SELECTION IN FINITE POPULATIONS WITH MULTIPLE ALLELES. III. GENETIC DIVERGENCE WITH CENTRIPETAL SELECTION AND MUTATION

B. D. H. LATTER

Division of Animal Genetics, C.S.I.R.O., Sydney, N.S.W., Australia

Manuscript received April 28, 1971
Revised copy received December 6, 1971

ABSTRACT

Natural selection for an intermediate level of gene or enzyme activity has been shown to lead to a high frequency of heterotic polymorphisms in populations subject to mutation and random genetic drift. The model assumes a symmetrical spectrum of mutational variation, with the majority of variants having only minor effects on the probability of survival. Each mutational event produces a variant which is novel to the population. Allelic effects are assumed to be additive on the scale of enzyme activity, heterosis arising whenever a heterozygote has a mean level of activity closer to optimal than that of other genotypes in the population.——A new measure of genetic divergence between populations is proposed, which is readily interpreted genetically, and increases approximately linearly with time under centripetal selection, drift and mutation. The parameter is closely related to the rate of accumulation of mutational changes in a cistron over an evolutionary time span.——A survey of published data concerning polymorphic loci in man and Drosophila suggests than an alternative model, based on the superiority of *hybrid* molecules, is not of general importance. Thirteen loci giving rise to hybrid zones on electrophoresis have a mean heterozygote frequency of 0.22 ± .06, compared with a value of 0.23 ± .04 for 16 loci classified as producing no hybrid enzyme.

ELECTROPHORETIC techniques have recently provided valuable details of the extent of allozyme variability in natural populations of man (HARRIS 1969a), Drosophila (PRAKASH, LEWONTIN and HUBBY 1969; GILLESPIE and KOJIMA 1968; O'BRIEN and MACINTYRE 1969), mice (SELANDER and YANG 1969) and other organisms (MANWELL and BAKER 1970). A striking feature of the more extensive surveys is the similar pattern of variation detected at particular loci in widely separated populations (PRAKASH, LEWONTIN and HUBBY 1969; O'BRIEN and MACINTYRE 1969), suggesting that natural selection is involved in the determination of the observed allelic frequencies. Clines have been reported for a serum esterase in Catostomid fish populations (KOEHN and RASMUSSEN 1967; KOEHN 1969) and for 6-phosphogluconate dehydrogenase and larval alkaline phosphatase in *Drosophila melanogaster* (O'BRIEN and MACINTYRE 1969). GILLESPIE and KOJIMA (1968) have also provided preliminary evidence for the greater stability of allelic frequencies in *Drosophila ananassae* at loci coding for enzymes involved in energy metabolism, by comparison with non-specific enzymes with broad substrate specificities.

The *functional* significance of most electrophoretically detectable variation in natural populations is at present largely unknown. However, a number of studies have shown that naturally occurring enzyme variants may differ appreciably in their biochemical properties. At least 50 variants of glucose-6-phosphate dehydrogenase are known in man, distinguishable by a combination of electrophoretic and enzymatic properties (KIRKMAN, KIDSON and KENNEDY 1968; YOSHIDA 1969). The allelic variants of red cell acid phosphatase give rise to five common phenotypes in man which differ substantially in their mean levels of enzyme activity, and significant differences in thermostability have also been detected (HARRIS 1966; HARRIS, HOPKINSON and LUFFMAN, 1968): heterozygotes for the common alleles are intermediate in activity, and allelic effects are essentially additive (HARRIS 1966). The electrophoretic variants of alcohol dehydrogenase in *Drosophila melanogaster* have also been shown to differ in specific enzyme activity, in loss of activity on heating, and in the percentage increase in activity observed in the presence of ethanol: the heterozygote is intermediate in enzyme activity, and in the percent change in activity induced by heat or alcohol (RASMUSON, NILSON and RASMUSON 1966; GIBSON 1970).

The genetic model discussed in this paper involves biochemical variants which can be graded on a single scale of activity, with additive allelic effects. A fixed intermediate level of activity is supposed to be optimal, so that a heterozygote may be favoured by natural selection if its mean activity is closer to the optimum than that of other genotypes present in the population. This simple model is capable of accounting for a substantial proportion of the balanced polymorphisms observed in natural populations (LATTER 1970), and is compatible with present knowledge of the functional and adaptive significance of enzyme and protein variants.

Three aspects of population behaviour under centripetal selection are examined:— (i) the rate and pattern of long-term evolutionary change; (ii) the stability of allelic frequencies over shorter time periods; and (iii) the expected characteristics of present-day natural populations. Of particular interest are the expected levels of heterozygosity in surveys of loci subject to similar intensities of selection, and the predicted frequencies of heterotic polymorphisms.

THE MODEL AND METHODS

The genetic model has been discussed briefly in the final section of the previous paper of this series (LATTER 1970). The following assumptions are made: (i) inheritance is disomic, and mating of surviving individuals is at random; (ii) a very large number of alleles A_i, $i = 1, 2, \ldots, \infty$, can be produced by mutation at the autosomal locus concerned. each spontaneous mutational event generating an allele which is novel to the population; (iii) the effective population size, N, is constant from generation to generation, so that random changes in the allelic frequencies p_i have an expected variance of $p_i(1-p_i)/2N$; (iv) allelic effects are additive on a scale measuring the functional differences among genotypes which are of adaptive significance (e.g. temperature or pH optima, catalytic activity, substrate binding capacity, etc.), a heterozygote having a mean level of "activity" equal to the average of those of the corresponding homozygotes; (v) the allelic effect of a new mutant is $a_i + \delta a_i$, where a_i is that of the parent allele, and δa_i is a random normal value with mean zero and variance $\sigma^2{}_m$; and (vi) the relative selective value of a genotype A_iA_j is proportional to

$$w^*{}_{ij} = 1 - \tfrac{1}{2}C^* \, (d^2{}_{ij}/\sigma^2{}_m) \qquad (1)$$

TABLE 1

Definition of symbols and relationships

Symbol	Definition
N	Effective breeding population size.
μ	Rate of spontaneous mutation.
p_i	Frequency of allele A_i.
a_i	Allelic effect of A_i, coded so that $\Sigma p_i a_i = 0$.
$\sigma^2{}_m$	Variance of mutational changes in allelic effect.
$\bar{x}$	Population mean on the scale of allelic effects.
$\sigma^2{}_g$	Genetic variance on the scale of allelic effects.
$\sigma^2{}_e$	Non-genetic variance on the scale of allelic effects.
$\sigma^2{}_p$	Phenotypic variance on the scale of allelic effects.
$\sigma^2{}_f$	A parameter specifying the decline in fitness with deviation from optimum.†
C	Coefficient of centripetal selection.‡
C^*	Defined to be equal to $C\sigma^2{}_m/\sigma^2{}_p$.
n_a	Mean number of alleles at equilibrium.
H	Mean frequency of heterozygotes at equilibrium.

