Genet. Res., Camb. (1982), **39**, *pp.* 303–314
With 1 text-figure
Printed in Great Britain

Genetic differentiation of quantitative characters between populations or species

I. Mutation and random genetic drift

By RANAJIT CHAKRABORTY and MASATOSHI NEI

Center for Demographic and Population Genetics, University of Texas at Houston, Texas 77025

(*Received 2 June 1980 and in revised form 23 November 1981*)

SUMMARY

Introducing a new genetic model called the discrete allelic-state model, the evolutionary change of genetic variation of quantitative characters within and between populations is studied under the assumption of no selection. This model allows us to study the effects of mutation and random genetic drift in detail. It is shown that when the allelic effects on phenotype are additive, the rate of approach of the genetic variance within populations to the equilibrium value depends only on the effective population size. It is also shown that the distribution of genotypic value often deviates from normality particularly when the effective population size and the number of loci concerned are small. On the other hand, the interpopulational variance increases linearly with time, if the intrapopulational variance remains constant. Therefore, the ratio of interpopulational variance to intrapopulational variance can be used for testing the hypothesis of neutral evolution of quantitative characters.

The dynamics and maintenance of genetic variability of quantitative characters were first studied by Fisher (1922). Later, Wright (1931, 1937) extended Fisher's work to a great extent, considering the joint effect of mutation, selection, and random genetic drift. In these studies quantitative characters were assumed to be controlled by many loci with two alleles at each locus. In 1965 Kimura introduced a new genetic model in which the existence of multiple alleles with varying phenotypic effects was assumed but no consideration was made about the effect of genetic drift. In recent years there has been a revival of interest in Kimura's model, and a number of authors (e.g. Latter & Novitski, 1969; Latter, 1970; Lande, 1975, 1976; Cavalli-Sforza & Feldman, 1976) extended his model in various directions.

In these studies the main concern was to explain the amount of genetic variability maintained in a population, and little attention was paid to the genetic differentiation among populations or species. The exception is the work by Latter (1970) and Lande (1976), who studied the genetic variation within and between

populations under the joint effect of centripetal (optimum model) selection, mutation, and genetic drift. Using the method of Gaussian processes, Lande derived a simple formula for the interpopulational variance of quantitative characters in transient states. However, the genetic model he used is not very specific, and it is not clear what kind of change is really occurring in the gene pool of the population in his treatment. Furthermore, he made a number of simplifying assumptions, which would not hold in many situations. Latter's model was more realistic than Lande's but he did not give any general solution and used computer simulation to solve some special problems.

Generally speaking, it is very difficult to develop a mathematical model for the evolutionary change of quantitative characters. This is because most quantitative characters are affected by both genetic and environmental factors, and environmental factors alone can cause a linear temporal change which mimics a genetic change. A good example is the increase in human stature in the last 100 years in many industrial countries (Cavalli-Sforza & Bodmer, 1971). The linear temporal change in the cephalic index of Japanese skulls observed in the last 500 years is also apparently due to environmental factors (Suzuki, 1960). If this type of change occurs, it is very difficult to develop a meaningful mathematical model. Another problem is that the genetic change of quantitative characters is often triggered by environmental change, and in practice, it is very difficult to know how often such an environmental change occurred in the past.

In certain characters, however, the effect of environmental factors seems to be relatively simple. For example, skin pigmentation in man is clearly related to adaptation to sunlight, so that if we know the sunlight intensity for any given population we may be able to model the evolutionary change of pigmentation. An even simpler character in man is dermal ridge count, the variation of which does not seem to be directly related to fitness except that of chromosomally aberrant individuals. The heritability of this character has been estimated to be almost 100 % (Holt, 1961). The variation in sternopleural or abdominal bristle number in Drosophila also seems to be as trivial as human dermal ridge count and apparently is not related to fitness except for extreme phenotypes at both ends (Clayton, Morris & Robertson, 1957; Robertson, 1967).

