# SELECTION IN FINITE POPULATIONS WITH MULTIPLE ALLELES. II. CENTRIPETAL SELECTION, MUTATION, AND ISOALLELIC VARIATION[1]

B. D. H. LATTER[2]

*Iowa State University, Ames, Iowa*

Received July 22, 1969

IN this series of papers the behavior of multiple allelic systems in populations subject to selection, genetic sampling, and mutation is to be systematically explored. The first paper (LATTER and NOVITSKI 1969) dealt with the effects of directional selection for a quantitative character in small populations, given an infinite base population with multiallelic variation similar to that described by KIMURA (1965). KIMURA's model postulates (i) that mutation can give rise to a very large number of alleles at each locus influencing the expression of the quantitative trait; (ii) that the genes are additive in effect on the character; and (iii) that the optimal phenotype is fixed, with fitness decreasing in proportion to the squared deviation of an individual's genotypic value from the optimum. KIMURA has shown that with small mutational changes, the allelic effects at a given locus are normally distributed at equilibrium in a large population. In this paper computer simulation techniques are used to extend KIMURA's model to populations of finite size, where the expected number of alleles segregating per locus is not necessarily large.

Contributions to the theory of centripetal selection have been made by many authors, including FISHER (1930), WRIGHT (1935), HALDANE (1954), ROBERTSON (1956), KOJIMA (1959), LATTER (1960), LEWONTIN (1964), JAIN and ALLARD (1965) and SINGH and LEWONTIN (1966). The emphasis of these studies has been the examination of deterministic equilibria for a variety of genetic models; in no case have the joint effects of genetic sampling and mutation been considered.

KIMURA and CROW (1964), EWENS (1964), WRIGHT (1966) and KIMURA (1968), have discussed the maintenance of isoallelic variants in finite populations in the *absence* of selection. The level of heterozygosity in such a population at equilibrium is expected to be approximately

$$1-F = \frac{4N\mu}{1+4N\mu}$$

where $N$ is the effective population size and $\mu$ the mutation rate, the number of possible allelic states being assumed to be very large (KIMURA and CROW 1964). EWENS (1964), WRIGHT (1966) and KIMURA (1968) have given approximate

[1] Journal paper No. J-6589, Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa. Project No. 1669. Supported by National Institute of Health Grant No. GM 13827.

[2] On leave from Division of Plant Industry, CSIRO, Canberra, A.C.T., Australia.

algebraic expressions for the number of alleles expected to be segregating at equilibrium, showing the number to be a function of both $N$ and $N\mu$. KIMURA has also presented a set of computer simulation results for comparison with the predicted values.

The objectives of the present study are:

(i) to examine the effects of centripetal selection on a multiallelic system in finite populations;

(ii) to draw comparisons with the infinite population theory of KIMURA (1965) on the one hand, and that developed for neutral isoallelic variation in finite populations on the other; and

(iii) to discuss the relevance of the results to recent studies of enzyme polymorphisms in Drosophila, man, and mice (SHAW 1965; LEWONTIN 1967; HARRIS 1969; SELANDER and YANG 1969.)

## SELECTION FOR A FIXED INTERMEDIATE OPTIMUM IN LARGE POPULATIONS

Consider a metric character $x$ with the following probability density function among juveniles:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2_p}} \exp\left[-\frac{(x-\bar{x})^2}{2\sigma^2_p}\right] \tag{1}$$

where $\bar{x}$ denotes the mean and $\sigma^2_p$ the phenotypic variance.

The mean reproductive fitness $\phi(x)$ of individuals with phenotypic value $x$ will be supposed to decrease with deviation from the optimum according to the relation

$$\phi(x) = \exp\left[-\frac{x^2}{2\sigma^2_f}\right] \tag{2}$$

where the optimal value of $x$ is taken to be zero, with a relative fitness of unity. The scale constant $\sigma_f$ specifies the rate at which fitness declines with deviation of $x$ from the optimum. The mean fitness of the population, relative to that of the optimal phenotype, is then given by

$$\bar{w} = \int_{-\infty}^{+\infty} \phi(x)f(x)\,dx$$

$$= \frac{\sigma_f}{\sigma} \exp\left[-\frac{\bar{x}^2}{2\sigma^2}\right] \tag{3}$$

where $\sigma^2 = \sigma^2_p + \sigma^2_f$. If differences in reproductive fitness are a matter only of differential survival, the variable $x$ *after* selection is distributed with mean $\bar{x}\left(\frac{\sigma_f}{\sigma}\right)^2$ and variance $\sigma^2_p\left(\frac{\sigma_f}{\sigma}\right)^2$. The deviation of the mean from the optimal value, and the phenotypic variance of the trait, are therefore reduced by the same factor $\left(\frac{\sigma_p}{\sigma}\right)^2$. We will call this factor the *coefficient of centripetal selection*, denoted by

$$C = \sigma^2_p\,(\sigma^2_p + \sigma^2_f)^{-1} \tag{4}$$

with a scale of values ranging from zero (no selection) to unity (absolute selection of the optimal phenotype alone). The product $Ch^2$, where $h^2$ is the heritability of

the metric trait, corresponds operationally to the definition of *homeostatic strength* proposed by ROBERTSON (1956), viz., the proportional return of the mean in one generation, following a period of directional selection away from the optimal value.

In a population with mean at the optimum, the coefficient of centripetal selection is simply related to the parameter $I$ defined by HALDANE (1954) to measure the *intensity* of natural selection for the optimum. The expression relating the two parameters is

$$I = -\tfrac{1}{2} \log_e (1-C) \tag{5}$$

so that $C$ is approximately equal to $2I$ at low selection intensities.

*Selective values under centripetal selection*: Suppose the genotypic configuration $A_iA_j$ at a given locus is present in the population with frequency $p_ip_j$. If the total contribution of the locus to the phenotypic variance is small, we may consider the subpopulation of values of the variable $x$, for those individuals with configuration $A_iA_j$, to be normally distributed with mean $d_{ij}$ and variance $\sigma^2_p$.

The contribution of subpopulation $A_iA_j$ to the succeeding generation is then proportional to

$$w'_{ij} = \exp\left[-\tfrac{1}{2} C \frac{d^2_{ij}}{\sigma^2_p}\right] \tag{6}$$

from equation (3), which we may take to be approximately given by

$$w^*_{ij} = 1 - \tfrac{1}{2} C \frac{d^2_{ij}}{\sigma^2_p} \tag{7}$$

at low intensities of selection. The mean of these approximate selective values is

$$\begin{aligned} \bar{w}^* &= \sum_{i,j} p_ip_j w^*_{ij} \\ &= 1 - \tfrac{1}{2} C [\bar{x}^2 + \sigma^2_g] / \sigma^2_p \end{aligned} \tag{8}$$

where $\sigma^2_g$ denotes the total genotypic variance contributed by the A locus. The expected frequency of allele $A_i$ in the progeny of surviving individuals is then

$$\begin{aligned} p'_i &= [\sum_j p_ip_j w^*_{ij}] / \bar{w}^* \\ &= \frac{p_i}{\bar{w}^*} [1 - \tfrac{1}{2} C (\bar{x}^2 + 2\bar{x}\alpha_i + \kappa_i) / \sigma^2_p] \end{aligned} \tag{9}$$

where $\alpha_i = \sum_j p_j (d_{ij} - \bar{x})$ and $\kappa_i = \sum_j p_j (d_{ij} - \bar{x})^2$. For the particular case in which all allelic effects are additive, i.e., $d_{ij} = \bar{x} + a_i + a_j$, we have

$$p'_i = \frac{p_i}{\bar{w}^*} [1 - \tfrac{1}{2} C (\bar{x}^2 + 2\bar{x}a_i + a^2_i + \tfrac{1}{2} \sigma^2_g)/\sigma^2_p]. \tag{10}$$