† LATTER (1970); equation 2.
‡ Given by $C = \sigma^2{}_p/(\sigma^2{}_e + \sigma^2{}_f)$ for a single locus model (LATTER, 1970).

where d_{ij} is the deviation of the mean activity of A_iA_j from optimal. In a population homozygous for an allele of optimal activity, approximately two thirds of newly arising mutants would then have relative selective values greater than $1 - \frac{1}{2}C^*$ as heterozygotes, and selective values greater than $1 - 2C^*$ as homozygotes.

The computer techniques used are described and justified in full by LATTER (1970). The simulation of centripetal selection involves a transformation of the vector of allelic frequencies p_i,

$$p_i' = p_i\,[1 - \tfrac{1}{2}C^*(\bar{x}^2 + 2\bar{x}a_i + a_i^2 + \tfrac{1}{2}\sigma_g^2)]/\bar{w}^* \qquad (2)$$

where $\bar{x}$ is the mean of the population on the scale of allelic effects, σ_g^2 is the corresponding genotypic variance (Table 1), and $\bar{w}^* = 1 - \frac{1}{2}C^*(\bar{x}^2 + \sigma^2{}_g)$. The scale of measurement is defined by $\sigma^2{}_m = 1.0$. Drift in gene frequency due to finite population size is introduced each generation by the use of a pseudorandom number generator to draw a random sample from a multinomial distribution with parameters $2N$; p_i, $i = 1, 2, \ldots, n$, where n denotes the number of alleles segregating. The mean number of mutant alleles per generation is $2N\mu$, where μ is the spontaneous rate of mutation. The actual number of mutants is determined in each generation by sampling from a Poisson distribution with mean $2N\mu$. The mutational events are allocated at random to the existing alleles in the population, weighting according to their frequencies after selection and genetic sampling.

The measurement of genetic divergence: For a *neutral* locus subject only to random drift in gene frequency and continual mutation to novel alleles, the change δp_i in the frequency of an allele A_i over a period of t generations has expectation $-t\mu p_i$ for $t\mu \ll 1$ and $p_i \neq 0$. Denote the level of heterozygosity in a cross between the ancestral and derived populations, with frequencies p_i, $p_i{}^*$ respectively, by

$$H_B = 1 - \Sigma p_i p_i{}^* \qquad (3)$$

and the mean frequency of heterozygotes in the two populations by H. Then

$$H_B - H = \tfrac{1}{2}\Sigma(p_i - p_i{}^*)^2 \qquad (4)$$

The expected level of within-population heterozygosity in an *equilibrium* population under random mating has been shown by KIMURA and CROW (1964) to be

$$E(H) = 4N\mu/(1 + 4N\mu) \qquad (5)$$

where the expectation is most usefully interpreted as an average over independently segregating loci with the same mutation rate μ. If equation (5) is satisfied, we may write

$$E(H_B - H) = \tfrac{1}{2}E\,[\Sigma p_i{}^{*2} - \Sigma p_i{}^2\,(1 - 2t\mu)] \\ = t\mu\,[1 - E(H)] \qquad (6)$$

In such an equilibrium situation, we may therefore define a parameter, γ, given by

$$\gamma = E(H_B - H)/[1 - E(H)] \qquad (7)$$

which is equal to $t\mu$ for populations separated in time by t generations, provided only that $t\mu \ll 1$. Estimates of γ may be used as a convenient measure of *genetic divergence* over relatively short periods of time: the statistic has a range of values from zero to unity and measures directly the rate of accumulation of mutational changes at a locus, expected to be μ per generation (KIMURA 1969). Isolated contemporary populations are similarly expected to diverge at a rate given by $\gamma = 2t\mu$, if individual population sizes remain constant before and after the splitting process, and the parent population is in equilibrium under drift and mutation.

Choice of parameter values: It is well known that many aspects of the behaviour of finite populations subject to selection and mutation are dependent on the magnitude of the parameters Ns and $N\mu$ respectively, where s and μ denote the selection coefficient and mutation rate (WRIGHT 1931; LI 1955; KIMURA and CROW 1964). In this study the coefficient of centripetal selection C, or its derivative C^* (Table 1), determines the relative selective values of genotypes segregating in the population, so that the parameters of importance are NC^* and $N\mu$ (LATTER 1970).

Numerous experiments with higher organisms indicate that the average mutation rate per gene per gamete is of the order of 10^{-5} (DRAKE 1969). If mutational changes of minor effect are also included, however, the total rate of spontaneous mutation could be at least as high as 5×10^{-5} (MUKAI 1964; KING and JUKES 1969). Effective population size in *Drosophila pseudoobscura* is probably of the order of 1,000 or less (NEI 1968). Analyses of the major racial groups of man, based on a model of drift in gene frequency due to sampling effects alone, tentatively lead to a figure of 2,000–5,000 for the effective population size during the period of differentiation (CAVALLI-SFORZA, BARRAI and EDWARDS 1964; CAVALLI-SFORZA 1969). A value of $N\mu$ equal to 0.05 has therefore been chosen in most of the regimes simulated in this study, since it covers a range of parameter combinations from $N = 5{,}000$, $\mu = 10^{-5}$ to $N = 1{,}000$, $\mu = 5 \times 10^{-5}$ which are of potential interest in discussions of the genetics of natural and human populations.

Operationally it is not necessary to simulate populations with N as large as 5,000, in view of the overriding importance of the values of NC^* and $N\mu$ in the determination of population structure. In this paper $N = 500$ has been used throughout—computer time rapidly becomes a limiting factor at higher values of N. Extrapolation to $N = 1{,}000$ or $N = 5{,}000$ must involve a corresponding reduction by a factor of 2 or 10 respectively in the estimated total and inbred genetic loads at equilibrium (LATTER 1970).

