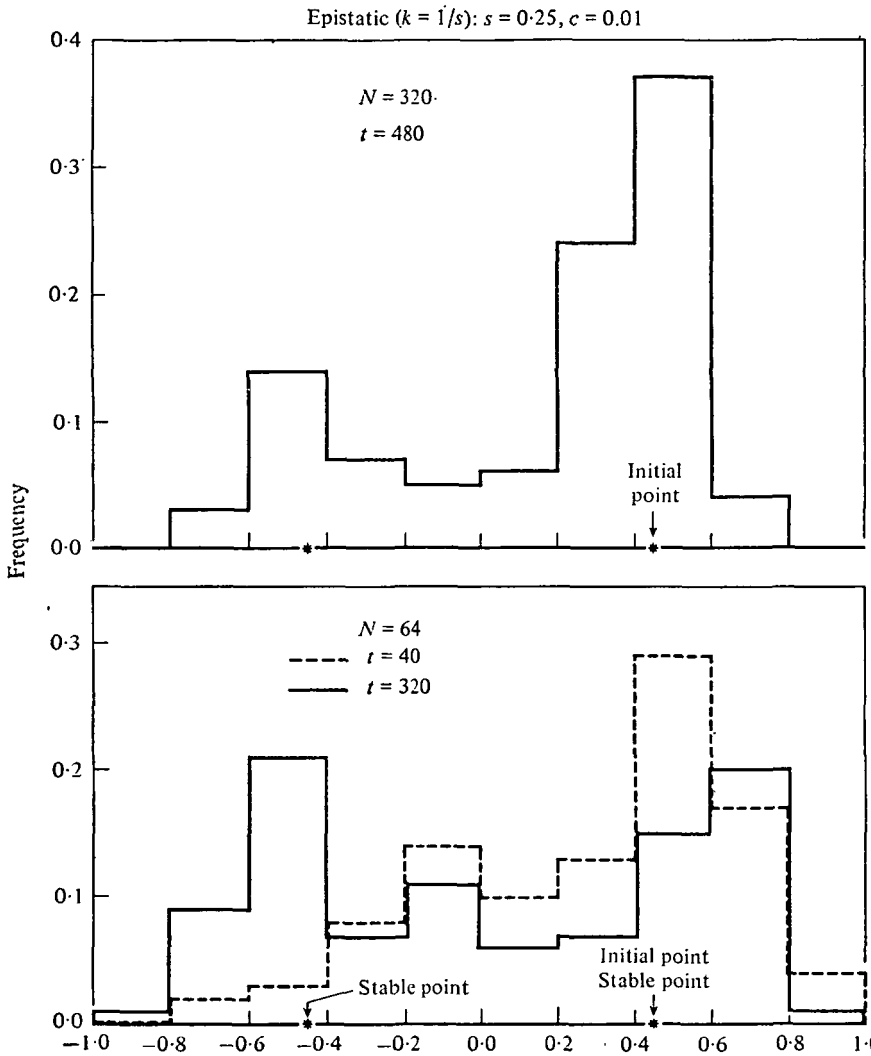

Fig. 4. As Fig. 2 but for 100 replicates of an epistatic model with $s = 0{\cdot}25$, $c = 0{\cdot}05$, and $N = 64$ or 320. Initially $r = \sqrt{0{\cdot}2}$ (stable point).

Diffusion approximation

In the models we have considered selection pressures and population sizes have been sufficiently large that the gene frequencies (p and q) have not departed far from their stable value of 0·5. The variation in D or r can, however, be large (Figs. 2–4), although the covariances of p with D and q with D are essentially zero (Felsenstein, 1974; Avery, 1978). It therefore seems reasonable to simplify the model by assuming p and q are fixed at one-half and considering the distribution of D alone. The diffusion approximation of this process now has only one dimension, and since fixation of the individual loci does not occur, a steady-state distribution is produced which can readily be solved. This univariate diffusion

https://doi.org/10.1017/S0016672300018140 Published online by Cambridge University Press

approximation has been used by Sved (1968), essentially for the additive model, and by Franklin & Lewontin (1970) for the multiplicative model. We shall explore it in more detail and obtain some insight into our simulation results.