For the purposes of computer simulation, equations (9) and (10) are the most useful expressions for the changes in allelic frequencies from generation to generation. For algebraic manipulations, however, a more satisfactory expression is

$$\Delta p_i = p_i [\tfrac{1}{2} C (\sigma^2_g - 2\bar{x}\alpha_i - \kappa_i)/\sigma^2_p]. \tag{11}$$

For a model involving additive allelic effects this becomes

$$\Delta p_i = p_i [\tfrac{1}{2} C (\tfrac{1}{2} \sigma^2_g - 2\bar{x}a_i - a^2_i)/\sigma^2_p] \tag{12}$$

so that the frequency of the allele $A_i$ will increase whenever

$$(\bar{x} + a_i)^2 < \sum_k p_k (\bar{x} + a_k)^2. \tag{13}$$

The predicted change in the mean of the population due to changes in allelic frequencies at the A locus, under the assumption of additive allelic effects, is given approximately by

$$\Delta G = 2 \sum_i a_i \Delta p_i$$
$$= C[\bar{x}(-\sigma^2_g) - \sum_i p_i a^3{}_i]/\sigma^2_p$$

from (12). Summing over all segregating loci gives

$$\Delta \bar{x} = -\bar{x} C h^2$$

where $h^2$ is the heritability of the character, provided the sum of the $\sum_i p_i a^3{}_i$ can be neglected, i.e., that there is no overall tendency to directional skewness of the distributions of allelic values.

*Single locus models*: Models involving a single locus under centripetal selection are of particular interest, since they may be used to represent natural selection for an optimal level of the catalytic activity of a specific enzyme. The foregoing theory can readily be applied to such models by an appropriate redefinition of the parameter $C$. With a single locus contributing all the genetic variance in the measured variable $x$, the relative selective value of the genotype $A_iA_j$ can be seen from equation (3) to be proportional to

$$w^*{}_{ij} = 1 - \tfrac{1}{2}\left[\frac{d^2{}_{ij}}{\sigma^2{}_e + \sigma^2{}_f}\right] \tag{14}$$

provided the intensity of selection is low, where $\sigma^2{}_e$ denotes the nongenetic variance in the character. Equations (7) to (13) are then valid for the single locus model, if $C$ is taken throughout to be equal to $C = \sigma^2{}_p (\sigma^2{}_e + \sigma^2{}_f)^{-1}$. Equation (14) will be a valid approximation provided the $d_{ij}$ are small by comparison with $\sigma_f$.

*Changes in genotypic variance due to selection*: Selection for a fixed intermediate optimum is expected to lead not only to the maintenance of the population mean in the vicinity of the optimum, but also to changes in the level of genotypic variability within the population. For a multiallelic locus with additive allelic effects, the change in genetic variance from one generation to the next due to selection is expected to be

$$\Delta (\sigma^2_g) = 2\sum_i \Delta p_i a^2{}_i - 2[\sum_i \Delta p_i a_i]^2$$

where $\Delta p_i$ is given by equation (12). Substitution leads to the general expression

$$\Delta (\sigma^2_g) = \frac{C}{\sigma^2_p}[(\mu^2{}_2 - 2\bar{x}\mu_3 - \mu_4) - \tfrac{1}{2}\frac{C}{\sigma^2_p}(2\bar{x}\mu_2 + \mu_3)^2]$$

where $\mu_r = \sum_i p_i a^r{}_i$. For the particular case in which the mean is at the optimum and the $a_i$ are normally distributed, we have $\bar{x} = 0$, $\mu_3 = 0$ and $\mu_4 = 3\mu^2{}_2$, so that

$$\Delta (\sigma^2_g) = -\sigma^2_g [\tfrac{1}{2} C\sigma^2_g/\sigma^2_p] \tag{15}$$

With a single locus model we can equate $\sigma^2_g/\sigma^2_p$ to $h^2$, the heritability of the metric trait, and the total phenotypic variance can be seen to be reduced by a fraction $\frac{1}{2} Ch^4$. For a model involving $n$ identical loci, the corresponding reduction is $(1/2n)Ch^4$, ignoring complications due to departures from gametic equilibrium.

It is clear from the foregoing that the changes in genotypic variance due to

selection will in general be complex, even for a single locus model, since the assumption of normality of allelic effects cannot hold strictly for more than one generation. In particular it is apparent that the changes depend on the number of loci contributing to the total genotypic variance in the metric trait, direct evidence to this effect having recently been provided by the computer simulation study of ALLEN and FRASER (1968).

## THEORY UNDERLYING THE COMPUTER SIMULATION PROCEDURES

The first objective of this study is a quantitative assessment of the impact of centripetal selection on a multiallelic locus in finite populations. An understanding of this phenomenon requires the examination of changes in at least four parameters as selection and random sampling proceed, viz.: (i) the number of alleles segregating; (ii) the mean level of heterozygosity; (iii) the mean genotypic variance contributed by the locus; and (iv) the drift variance, defined as the variance among means of replicate populations. The second objective is to describe quantitatively the accumulation of *mutational* variability in populations subject to centripetal selection, in terms of the same four parameters.

The only satisfactory technique for such a comprehensive study is that of computer simulation, aided wherever possible by algebraic treatment. We will begin by elaborating the theory given in the previous section to take account of the effects of finite population size, thereby providing a frame of reference for the numerical results obtained by computer simulation. The theory necessary for the simulation of mutation in finite populations under centripetal selection will then be presented, and an approximate formula derived for the equilibrium genotypic variance.

*Changes in variance in finite populations:* The theory presented in the preceding section involves no restrictions as to the number of loci contributing to variation in the quantitative trait under selection. However, for multilocus models the genotypic variance due to the A locus, $\sigma^2_g$, was assumed to be small relative to the total variance, $\sigma^2_p$, and the coefficient of centripetal selection, $C$, was assumed to be sufficiently small for equation (7) to be a satisfactory approximation. With single locus models only the latter restriction is necessary.

An approximate expression can be given for the expected total change in $\sigma^2_g$ in *finite* populations, following $t$ generations of weak centripetal selection and random sampling in a population with mean initially at the optimal value. The expected change in genetic variance in one generation is approximately

$$\Delta(\sigma^2_g) = -\sigma^2_g\left[\tfrac{1}{2}\,C\left(\frac{\sigma^2_g}{\sigma^2_p}\right) + \frac{1}{2N}\right] \qquad (16)$$

for low intensities of selection. We are here making use of the fact that the decline in heterozygosity due to random sampling is expected to be a fraction $1/2N$ per generation (KIMURA 1955): for a locus with additive allelic effects the genetic variance is expected to decline by the same fraction (LATTER and NOVITSKI 1969).

Equations (15) and (16) are based on the assumption of normality of the distribution of allelic effects $a_i$. If this requirement is satisfied in the initial population, and changes in the moments $\mu_3$ and $\mu_4$ with time are negligible, we may derive an approximate formula for the magnitude of $\sigma^2_g$ as a function of $t$, viz.,

$$\sigma^2_g = (g^2\sigma^2_p)e^{\frac{-t}{2N}}[1 + NCg^2\,(1 - e^{\frac{-t}{2N}})]^{-1} \tag{17}$$

where $g^2$ denotes the initial value of $\sigma^2_g/\sigma^2_p$. If time is measured in units of $t/N$, we may then expect the rate of decline in genetic variance to be dependent primarily on the value of the parameter combination $NCg^2$.

*Variability due to mutation:* The following model is to be used exclusively in this paper to simulate the production of new genetic variation in a population due to mutational changes. Following KIMURA and CROW (1964) and EWENS (1964), it will be assumed that the number of allelic states at the locus is sufficiently large for each new mutant to represent a novel allele. We will consider only loci with *additive* allelic effects on the metric trait under selection, the existing alleles $A_i$ having frequencies $p_i$, $i = 1, 2, \ldots, n$, and effects $a_i$, coded so that $\Sigma p_i a_i = 0$.