RESULTS

Long-term evolutionary changes under centripetal selection: The rate of amino acid replacement in a protein in the course of evolution is a fundamental parameter in molecular population genetics. (KIMURA 1969; FITCH and MARGOLIASH 1969). The rate of gene substitution, i.e. of the complete replacement of a predominant allele by a mutant derivative, is a closely related concept which ignores the number of mutational changes involved in the derivation of the mutant allele concerned. Meaningful estimates of both these parameters can be derived from computer populations spanning periods of the order of $100N$ generations, which is the routine time span involved in this study. For populations of effective breeding size in the range 1,000–5,000 this represents the passage of from 100,000 to 500,000 generations.

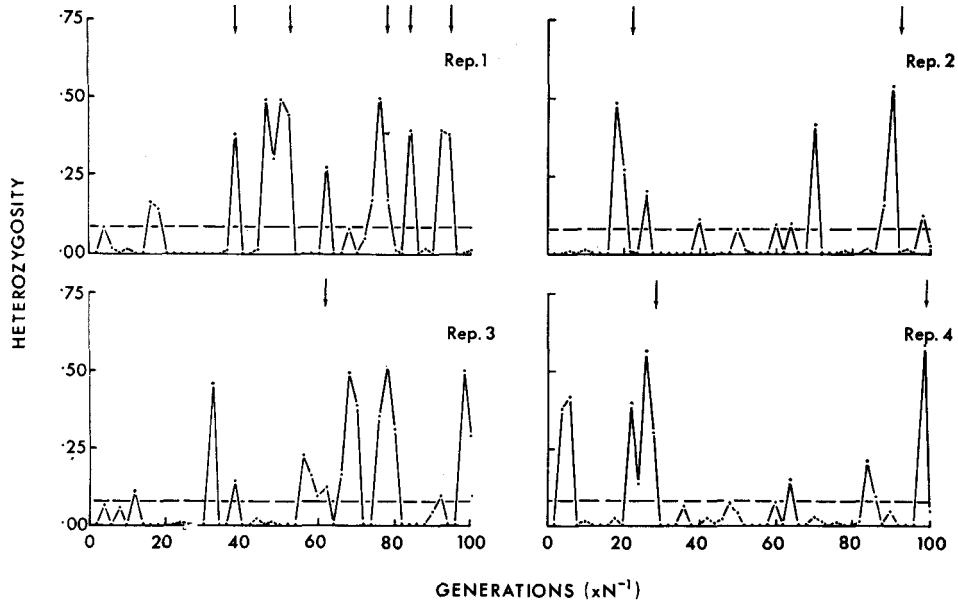

FIGURE 1.—Observed levels of heterozygosity at a neutral locus ($C^* = 0.00$) with $N\mu = 0.025$, $N = 500$, in four replicate populations each initially homozygous for a single allele. The dotted line indicates the mean value averaged over all replicates ($H = 0.083 \pm .015$). Arrows mark the completion of a single gene substitution at the locus. The rate of gene substitution is estimated from these data to be 0.025 per N generations.

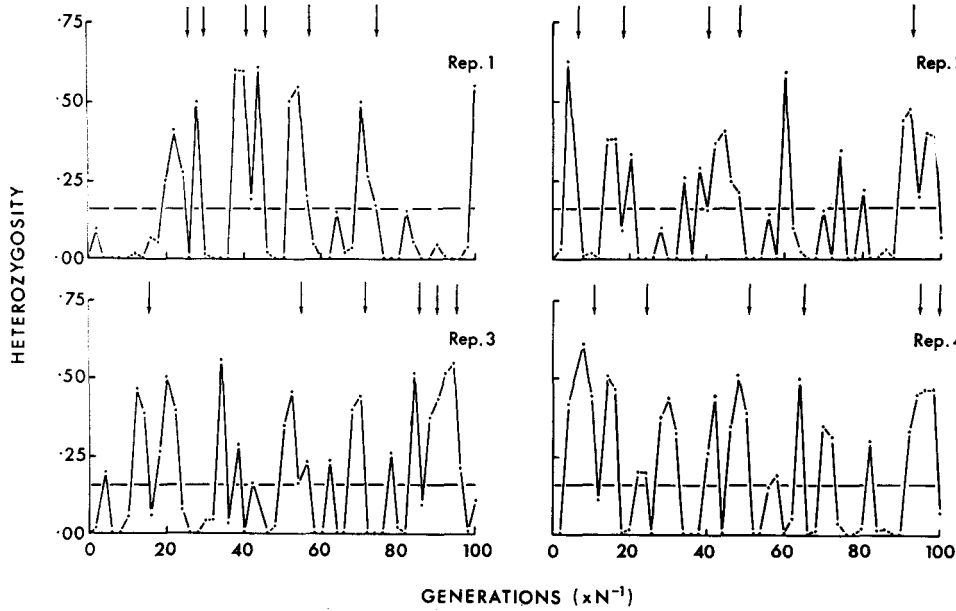

FIGURE 2.—Observed levels of heterozygosity at a neutral locus ($C^* = 0.00$) with $N\mu = 0.05$, $N = 500$ (c.f. Figure 1). The mean over the four replicates was $H = 0.164 \pm .014$. The rate of gene substitution observed was 0.058 per N generations.

Figures 1 and 2 present the salient features of long-term evolutionary change at *neutral* loci with $N\mu$ equal to 0.025 and 0.050 respectively. The observed level of heterozygosity is plotted at $2N$-generation intervals, and arrows indicate the completion of a single gene substitution, i.e. the generation at which a mutant allele first reaches a frequency of 100%. The rate of amino acid replacement (or the rate of accumulation of mutational changes) is in each case close to the expected value of $N\mu$ replacements per N generation for a neutral locus (KIMURA 1969); and the equilibrium values of H (mean heterozygosity over the intervals $10N$–$60N$; $70N$–$120N$) are in close agreement with the prediction $H = 4N\mu/(1 + 4N\mu)$ given by KIMURA and CROW (1964).

Of particular interest is the distribution of heterozygote frequency observed over time in these simulated populations, since present-day populations may be expected to show similar distributions of frequency in surveys of independently segregating loci. The heterozygote frequency distribution is markedly skew for equilibrium values of $H \leqslant 0.10$, long periods at very low levels of heterozygosity being interspersed with comparatively short intervals at extremely high values (Figure 1).