In view of the existence of these simple characters we have initiated a mathematical study of the evolutionary change of quantitative characters. Our model is more specific than Lande's about the production of mutation and gene frequency changes. In the absence of selection the mathematical treatment of our model is relatively simple, and we can make a rigorous theoretical study about the evolutionary change of quantitative characters. The main purpose of this paper is to report the results of our study for this simple case. Strictly speaking, of course, any quantitative character would be subject to some sort of selection, but the following theory would be applicable to any weakly selected character at least temporarily. Furthermore, it can be used for testing the effect of natural selection. In this paper special attention will be paid to the relative magnitude of genetic

variation within and between populations when the populations have been reproductively isolated for a long time. The effect of selection will be studied in the following paper.

THE MODEL

In the genetic model used in early studies it was generally assumed that only one pair of alleles exist at each locus and mutation occurs forward and backward between them (Wright, 1931). This model may be called the *two-allele model*. Kimura (1965) introduced a new model, in which there are an infinite number of allelic states at each locus and the phenotypic effects of these alleles are continuously distributed. The phenotypic effects of mutant alleles from a given allelic state are assumed to be normally distributed around the phenotypic effect of the original allele. Since he considered an infinitely large population, he assumed that the number of alleles existing in a population is effectively infinite. This model was apparently motivated by the fact that at the molecular level the number of *possible* alleles at a locus is extremely large (Kimura & Crow, 1964). We call this model the *continuous allelic-state model*. Lande (1975, 1976) used this model for both infinite and finite populations.

In finite populations, however, the number of *existing* alleles in a population can be very small even if the number of *possible* alleles is practically infinite. Indeed, unless population size and mutation rate are extremely large, a substantial proportion of loci are expected to be monomorphic, and polymorphic loci will have a relatively small number of alleles (Kimura & Ohta, 1971). Statistical analyses of biochemical data for structural genes support this view (Fuerst, Chakraborty & Nei, 1977; Chakraborty, Fuerst & Nei, 1980), though we are not sure whether these structural genes are really concerned with quantitative characters or not. Clearly, we need a more realistic model in which the above property is taken into account. We also note that mutation is a discrete change of gene, so that a discrete distribution of mutational effects is likely to be more realistic than a continuous distribution. In view of this situation we propose the following genetic model, which we call the *discrete allelic-state model*. This model allows us to study the effects of mutation and genetic drift more rigorously than the continuous allelic-state model does.

Consider a character which is controlled by n loci and assume that at each locus there is an infinite number of *possible* allelic states. We assume that the phenotypic effects of the alleles are discrete as given in Figure 1. In this diagram A_i represents an allele occupying state i and having an allelic effect of ia. We assume that all allelic effects are additive with no dominance and no epistasis and that once A_i mutates, it changes to allelic state $i+r$ with the binomial probability

$$\alpha_r = \alpha_{-r} = \binom{2m}{m-r} (\tfrac{1}{2})^{2m} \quad \text{for} \quad 0 \leqslant r \leqslant m \tag{1a}$$

$$\alpha_r = 0 \quad \text{for} \quad r > m. \tag{1b}$$

https://doi.org/10.1017/S0016672300020978 Published online by Cambridge University Press

Note that the case of $r = 0$ does not really represent new mutation, but for the mathematical convenience we call it mutation and denote the total mutation rate by v. In the conventional definition the real mutation rate is $v' = (1-\alpha_0)\, v$. We also note that the mean and variance of the above distribution is 0 and $m/2$, respectively. Therefore, the increment of the variance of allelic effect by mutation

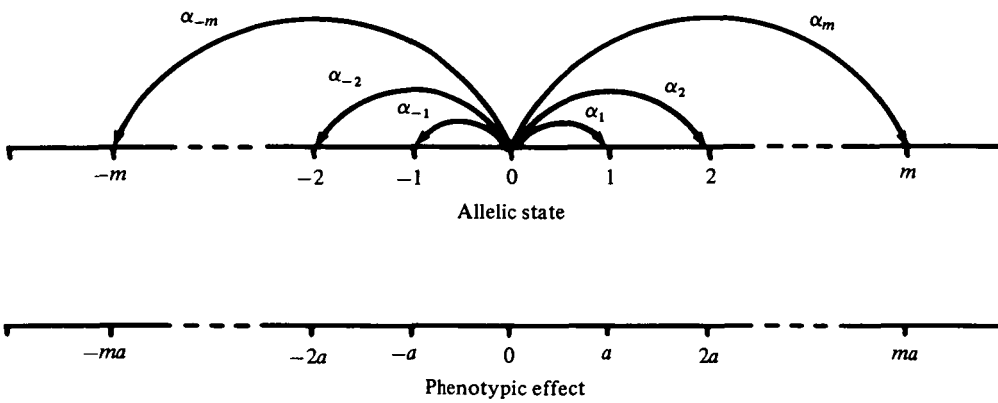

Fig. 1. Discrete allelic-state model used in this paper. In this model allele A_i mutates to A_{i+r} with probability a_r ($= a_{-r}$). Allele A_i has a phenotypic effect of ai.