Denote the probability of a mutational event by $\mu$, assumed to be the same for all alleles $A_i$. Then if $\sigma^2_g$, $\sigma^2_g{}^*$ denote the additive genetic variance due to the A locus before, and after, respectively, the occurrence of mutation in any specified generation, we have

$$\sigma^2_g = 2[\sum_i p_i a^2{}_i]$$

and

$$\begin{aligned}\sigma^2_g{}^* &= 2[\sum_i p_i(1-\mu)a^2{}_i + \sum_i \mu p_i\,(a^2{}_i + \sigma^2{}_m)] \\ &= \sigma^2_g + 2\mu\sigma^2{}_m\end{aligned} \tag{18}$$

where the changes in allelic effect $\delta a_i$ due to mutation are independent of $a_i$, and have expectation zero and variance $\sigma^2{}_m$. If the allelic effects $a_i$ are normally distributed, equations (16) and (18) may be combined for a population with mean at the optimum to give the following rough approximation for a finite population:

$$\Delta(\sigma^2_g) = -\sigma^2_g\left[\tfrac{1}{2}\,C\left(\frac{\sigma^2_g}{\sigma^2_p}\right) + \frac{1}{2N}\right] + 2\mu\sigma^2{}_m \tag{19}$$

where it is assumed for simplicity that the sequence of events in each generation is (i) measurement, (ii) selection, (iii) random sampling, and (iv) mutation.

The equilibrium variance due to this locus, $\hat{\sigma}^2_g$, is then expected to be given by the relationship

$$\frac{\hat{\sigma}^2_g}{2\sigma^2{}_m} = \frac{1}{4NC^*}\left[\sqrt{1 + (4NC^*)\,(4N\mu)} - 1\right] \tag{20}$$

where $C^* = C\sigma^2{}_m/\sigma^2{}_p$,

provided the distribution of allelic effects remains close to normal throughout the approach to equilibrium. This can be expected to be a valid assumption only for small values of $C^*$, i.e., for weak selection. For $C^* = 0$ or $(4NC^*)$ $(4N\mu)$ small by comparison with unity, equation (20) becomes

$$\hat{\sigma}^2{}_g = 4N\mu\sigma^2{}_m \tag{21}$$

As $N$ tends to infinity, equation (20) becomes identical with the solution given by KIMURA (1965) for a population with mean at the optimum, viz.,

$$\frac{\hat{\sigma}^2{}_g}{2\sigma^2{}_m} = \sqrt{\frac{4N\mu}{4NC^*}} = \sqrt{\frac{\mu}{C^*}} \tag{22}$$

## CENTRIPETAL SELECTION IN THE ABSENCE OF MUTATION

In this section we present numerical results for *single locus models* of centripetal selection in populations of effective size $N$, in the *absence* of mutation, for a range of values of the parameter combination $NCg^2$. In each population a series of $2N$ alleles is supposed to be segregating initially, each with frequency $p_i = 1/2N$, and allelic effects $a_i$ are sampled at random from a normal distribution with mean at the optimum (zero) and variance $\frac{1}{2}g^2$. A separate sample of $a_i$ values is chosen for each replicate run. The regimes concerned are listed in Table 1.

The computer simulation procedure involves two steps each generation: (i) a transformation of the vector of allelic frequencies, making use of equation (10), to simulate the effects of centripetal selection; and (ii) random sampling from a multinomial distribution with parameters $2N$; $p'_i$, $i = 1, 2, \ldots n$, where $n$ denotes the number of alleles segregating, and the $p'_i$ denote the transformed allelic frequencies.

The statistics for regimes *without* selection are summarized in Table 2. Over the range of population sizes from $N = 10$–$100$, the mean number of alleles segregating after $N$, $2N$, and $3N$ generations can be seen to depend on the value of $N$, since the initial number of alleles is directly proportional to $N$. However, after $3N$ generations the differences in allele number among the regimes are small. The levels of heterozygosity, additive genetic variance and drift variance are in agreement throughout the table with expectations of $1 - F = [1 - (1/2N)]^t$, $(1 - F)g^2$ and $2Fg^2$, respectively. This provides a check on many aspects of the computer program.

Table 3 sets out corresponding data for those regimes involving centripetal selection in addition to random sampling, the order in the table being that of increasing values of $NCg^2$ from 0.5 to 20.0. Three important conclusions can be drawn from the data of Tables 2 and 3:

1. Comparisons of regimes with the same value of $NCg^2$, but differing in the individual component parameters, support the view that population behavior is largely determined by the value of $NCg^2$. Of 36 such differences throughout Table 3, only one is statistically significant at the .05 level, and a second comparison is on the borderline.

TABLE 1

*Regimes involved in the study of elimination of allelic variation by centripetal selection in the absence of mutation*

| Designation of regime* | Population size ($N$) | Coefficient of selection ($C$) | Initial genetic variance ($g^2$)† | Value of $NCg^2$ |
|---|---|---|---|---|
| D(10) | 10 | 0.00 | 1.0 | 0.0 |
| D(20) | 20 | 0.00 | 1.0 | 0.0 |
| D(50) | 50 | 0.00 | 1.0 | 0.0 |
| D(100) | 100 | 0.00 | 1.0 | 0.0 |
| DS(10,0.5) | 10 | 0.10 | 0.5 | 0.5 |
| DS(20,1.0)a | 20 | 0.50 | 0.1 | 1.0 |
| DS(20,1.0)b | 20 | 0.05 | 1.0 | 1.0 |
| DS(50,5.0) | 50 | 0.10 | 1.0 | 5.0 |
| DS(100,5.0) | 100 | 0.50 | 0.1 | 5.0 |
| DS(50,10.0) | 50 | 0.20 | 1.0 | 10.0 |
| DS(100,10.0) | 100 | 0.10 | 1.0 | 10.0 |
| DS(100,20.0) | 100 | 0.20 | 1.0 | 20.0 |

* D($N$) denotes a regime of genetic drift alone, with effective population size $N$; DS ($N$,$z$) denotes a regime involving drift and centripetal selection, with population size $N$ and $z = NCg^2$.
† Expressed as a fraction of the initial phenotypic variance.

2. Figure 1 illustrates the response to increases in the value of $NCg^2$ shown by the four observed statistics, viz., mean number of alleles segregating, mean heterozygosity, mean genetic variance within populations, and the variance among means of replicate populations, i.e., drift variance. Observations for regimes with the same value of $NCg^2$ have been averaged in preparing the figure from Tables 2 and 3. Only the data following $2N$ generations of selection are illustrated, but the comparisons are similar at any point throughout the $3N$ generations of selection which have been studied.

At values of $NCg^2 > 1.0$, centripetal selection can be seen to lead to a reduction of well over 50% in the drift variance after $2N$ generations, and the genetic variance due to segregation within the populations is comparably reduced with $NCg^2 > 5.0$. However, both the number of alleles segregating and the mean level of heterozygosity are scarcely affected by regimes with values of $NCg^2$ up to 20.0. Observations of mean levels of heterozygosity, or of numbers of alleles segregating, are therefore unlikely to be sensitive indicators of the selective forces operating in a finite population.