The effects of low intensities of centripetal selection on population behaviour can be deduced from Figures 3, 4 and the data of Tables 2–4. The populations concerned were each initially homozygous for a different allele with effect randomly chosen in the range $\pm(0.005/C^*)^{1/2}$, to give an initial population fitness

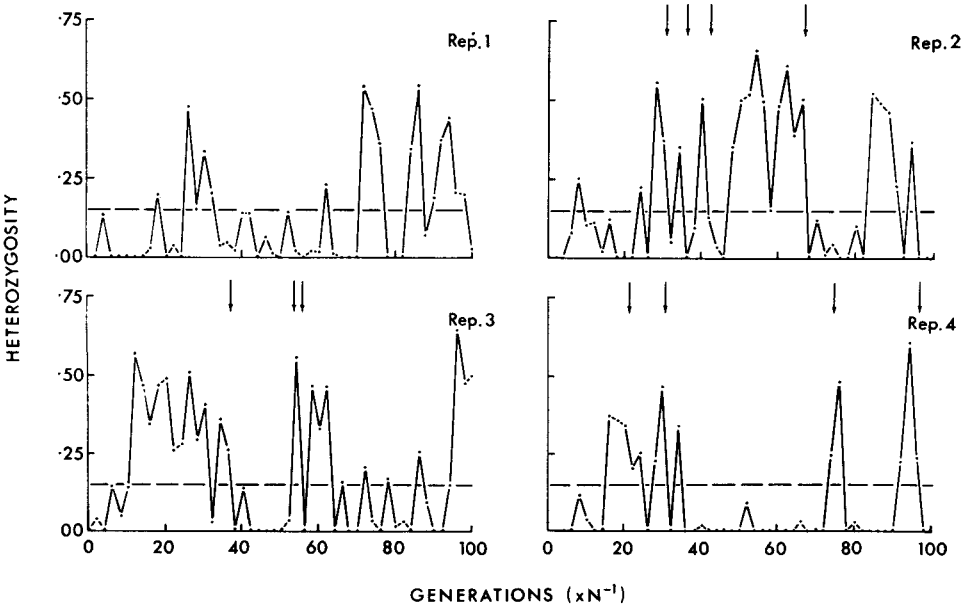

FIGURE 3.—Levels of heterozygosity and completed gene substitutions observed in four populations subjected to centripetal selection ($C^* = 0.005$) favouring an optimal level of gene or enzyme activity. $N = 500$, $N\mu = 0.05$. Comparisons with the populations of figure 2 are given in Tables 3–8.

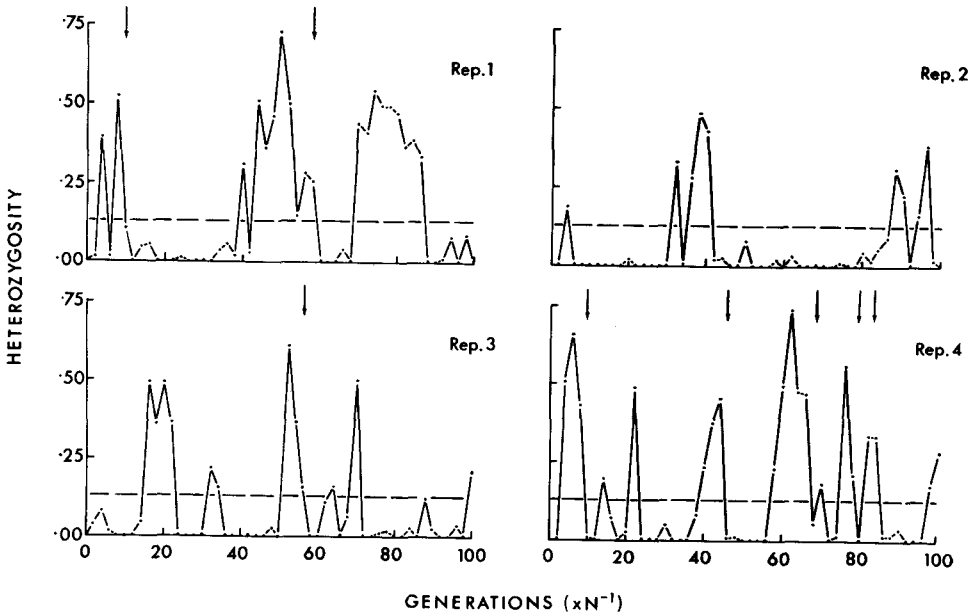

FIGURE 4.—As for Figure 3, with $C^* = 0.010$.

of at least 0.99. The scale of measurement of gene or enzyme activity has been defined to have origin at the optimum and unit equal to the mutational variance σ_m^2. The rates of gene substitution and accumulation of mutational changes can be seen from Table 2 to be significantly reduced by roughly 75% with an intensity of centripetal selection corresponding to $C^* = 0.05$. Averaging over the four values of $C^* = 0.005$, 0.01, 0.02 and 0.05, the observed rate of amino acid replacement is 0.024 ± .005 per N generations, by comparison with the theoretical value for a *neutral* locus of 0.05 (i.e. $N\mu$).

TABLE 2

Observed rates of gene substitution and amino-acid replacement in simulated populations with $N = 500$, $N\mu = 0.05$. *Standard errors are based on between-replicate variation*

Intensity of centripetal selection C^*	Substitutions per N generations†	
	Gene substitutions	Amino acid replacements
0.000	0.058 ± .002	0.063 ± .003
0.005	0.028 ± .010	0.031 ± .011
0.010	0.020 ± .011	0.024 ± .013
0.020	0.028 ± .005	0.028 ± .005
0.050	0.011 ± .004	0.012 ± .005

† Based on the number of substitutions observed in four replicate populations over a $100N$-generation period. The initial populations were each homozygous for a randomly chosen allele with effect in the range $\pm \sqrt{\frac{0.005}{C^*}}$.

Comparison of Figures 2, 3 and 4, with $C^* = 0.000$, 0.005 and 0.010 respectively, and $N\mu = 0.05$, suggests the importance of the following two phenomena: (i) the periods at *low* levels of heterozygosity are substantially increased in duration by centripetal selection, due to the retention of an allele of near-optimal activity and continued rejection of mutant alleles; and (ii) the duration of periods of *above-average* levels of heterozygosity is also increased by centripetal selection, due to selection for the heterozygote in polymorphisms involving alleles of above and below optimal activity (WRIGHT 1935).

