Multivariate normal genetic models with a finite number of loci

Joseph Felsenstein
DEPARTMENT OF GENETICS
UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195

ABSTRACT

A genetic model due to Russell Lande is described. The model assumes a finite number of loci at each of which there are an infinite number of alleles whose effects on the phenotype are normally distributed. Analytic and numerical results using this model depend on the allele effects remaining multivariate normally distributed. This is almost never exactly true, but may often be a good approximation. Numerical results for several kinds of natural selection are discussed. A model involving overdominance is presented which seems to exhibit the Franklin-Lewontin crystallization effect. A model is presented of the maintenance of genetic variation by a cline along which there is linear change in the optimum phenotype under optimizing selection. The equilibrium has been found analytically for the case of an infinite cline. Remarkably, there is no linkage disequilibrium maintained at equilibrium in this case.

INTRODUCTION

The use of normal distributions for quantitative characters goes back to the Biometricians, and is most visible in selection theory (e.g. Cochran, 1951). The last twenty years of work on

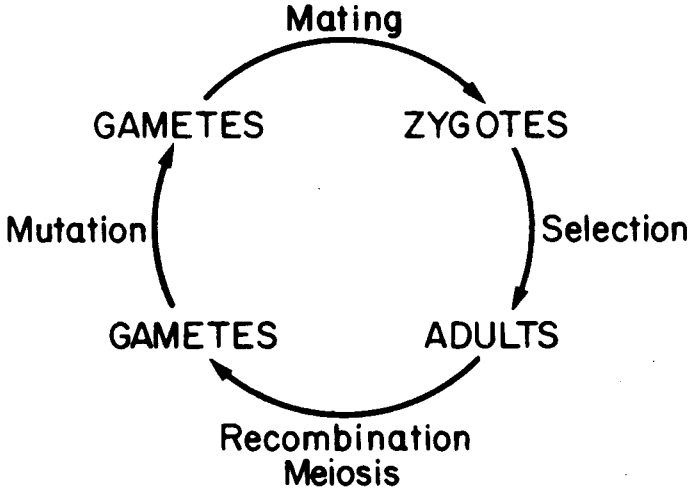
Proceedings of the International Conference on QUANTITATIVE GENETICS

the theory of polygenic inheritance has concentrated on phenomena for which normal distributions are inadequate, such as selection limits and effects of linkage (for which the reader should consult the papers by Robertson and Karlin in this volume). Lately there has been a revival of interest in the normal theory, in connection with the application of quantitative genetics to evolutionary ecology and to the study of cultural evolution. Slatkin (1970) and Bulmer (1971a, b, 1972, 1973) have examined the effect of normalizing (optimizing) selection on genetic variability. Roughgarden (1972, 1974a, b), Bulmer (1974a, b), and Slatkin and Lande (1976) have examined frequency-dependent selection and the evolution of niche width. Lande (1976) has considered the rate of evolutionary change induced by a moving optimum when there is optimizing selection. Bulmer (1971c, d) has considered the maintenance of genetic variability by geographic variation. Cavalli-Sforza and Feldman (1976) have considered the equilibrium between mutation and normalizing selection. Feldman (this volume) incorporates phenotypic transmission (cultural inheritance) into this scheme.

The use of normal distributions in all of these papers is motivated by the assumption that they are good approximations to a large and relevant class of schemes for the genetic determination of quantitative characters, in particular models with many loci, each of small effect, and with the phenotype the sum of the effects of the individual loci. There has been almost no investigation of the accuracy of this approximation (but see Wiorkowski, 1972). One would like to have genetic models in which a normal distribution of phenotypes is achieved and maintained exactly. This has frequently been achieved when only a single locus is considered. Kimura (1965), in the first of the modern wave of normal-distribution papers, considered the equilibrium between optimizing selection and mutation. Eshel (1971, 1972, 1973) has considered the rate of evolution in models where there is a

THE MULTIVARIATE NORMAL MODEL - SELECTION

We consider an infinitely large diploid population with discrete generations. The life cycle of the organism will consist of random mating, followed by natural selection among the diploid zygotes, followed by production of gametes by meiosis:



There are n loci. At each locus there are an infinite number of possible alleles. With each allele we associate a real number x, which is to be interpreted as the contribution of that allele to the phenotype. At each locus the values x have a distribution, which we take to be a normal distribution. Note that gene frequencies do not exist in this scheme. A gamete can be represented by a vector of n real numbers, which are the alleles at loci 1 through n: $(x_1, x_2, \ldots x_n)$. Let us begin by assuming that, at generation t, the gametes in the gamete pool are in a multivariate normal distribution with mean vector m and covariance matrix V. We wish to know whether the gametes of the next generation will still be in a multivariate normal distribution. We must determine what happens at each of four stages: formation of zygotes, selection, meiosis, and mutation.

distribution of changes in fitness after mutation. His results are general, including normals as a special case. Bodmer and Cavalli-Sforza (1972) have simulated such a model in a finite population with normal distributions of mutation effect. Latter (1970, 1972) has simulated finite populations with optimizing selection and normal distribution of mutational effect. Finite populations will depart from normal distribution of the phenotype, so that analytical results are usually not available. The exception is the diffusion approximations to the directional selection case developed by Guess (1974).

When we assume more than a single locus, we must somehow incorporate the effects of linkage into the mathematics. Latter (1969) has simulated directional selection in a model with a finite number of loci each with infinitely many alleles falling in a normal distribution of allelic effects. Bulmer (1971a) has investigated the generation of linkage disequilibrium in a model with infinitely many loosely-linked loci, finding that there is a finite effect of disequilibrium, which can be incorporated into the normal-distribution approximately even though we are not keeping track of individual loci. In a later (1974b) paper, he approximated the effects of linkage by an analysis which is approximate unless the pairwise recombination fractions among all loci are the same. Cavalli-Sforza and Feldman (1976) have, in effect, ignored linkage disequilibrium when computing the equilibrium between mutation and normalizing selection.

If one set up a model involving a finite number of loci, each with a normal distribution of allele effects, it would be natural to ask under what conditions the joint distribution of allele effects at all loci would remain multivariate normal. This model has been posed, and the question answered, in a beautiful paper by Lande (1975). In the present paper I will present and slightly extend Lande's results.

(a) Formation of Zygotes

Each zygote will be composed of two gametes. We can number the genes in the maternal gamete 1 through n, and those in the paternal gamete n+1 through 2n. Thus we can represent an individual by a vector $\underline{x}$ of length 2n, whose distribution we wish to know. We are assuming that there are no differences between the sexes, so that the maternal and paternal gametes each have distribution $N(\underline{m}, \underline{V})$. Since random mating of individuals is equivalent (in this case) to random combination of gametes, the two halves of the vector $\underline{x}$ each are drawn from $N(\underline{m}, \underline{V})$ independently. There is (at this stage in the life cycle) no covariance between x_i and x_j if $i \leq n < j$. Clearly the distribution of $\underline{x}$ is multivariate normal.

$$\underline{x} \frown N\left(\begin{bmatrix} \underline{m} \\ \underline{m} \end{bmatrix}, \begin{bmatrix} \underline{V} & \underline{0} \\ \underline{0} & \underline{V} \end{bmatrix} \right) \qquad (1)$$

$\underline{m}$ being a column vector, and $\underline{0}$ being the n x n matrix of zeros. In what follows, I will use $\underline{V}$ sometimes for the 2n x 2n covariance matrix of genotypes, and sometimes for the n x n covariance matrix of gametes. It should be clear from the context which is intended.

Note that if, in the initial generation, each gene had a normal distribution independently of each other gene, we would be in a state which could only be called linkage equilibrium. In that state $\underline{V}$ would be a diagonal matrix. We shall consider this to be the definition of linkage equilibrium in the present context.