It is convenient to define a new variable, $y = 2x_1$, i.e. twice the frequency of the AB gamete or since $p = q = 0{\cdot}5$, y equals the sum of the frequencies of coupling gametes, and then $D = y/2 - 0{\cdot}25$ and $r = 2y - 1$. Sved (1968) and Franklin & Lewontin (1970) made the equivalent transformation, but used the sum of frequencies of the repulsion gametes. The range of y values of $0 \leqslant y \leqslant 1$ and this reparametrization enables diffusion equation results for single loci to be used. The quantities $1/N$, ks^2 and c are assumed to be small and of similar order of magnitude, but the selective coefficient at single loci, s, is assumed of greater order than $1/N$ so segregation is maintained. Therefore an epistatic model such as the one we have used where $k = 1/s$ cannot be described by the following diffusion approximation. Ignoring higher order terms in formulae of Avery (1978), the mean ($M_{\delta y}$) and variance ($V_{\delta y}$) of change in y per generation are

$$M_{\delta y} = [-A(y-\tfrac{1}{2}) + By(1-y)\,(y-\tfrac{1}{2})]/(2N), \quad V_{\delta y} = (1-y)/(2N), \tag{7}$$

where

$$A = 2Nc/(1-s), \quad B = 2Nks^2/(1-s). \tag{8}$$

At the steady state, the density function $\phi(y)$ is given by

$$\phi(y) = (K/V_{\delta y}) \exp\left(2\int (M_{\delta y}/V_{\delta y})\,dy\right) \tag{9}$$

(Wright, 1938; e.g. also Crow & Kimura, 1970, p. 434), where K is a normalizing constant such that $\int_0^1 \phi(y)\,dy = 1$. From (7) and (9)

$$\phi(y) = K[y(1-y)]^{A-1}\, e^{-By(1-y)}. \tag{10}$$

(This distribution is the same as that for a single locus with equal forward and reverse mutation, homozygotes of equal fitness and heterozygote superiority if $B < 0$, neutrality if $B = 0$ or heterozygote inferiority if $B > 0$.) For $B = 0$, $\phi(y)$ has the Beta distribution with $K = \Gamma(2A)/\Gamma^2(A)$, where $\Gamma(\,)$ is the gamma function; and for $B \neq 0$, it can be shown that

$$K = \left[\sum_{i=0}^{\infty} \frac{(-B)^i\,\Gamma^2(A+i)}{i!\,\Gamma(2A+2i)}\right]^{-1}$$

For additive gene action ($B = 0$) the frequency distribution has three alternative forms (Fig. 5). If N and c are sufficiently small that $A < 1$, $\phi(y)$ is U-shaped, going to infinity at the bounds of $y = 0$ or 1, equivalent to $r = \pm 1$. If $A = 1$ the distribution is uniform, while if $A > 1$ the distribution is unimodal and symmetric about $r = 0$ ($y = \frac{1}{2}$), the stable point in infinite populations. The example simulated in Fig. 2, with $s = 0{\cdot}25$, $c = 0{\cdot}01$ and $N = 64$ corresponds to $A = 1{\cdot}71$. The diffusion approximation predicts a slightly humped distribution (Fig. 5) and this was observed in Fig. 2.

Without additivity ($B \neq 0$), the results are more complicated, the distribution

can be U-shaped, unimodal or bimodal. Positions of maxima and minima can be obtained by differentiation of $\phi(y)$ or, more easily, log $\phi(y)$. The solutions are as follows assuming $B > 0$ (e.g. multiplicative case):

(*a*) If $0 < 4(A-1)/B < 1$ there is a minimum at $y = \frac{1}{2}$, i.e. $r = 0$ and there are maxima at $y(1-y) = (A-1)/B$, i.e.

$$r = \pm\sqrt{[1-4(A-1)/B]} = \pm\sqrt{[1-4c/(ks^2)+2(1-s)/(Nks^2)]}. \qquad (11)$$