Table 3 summarizes the available data on the duration of polymorphisms with the heterozygote closer to optimal activity than either of the corresponding homozygotes, by comparison with those in which a homozygote has the highest selective value. A polymorphism has been defined in this study as the occurrence of a variant at a frequency of 0.05 or greater for at least $0.2N$ generations (LATTER 1970). Such an operational definition confines attention to those alleles which can readily be detected in surveys involving samples of the order of 100 individuals, and which are in addition likely to be represented in each isolate of a recently subdivided population.

The mean duration of such polymorphisms at a neutral locus ($C^* = 0.000$; $N\mu = 0.05$) was observed to be $1.24N$ generations, with a variance of 1.52 (Table 3). Under regimes of centripetal selection with C^* in the range 0.005–0.05, the overall average duration of polymorphisms is unaltered, but the *variance* is greatly increased by comparison with that at a neutral locus (last column, Table 3). Polymorphisms maintained by selection for the heterozygote have a mean duration which appears to be essentially independent of C^* over the tested range, averaging $2.14N$ generations: the remaining polymorphisms in which selection favoured one of the two homozygotes, had a mean duration of only $0.84N$. Overdominance is therefore a potentially significant feature of the model under examination, despite the fact that mutant alleles must ultimately arise with homozygote selective values greater than that of the current overdominant heterozygote.

The nature of isoallelic variation: We turn next to an examination of the *probability* of a polymorphism existing at a randomly chosen locus at the time of survey, based on the model of centripetal selection for an optimal level of gene or

TABLE 3

The duration of polymorphisms observed in simulated populations with $N = 500$, $N\mu = 0.05$

Intensity of selection C^*	Duration of polymorphisms† with Superior heterozygote Mean	Variance	Superior homozygote Mean	Variance	Total Mean	Variance
0.000	—		—		1.24 ± .09	1.52
0.005	2.03 ± .40	7.88	0.88 ± .10	1.31	1.22 ± .14	3.46
0.010	2.43 ± .68	12.49	0.84 ± .12	1.38	1.21 ± .19	4.34
0.020	1.92 ± .33	4.15	0.90 ± .13	1.23	1.26 ± .15	2.50
0.050	2.18 ± .56	5.56	0.72 ± .12	0.68	1.12 ± .19	2.38

† A polymorphism has been operationally defined as the occurrence of a variant at a frequency of 0.05 or greater, with a duration of at least $0.2N$ generations. The duration is expressed in each case as a multiple of N.

TABLE 4

Incidence of polymorphisms in simulated populations with $N = 500$, $N\mu = 0.05$

Intensity of selection C^*	Probability of polymorphism	Heterotic polymorphisms† Mean duration (N)	Mean heterozygote superiority (percent)	Probability
0.000	0.426 ± .032	—	—	—
0.005	0.388 ± .058	3.59 ± 0.66	0.30 ± .05	0.210 ± .063
0.010	0.312 ± .043	4.08 ± 1.05	0.25 ± .05	0.147 ± .053
0.020	0.251 ± .042	2.88 ± 0.46	0.32 ± .06	0.138 ± .039
0.050	0.175 ± .046	3.29 ± 0.73	0.26 ± .11	0.080 ± .039

† Heterotic polymorphisms are those in which the heterozygote concerned is superior in fitness to both homozygotes, which have persisted in the population for at least N generations: this operational definition effectively restricts the classification to those overdominant polymorphisms which become established in the population.

enzyme activity. The data of Table 4 show that the total probability of polymorphism is reduced at high values of C^*. However, in a survey of independently segregating loci, roughly one half of the polymorphisms detected would be expected to be "heterotic polymorphisms" as defined in Table 4; *viz.* polymorphisms maintained by overdominance for N generations or more, and hence classifiable operationally as *established* overdominant polymorphisms. The mean duration and mean heterozygote superiority of the heterotic polymorphisms are both largely independent of the intensity of selection over the tested range of values of C^* (Table 4): the duration as a multiple of N was 3.46 ± .38 N generations, and the mean superiority in fitness of the heterozygotes was 0.28% ± .04% relative to that of the homozygotes.

The equilibrium statistics of Table 5 show the mean number of alleles segregating (n_a) and the mean level of heterozygosity (H) to decrease with increasing selection intensity, though the effect is slight for $C^* \leqslant 0.01$. The observed values of H, with C^* in the range 0.005–0.05, correspond to the levels of heterozygosity found for electrophoretic variants in natural populations (PRAKASH, LEWONTIN and HUBBY 1969; HARRIS 1969a; SELANDER and YANG 1969). Of particular interest, then, are the data concerning the total genetic load due to departure from optimal activity, *viz.* $L = \frac{1}{2}C^*(\bar{x}^2 + \sigma_g^2)$, where $\bar{x}$ denotes the deviation of the population mean from the optimum, and σ_g^2 the genetic variance in activity about the mean (LATTER 1970). The available estimates suggest that L is largely independent of C^* over the tested range, though the standard errors are large. The same is true of estimates of the inbred load carried by an equilibrium population, *viz.* $L_I = \frac{1}{2}C^* \sigma_g^2$.

The *distribution* of heterozygote frequency observed in the simulated populations (Table 6) is of interest for two reasons. On the one hand, direct comparisons can be made with the distribution observed over loci in a survey of present-day populations: major discrepancies may indicate that the model under discussion is inappropriate, or that important facets of population behaviour are being ignored. On the other hand, an understanding of the shape of the distributions is essential in discussions of the relationship between observed rates of amino-

TABLE 5

Equilibrium statistics for simulated populations with N = 500, Nμ = 0.05. *The figures are means over four replicate populations, based on the survey periods* 10N–60N, 70N–120N

Intensity of selection C^*	Equilibrium statistics†					
	n_a	H	σ^2_g	$\bar{x}^2$	$L(\times 10^4)$	$L_I(\times 10^4)$
0.000	2.48 ± .05	0.164 ± .014	0.149 ± .024	—	—	—
0.005	2.38 ± .08	0.154 ± .022	0.118 ± .038	0.549 ± .113	16.7 ± 3.0	2.95 ± .95
0.010	2.26 ± .09	0.129 ± .021	0.056 ± .016	0.128 ± .047	9.2 ± 2.5	2.80 ± .80
0.020	2.19 ± .08	0.099 ± .021	0.028 ± .006	0.168 ± .036	19.6 ± 3.7	2.80 ± .60
0.050	1.83 ± .07	0.065 ± .021	0.007 ± .001	0.029 ± .012	9.0 ± 3.0	1.75 ± .28