(b) Selection

If the fitness of an individual is the function $w(\underline{x})$ of its genotype vector, and if we take $f(\underline{x})$ as the density function of the random variable $\underline{x}$, then after selection, the survivors will be found to have the density

$$g(\underline{x}) = f(\underline{x})\, w(\underline{x}) \Big/ \int_{R^n} f(\underline{x})\, w(\underline{x})\, d\underline{x}. \qquad (2)$$

The denominator of the right side of (2) is the mean fitness $\bar{w}$ of individuals in the original distribution.

We wish to find a fitness function which will leave the survivors still in a multivariate normal distribution. The particular form we will use is

$$w(\underline{x}) = \exp[\ \underline{c}'\ \underline{x} - \tfrac{1}{2}\ (\underline{x} - \underline{a})'\ \underline{B}\ (\underline{x} - \underline{a})\ ], \qquad (3)$$

where $\underline{c}$ and $\underline{a}$ are vectors of length 2n, and $\underline{B}$ is a symmetric 2n x 2n matrix. As we shall see shortly, this encompasses a number of interesting types of selection. One can readily show, using equation (2), that if $g(\underline{x})$ is to be a multivariate normal density, $w(\underline{x})$ must be of the form (3), with the trivial exception that it might be further multiplied by a constant. The converse is not true: for some $\underline{a}$, $\underline{B}$, and $\underline{c}$, $g(\underline{x})$ is not a multivariate normal density — in fact, is not a density at all.

If we specify $f(\underline{x})$ and $g(\underline{x})$ as:

$$f(\underline{x}) = (2\pi)^{-n}\ |\ \underline{V}_1\ |^{-\frac{1}{2}} \exp[\ -\tfrac{1}{2}(\underline{x}-\underline{m}_1)'\ \underline{V}_1^{-1}\ (\underline{x}-\underline{m}_1)\ ] \qquad (4)$$

and

$$g(\underline{x}) = (2\pi)^{-n}\ |\ \underline{V}_2\ |^{-\frac{1}{2}} \exp[\ -\tfrac{1}{2}(\underline{x}-\underline{m}_2)'\ \underline{V}_2^{-1}\ (\underline{x}-\underline{m}_2)\ ], \qquad (5)$$

and keep in mind that the denominator of (2) is a constant, by equating like terms in (2) we can compute that:

$$\underline{V}_2^{-1} = \underline{V}_1^{-1} + \underline{B} \qquad (6)$$

and

$$\underline{V}_2^{-1}\ \underline{m}_2 = \underline{V}_1^{-1}\ \underline{m}_1 + \underline{B}\ \underline{a} + \underline{c}. \qquad (7)$$

These expressions can be used to compute the effect of selection on the means and covariances of the distribution of $\underline{x}$. If $\underline{V}_2$ is not a positive definite matrix, then the integral in the denominator of (2) diverges, and no density function $g(\underline{x})$ exists.

(c) Recombination

If we have n loci, there will be 2^n different recombination classes possible. Suppose that the k-th of these has probability

R_k. It might, for example, consist of the maternally-derived genes at loci 1 through 10, the paternally-derived genes at loci 11 and 12, and the maternally-derived genes at loci 13 through 16. The 16-tuple corresponding to this recombinant gamete would be $(x_1, \ldots, x_{10}, x_{27}, x_{28}, x_{13}, \ldots, x_{16})$. If we consider all the gametes in the population formed by this particular type of recombination, these will follow a multivariate normal distribution, since they are a subset of the 2n variables which are themselves multivariate normal. The mean vector of this particular recombinant class is simply the appropriate subset (ordered as above) of the mean vector $\underline{m}_2$ following selection. The covariance matrix of this recombinant class can be obtained by taking the appropriate subset of rows and the same subset of columns of $\underline{V}_2$.