is $vma^2/2$ per generation. Using this genetic model, we shall now study the genetic variation within and between populations. At the gene level the present model is an extension of Ohta and Kimura's (1973) stepwise mutation model for electrophoretic variation to the case of m steps. In this paper, however, we are concerned with the genetic variation of quantitative characters rather than with electrophoretic variation.

GENETIC VARIATION WITHIN POPULATIONS

We first consider the genetic variation within populations. Let x_l and x_l' be the allelic effects of the genes at the lth locus in an individual. If we neglect the environmental effect, the phenotypic value of this individual is given by

$$X = g + \sum_{l=1}^{n} (x_l + x_l'), \tag{2}$$

where g is the effect of genetic background. In this paper we assume that there is no interaction between genotype and environment, and consider only the genetic variance. If one wants to have the phenotypic variance, he should simply add the environmental variance to the genetic variance. We also assume that all loci have the same mutational pressure and are subject to random genetic drift

in the same fashion and there is no linkage disequilibrium among loci. Then, the genetic variance of the character in the population is

$$V_a(X) = \sum_{l=1}^{n} [V(x_l) + V(x_l')]. \tag{3}$$

Under our assumption the expectations of $V(x_l)$ and $V(x_l')$ over the stochastic process are equal to $V/2$ for all loci, where V is the expected genetic variance for one locus. Therefore,

$$E[V_a(X)] \equiv V(X) = nV. \tag{4}$$

Thus, if we know V, $V(X)$ may be obtained. In the following we denote the effective population size by N and assume that random mating occurs in every generation.

Let p_i be the frequency of allele A_i at a locus in the population. Then, the variance for this locus is $V_a = 2[\Sigma p_i(ia)^2 - (\Sigma p_i ia)^2] = 2a^2[\Sigma p_i i^2 - (\Sigma p_i i)^2]$, in which Σ indicates summation over all alleles. Therefore, if we know the distribution of p_i, we will be able to compute the genetic variance of quantitative characters. However, under the present assumption the distribution of p_i never reaches a stable form but wanders back and forth on the scale of allelic states, as in the case of the stepwise mutation model of electrophoretic mobility (Ohta & Kimura, 1973). Yet, it is possible to determine the variance. A simple way to determine the variance in this case is to consider the difference between the allelic affects of two randomly chosen genes and compute its variance. Let x and x' be the allelic effects of two randomly chosen alleles. Then, the variance of $x-x'$ is V and thus equal to the genetic variance at this locus. The variance of $x-x'$ can be obtained by using the same technique as that for obtaining the expected homozygosity for the stepwise mutation model.

Let us first consider the effect of mutation. We assume that the number of mutations occurring at a locus in a generation follows the Poisson distribution with mean v and the allelic state of a gene in the initial generation is 0. Then, the probability $[Q_i(t)]$ of an allele being in state i in the generation t is given by the following probability generating function (*pgf*) (see Chakraborty & Nei, 1976):

$$Q(s,t) = \sum_{i=-\infty}^{\infty} Q_i(t)\, s^i = \exp\left[vt\left\{\left(\frac{1+2s+s^2}{4s}\right)^m - 1\right\}\right]. \tag{5}$$

This formula is obtained by noting that the *pgf* for the binomial distribution in (1) is $[(1+2s+s^2)/4s]^m$.