† n_a denotes the mean number of alleles segregating; H the mean level of heterozygosity; σ^2_g the within-population genetic variance; $\bar{x}^2$ the mean squared deviation of the population mean from the optimum; L is the total genetic load and L_I the inbred load, i.e. $L = \frac{1}{2}C^*(\bar{x}^2 + \sigma^2_g)$; $L_I = \frac{1}{2}C^*\sigma^2_g$.

acid replacement in the course of evolution, and current estimates of population variability (FITCH and MARGOLIASH 1969). Apart from the reduction in mean heterozygosity with increasing C^*, the frequency distributions of Table 6 are essentially similar, being markedly skew, with a very high probability of frequencies less than 0.05

Calculations of the probability of heterozygosity at a *particular* locus in a present-day population may be misleading if based only on the *expected* level of heterozygosity in an equilibrium population. While the observed rate of amino acid replacement in a protein in the course of evolution may be used to deduce a value of μ, the "neutral" mutation rate, and the corresponding value of $N\mu$ used to predict H, it should not surprise us if the actual frequency of heterozygotes is very different from expectation, as in the case of fibrinopeptide A discussed by FITCH and MARGOLIASH (1969). For a locus with a value of H of the order of 0.15, for instance, the data of Table 6 show a chance greater than 50% that a present-day population would be found to have a level of heterozygosity less than 0.05.

TABLE 6

The distribution of heterozygote frequency in simulated populations with N = 500, Nμ = 0.05

Frequency interval	Distribution of heterozygote frequency as a function of C^*				
	0.000	0.005	0.010	0.020	0.050
0.00 – 0.05	0.486	0.518	0.584	0.644	0.755
0.05 – 0.15	0.127	0.129	0.125	0.118	0.081
0.15 – 0.25	0.101	0.082	0.062	0.062	0.046
0.25 – 0.35	0.080	0.068	0.057	0.051	0.037
0.35 – 0.45	0.078	0.070	0.062	0.049	0.030
0.45 – 0.55	0.091	0.100	0.085	0.061	0.048
0.55 – 0.65	0.033	0.026	0.018	0.014	0.003
0.65 – 0.75	0.005	0.006	0.007	0.002	0.000
Mean (H)	0.164	0.154	0.129	0.099	0.065

TABLE 7

Mean values of H_B—H *and* γ *in simulated populations with* N = 500, Nμ = 0.05. *The parameter* γ, *defined by equation 7, is a direct measure of the degree of genetic divergence*

Intensity of selection C^*	H_B—H in populations separated by†			
	N generations	$2N$ generations	$3N$ generations	$4N$ generations
0.000	0.047 ± .012	0.074 ± .020	0.104 ± .029	0.214 ± .056
0.005	0.026 ± .006	0.047 ± .015	0.108 ± .032	0.103 ± .039
0.010	0.017 ± .005	0.057 ± .027	0.078 ± .025	0.052 ± .027
0.020	0.016 ± .006	0.038 ± .025	0.094 ± .039	0.086 ± .015
0.050	0.021 ± .013	0.040 ± .021	0.018 ± .012	0.090 ± .053
Mean γ	0.029 ± .005	0.059 ± .011	0.093 ± .015	0.126 ± .021

† Each 10N generation interval has been subdivided to give non-overlapping periods of duration N, $2N$, $3N$ and $4N$ generations. The estimate of H_B—H is $\frac{1}{2}\sum_i (p_i - p_i^*)^2$, where p_i, p_i^* denote the frequency of allele A_i at the beginning and end of the period concerned.

Rate of genetic divergence: The theory leading to equations (6) and (7) is concerned exclusively with neutral loci. However, the data in Tables 7 and 8 indicate that the measure of genetic divergence, γ, may be equally useful in specifying the rate of genetic divergence in populations subject to centripetal selection in addition to drift and mutation. The statistic can be seen to increase linearly with time, and is closely related to the rate of amino acid replacement (i.e. accumulation of mutational changes) observed throughout the long-term evolutionary history of the populations concerned (Table 8).

The *angular* distance between pairs of populations, $2\theta/\pi$, defined by CAVALLI-SFORZA and EDWARDS (1967) for a pure model, is based on the following relationship

$$\cos\theta = \sum_i (p_i p_i^*)^{\frac{1}{2}} \qquad (8)$$

Estimates of $2\theta/\pi$ have been calculated for comparison with the values of γ in Table 7, and are presented in Table 9. The angular measure can be seen not to change linearly with time, and gives estimates of genetic distance over periods

TABLE 8

The rates of genetic divergence and of amino acid replacement in simulated populations with N = 500, Nμ = 0.05

Intensity of selection C^*	Genetic divergence per N generations (γ)	Rate of amino acid replacement per N generations†
0.000	0.051 ± .007	0.063 ± .003
0.005	0.033 ± .005	0.031 ± .011
0.010	0.024 ± .005	0.024 ± .013
0.020	0.024 ± .005	0.028 ± .005
0.050	0.019 ± .006	0.012 ± .005
Means	0.030 ± .003	0.032 ± .004

† c.f. Table 2.

TABLE 9

Mean values of angular distance, as defined by Cavalli-Sforza and Edwards (1967), derived from the same allelic frequencies as the statistics of Table 7

Intensity of selection C^*	$\frac{2\theta}{\pi}$ in populations separated by N generations	$2N$ generations	$3N$ generations	$4N$ generations
0.000	0.229 ± .014	0.226 ± .024	0.254 ± .032	0.393 ± .066
0.005	0.167 ± .027	0.187 ± .022	0.261 ± .024	0.275 ± .043
0.010	0.135 ± .015	0.187 ± .042	0.219 ± .035	0.182 ± .042
0.020	0.125 ± .013	0.149 ± .029	0.223 ± .031	0.220 ± .015
0.050	0.137 ± .014	0.130 ± .025	0.108 ± .022	0.199 ± .066
Mean $\frac{2\theta}{\pi}$	0.159 ± .008	0.176 ± .013	0.213 ± .013	0.254 ± .022

of short duration which substantially overestimate the true rate of gene substitution.