## TABLE 2

*Observed statistics after* N, 2N, *and* 3N *generations of genetic sampling in the absence of selection or mutation, based on 200 replicates: each population started with* 2N *alleles at frequencies* $p_i = 1/2N$

| No. of generations | Regime* | No. of alleles segregating in replicate populations: Frequency distribution 1 | 2 | 3 | 4 | 5 | 6 | 7+ | Mean ($n_a$) | Heterozygosity (H) | Genetic Variance ($\sigma^2_g$) | Drift Variance† ($\bar{x}^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N | D(10) | 1 | 21 | 63 | 79 | 27 | 8 | 1 | 3.69 ± .07 | 0.59 ± .01 | 0.67 ± .04 | 0.76 ± .07 |
| | D(20) | 1 | 18 | 49 | 78 | 40 | 14 | . | 3.90 ± .07 | 0.59 ± .02 | 0.56 ± .05 | 0.81 ± .07 |
| | D(50) | . | 9 | 49 | 63 | 50 | 27 | 2 | 4.22 ± .08 | 0.61 ± .01 | 0.55 ± .04 | 0.70 ± .05 |
| | D(100) | . | 6 | 44 | 71 | 51 | 23 | 5 | 4.28 ± .08 | 0.62 ± .01 | 0.60 ± .05 | 0.87 ± .08 |
| 2N | D(10) | 49 | 102 | 44 | 5 | . | . | . | 2.02 ± .05 | 0.32 ± .02 | 0.37 ± .04 | 1.31 ± .15 |
| | D(20) | 31 | 102 | 59 | 8 | . | . | . | 2.22 ± .05 | 0.35 ± .02 | 0.35 ± .03 | 1.27 ± .14 |
| | D(50) | 27 | 106 | 56 | 11 | . | . | . | 2.25 ± .05 | 0.36 ± .01 | 0.31 ± .04 | 1.30 ± .18 |
| | D(100) | 24 | 96 | 62 | 15 | 3 | . | . | 2.38 ± .06 | 0.38 ± .01 | 0.35 ± .05 | 1.41 ± .19 |
| 3N | D(10) | 94 | 93 | 13 | . | . | . | . | 1.60 ± .04 | 0.19 ± .01 | 0.22 ± .03 | 1.44 ± .16 |
| | D(20) | 90 | 86 | 24 | . | . | . | . | 1.67 ± .05 | 0.21 ± .02 | 0.21 ± .03 | 1.58 ± .17 |
| | D(50) | 82 | 100 | 17 | 1 | . | . | . | 1.68 ± .05 | 0.20 ± .02 | 0.17 ± .03 | 1.48 ± .19 |
| | D(100) | 71 | 105 | 24 | . | . | . | . | 1.76 ± .05 | 0.23 ± .02 | 0.22 ± .04 | 1.63 ± .18 |

* Regimes designated as in Table 1.

† Calculated as the variance among means of replicate populations: variance expressed as proportions of the initial genetic variance.

3. The frequency distributions of number of alleles segregating in replicate populations under centripetal selection (Table 3) are not detectably different in this study from those under genetic sampling alone (Table 2). It is well known that stable equilibria are possible under centripetal selection in large populations

with a single locus segregating, the two-allele configuration being stable whenever the heterozygote is closer in genotypic value to the optimum than either of the two homozygous genotypes (ROBERTSON 1956). We can infer from the results presented here that this phenomenon has not led to any appreciable

TABLE 3

*Observed statistics after* N, 2N *and* 3N *generations of centripetal selection and genetic sampling without mutation, based on 200 replicates: each population started with* 2N *alleles at frequencies* $p_i = 1/2N$

| No. of generations | Regime* | No. of alleles segregating in replicate populations — Frequency distribution | | | | | | | | Heterozygosity (H) | Genetic Variance ($\sigma_g^2$) | Drift Variance† ($\bar{x}^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7+ | Mean ($n_a$) | | | |
| N | DS(10,0.5) | 2 | 16 | 60 | 78 | 34 | 7 | 3 | 3.80 ± .07 | 0.60 ± .01 | 0.55 ± .04 | 0.57 ± .04 |
| | DS(20,1.0)a | 1 | 15 | 54 | 69 | 47 | 12 | 2 | 3.95 ± .08 | 0.60 ± .01 | 0.45 ± .03 | 0.42 ± .05 |
| | DS(20,1.0)b | . | 13 | 50 | 74 | 47 | 14 | 2 | 4.02 ± .07 | 0.59 ± .01 | 0.47 ± .03 | 0.38 ± .05 |
| | DS(50,5.0) | . | 13 | 45 | 70 | 45 | 21 | 6 | 4.18 ± .08 | 0.58 ± .02 | 0.24 ± .03 | 0.16 ± .03 |
| | DS(100,5.0) | 4 | 22 | 32 | 61 | 55 | 23 | 3 | 4.11 ± .09 | 0.58 ± .02 | 0.21 ± .03 | 0.13 ± .03 |
| | DS(50,10.0) | 1 | 16 | 62 | 74 | 34 | 12 | 1 | 3.82 ± .07 | 0.58 ± .01 | 0.15 ± .01 | 0.07 ± .01 |
| | DS(100,10.0) | 2 | 14 | 45 | 67 | 48 | 20 | 4 | 4.10 ± .08 | 0.57 ± .01 | 0.15 ± .01 | 0.07 ± .01 |
| | DS(100,20.0) | 1 | 14 | 65 | 69 | 35 | 15 | 1 | 3.86 ± .08 | 0.56 ± .01 | 0.07 ± .01 | 0.04 ± .01 |
| 2N | DS(10,0.5) | 37 | 97 | 57 | 8 | 1 | . | . | 2.20 ± .06 | 0.35 ± .01 | 0.33 ± .03 | 0.80 ± .05 |
| | DS(20,1.0)a | 30 | 90 | 72 | 8 | . | . | . | 2.29 ± .05 | 0.37 ± .01 | 0.25 ± .02 | 0.52 ± .06 |
| | DS(20,1.0)b | 26 | 89 | 75 | 10 | . | . | . | 2.35 ± .05 | 0.38 ± .01 | 0.26 ± .03 | 0.48 ± .06 |
| | DS(50,5.0) | 31 | 101 | 55 | 12 | 1 | | | 2.26 ± .06 | 0.35 ± .01 | 0.13 ± .02 | 0.17 ± .02 |
| | DS(100,5.0) | 32 | 92 | 61 | 14 | 1 | . | . | 2.30 ± .06 | 0.35 ± .01 | 0.11 ± .01 | 0.15 ± .02 |
| | DS(50,10.0) | 36 | 100 | 50 | 14 | . | . | . | 2.21 ± .06 | 0.33 ± .01 | 0.08 ± .01 | 0.08 ± .01 |
| | DS(100,10.0) | 34 | 91 | 65 | 10 | . | . | . | 2.26 ± .06 | 0.34 ± .02 | 0.07 ± .01 | 0.10 ± .01 |
| | DS(100,20.0) | 38 | 89 | 65 | 8 | . | . | . | 2.22 ± .06 | 0.34 ± .01 | 0.06 ± .01 | 0.04 ± .01 |
| 3N | DS(10,0.5) | 92 | 91 | 15 | 2 | . | . | . | 1.64 ± .05 | 0.21 ± .01 | 0.20 ± .03 | 1.01 ± .06 |
| | DS(20,1.0)a | 76 | 111 | 13 | . | . | . | . | 1.68 ± .04 | 0.22 ± .01 | 0.15 ± .02 | 0.53 ± .07 |
| | DS(20,1.0)b | 76 | 107 | 17 | . | . | . | . | 1.70 ± .04 | 0.22 ± .02 | 0.14 ± .02 | 0.58 ± .07 |
| | DS(50,5.0) | 82 | 102 | 16 | . | . | . | . | 1.67 ± .04 | 0.23 ± .02 | 0.09 ± .01 | 0.18 ± .02 |
| | DS(100,5.0) | 90 | 88 | 21 | 1 | . | . | . | 1.66 ± .05 | 0.21 ± .01 | 0.06 ± .01 | 0.18 ± .02 |
| | DS(50,10.0) | 88 | 95 | 17 | . | . | . | . | 1.64 ± .05 | 0.21 ± .02 | 0.05 ± .01 | 0.10 ± .01 |
| | DS(100,10.0) | 93 | 89 | 18 | . | . | . | . | 1.62 ± .05 | 0.19 ± .02 | 0.04 ± .01 | 0.12 ± .02 |
| | DS(100,20.0) | 83 | 104 | 13 | . | . | . | . | 1.65 ± .04 | 0.20 ± .01 | 0.02 ± .01 | 0.04 ± .01 |

*, † See footnotes to Table 2.