In all of the cases which we shall consider, there is a particular symmetry property of the selection. Selection does not discriminate, in these cases, between which gene is maternally- and which paternally-derived. So for the vector $\underline{c}$,

$$c_i = c_{i+n} \tag{8a}$$

and for the matrix $\underline{B}$,

$$b_{ij} = b_{i+n,j+n} = b_{i,j+n} = b_{i+n,j} \tag{8b}$$

It is not difficult to show that this implies that if $\underline{V}_1$ (before selection) is of the form

$$\begin{bmatrix} \underline{V}_{11} & \underline{V}_{12} \\ \underline{V}_{12} & \underline{V}_{11} \end{bmatrix},$$

$\underline{V}_2$ will also show the same symmetry properties. It will also be true that since before selection $m_i = m_{i+n}$ (by random mating), this will continue to hold after selection.

The upshot of all of this is that if we consider all 2^n of the recombinant classes, (i) all of them have the same mean vector, and (ii) all of them have the same variances v_{ii}. This would seem to offer hope that all might follow the same multivariate normal distribution. But these hopes are dashed when

we consider the covariances. Consider the case of two loci (n = 2). Suppose that

$$\underline{V}_2 = \left[\begin{array}{cc|cc} v_{11} & v_{12} & v_{13} & v_{14} \\ v_{12} & v_{22} & v_{14} & v_{13} \\ \hline v_{13} & v_{14} & v_{11} & v_{12} \\ v_{14} & v_{13} & v_{12} & v_{22} \end{array}\right] . \qquad (9)$$

There are four possible recombinant classes, but in the symmetrical cases which we consider, these reduce to two: recombinants and nonrecombinants. Each class of gametes is multivariate normal with the same mean vector, but different covariance matrices. A fraction (1-r) of the gametes have covariance matrix

$$\begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix} ,$$

while r of them have covariances:

$$\begin{bmatrix} v_{11} & v_{14} \\ v_{14} & v_{22} \end{bmatrix} .$$

Thus the distribution of gametes is a mixture of two bivariate normals, with the same means and variances but with different covariances.

Under what conditions will such mixtures be themselves multivariate normal? We can investigate this by considering the characteristic function of the mixture. The characteristic function of a multivariate normal distribution with mean vector $\underline{m}$ and covariance matrix $\underline{V}$ is the complex-valued function

$$F(\underline{\theta}) = E[\, \exp(i\, \underline{\theta}'\, \underline{x})\,] = \exp(\, i\, \underline{\theta}'\, \underline{m} - \tfrac{1}{2}\, \underline{\theta}'\, \underline{V}\, \theta\,), \qquad (10)$$

$\underline{\theta}$ being a vector $(\theta_1, \ldots, \theta_n)$. The characteristic function of the mixture of gamete subpopulations is

$$G(\underline{\theta}) = \sum_r R_r \exp\, (i\, \underline{\theta}'\, \underline{m}^{(r)} - \tfrac{1}{2}\, \underline{\theta}'\, \underline{V}^{(r)}\, \underline{\theta}) \qquad (11)$$

where $\underline{m}^{(r)}$ and $\underline{V}^{(r)}$ are respectively the mean vector and covar-

iance matrix of the r-th recombination class. Since distributions are uniquely determined by their characteristic functions, the distribution of gametes will be multivariate normal if and only if a mean vector $\underline{m}$ and covariance matrix $\underline{V}$ exist such that (11) is of form (10). In fact, the mean vectors of all of the recombinant classes are the same, so $\underline{m}^{(r)} = \underline{m}$ for all r. Since all the recombination classes have the same means, the overall covariance matrix is simply the mean of the covariance matrices $\underline{V}^{(r)}$.

The question of multivariate normality of the gamete population thus reduces to whether

$$\exp\Big(\sum_r R_r\, \underline{\theta}'\, \underline{V}^{(r)}\, \underline{\theta}\Big) = \sum_r R_r \exp\big(\underline{\theta}'\, \underline{V}^{(r)}\, \underline{\theta}\big). \qquad (12)$$

for <u>all</u> vectors $\underline{\theta}$. It is not difficult to show that this almost never holds true. Consider only those $\underline{\theta}$ in which the j-th and k-th elements are the only nonzero ones, i.e. $\underline{\theta} = (0, 0, \ldots, 0, \theta_j, 0, \ldots, 0, \theta_k, 0, \ldots, 0)$. If (12) holds for all $\underline{\theta}$, it must hold for these in particular. For all values of θ_j and θ_k, it would have to be true that