DISCUSSION

The evolutionary model explored in this paper has been shown to have many of the properties required to accommodate both the experimental data concerning the extent and stability of isoallelic variation in present-day populations, and the estimated rates of evolutionary change based on amino acid sequence data. A value of $N\mu = 0.05$ has been used in most of the simulated populations, based on current estimates of total mutation rates and effective breeding population size in Drosophila and man. A population size of $N = 500$ with $\mu = 10^{-4}$ has been chosen for reasons of economy, though the results can be considered as equally applicable to a range of parameter combinations such as $N = 1{,}000$, $\mu = 5 \times 10^{-5}$; $N = 5{,}000$, $\mu = 10^{-5}$; etc., with appropriate adjustment of the estimates of genetic load.

Centripetal selection for a fixed optimal level of gene or enzyme activity, superimposed on the effects of drift and mutation, has been shown to lead to the following changes in the genetic structure of a random breeding population:

(i) The stability of allelic frequencies is greatly increased by selection intensities in the range of $0.005 \leqslant C^* \leqslant 0.05$: the enhanced stability can be expressed in terms of the rate of long-term evolutionary change observed, or in terms of genetic divergence measured over periods of the order of N generations (Table 8).

(ii) The *mean* duration of all polymorphisms arising in a population is virtually unaffected by centripetal selection over the range of intensities investigated: however, those in which a homozygote has highest selective value are *decreased* in duration, while those showing heterozygote superiority have a substantially *longer* life than average (Table 3).

(iii) The probability that a randomly chosen locus will be polymorphic at the time of sampling is reduced at high intensities of centripetal selection, but is of

the order of 0.28 for $0.005 \leqslant C^* \leqslant 0.05$. Approximately half of these polymorphisms can be expected to be maintained by selection for the heterozygote over periods in excess of N generations (Table 4): the mean duration of this group of stable polymorphisms is close to 3.5 N generations.

(iv) The mean frequency of heterozygotes decreases with increasing selection intensity, averaging 0.11 for $0.005 \leqslant C^* \leqslant 0.05$ (Table 5): if the chosen value of $N\mu$ proves to be an appropriate figure when improved estimates of effective population size and total mutation rates are available, the tested range of values of C^* appears to be of the right order to account for isoallelic variation found in surveys of natural populations.

KING and JUKES (1969) and KIMURA (1969) have argued that most evolutionary changes in the amino acid sequence of a protein must be due to the passive fixation of effectively neutral mutations. However, a model allowing for only two types of mutations, *viz.* deleterious changes which are eliminated by natural selection, and effectively neutral mutations which may be fixed by chance in a finite population, does not appear to explain the observed patterns of allozyme frequencies in geographical surveys of natural populations (PRAKASH, LEWONTIN and HUBBY 1969; MAYNARD SMITH 1970). The model examined in this paper has been shown to introduce the additional phenomenon of selection for the heterozygote at a high proportion of loci, and provides for the maintenance of stable patterns of allelic frequencies at polymorphic loci over extended periods. It is potentially applicable to populations which have been geographically isolated for something like $2N$ generations or less, where N denotes the effective size of each population. The effects of migration on the stability of allele frequencies with centripetal selection have not yet been examined.

One of the attractive features of the model of KING and JUKES (1969) and KIMURA (1969) is the deduction that the rate of amino acid substitution is expected to equal the rate of mutation to neutral alleles: for constant μ and constant generation interval, therefore, the number of mutational changes in a protein is predicted to be proportional to evolutionary time. The centripetal selection model has been shown to possess similar properties of regularity, provided NC^* also remains constant over the period of observation (Table 8). The rate of genetic change measured by the distance parameter γ over relatively short periods has been shown not to differ significantly from the long-term rate of accumulation of mutational changes in the simulated populations of this study. The use of γ as a measure of differentiation among populations should therefore make possible the accurate comparison of rates of recent and long-term evolution.

We turn now to the question of genetic load under centripetal selection, drift and mutation. The data of Table 5 suggest that, for constant N, both the total load, L, and the inbred load, L_I, are largely independent of the intensity of centripetal selection over the range of values tested, *viz.* $0.005 \leqslant C^* \leqslant 0.05$. We may therefore use the mean values, *viz.* $L = 13.6 \times 10^{-4} \pm 1.5 \times 10^{-4}$ and $L_I = 2.58 \times 10^{-4} \pm 0.35 \times 10^{-4}$ in predicting average levels of the genetic load. Provided effective population sizes in Drosophila are of the order of 500–1,000, the calculations summarized in Table 10 indicate that 10,000–20,000 loci would have a

TABLE 10

Predicted values of total genetic load per locus in a random breeding population, and of the inbred load per locus, for populations with $N\mu = 0.05$

Population size N	Intensity of selection C^*	Total load per locus $L(\times 10^4)$	Inbred load per locus $L_I(\times 10^4)$	Number of loci†
500	0.0050 – 0.0500	13.6 ± 1.5	2.58 ± .35	8,900
1,000	0.0025 – 0.0250	6.8 ± 0.8	1.29 ± .18	17,800
2,000	0.0012 – 0.0125	3.4 ± 0.4	0.65 ± .09	35,600
5,000	0.0005 – 0.0050	1.4 ± 0.2	0.26 ± .04	89,000

† Number of independently segregating loci necessary to give a summed inbred load equal to 2.30 ± 0.13 (LATTER 1970). The standard error in each case is approximately 15% of the estimated number of loci.

combined inbred load equal to that observed under competitive conditions in *D. melanogaster* (LATTER and ROBERTSON 1962; LATTER 1970). The actual number of loci in Drosophila is not known with any precision, but it is almost certainly less than 40,000, and must be greater than 3,000 (KING and JUKES 1969). The consequences of the model are therefore roughly in accord with expectation.