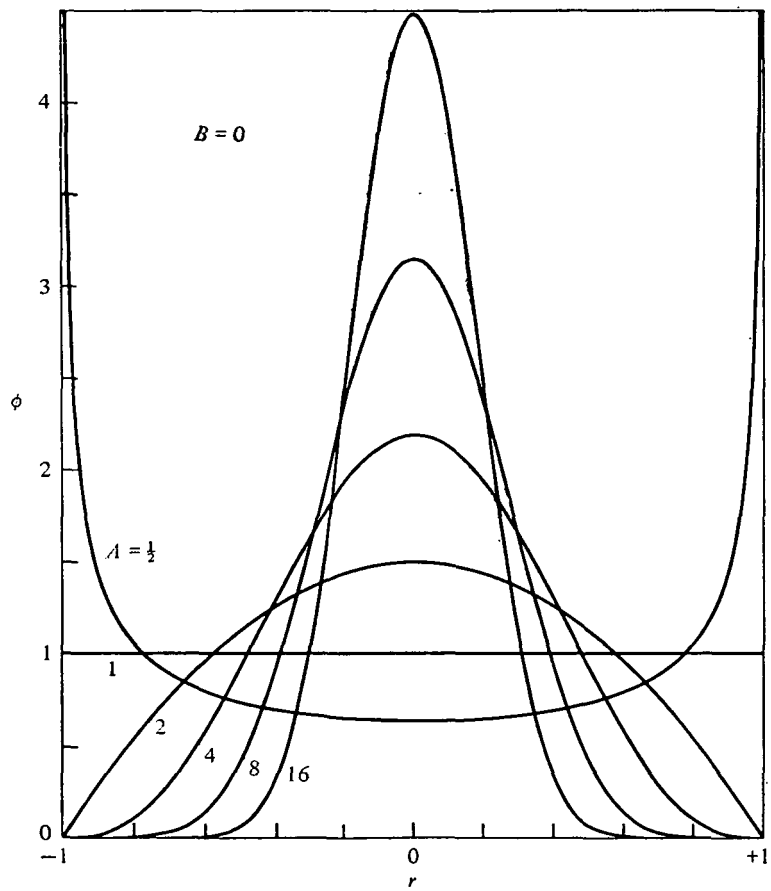

Fig. 5. Distribution of r from the diffusion approximation for additive model ($k = B = 0$). Curves are given for increasing values of $A = 2Nc/(1-s)$, i.e. changing population size and/or recombination fraction.

For large values of N these maxima correspond to the stable points in infinite population given by (1) which exist when $ks^2 > 4c$, but now the distribution can be bimodal even when this condition is not satisfied.

(*b*) If $4(A-1)/B \leqslant 0$, or simply $A \leqslant 1$, there is a U-shaped distribution with a minimum at $r = 0$.

(*c*) If $4(A-1)/B \geqslant 1$ there is a unimodal distribution with a maximum at $r = 0$.

https://doi.org/10.1017/S0016672300018140 Published online by Cambridge University Press

Examples of cases where $B > 0$ are given in Figs. 6–8. These illustrate the effect of increasing recombination fraction when N, s and k are fixed, so that A increases while B remains fixed (Fig. 6), and of increasing population size when s, k and c are fixed (Figs. 7 and 8). Figure 8 gives an example where there is a bimodal distribution as N is small, reducing to a unimodal distribution as N becomes very large. The example simulated in Fig. 3 for the multiplicative model corresponds to $A = 1{\cdot}71$, $B = 10{\cdot}67$ with $4(A-1)/B = 0{\cdot}265$, so two modes at