Let us now consider the probability that two alleles which had the same allelic state in generation 0 differ by k steps in allelic state in generation t, given that both alleles have survived up to generation t. Since the distribution of mutational effects is symmetric with respect to state 0, the *pgf* for this probability is obviously given by $R(s,t) = Q^2(s,t)$ (Nei & Chakraborty, 1973; Chakraborty & Nei, 1977). In practice, however, only $2N$ gametes are sampled in every generation at the time of reproduction, and thus some types of alleles may be lost from the population. Furthermore, the probability that two randomly chosen alleles differ by k steps

in generation t should be computed by considering the contribution from all generations in the past as well as the probability distribution of allelic states in the initial population. If we follow Wehrhahn (1975) and Chakraborty & Nei (1976, 1977), this probability is given by the following generating function.

$$P(s,t) = \int_0^t \frac{1}{2N} e^{-r/2N} R(s,r)\, dr + P(s,0)\, e^{-t/2N} R(s,t)$$
$$= -\frac{1}{2Na(s)} + \left[P(s,0) + \frac{1}{2Na(s)}\right] e^{a(s)t}, \tag{6}$$

where $a(s) = -\lambda + 2v[(1+2s+s^2)/4s]^m$ and $\lambda = 2v + \frac{1}{2N}$.

At equilibrium, i.e. when $t = \infty$, we have

$$P(s,\infty) = -\frac{1}{2Na(s)} \tag{7}$$

Therefore, the genetic variance at a locus in equilibrium is

$$V_\infty = 2Nmva^2, \tag{8}$$

since $P''(1,\infty) + [P'(1,\infty)] - [P'(1,\infty)]^2 = 2Nmv$. $V(X)$ in (4) is therefore nV_∞. Note that (8) becomes identical with Latter's (1970) equivalent forumla, $V_\infty = 4Nv'ma^2/2$, when $m = \infty$, as expected. When m is small, the equilibrium variance is larger than expected from his formula.

Note that the mean of allelic difference $[(P'(1,t)]$ is always zero. On the other hand, the expected genetic variance, V_t, for a locus in transient states may be obtained from equation (6). It becomes

$$V_t = V_\infty + [V_0 - V_\infty]\, e^{-t/2N}, \tag{9}$$

where V_0 is the genetic variance at time 0. It is obvious that the total genetic variance in generation t can be written as $V_t(X) = V_\infty(X) + [V_0(X) - V_\infty(X)]e^{-t/2N}$.

One important parameter in population genetics is heterozygosity, i.e. the expected proportion of heterozygous individuals under random mating. The expected heterozygosity for the present model may be obtained by $H_t = 1 - P_0(t)$, where $P_0(t)$ is the probability of two randomly chosen genes having the same allelic state and given by the coefficient of s^0 in (6). $P_0(t)$ is a complicated function of mutation rate and population size, but the expected heterozygosity at steady state is given by

$$H = 1 - \frac{1}{1+M} \sum_{r=0}^{\infty} \left(\frac{M}{1+M}\right)^r \binom{2mr}{mr} \Big/ 4^{mr}, \tag{10}$$

where $M = 4Nv$. It is noted that in the case of $m = 1$, the model becomes identical with Ohta and Kimura's stepwise mutation model if we redefine the mutation rate by $v' = (1-\alpha_0)v = v/2$. Indeed, in this case H in (10) becomes $1-(1+8Nv')^{-\frac{1}{2}}$, which is identical with Ohta and Kimura's formula obtained by a different method. When $m \neq 1$, the expected heterozygosity increases with increasing m, as expected (see Table 1).

In the above we have studied the expected variance of quantitative characters. In practice, however, the genetic variance in a population may vary from time to

time by genetic drift even in steady state. It is therefore interesting to know the coefficient of variation of genetic variance. The variance $[V\{V_a(X)\}]$ of genetic variance in an equilibrium population is given by (A5) in the Appendix. Therefore, the coefficient of variation of variance is

$$c.v. = [V\{V_a(X)\}]^{\frac{1}{2}}/V_\infty(X) = \left[\left\{2+\frac{3}{16Nv}\left(3-\frac{1}{m}\right)\right\}\Big/(Nn)\right]^{\frac{1}{2}}. \tag{11}$$

Table 1. *Average heterozygosity (H) and the kurtosis (γ_2) of genotypic value in an equilibrium population when a quantitative character is controlled by n loci*