$$\exp\Big[\sum_r R_r\, (\theta_j^{\,2}\, v_{jj}^{(r)} + \theta_k^{\,2}\, v_{kk}^{(r)} - 2\,\theta_j\,\theta_k\, v_{jk}^{(r)})\Big]$$

$$= \sum_r R_r \exp\,(\theta_j^{\,2}\, v_{jj}^{(r)} + \theta_k^{\,2}\, v_{kk}^{(r)} - 2\theta_j\,\theta_k\, v_{jk}^{(r)}\,). \qquad (13)$$

All of the $v_{jj}^{(r)}$ and $v_{kk}^{(r)}$ are equal (respectively) to v_{jj} and v_{kk}. $\exp(x)$ is a convex function, so by Jensen's Inequality (13) is true only if $-2\theta_j\,\theta_k\, v_{jk}^{(r)}$ is equal for all those r for which R_r is nonzero. But $v_{jk}^{(r)}$ can have only two possible values. It was selected from either v_{jk} or $v_{j,k+n}$ in the original covariance matrix of diploid zygotes, $\underline{V}_2$.

We have the following result: a necessary condition for the population of gametes be multivariate normally distributed is that either (a) there is only one recombination class (i.e., no recombination at all), or (b) that $v_{jk} = v_{j,k+n}$ in $\underline{V}_2$ (after selection). It follows that recombination generally causes the

distribution of gametes to depart from multivariate normality. In assuming that normality continues to apply, we are making an approximation. It will be accurate if all of the $\underline{V}^{(r)}$ are nearly equal, but as far as I know its accuracy has not been seriously investigated.

To find the parameters of the multivariate normal approximation to the distribution of gametes is not difficult. We need not compute all of the R_r and all of the $\underline{V}^{(r)}$. No computation of the mean vector is necessary, since under our symmetry conditions, it does not change. The gamete mean vector is simply the first n (or the last n) elements of the diploid genotype mean vector. An element of the new gamete covariance matrix, v_{jk}', is a weighted average of the $v_{jk}^{(r)}$. But the latter are each either v_{jk} or $v_{j,k+n}$, depending on whether recombination class $\underline{i}$ is nonrecombinant or recombinant for loci j and k. Thus we need only know the recombination fraction between loci j and k, r_{jk}, and can compute the gamete covariances as

$$v_{jk}' = (1 - r_{jk})\, v_{jk} + r_{jk}\, v_{j,k+n}. \qquad (14)$$

The approximation we are using is to take as the distribution of gametes the multivariate normal distribution implied by the means and covariances of the gametes.

The essential conclusions of this section will be found in Lande's paper.

(d) Mutation

We may also wish to have mutation occur among the gametes. We shall assume that to each gene x_i is added an independent change in the identity of the allele. The change, due to mutation, is assumed to have mean zero and variance u_i at the i-th locus. If all of these mutational events are independent, clearly the gametes after mutation will each be the sum of two vectors, $\underline{x} + \underline{y}$, where $\underline{x}$ is drawn from the pre-mutation gamete distribution and $\underline{y}$ is the change due to mutation. Since both are multivariate

normal (the former by our recombination approximation) so is their sum. This type of mutation at the gamete stage leaves the mean vector unchanged, but adds to the covariance matrix a diagonal matrix (here called $\underline{U}$).

To carry a population through random mating, selection, recombination and mutation, we start with the mean vector $\underline{m}$ and covariance matrix $\underline{V}$ of the gametes. We then:

(i) Create the mean vector and covariance matrix of the diploid genotypes as shown in (1).

(ii) Compute the means and covariances in the diploid survivors of selection using (6) and (7).

(iii) Compute the covariance matrix of gametes after recombination, using (14). This is an n x n matrix. The mean vector of gametes is simply the first n elements of the mean vector of diploid genotypes (after selection).

(iv) Compute the covariance matrix after mutation by adding $\underline{U}$ to the matrix (i.e., $v_{ii}' = v_{ii} + u_i$ for all i).