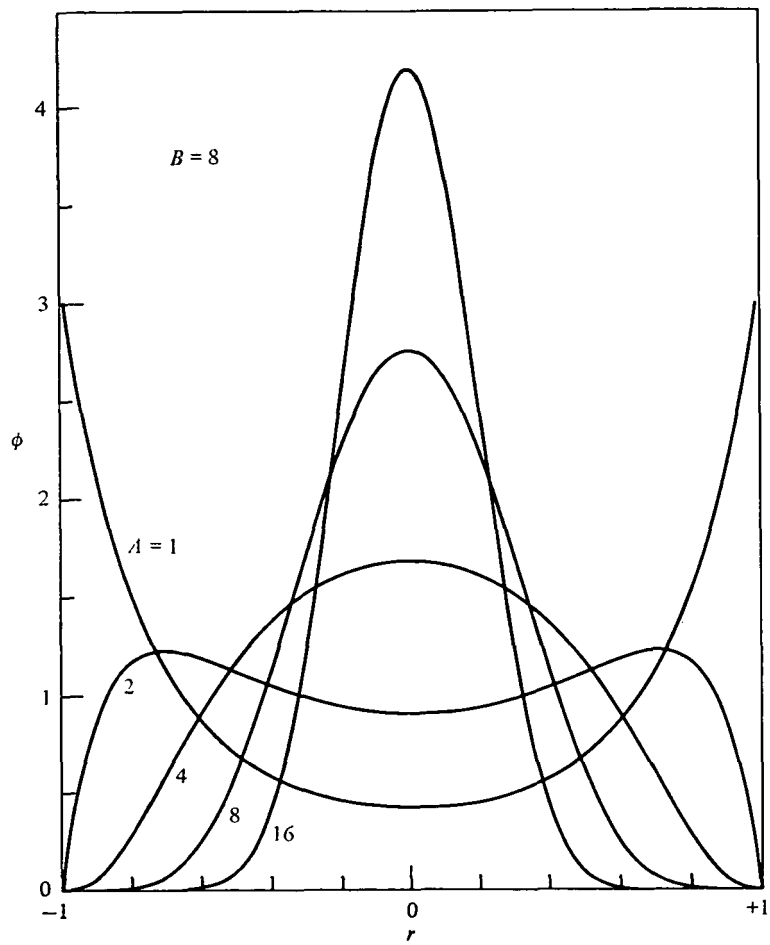

Fig. 6. Distribution of r from the diffusion approximation for model where $k > 0$. Curves are given for fixed $B = 8$ and A increasing, i.e. changing recombination fraction.

$r = \pm 0{\cdot}86$ are predicted from the diffusion equation. The results shown in Fig. 3 fit this prediction as far as the lack of replication enables us to tell.

It is noticeable in Fig. 7 that even when N is very large and the distribution is bimodal, there is still an appreciable probability that D or r takes values near zero, or y values near one-half. Letting $\phi(\max)$ denote the density at the modes,

https://doi.org/10.1017/S0016672300018140 Published online by Cambridge University Press

it can be shown from (10) that

$$\frac{\phi(0{\cdot}5)}{\phi(\max)} = \left[\frac{B}{4(A-1)}\right]^{A-1} e^{-B/4+A-1}. \tag{12}$$

Examples computed from (12) are given in Table 4, and for $A = 1{\cdot}71$, $B = 10{\cdot}67$ (Fig. 3), $\phi(0{\cdot}5)/\phi(\max) = 0{\cdot}360$. Presumably as a consequence of the unstable point which exists at $r = 0$, there is a substantial movement of populations from positive to negative disequilibrium and vice versa, even though either is stable in an infinitely large population.

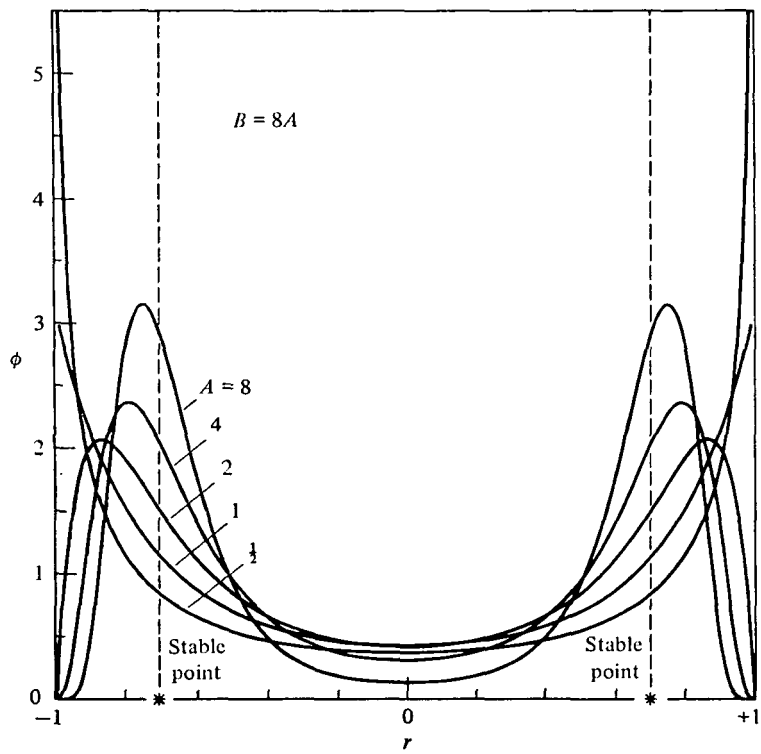

Fig. 7. Distribution of r from the diffusion approximation for model where $k > 0$. Curves are given for $B = 8A$, and A and B increasing together, i.e. changing population size. In this model the stable points in infinite population have $r = \pm 1/\sqrt{2}$.