The foregoing discussion indicates that the *intermediate* activity of heterozygotes will have important consequences in terms of population structure, if selection favours an optimal level of gene activity. It is relevant here to consider briefly an alternative model based on the ability of some heterozygotes to form *hybrid* molecules. Approximately half of the enzymes studied by electrophoretic methods in heterozygous individuals have been shown to form hybrid bands, indicating the presence of a heteropolymer in addition to the parental homopolymers (SHAW 1965; HARRIS 1969b). Is there any evidence to suggest that hybrid molecules confer a selective advantage on their carriers in natural populations, providing the selective mechanisms for the maintenance of stable balanced polymorphisms? Current knowledge of the evolution of proteins with biquaternary structure, such as haemoglobin and lactate dehydrogenase, indicates that they are coded for by the equivalent of *fixed heterozygotes*, produced by gene duplication and subsequent co-evolution of the two loci concerned (MARKERT 1968). It may therefore be surmised that some hybrid molecules formed by heterozygotes will be found to have an adaptive advantage by comparison with the corresponding molecules produced by homozygotes. However, a survey of polymorphic loci in man (HARRIS 1969a), and in North American populations of *Drosophila pseudoobscura* (PARKASH, LEWONTIN and HUBBY 2969) and *D. melanogaster* (O'BRIEN and MACINTYRE 1969) fails to show any conspicuous difference in the mean level of heterozygosity at loci with and without hybrid bands. Thirteen loci giving rise to hybrid zones on electrophoresis have a mean frequency of heterozygotes of 0.22 ± .06, compared with a figure of 0.23 ± .04 for 16 loci classified as producing no hybrid enzyme.

LITERATURE CITED

CAVALLI-SFORZA, L. L., 1969 Human diversity. Proc. XII Intern. Congr. Genetics **3**: 405–416.

CAVALLI-SFORZA, L. L., I. BARRAI and A. W. F. EDWARDS, 1964 Analysis of human evolution under random genetic drift. Cold Spring Harbor Symp. Quant. Biol. **24**: 9–20.

CAVALLI-SFORZA, L. L. and A. W. F. EDWARDS, 1967 Phylogenetic analysis: models and estimation procedures. Evolution **21**: 550–570.

DRAKE, J. W., 1969 *An Introduction to the Molecular Basis of Mutation.* Holden-Day, San Francisco.

FITCH, W. M. and E. MARGOLIASH, 1969 The usefulness of amino acid and nucleotide sequences in evolutionary studies. Evolutionary Biol. **4**: 67–109.

GIBSON, J., 1970 Enzyme flexibility in *Drosophila melanogaster*. Nature **227**: 959–960.

GILLESPIE, J. H. and K. KOJIMA, 1968 The degree of polymorphisms in enzymes involved in energy production compared to that in non-specific enzymes in two *Drosophila ananassae* populations. Proc. Natl. Acad. Sci. U.S. **61**: 582–585.

HARRIS, H., 1966 Enzyme polymorphisms in man. Proc. Roy. Soc. Lond. B **164**: 298–310. ——, 1969a Enzyme and protein polymorphism in human populations. Brit. Med. Bull. **25**: 5–13. ——, 1969b Genes and isozymes. Proc. Roy. Soc. Lond. B **174**: 1–31.

HARRIS, H., D. A. HOPKINSON and J. LUFFMAN, 1968 Enzyme diversity in human populations. Ann. N.Y. Acad. Sci. **151**: 232–242.

KIMURA, M., 1969 The rate of molecular evolution considered from the standpoint of population genetics. Proc. Nat. Acad. Sci. U.S. **63**: 1181–1188.

KIMURA, M. and J. F. CROW, 1964 The number of alleles that can be maintained in a finite population. Genetics **49**: 725–738.

KING, J. L. and T. H. JUKES, 1969 Non-Darwinian evolution. Science **164**: 788–798.

KIRKMAN, H. N., C. KIDSON and M. KENNEDY, 1968 In *Hereditary Disorders of Erythrocyte Metabolism.* (ed. Beutler) pp. 126–145. Grune and Stratton, New York.

KOEHN, R. K., 1969 Esterase heterogeneity: dynamics of a polymorphism. Science **163**: 943–944.

KOEHN, R. K. and D. I. RASMUSSEN, 1967 Polymorphic and monomorphic serum esterase heterogeneity in catostomid fish populations. Biochem. Genet. **1**: 131–144.

LATTER, B. D. H., 1970 Selection in finite populations with multiple alleles. II. Centripetal selection, mutation and isoallelic variation. Genetics **66**: 165–186.

LATTER, B. D. H. and A. ROBERTSON, 1962 The effects of inbreeding and artificial selection on reproductive fitness. Genet. Res. **3**: 110–138.

LI, C. C., 1955 *Population Genetics.* University of Chicago Press.

MANWELL, C. and C. M. A. BAKER, 1970 *Molecular Biology and the Origin of Species.* Sidgwick and Jackson, London.

MARKERT, C. L., 1968 The molecular basis for isozymes. Ann. N. Y. Acad. Sci. **151**: 14–40.

MAYNARD SMITH, J., 1970 Population size, polymorphism and the rate of non-Darwinian evolution. Am. Naturalist **104**: 231–237.

MUKAI, T., 1964 The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. Genetics **50**: 1–19.

NEI, M., 1968 The frequency distribution of lethal chromosomes in finite populations. Proc. Nat. Acad. Sci. U.S. **60**: 517–524.

O'BRIEN, S. J. and R. J. MACINTYRE, 1969 An analysis of gene–enzyme variability in natural populations of *Drosophila melanogaster* and *D. simulans*. Am. Naturalist **103**: 97–113.

PRAKASH, S., R. C. LEWONTIN and J. L. HUBBY, 1969 A molecular approach to the study of genic heterozygosity in natural populations. IV. Patterns of genic variation in central, marginal and isolated populations of *Drosophila pseudoobscura*. Genetics **61**: 841–858.

RASMUSON, B., L. R. NILSON and M. RASMUSON, 1966 Effects of heterozygosity on alcohol dehydrogenase (Adh) activity in *Drosophila melanogaster*. Hereditas **56**: 311–316.

SELANDER, R. K. and S. Y. YANG, 1969 Protein polymorphism and genic heterozygosity in a wild population of the house mouse (*Mus musculus*) Genetics **63**: 653–667.

SHAW, C. R., 1965 Electrophoretic variation in enzymes. Science **149**: 936–943.

WRIGHT, S., 1931 Evolution in Mendelian populations. Genetics **16**: 97–159. ——, 1935 Evolution in populations in approximate equilibrium. J. Genet. **30**: 257–266.

YOSHIDA, A., 1969 Genetic variants of human glucose-6-phosphate dehydrogenase. Japan J. Genetics **44**, Suppl. **1**: 258–265.