	$n=1$			$n=5$			$n=10$		
	$m=1$	$m=2$	$m=5$	$m=1$	$m=2$	$m=5$	$m=1$	$m=2$	$m=5$
$4Nv=0{\cdot}05$									
H	0·024	0·030	0·036	0·024	0·030	0·036	0·024	0·030	0·036
γ_2	30·00	37·50	42·00	6·00	7·50	8·40	3·00	3·75	4·20
$4Nv=0{\cdot}10$									
H	0·047	0·058	0·069	0·047	0·058	0·069	0·047	0·058	0·069
γ_2	15·00	18·75	21·00	3·00	3·75	4·20	1·50	1·88	2·10
$4Nv=0{\cdot}50$									
H	0·184	0·222	0·261	0·184	0·222	0·261	0·184	0·222	0·261
γ_2	3·00	3·75	4·20	0·60	0·75	0·84	0·30	0·38	0·42
$4Nv=1{\cdot}0$									
H	0·293	0·347	0·400	0·293	0·347	0·400	0·293	0·347	0·400
γ_2	1·50	1·88	2·10	0·30	0·38	0·42	0·15	0·19	0·21

This indicates that the variation of variance among independent populations is rather small as long as N is substantially large. For example, if $4Nv = 0{\cdot}1$, $N = 1000$, $m = 1$, and $n = 10$, *c.v.* is 0·04.

In theoretical studies of quantitative genetics it is often assumed that the genotypic value is normally distributed. By using the two-allele model with mutation, however, Nei & Imaizumi (1966) have shown that the distribution of genotypic value may deviate substantially from normality unless the population size and the number of loci are large. In the present model the skewness in an equilibrium population is always 0, but the kurtosis may vary with the parameters specified. In a given population the kurtosis of genotypic value is given by

$$\gamma_{2(a)}(X) = \frac{\mu_{4(a)}(X)}{V_a^2(X)} - 3,$$

where $\mu_{4(a)}(X)$ is the fourth moment of X in a particular population. The expectation of $\gamma_{2(a)}(X)$ over the stochastic process is

$$\gamma_2(X) = E\left[\frac{\mu_{4(a)}(X)}{V_a^2(X)}\right] - 3 \approx \frac{\mu_a(X)}{V^2(X)} - 3, \tag{12}$$

https://doi.org/10.1017/S0016672300020978 Published online by Cambridge University Press

approximately in a sufficiently large population, where $\mu_4(X)$ is the fourth moment of X over the stochastic process. This approximation holds, because the coefficient of variation of variance is small, as shown earlier. At any rate, $\mu_4(X)$ can be determined by the method given in the Appendix. In an equilibrium population, therefore, we have

$$\gamma_2(X) = \frac{1}{n} 3(3m-1)/16Nmv. \tag{13}$$

The above formula indicates that the distribution of genotypic value is leptokurtic if n and Nv are small, but as these values increase, the distribution gradually reaches normality. Some numerical values of $\gamma_2^{(\infty)}(X)$ are given in Table 1. It is clear that when $4Nv$ is 0·05 with the expected heterozygosity of 0·024, $\gamma_2^{(\infty)}(X)$ is rather large even if n is as large as 10. We note that in many mammalian species the average heterozygosity for protein loci is of this order of magnitude (Fuerst *et al.* 1977). Therefore, if the loci controlling a quantitative character have the same degree of heterozygosity as that of protein loci, then the normal distribution of genotypic value may not be guaranteed even if the number of loci concerned is as large as 10. This is particularly so in a population of small effective population size. However, if the average heterozygosity is as high as 0·2 and the number of loci is about 5 or more, the distribution will be approximately normal.

It is somewhat counter-intuitive that in Table 1 γ_2 increases as the number of mutational steps (m) increases for given values of n, N, and v, though the distribution $\{\alpha_r\}$ tends to be normal with increasing m. This perplexing result is apparently due to the fact that the distribution $\{\alpha_r\}$ includes the case of no mutation, i.e. $r = 0$. It is noted that when $m = 1$, for example, allele A_i mutates to only A_{i-1} and A_{i+1}, and if we neglect the case of no mutation, the distribution is actually rectangular rather than binomial, and thus the immediate effect of mutation on $\gamma_2(X)$ is negative. As m increases, however, the value of α_0 decreases and the distribution $\{\alpha_r\}$ gradually becomes normal. Consequently, the effect of mutation on $\gamma_2(X)$ increases to 0 as m increases. The effect of finite population size is to produce a positive value of $\gamma_2(X)$. Thus, the total effect will be positive and will increase with m in a finite population.