When the fitness parameter k is negative, there is always stability of linkage equilibrium in an infinitely large population. However, in a finite population various distributions of r can be generated for $B < 0$.

(*a*) If $0 < 4(A-1)/B < 1$, the distribution is W-shaped, with a maximum at $r = 0$ and minima at $r = \pm\sqrt{[1-4(A-1)/B]}$. An example is shown in Fig. 9, in which population size is varied, the W-shaped distribution being obtained when $A = \frac{1}{2}$, $B = -4$.

https://doi.org/10.1017/S0016672300018140 Published online by Cambridge University Press

(*b*) If $4(A-1)/B \leqslant 0$, i.e. $A \geqslant 1$ there is a unimodal distribution with a maximum at $r = 0$.

(*c*) If $4(A-1)/B \geqslant 1$ there is a U-shaped distribution with a minimum at $r = 0$.

The results from the diffusion equation give a useful picture of the distributions likely to be obtained, although predictions of variance of r are less satisfactory. Whilst the method cannot formally be used for our epistatic model, where $k = 1/s$, it should give a guide to the distribution of D or r when there is rather less epistasis. It is clear that the distribution can take a variety of forms.

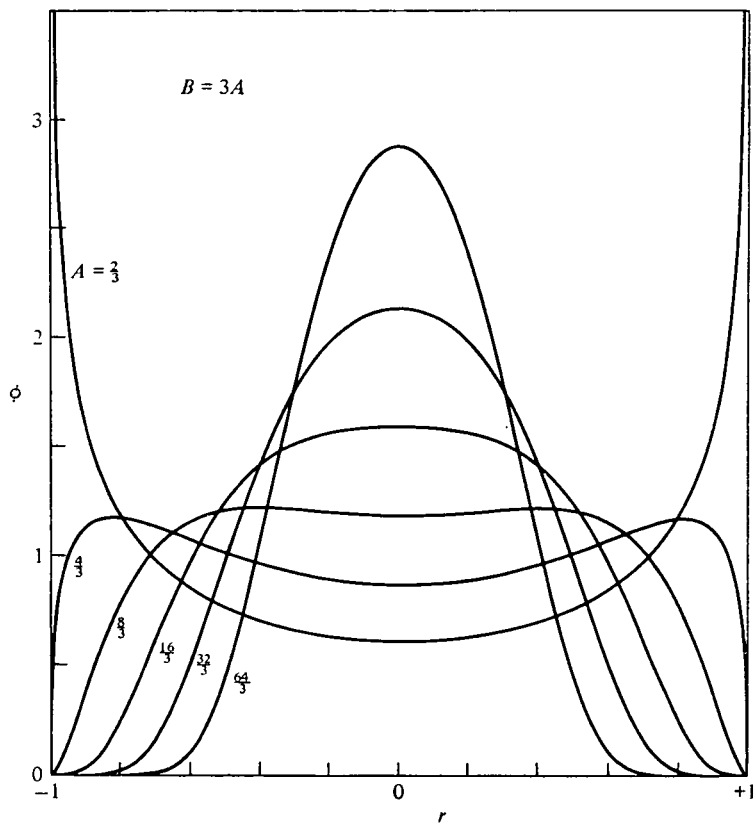

Fig. 8. As Fig. 7, but $B = 3A$, so in infinite population $r = 0$.

5. DISCUSSION

The first issue we considered was the rate of return to the stable point in populations greatly perturbed from it, say by having few founders. Despite strong selective pressures and cases of stability with considerable disequilibrium, the rate of approach was slow for the cases of tight linkage we considered. Of course, with more loosely linked genes the rates of approach would have been more rapid, but we have concentrated on tight linkage since only then can there be stable linkage disequilibria with feasible models of selection effects. Equation (6), derived here as the correlation of D values in successive generations in finite

https://doi.org/10.1017/S0016672300018140 Published online by Cambridge University Press

Table 4. *Ratios of probability density at $r = 0$ (i.e. $y = \frac{1}{2}$) to that at maximum for examples of* W*-shaped distributions. Values of A and B correspond to changes in N*

Examples of Fig. 7: $B/A = ks^2/c = 8$

A	1·5	2	3	4	6	8	12	16	24
B	12	16	24	32	48	64	96	128	192
$\phi(\frac{1}{2})/\phi$ (max)	0·201	0·199	0·165	0·128	0·073	0·040	0·012	0·0036	0·0003