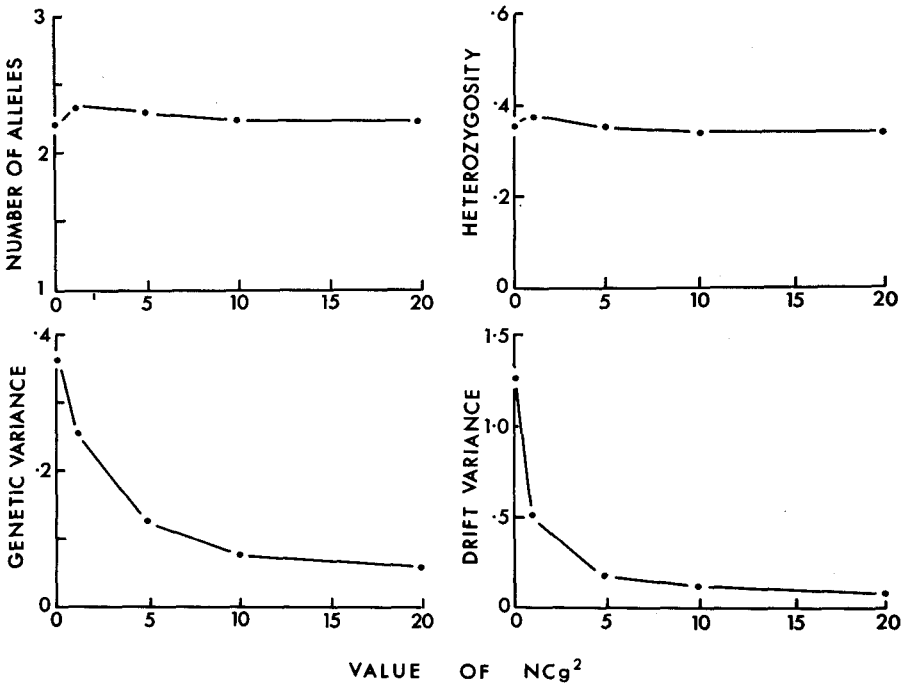


FIGURE 1.—The sensitivity of four population parameters to differences in the value of $NCg^2$, based on the data of Tables 2 and 3 for $2N$ generations of selection in the absence of mutation. $N$ denotes breeding population size, $C$ is the coefficient of centripetal selection, and $g^2$ the initial contribution of the locus to the total phenotypic variance.

retention of allelic variation with $NCg^2 < 20$, given initially a set of multiple alleles with effects normally distributed about the optimum.

It is of some interest to check on the accuracy of equation (17) in predicting changes in the level of genetic variance due to alleles with initially normally distributed effects. Table 4 shows the predicted and observed values after $N$ gen-

TABLE 4

*Comparison of observed levels of genetic variance within populations with those predicted by equation (17). The data refer to a period of* N *generations of centripetal selection in the absence of mutation, there being initially* 2N *alleles with effects normally distributed about the optimum*

| Value of $NCg^2$ | Genetic variance observed* | Predicted variance* |
|---|---|---|
| 0.5 | 0.55 ± .04 | 0.51 |
| 1.0 | 0.46 ± .02 | 0.44 |
| 5.0 | 0.22 ± .02 | 0.20 |
| 10.0 | 0.15 ± .01 | 0.12 |
| 20.0 | 0.07 ± .01 | 0.07 |

* Expressed as a fraction of the initial genetic variance, $g^2$.

erations of selection. For the range of values of $NCg^2$ tested, the predicted values somewhat underestimate the residual genetic variability, since equation (17) ignores changes with time in the measure of kurtosis of the distribution of allelic effects. After $2N$ and $3N$ generations of selection in the absence of mutation, the predicted values are quite appreciably less than those observed.

## POPULATIONS IN MUTATION-SELECTION EQUILIBRIUM

Our objective in this final section is the simulation of essentially equilibrium populations under centripetal selection, genetic sampling and mutation, dealing only with single locus models. The expected number of mutant alleles per generation is $2N\mu$, and the actual number has been determined each generation as a random Poisson variate with the same expectation. The mutational events have been allocated at random to the existing alleles, with probabilities equal to the allelic frequencies after selection and genetic sampling. The effect of a new mutant has been determined as $a_i + \delta a_i$, where $a_i$ is that of the parent allele concerned, and $\delta a_i$ is a random normal value with zero mean and variance $\sigma^2{}_m$. Selection and random sampling have been simulated in the manner indicated in the previous section. The program has been checked to establish that new mutants *are* introduced into the population according to Poisson expectations, and that the mean probabilities of extinction for neutral mutations ($C = 0.0$) are those predicted by existing theory (Table 5).

The regimes of Table 6 have been chosen to represent a series of populations with values of $N$, the breeding population size, ranging from 100 to 1,000, with

TABLE 5

*Observed and expected numbers of new mutants per generation, and the probabilities of extinction observed for neutral mutations:* $N = 500$, $2N\mu = 1$

| No. of mutants | Frequency Observed | Frequency Poisson | Generation | Probability of extinction Observed* | Probability of extinction Theoretical* |
|---|---|---|---|---|---|
| 0 | 366 | 368 | 1 | 0.353 | 0.368 |
| 1 | 339 | 368 | 3 | 0.601 | 0.626 |
| 2 | 203 | 184 | 7 | 0.784 | 0.790 |
| 3 | 73 | 61 | 15 | 0.894 | 0.887 |
| 4 | 14 | 15 | 31 | 0.947 | 0.941 |
| $\geq 5$ | 5 | 4 | 63 | 0.975 | 0.970 |
| | 1,000 | 1,000 | | | |

* Observed probabilities based on 1000 observations; theoretical probabilities following FISHER (1922).

the parameter combinations $NC^*$ and $N\mu$ held constant. All populations began with a single allele of optimal or suboptimal effect, and a period of at least $10N$ generations was allowed for mutational variability to accumulate under centripetal selection before the survey period began. The "equilibrium" populations have been characterized in terms of the four parameters studied in the previous section, viz., the number of alleles segregating, $n_a$; the level of heterozygosity, $H$; the within-population variance, $\sigma^2_g$; and the drift variance, $\bar{x}^2$, estimated as the mean squared deviation of the population mean from the optimum. The total genetic load has also been estimated by the formula given in the footnote to Table 6.

The average number of alleles segregating in these populations is virtually identical in each case with the number predicted by WRIGHT (1966) for multiple alleles *without* selection, i.e.,

$$n_a = \left(1 + \frac{4N\mu}{4Nv}\right)\left[1 - \frac{\Gamma(4N\mu + 4Nv)}{\Gamma(4N\mu)\,\Gamma(1 + 4Nv)}\left(\frac{1}{2N}\right)^{4Nv}\right] \qquad (23)$$

where $4Nv$ has been taken to be 0.0001 for purposes of calculation. In the present instance, with $4N\mu = 1.0$, a simpler prediction equation is

$$n_a = \log_e (2N) \qquad (\text{for } 4N\mu = 1.0)$$

as given by EWENS (1964) and KIMURA (1968). The predicted values of $n_a$ for populations of the same size as those in Table 6 are 5.30, 6.21, 6.91, and 7.60, respectively. We must therefore conclude that the mean number of alleles segregating in these equilibrium populations is little affected by the centripetal selection imposed (Figure 1).