GENETIC VARIATION BETWEEN POPULATIONS

To study the amount of genetic variation between populations, let us consider the case where a population splits into a large number of completely isolated populations of equal size at an evolutionary time and thereafter no migration occurs among these populations. In this case we may write the genotypic value of the jth individual in the kth population as follows:

$$x_{kj} = g + \sum_{l=1}^{n} \{2y_{kl} + (x_{lj} + x'_{lj})\}, \tag{14}$$

neglecting the environmental effect, where g is the effect of overall genetic background, $2y_{kl}$ the deviation of the kth population mean from the overall

population mean for the lth locus, and x_{lj}, x'_{lj} the allelic effects. The variance of x_{kj} over all populations is then given by

$$V(x_{kj}) = \sum_{l=1}^{n} [4V(y_{kl}) + \{V(x_{lj}) + V(x'_{lj})\}],$$

which reduces to

$$V(x_{kj}) = 4nV(y_k) + 2nV(x_j), \tag{15}$$

in our case. In this equation $4nV(y_k)$ and $2nV(x_j)$ are the expected interpopulational variance $[B(X)]$ and intrapopulational variance $[V(X)]$, respectively. To obtain the interpopulational variance, let us consider two randomly chosen alleles one from each of two different populations. For a particular locus their effects may be represented by

$$z_{kj} = y_k + x_j$$

and

$$z_{k'j'} = y_{k'} + x_{j'},$$

neglecting the effect of genetic background and the environment effect. The variance of the difference between these two alleles is then given by

$$D = V(z_{kj} - z_{k'j'}) = 2V(y_k) + 2V(x_j). \tag{16}$$

Thus, the interpopulational variance is

$$B(X) = 2n(D - V). \tag{17}$$

The values (D_t and B_t) of D and B in the tth generation can be obtained by the same method as that for obtaining the genetic distance for the stepwise mutation model (Li, 1976; Chakraborty & Nei, 1977). Let $W_0(s) = \Sigma Q_i s^i$ represent the probability generating function of the allelic state difference at the time of population split, and $D_k(t)$ be the probability that in generation t two randomly chosen alleles, one from each of the two populations, differ from each other by k states. Then the *pgf* $[D(s,t)]$ of $\{D_k(t)\}$ is given by

$$D(s,t) = W_0(s)\, R(s,t). \tag{18}$$

The expectation of the distribution of $\{D_k(t)\}$ is

$$D'(s,t)|_{s=1} = W'_0(1) + R'(1,t).$$

In the present case $R(s,t) = \exp[2vt\{((1+s)^2/4s)^m - 1\}]$, so that $R'(1,t)$ is zero. The variance of allelic-state difference in generation t is then given by

$$\begin{aligned} &D''(1,t) + D'(1,t) - [D'[1,t)]^2 \\ &\quad = W''_0(1) + W'_0(1)[1 - W'_0(1)] + R''(1,t) + R'(1,t)[1 - R'(1,t)] \\ &\quad = \text{variance of the distribution } \{Q_k\} \\ &\qquad + \text{variance of the distribution } \{R_k(t)\}. \end{aligned} \tag{19}$$

The distribution of $R_k(t)$ has a variance equal to mvt. Thus, if we denote by V_0 the intrapopulational variance in generation 0, D_t is given by

$$D_t = V_0 + mva^2t.$$

https://doi.org/10.1017/S0016672300020978 Published online by Cambridge University Press

Therefore, the intrapopulational variance is

$$\begin{aligned} B_t(X) &= 2n(D_t - V_t) \\ &= 2n[mva^2t + (V_0 - V_\infty)(1 - e^{-t/2N})] \\ &= V_\infty(X)t/N + (V_0(X) - V_\infty(X))(1 - e^{-t/2N}). \end{aligned} \tag{20}$$

If the initial population is monomorphic, i.e. $V_0 = 0$. B_t is given by

$$B_t(X) = 2na^2[mvt - 2Nmv(1 - e^{-t/2N})], \tag{20a}$$

whereas if $V_0 = V_\infty$,

$$B_t(X) = 2nmva^2t = V_\infty(X)\,t/N. \tag{20b}$$

Formula (20b) is equivalent to the formula obtained by Lande (1976) with the method of Gaussian process.