Examples of Fig. 3: $k = 1$, $s = 0{\cdot}25$, $c = 0{\cdot}01$ ($B/A = ks^2/c = 6{\cdot}25$)

N	48	64	96	128	192	256	384	512	768
A	1·28	1·71	2·56	3·41	5·12	6·83	10·24	13·6	20·4
B	8·0	10·7	16·0	21·3	32·0	42·7	64·0	85·3	128
$\phi(\frac{1}{2})/\phi$ (max)	0·311	0·360	0·379	0·366	0·318	0·268	0·185	0·126	0·058

populations, also gives the proportional decline in the deviation of D from its stable point in infinite populations when p and q are close to 0·5 and this, as other workers previously and Table 3 have illustrated, tends to $1-c$ if selective values are very small.

The examples given in Fig. 1 are for symmetric models and in such cases the zone of attraction and general behaviour of the stable points can be calculated. Karlin & Carmelli (1975) sampled fitness sets randomly, and argued that epistasis

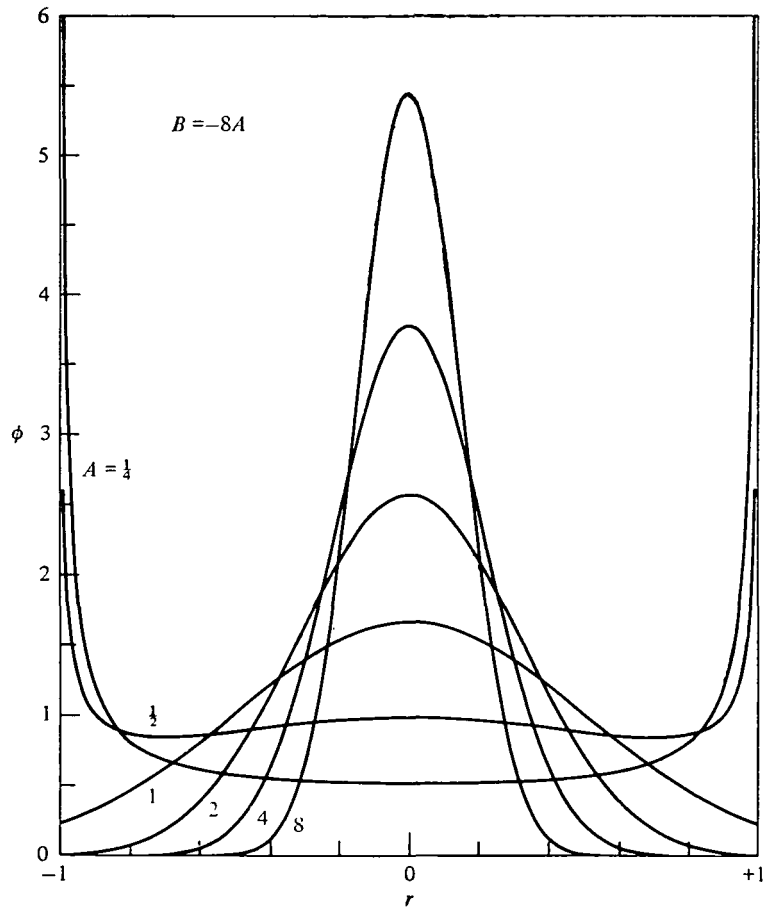

Fig. 9. As Fig. 7, but $k < 0$ with $B = -8A$, so in infinite population $r = 0$.

and thus linkage disequilibrium was likely to be the rule rather than the exception. One example of a fitness matrix by Karlin & Carmelli which has heterozygote superiority at both loci is as follows:

	BB	*Bb*	*bb*
AA	0·683474	0·365552	0·166198
Aa	0·443646	0·804193	0·476447
aa	0·493213	0·678145	0·442100

https://doi.org/10.1017/S0016672300018140 Published online by Cambridge University Press