The observed mean values of $H$, $\sigma^2_g$ and $\bar{x}^2$ in Table 6 do not change significantly as population size is increased with $NC^*$ and $N\mu$ constant, the overall mean values being $H = 0.419 \pm .015$, $\sigma^2_g = 0.287 \pm .043$ and $\bar{x}^2 = 0.195 \pm .022$. The corresponding predicted values of $H$ and $\sigma^2_g$ for isoallelic variation in the *absence* of selection are 0.500 (CROW and KIMURA 1964) and 1.000 (equation (21)). Both parameters are therefore significantly reduced by centripetal selection, but the reduction in variance is appreciably greater than that in heterozygosity. These simulated "equilibrium" populations therefore reinforce the conclusions of the previous section as regards the differential sensitivity of $n_a$, $H$, and $\sigma^2_g$ to centripetal selection.

The statistics of Table 7 describe the impact of centripetal selection on the simulated populations in more detail. The single locus model we are exploring inevitably gives rise to pairs of alleles for which the heterozygote is superior in fitness to the two homozygous genotypes concerned (ROBERTSON 1956). Our aim is to determine the importance of this phenomenon in maintaining segregation over appreciable periods of time: the statistics for heterotic polymorphisms in Table 7 have therefore been restricted arbitrarily to those which persisted in the population for at least $N$ generations.

The mean selective advantage of the heterozygote in these polymorphisms can be seen to be inversely related to $N$, as was to be expected (KIMURA 1968).

The average duration of a heterotic polymorphism is approximately 2–3$N$ generations, and the overall probability of the population being in a state of heterotic polymorphism at any given time is roughly 0.16 for these regimes. The latter two statistics appear to be independent of population size for fixed values of $NC^*$

TABLE 6

*Statistics observed at equilibrium in populations of size* N. *The mutation rate is denoted by* $\mu$, *and the coefficient of centripetal selection by* C. *Four replicate populations were simulated for each regime, and standard errors have been calculated on the basis of the observed between-replicate variation. Each population was initially homozygous for a single allele*

| Regime[1] | | | Survey Period | Equilibrium statistics[2] | | | | |
|---|---|---|---|---|---|---|---|---|
| N | 4NC* | 4Nμ | | $n_a$ | H | $\sigma^2_g$ | $\bar{x}^2$ | L (x $10^4$) |
| 100 | 10.0 | 1.00 | 30N - 100N | 5.18 ± .18 | 0.380 ± .029 | 0.237 ± .033 | 0.156 ± .027 | 49.1 ± 8.0 |
| 250 | 10.0 | 1.00 | 10N - 40N | 6.17 ± .24 | 0.404 ± .034 | 0.287 ± .029 | 0.244 ± .029 | 26.5 ± 2.0 |
| 500 | 10.0 | 1.00 | 10N - 30N | 7.03 ± .10 | 0.453 ± .033 | 0.420 ± .158 | 0.250 ± .066 | 16.8 ± 4.3 |
| 1000 | 10.0 | 1.00 | 10N - 20N | 7.62 ± .17 | 0.438 ± .019 | 0.204 ± .052 | 0.130 ± .044 | 4.2 ± 0.9 |

[1] The parameter $C^*$ denotes $C\sigma^2_m/\sigma^2_p$, which is equal to $\sigma^2_m(\sigma^2_e + \sigma^2_f)^{-1}$ for a single locus model; $\sigma^2_m$ has been set equal to unity throughout, without any loss of generality, being equivalent only to a definition of the scale of measurement.

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

TABLE 7

*Incidence of heterotic polymorphisms and observed rates of amino acid replacement in the simulated populations of Table 6*

| Regime[1] | Heterotic Polymorphisms[2] | | | Rate of amino-acid replacement/N gens[3] | |
|---|---|---|---|---|---|
| N | Mean duration (N) | Mean heterozygote superiority (%) | Probability | Maximum | Minimum |
| 100 | 2.59 ± 0.55 | 0.92 ± .24 | 0.11 ± .04 | 0.08 ± .02 | 0.07 ± .01 |
| 250 | 2.87 ± 0.48 | 0.36 ± .08 | 0.20 ± .09 | 0.12 ± .03 | 0.08 ± .01 |
| 500 | 2.05 ± 0.25 | 0.42 ± .12 | 0.21 ± .07 | 0.16 ± .03 | 0.14 ± .01 |
| 1000 | 3.95 ± 2.75 | 0.03 ± .03 | 0.12 ± .10 | 0.13 ± .04 | 0.08 ± .03 |
| Means | 2.86 ± .71 | - | 0.16 ± .04 | 0.12 ± .02 | 0.09 ± .01 |

[1] For all regimes $4NC^* = 10.0, 4N\mu = 1.00$ (cf. Table 6).
[2] These data refer to polymorphisms maintained by selection for the heterozygote, and persisting in the population for a period of at least $N$ generations.
[3] For the purposes of this calculation, all mutations specified by $\mu$ have been assumed to involve a single amino acid change. The figures give the observed average number of amino acid replacements in the predominant allele over a period of $N$ generations. Maximum and minimum estimates are given because of incomplete replacement at the termination of the computer runs.

and $N\mu$, though the available data are as yet too few to test the statement critically.

On an evolutionary time scale, the parameter of greatest interest is the *rate of amino acid replacement* in the protein corresponding to the predominant

allele, expected to be $N\mu$ over a period of $N$ generations for isoallelic variation in the absence of selection (KIMURA 1969). The parameter $\mu$ here refers to the rate of spontaneous mutation leading to single amino acid changes in the protein concerned. The rate of amino acid replacement in the simulated populations has been calculated from the number of mutational changes differentiating the final predominant allele from that in the initial population. The observed values appear from the limited data of Table 7 to be independent of $N$ for given values of $NC^*$ and $N\mu$, averaging less than 0.12 replacements per $N$ generations. The value of $N\mu$ for these regimes is 0.25. In populations of breeding size $N = 5000$, therefore, with a mutation rate of $5 \times 10^{-5}$, we would expect an intensity of centripetal selection of $C^* = 5 \times 10^{-4}$ to be sufficient to at least halve the rate of amino acid replacement in the course of evolution.

*Allelic variation in natural populations:* Surveys of naturally occurring electrophoretic variability in *Drosophila pseudoobscura* consistently show approximately one-third of loci to be polymorphic, with mean levels of heterozygosity close to 0.12 (LEWONTIN and HUBBY 1966; PRAKASH, LEWONTIN and HUBBY 1968). The data of O'BRIEN and MACINTYRE (1969) indicate somewhat higher values in *D. melanogaster*, and appreciably less variability in *D. simulans*. HARRIS (1969) has summarized the results of human population surveys for electrophoretic variants in 20 randomly chosen enzymes, showing one-third of the loci to be polymorphic in European and African populations, with a mean level of heterozygosity of 0.072. For blood group loci in the English population, the mean frequency of heterozygosis has been shown by LEWONTIN (1967) to be close to 0.15.