RATIO OF INTERPOPULATIONAL VARIANCE TO INTRAPOPULATIONAL VARIANCE

With proper experimental designs it is possible to estimate the interpopulational genetic variance as well as the intrapopulational variance. Under certain circumstances, comparison of these variance components will provide some information about the evolutionary forces of the character under investigation. We have seen that when $V_0(X) = V_\infty(X)$ the interpopulational variance increases linearly, i.e. $B_t(X) = V_\infty(X)\,t/N$ and $V_t(X) = V_0(X)$. Therefore, we have

$$B_t(X)/V_t(X) = t/N. \tag{21}$$

This property can be used for testing the hypothesis of neutral evolution. In practice, the intrapopulational variance includes the environmental variance, and it is not always easy to separate this component from the genetic variance. However, even if the environmental variance is included, the $B_t(X)/V_t(X)$ ratio still remains proportional to t.

When $V_0(X)$ is not equal to $V_\infty(X)$, this property does not hold even under neutral evolution. The general expression of the $B_t(X)/V_t(X)$ ratio is somewhat complicated. However, if $t \ll 2N$, the ratio may be written as

$$\frac{B_t(X)}{V_t(X)} = \frac{[V_0(X) + V_\infty(X)]\,t}{2V_0(X)\,N + [V_\infty(X) - V_0(X)]\,t}, \tag{22}$$

approximately. Therefore, if $2V_0(X)\,N \gg [V_\infty(X) - V_0(X)]\,t$, the ratio again increases linearly but the coefficient is no longer $1/N$. In the evolutionary process new populations are often started from a small number of individuals. In this case $V_0(X)$ may be close to 0. If $V_0(X) = 0$, the $B_t(X)/V_t(X)$ ratio remains 1 as long as $t \ll 2N$.

This study was supported by research grants from the U.S. National Institute of Health and the U.S. National Science Foundation.

REFERENCES

CAVALLI-SFORZA, L. L. & BODMER, W. F. (1971). *The Genetics of Human Populations*. San Francisco: W. H. Freeman.

CAVALLI-SFORZA, L. L. & FELDMAN, M. W. (1976). Evolution of continuous variation: direct approach through joint distribution of genotypes and phenotypes. *Proceedings of the National Academy of Sciences, U.S.A.* **73**, 1689–1692.

CHAKRABORTY, R. & NEI, M. (1976). Hidden genetic variability within electromorphs in finite populations. *Genetics* **84**, 385–393.

CHAKRABORTY, R. & NEI, M. (1977). Bottleneck effect on average heterozygosity and genetic distance with the stepwise mutation model. *Evolution* **31**, 347–356.

CHAKRABORTY, R., FUERST, P. A. & NEI, M. (1980). Statistical studies on protein polymorphism in natural populations. III. Distribution of allele frequencies and the number of alleles per locus. *Genetics* **94**, 1039–1063.

CLAYTON, G. A., MORRIS, J. A. & ROBERTSON, A. (1957). An experimental check on quantitative genetical theory. I. Short-term responses to selection. *Journal of Genetics* **55**, 131–151.

FISHER, R. A. (1922). On the dominance ratio. *Proceedings of the Royal Society of Edinburgh* **42**, 321–341.

FUERST, P. A., CHAKRABORTY, R. & NEI, M. (1977). Statistical studies on protein polymorphism in natural populations. I. Distribution of single locus heterozygosity. *Genetics* **86**, 455–483.

HOLT, S. B. (1961). Quantitative genetics of finger-print patterns. *British Medical Bulletin* **17**, 247–249.

KENDALL, M. G. (1947). *The Advanced Theory of Statistics, Vol. 1, 3rd ed.* London: Charles Griffin & Co., Ltd.

KIMURA, M. (1965). A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proceedings of the National Academy of Sciences, U.S.A.* **54**, 731–736.