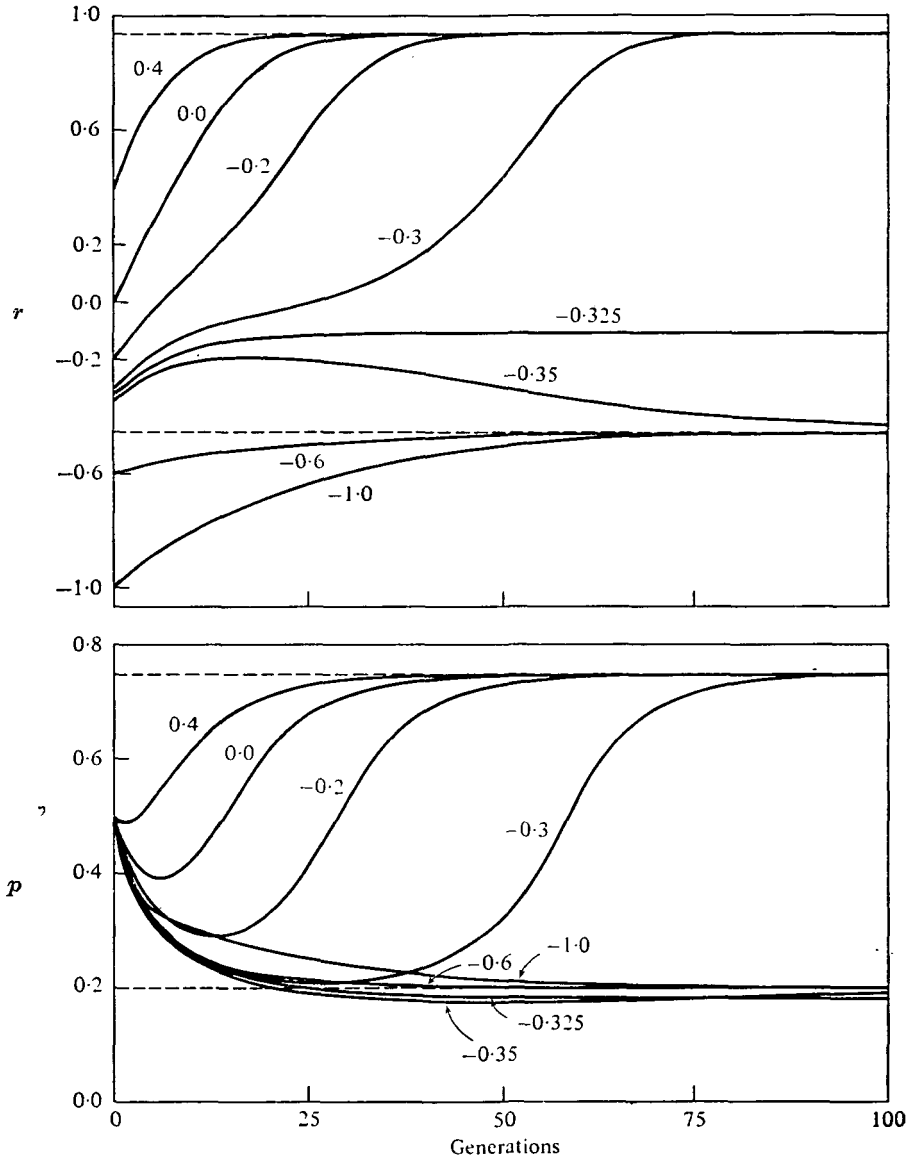

Fig. 10. Values of r and p plotted against generations for Karlin & Carmelli's model (see text) with $c = 0{\cdot}02$ in infinite population. Initially $p = q = 0{\cdot}5$ and a range of r values (shown on graphs) are given. The dotted lines show stable points.

Karlin & Carmelli showed that two polymorphic stable points existed with tight recombination, and one otherwise. For $c = 0{\cdot}02$, the stable points are $p = 0{\cdot}196522$, $q = 0{\cdot}569544$ and $D = -0{\cdot}089119$, giving $r = -0{\cdot}452949$ and $p = 0{\cdot}749444$, $q = 0{\cdot}754393$ and $D = 0{\cdot}174870$, giving $r = 0{\cdot}937504$. In such a case the zones of attraction are not easily computed. The reason we have introduced this complicated model is, however, to illustrate more strikingly than does Fig. 1 the peculiar trajectories by which the stable point can be reached, and the long period taken

https://doi.org/10.1017/S0016672300018140 Published online by Cambridge University Press

for the population to appear to 'decide' on which stable point it is moving towards. Results using a deterministic model are given in Fig. 10; in each case the initial gene frequencies are taken as 0·5 but a range of initial r ($= 4D$) values are given.