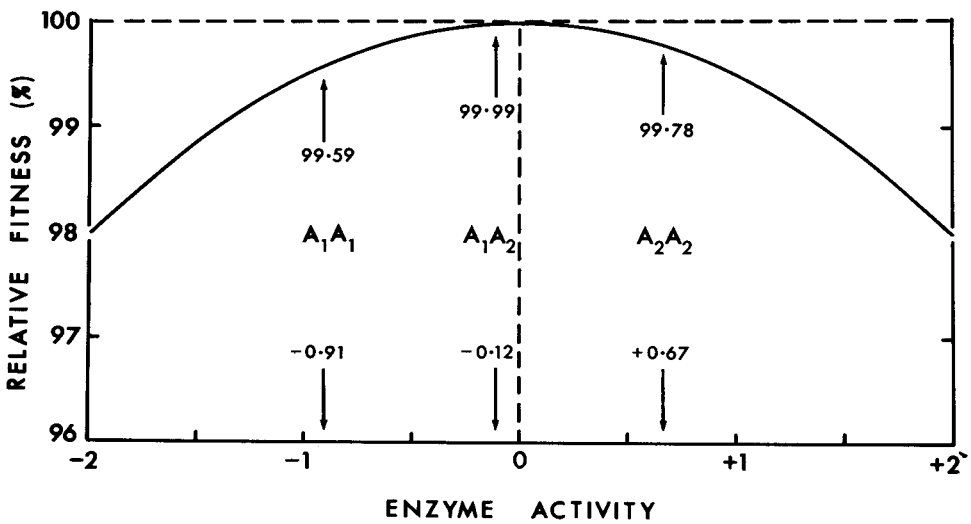


FIGURE 2.—Selective advantage of the heterozygote for a pair of alleles with additive effects on the scale of enzyme activity. Numerical values refer to a polymorphism arising in the regime of Table 8, with a mean heterozygote superiority in fitness of 0.31%. Units of enzyme activity are defined by reference to the magnitude of the mutational variance, $\sigma^2_m = 1.0$, and expressed as deviations from the optimum.

These observed levels of polymorphism and heterozygosity can readily be duplicated by simulation based on the single locus model which has been studied in this paper. The model in its simplest form assumes additive allelic effects, a heterozygote having a level of activity equal to the mean of the activities of the corresponding homozygotes (Figure 2). The rate of decline in fitness with deviation from optimal activity is specified by the magnitude of the parameter $C^*$ (Table 6), and the rate of mutation to novel alleles is denoted by $\mu$. The data of Table 8 indicate that values of $NC^* = 5.0$, and $N\mu = 0.05$, lead to an estimated mean level of heterozygosity of $0.122 \pm .016$, and an overall probability of polymorphism of $0.300 \pm .026$ at equilibrium.

These two statistics are accurately estimated, and come very close to the observed values in populations of Drosophila and man. The expected number of alleles segregating in a population of size 500 is 2.32 based on equation (23), in excellent agreement with the value of $2.16 \pm .09$ in Table 8. The level of heterozygosity observed is roughly 75% of that expected for completely neutral isoallelic variants.

The stability of the polymorphisms arising by mutation under this regime can also be gauged from the statistics in Table 8. Approximately 50% of the polymorphisms detected in a contemporary population would be expected to be heterotic on the basis of this model, with a duration of roughly 2.5–5.0 $N$ generations in the life of the population. With a breeding population size of 500, the mean superiority in fitness of the heterozygotes in these heterotic polymorphisms is predicted to be $0.25 \pm .08\%$.

## DISCUSSION

The behavior of simulated populations under selection for a fixed optimal level of gene activity, with continual spontaneous mutation to novel alleles, has been interpreted in this study largely in terms of the parameters $N\mu$ and $NC^*$, where $C^* = C\sigma^2_m/\sigma^2_p$. These arise in the derivation of an algebraic expression for the equilibrium genetic variance in a population of breeding size $N$, with mutation rate $\mu$, coefficient of centripetal selection $C$ and mutational variance $\sigma^2_m$ (equation (20)). Extensive tests of the effects of centripetal selection in finite populations in the *absence* of mutation have shown that the parameter $NC$ accounts for observed changes in both genetic variance within populations ($\sigma^2_g$), and the variance of replicate population means ($\bar{x}^2$). Variation in $N$ for constant $NC$ has been shown to be without detectable effect on these statistics (Table 3). The more limited data for populations in equilibrium under a regime of centripetal selection and mutation, where $N$ has been varied with $N\mu$ and $NC^*$ held constant, point to the same conclusion (Table 6). Both sets of data clearly show the relative insensitivity to centripetal selection of the mean number of alleles segregating, $n_a$, and to a lesser extent mean heterozygosity, $H$, by comparison with the two variance parameters $\sigma^2_g$ and $\bar{x}^2$ (Figure 1).

Equation (23) can therefore be used to predict the mean number of alleles segregating in equilibrium populations at intensities of centripetal selection such as those involved in the simulation experiments reported in this paper. The data

TABLE 8

*Equilibrium statistics for simulated populations with* $N = 500$, $C^* = 0.01$, $\mu = 10^{-4}$. *The figures are means based on four replicate populations, each derived from a single allele, averaged over a period of 50N generations. Standard errors are based on the observed variation between replicates*

| Population variability | | | Polymorphisms[1] | Probability | Mean duration ($\bar{N}$) |
|---|---|---|---|---|---|
| Number of alleles | ($\bar{n}_a$) | 2.16 ± .09 | Heterotic | 0.146 ± .052 | 3.74 ± 1.21 |
| Heterozygosity | ($\bar{H}$) | 0.122 ± .016 | Heterotic or transient | 0.173 ± .053 | 3.50 ± 0.98 |
| Genetic variance | ($\sigma_g^2$) | 0.048 ± .017 | Heterotic, transient or chance | 0.300 ± .026 | 1.05 ± 0.25 |
| Drift variance | ($\bar{\chi}^2$) | 0.133 ± .072 | | | |

[1] A polymorphism has been defined as the occurrence of a variant at a frequency of 0.05 or greater, which persists in the population for at least $0.2N$ generations. Heterotic polymorphisms are those maintained by selection for the heterozygote for at least $N$ generations. Transient refers to non-heterotic polymorphisms involving the substitution of one allele for another in the population.

of Tables 6 and 7 indicate that the formula predicting the mean level of heterozygosity in the case of neutral isoalleles, viz., $H = 4N\mu/(1 + 4N\mu)$, provides a useful upper limit to the level expected under centripetal selection. The comparisons of Table 9 also suggest that the solution of equation (20) can be treated as

TABLE 9

*Comparisons of the equilibrium genetic variance ($\sigma_g^2$) observed in simulated populations, and predictions given by equations (20), (21), and (22)*

| Regime | | | Observed within population genetic variance ($\sigma_g^2$) [1] | Calculated within-population genetic variance [2] | | |
|---|---|---|---|---|---|---|
| N | $4NC^*$ | $4N\mu$ | | Equation 20 | Equation 21 | Equation 22 |
| 100 | 10 | 1.0 | 0.237 ± .033 | 0.464 | 1.000 | 0.632 |
| 250 | 10 | 1.0 | 0.287 ± .029 | 0.464 | 1.000 | 0.632 |
| 500 | 10 | 1.0 | 0.420 ± .158 | 0.464 | 1.000 | 0.632 |
| 1000 | 10 | 1.0 | 0.204 ± .052 | 0.464 | 1.000 | 0.632 |
| 500 | 20 | 0.2 | 0.048 ± .017 | 0.124 | 0.200 | 0.200 |

[1] Data of Tables 6 and 8.
[2] Equation (21) predicts the genetic variance in a finite population at equilibrium in the *absence* of selection : equation (22) predicts that in an *infinite* population with the values of $C^*$ and $\mu$ concerned.

an upper limit for the average genetic variance in an equilibrium population under centripetal selection and mutation, being considerably more useful than either the prediction for an infinite population (equation (22)), or that for neutral genetic variation (equation (21)). This enables us to calculate a probable upper

limit to the *inbred load*, $L_I$, in such an equilibrium population, expected to be

$$L_I = \tfrac{1}{2}C^* \sigma^2_g \quad (\text{for } \sigma^2_m = 1.0) \tag{24}$$

It should be noted that equations (20) and (24) imply that the inbred load in a small population at equilibrium will be less than that in a large population with the same value of $C^*$ and $\mu$. Preliminary simulation results (LATTER, unpublished) bear out this prediction. However, the *total load*, $L = \tfrac{1}{2}C^* (\bar{x}^2 + \sigma^2_g)$, is *greater* in the small populations than in the large, due to the increased drift in the population mean away from the optimal level of activity. KIMURA, MARUYAMA and CROW (1963) have shown the mutational load in small populations to be usually greater than that in a large population, using a conventional two-allele model with forward and back mutation rates.