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

KIMURA, M. & OHTA, T. (1971). *Theoretical Aspects of Population Genetics*. Princeton, New Jersey: Princeton University Press.

LANDE, R. (1975). The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genetical Research* **26**, 221–235.

LANDE, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**, 314–334.

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. & NOVITSKI, C. E. (1969). Selection in finite populations with multiple alleles. I. Limits to directional selection. *Genetics* **62**, 859–876.

LI, W.-H. (1976). Electrophoretic identity of proteins in a finite population and genetic distance between taxa. *Genetical Research* **28**, 119–127.

NEI, M. & CHAKRABORTY, R. (1973). Genetic distance and electrophoretic identity of proteins between taxa. *Journal of Molecular Evolution* **2**, 323–328.

NEI, M. & IMAIZUMI, Y. (1966). Effects of restricted population size and increase in mutation rate on the genetic variation of quantitative characters. *Genetics* **54**, 763–782.

OHTA, T. & KIMURA, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research* **22**, 201–204.

ROBERTSON, A. (1967). The nature of quantitative genetic variation. *Heritage from Mendel* (ed. R. A. Brink), pp. 265–280. Madison, Wisconsin: University of Wisconsin Press.

SUZUKI, H. (1960). Temporal changes of anthropometric characters of the Japanese. *Proceedings of the Darwin Centennial Symposium on Evolution* (ed. H. Oka), pp. 140–149. Tokyo: Japan Society for the Promotion of Science.

WEHRHAHN, C. F. (1975). The evolution of selectively similar electrophoretically detectable alleles in finite natural populations. *Genetics* **80**, 375–394.

WRIGHT, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.

WRIGHT, S. (1937). The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences, U.S.A.* **23**, 307–320.

APPENDIX

We intend to compute the fourth moment and the variance of variance of

$$X = g + \sum_{l=1}^{n} (x_l + x_l')$$

given in (2). We know that the observed variance of X is the sum of variances, V_a, over n loci, and thus $E[V_a(X)] = nV$ where V is the expected variance of $x+x'$ at a locus as given in (4). The variance of variance of X is given by $V[V_a(x)] = nV(V_a)$, since the n loci are assumed to be independent. Following Kendall (1947, pp. 206–207), the variance of variance at a locus is approximately given by

$$V(V_a) = \frac{1}{N}(\mu_4 - V^2), \tag{A 1}$$

where μ_4 is the expected fourth moment of $x+x'$ for a locus.

On the other hand, it can be shown that the expected fourth moment of X is

$$\mu_4(X) = n\mu_4 + 3n(n-1)\,V^2. \tag{A 2}$$

Thus, if we know μ_4, the variance of variance of X, $V[V_a(X)]$ as well as $\mu_4(X)$ can be determined. In the following we assume that the distribution of X (and x') is symmetric, so that the first and third moments of $x+x'$ are 0. Under this assumption, $\mu_4 = E(x+x')^4 = E(x-x')^4$. Therefore, μ_4 can be determined by using the *pgf* $P(s,t)$ in (6).

The computation of μ_4 is tedious but straightforward. This moment in generation t becomes

$$\mu_4^{(t)} = \mu_4^{(\infty)} + [(\mu_4^{(0)} - \mu_4^{(\infty)}) + 6mvta^2(V_0 - V_\infty)]\,e^{-t/2N}, \tag{A 3}$$

where

$$\mu_4^{(\infty)} = Nmva^4[12Nmv + \tfrac{3}{4}(3m-1)]. \tag{A 4}$$

On the other hand, the variance of variance of X at steady state is given by

$$V[V_a(X)] = nmva^4[8Nmv + \tfrac{3}{4}(3m-1)]. \tag{A 5}$$

The expected kurtosis is

$$\begin{aligned}\gamma_2(X) &\approx \frac{\mu_4(X)}{V^2(X)} \\ &= \frac{n\mu_4 + 3n(n-1)\,V^2}{n^2V^2} - 3 \\ &= \frac{1}{n}\left(\frac{\mu_4}{V^2} - 3\right) \\ &= \frac{1}{n}\left\{\frac{3(3m-1)}{16Nmv}\right\}.\end{aligned} \tag{A 6}$$

https://doi.org/10.1017/S0016672300020978 Published online by Cambridge University Press