The examples shown in Fig. 10 illustrate that observations taken over many generations can often give little indication of the final state of the population, and could be consistent with many real stable points, including non-polymorphic ones.

The variation in linkage disequilibrium with drift and no selection has been extensively studied theoretically, following Hill & Robertson (1968). Joint studies of drift and selection have been limited by analytical difficulties except for populations distributed closely round the stable point (Felsenstein, 1974; Avery, 1978). The main observation we have made here (Figs. 2–9) is that, except when populations are very large, implying no seasonal or sporadic bottlenecks of size, a bimodal or wide distribution of linkage disequilibrium can be observed. Even though there is a stable point with quite substantial linkage disequilibrium, populations may pass through $D = 0$ and move from positive to negative linkage disequilibrium, or vice versa. Thus populations subdivided for many generations may show disequilibrium of opposite sign even though the selective forces in each are the same.

Selection and drift are only two of the possible causes of linkage disequilibrium. Other possible causes are migration and population admixture, or directional selection at a locus causing disequilibrium between it and a linked neutral locus or between an adjacent pair of neutral loci ('hitch hiking' effects). Either of these alternatives can, for example, be used to explain disequilibrium in the human histocompatibility system, HLA (Thomson, 1977).

Hedrick *et al.* (1978) reviewed experimental tests for the presence of linkage disequilibrium, and except for the HLA system and genes associated with inversions, found that only rarely has it been observed. Even if linkage disequilibrium is found, our results show that interpretation is bound to be difficult. Langley (1977) has discussed how to test whether the data are consistent with neutrality, but such tests lack the sophistication of those available for testing neutrality at single loci, such as that of Ewens (see review by Ewens, 1977). Since it is only with very tight linkage that stable polymorphisms having linkage disequilibrium are possible, any sampling effects take very many generations to be eliminated. Thus we think it unlikely that observations on disequilibria among linked loci are likely to give much information about any selective forces that may be present.

We are grateful to the Science Research Council for financial support.

REFERENCES

AVERY, P. J. (1978). The effects of finite population size on models of linked overdominant loci. *Genetical Research* **31**, 239–254.

BODMER, W. F. & FELSENSTEIN, J. (1967). Linkage and selection: Theoretical analysis of the deterministic two-locus random-mating model. *Genetics* **57**, 237–265.

CLEGG, M. T. (1978). Dynamics of correlated genetic systems: II. Simulation studies of chromosomal segments under selection. *Theoretical Population Biology* **13**, 1–23.

CROW, J. F. & KIMURA, M. (1970). *An Introduction to Population Genetics Theory*. New York, Evanston and London: Harper and Row.

EWENS, W. J. (1977). Population genetics theory in relation to the neutralist selectionist controversy. *Advances in Human Genetics* **8**, 67–134.

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.

HEDRICK, P., JAIN, S. & HOLDEN, L. (1978). Multi-locus systems in evolution. (In the Press.)

HILL, W. G. & ROBERTSON, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**, 226–231.

KARLIN, S. (1975). General two-locus selection models: some objectives, results and interpretations. *Theoretical Population Biology* **7**, 364–398.

KARLIN, S. & CARMELLI, D. (1975). Numerical studies on two loci selection models with general viabilities. *Theoretical Population Biology* **7**, 399–421.

LANGLEY, C. H. (1977). Non-random association between allozymes in natural populations of *Drosophila melanogaster*. *Measuring Selection in Natural Populations* (ed. F. B. Christiansen and T. M. Fenchel), pp. 265–274. Lecture Notes in Biomathematics, no. 19. Berlin, Heidelberg and New York: Springer-Verlag.

LEWONTIN, R. C. (1974). *The Genetic Basis of Evolutionary Change*. New York and London: Columbia University Press.

LEWONTIN, R. C. & KOJIMA, K. (1960). The evolutionary dynamics of complex polymorphisms. *Evolution* **14**, 458–472.

SVED, J. A. (1968). The stability of linked systems of loci with a small population size. *Genetics* **59**, 543–563.

THOMSON, G. (1977). The effect of a selected locus on linked neutral loci. *Genetics* **85**, 753–788.

WRIGHT, S. (1938). The distribution of gene frequencies under irreversible mutation. *Proceedings of the National Academy of Sciences, U.S.A.* **24**, 253–259.