It has been demonstrated in the final section of this paper that equilibrium computer populations with $N = 500$, $NC^* = 5$ and $N\mu = 0.05$ come remarkably close to simulating the pattern of allelic variation in natural populations of man and Drosophila, recently discovered by electrophoretic techniques. The most extensive data are those of PRAKASH, LEWONTIN and HUBBY (1969) for *D. pseudoobscura*, showing an average level of heterozygosity of 12% in North American populations of the species, with an average of 42% of loci showing polymorphism. If the two loci associated with inversions in the third chromosome are excluded, the figures become 11% and 36%, respectively (cf. Table 8). We may interpret the single locus model of centripetal selection used in this paper in terms of natural selection for an optimal level of the catalytic activity of a given enzyme, with spontaneous mutation to alleles of above and below optimal activity. Such a model is almost certainly too simple, and variations on the same theme remain to be explored, particularly those involving (i) low levels of migration between neighboring populations; and (ii) selection for optima which change from generation to generation in a random or cyclic fashion.

There are two properties of the mutation-centripetal selection model which make it particularly relevant to the survey data on electrophoretic variants which are now being collected in man, Drosophila and mice. On the one hand, we have seen that *heterotic* polymorphisms are found in the populations of Table 8 with a probability of the order of 0.15, accounting for roughly half of the polymorphisms observed. These heterotic polymorphisms are maintained over something like 2.5–5.0 $N$ generations on the average, by a mean heterozygote advantage of approximately 0.25%. Separated populations may therefore be expected to maintain a stable pattern of allelic frequencies at many loci over considerable periods of time, and a small degree of migration between such populations would reinforce this tendency to stability. More accurate estimates of these parameters must be provided by future simulation experiments, together with a more thorough exploration of the process of population differentiation.

The second important feature of the model is that it appears to provide an answer to the dilemma posed by SVED, REED and BODMER (1967), concerning the expected drop in fitness on inbreeding in a natural population. They point out that if large numbers of loci are maintained as heterotic polymorphisms, the decline in fitness with inbreeding should be extremely rapid. With a multiallelic

model involving centripetal selection and mutation, we have seen that the regime of Table 8 leads to a probability of approximately one-third that a given locus will be found to be polymorphic at any point in time. Of these, roughly one-half are expected to be heterotic, and the mean inbred load per locus can be estimated from equation (24) to be $L_l = 2.40 \times 10^{-4} \pm 0.85 \times 10^{-4}$. If we take the rate of inbreeding depression in Drosophila to be $(-2.88 \pm 0.16)$ $F$, as calculated from Table 3 of LATTER and ROBERTSON (1962) for the Kaduna cage population of *D. melanogaster*, the number of loci concerned can be estimated to be 12,000 ± 4,300. If the estimate if inbreeding depression is based only on the fitness of surviving inbred lines in LATTER and ROBERTSON's experiment, the figure becomes $(-2.30 \pm 0.13)$ $F$, and the estimated number of loci is 9,600 ± 3,400.

The studies of O'BRIEN and MACINTYRE (1969) suggest that the Kaduna cage population is no less polymorphic than natural populations of *D. melanogaster*, but the average level of heterozygosity in this species may be higher than that in *D. pseudoobscura*. The number of loci suggested by our simple calculation is therefore likely to be an overestimate, and is as yet imprecise. The figure nevertheless appears to be of the right order, as judged by current estimates of the number of loci in Drosophila (LEWONTIN and HUBBY 1966), and indicates that further exploration of the model is warranted.

## SUMMARY

The maintenance of isoallelic variation under centripetal selection in finite populations has been studied by means of computer simulation. A single locus model of natural selection for an optimal level of gene activity has been used, with continual mutation to alleles of above or below optimal activity, not previously represented in the population.——The mean number of alleles segregating at the locus, and the mean level of heterozygosity, have been shown to be far less sensitive to centripetal selection than parameters measuring genetic variability between and within populations. The variance parameters in equilibrium populations have been shown to depend primarily on the values of the parameter combinations $NC$ and $N\mu$, where $N$ denotes breeding population size, $C$ is the coefficient of centripetal selection defined in this paper, and $\mu$ is the rate of mutation to novel alleles. A prediction equation involving $NC$ and $N\mu$ has been derived which gives an upper limit to the expected genetic variance within populations, and consequently the inbred load involved.——Populations in mutation-selection equilibrium have been analyzed to determine the mean levels of heterozygosity, and the duration and frequency of heterotic polymorphisms. The computer results have been compared with experimental data on the frequencies of electrophoretic enzyme variants in natural populations of *Drosophila pseudoobscura* and man, and with an estimate of the inbred load measured under competitive conditions in *D. melanogaster*.

## LITERATURE CITED

ALLEN, R. and A. S. FRASER, 1968 Simulation of genetic systems. XI. Normalizing selection. Theoret. Appl. Genet. **38**: 223–225.

EWENS, W. J., 1964 The maintenance of alleles by mutation. Genetics **50**: 891–898.

FISHER, R. A., 1922 On the dominance ratio. Proc. Roy. Soc. Edinburgh B **42**: 321–341. ——, 1930 *The Genetical Theory of Natural Selection.* Oxford University Press, Oxford.

HALDANE, J. B. S., 1954 The measurement of natural selection. Proc. 9th Intern. Congr. Genet. **1**: 480–487.

HARRIS, H., 1969 Enzyme and protein polymorphism in human populations. Brit. Med. Bull. **25**: 5–13.

JAIN, S. K. and R. W. ALLARD, 1965 The nature and stability of equilibria under optimizing selection. Proc. Natl. Acad. Sci. U.S. **54**: 1436–1443.

KIMURA, M., 1955 Random genetic drift in multi-allelic locus. Evolution **9**: 419–435. ——, 1965 A stochastic model concerning the maintenance of genetic variability in quantitative characters. Proc. Natl. Acad. Sci. U.S. **54**: 731–736. ——, 1968 Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet. Res. **11**: 247–269. ——, 1969 The rate of molecular evolution considered from the standpoint of population genetics. Proc. Natl. 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.

KIMURA, M., T. MARUYAMA and J. F. CROW, 1963 The mutation load in small populations. Genetics **48**: 1303–1312.

KOJIMA, K., 1959 Stable equilibria for the optimum model. Proc. Natl. Acad. Sci. U.S. **45**: 989–993.

LATTER, B. D. H., 1960 Natural selection for an intermediate optimum. Australian J. Biol. Sci. **13**: 30–35.

LATTER, B. D. H. and C. E. NOVITSKI, 1969 Selection in finite populations with multiple alleles. I. Limits to directional selection. Genetics **62**: 859–876.

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

LEWONTIN, R. C., 1964 The interaction of selection and linkage. II. Optimum models. Genetics **50**: 757–782. ——, 1967 Population genetics. Ann. Rev. Genet. **1**: 37–70.

LEWONTIN, R. C. and J. L. HUBBY, 1966 A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura.* Genetics **54**: 595–609.

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.

ROBERTSON, A., 1956 The effect of selection against extreme deviants based on deviation or on homozygosis. J. Genet. **54**: 236–248.

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.

SINGH, M. and R. C. LEWONTIN, 1966 Stable equilibria under optimizing selection. Proc. Natl. Acad. Sci. U.S. **56**: 1345–1348.

SVED, J. A., T. E. REED and W. F. BODMER, 1967 The number of balanced polymorphisms that can be maintained in a natural population. Genetics **55**: 469–481.

WRIGHT, S., 1935 The analysis of variance and the correlation between relatives with respect to deviations from an optimum. J. Genet. **30**: 243–256. ——, 1966 Polyallelic random drift in relation to evolution. Proc. Natl. Acad. Sci. U.S. **55**: 1074–1081.