RESULTS WITH VARIOUS KINDS OF SELECTION

Multiplicative Selection

If we restrict attention to types of selection which are within the framework of equation (3), the simplest case is when $\underline{a} = \underline{0}$, $\underline{B} = \underline{0}$, but $\underline{c}$ is nonzero. In this case, fitness is a product of fitnesses at the individual loci, with no dominance within loci. If $c_i = c_{i+n}$, then the fitness function is

$$w(\underline{x}) = \exp[\sum_i c_i(x_i + x_{i+n})] = \prod_{i=1}^{n} \exp[c_i(x_i + x_{i+n})] \quad (15)$$

It is easy to see what happens in such a case. Under selection, (6) shows that there is no change in the covariance matrix $\underline{V}_1$. However, by equations (1) and (14), recombination will cause all but the diagonal elements of $\underline{V}_1$ to become zero, provided that there is some recombination between all pairs of loci. If there is no further mutation, equation (7) shows that $\underline{m}$ will ultimately be changing by the constant amount $\underline{V}_1\ \underline{c}$ each generation. If there

is mutation, the diagonal elements of $\underline{V}_1$ will increase linearly, and thus the amount of change of $\underline{m}$ will increase continually once $\underline{V}$ is essentially diagonal. We thus have no asymptote or equilibrium of $\underline{m}$, and no recombination effects, except possibly effects due to its breakdown of initial linkage disequilibrium.

Optimizing Selection

Suppose that selection favors phenotypes close to some optimum value A. If the genotypic contribution to the phenotype is simply the sum of allele effects $G = \Sigma x_i$, and if the environmental contribution E is normally distributed with variance V_E and is independent of G, then the fitness of $\underline{x}$ is

$$w(\underline{x}) = \int (2\pi V_E)^{-1/2} \exp[-E^2/(2V_E)] W(G + E) \, dE. \qquad (16)$$

If the logarithm of fitness declines quadratically with deviation from the optimum, we can find a positive constant V_S such that

$$W(P) = \exp[- (P - A)^2/(2V_S)], \qquad (17)$$

which when substituted into (16) gives

$$w(\underline{x}) = (1 + V_E/V_S)^{-1/2} \exp [- (G - A)^2/(2V_S + 2V_E)] \qquad (18)$$

Since $G = \Sigma x_i = \underline{1}'\underline{x}$, this is (except for the leading constant), of form (3) with $\underline{c} = 0$,

$$\underline{a} = [A/(2n)] \, \underline{1} \qquad (19a)$$

and

$$\underline{B} = [1/(V_S + V_E)] \, \underline{J}, \qquad (19b)$$

where $\underline{J} = \underline{1} \, \underline{1}'$ is a matrix all of whose elements are ones.

This type of selection, balanced against mutation, has been extensively examined by Lande (1975). An equilibrium is reached in which the mean phenotype is at the optimum, genetic variance being maintained by a balance between mutation and selection. At this equilibrium there is negative (repulsion) linkage disequilibrium between different loci. The reader should consult Lande's paper for details, including some interesting comments on the opportunity for divergence of populations by random

genetic drift under this kind of selection-mutation balance.

Although investigation of the convergence of this case to the equilibrium has been incomplete, I have done numerical iterations of the mean vector and covariance matrix for a number of different parameter values, and have always found them to approach the equilibrium given by Lande. An interesting feature of Lande's solution is that the total genetic variance of the character does not depend on the pattern or amount of recombination. In the case where all mutation rates are equal, Lande's equation reduces to

$$V_G = 2n[u(V_S + V_E + n^2u)]^{1/2} + 2nu^{1/2} \qquad (20)$$

A change in the recombination fractions redistributes this variance among the variances v_{ii} and the covariances v_{ij}, but does not change the total $V_G = \underline{1}'\underline{V}\underline{1}$. If the mutation rates are made to approach zero, the equilibrium genetic variance also approaches zero, as can be seen in equation (20).

Frequency-Dependent Selection

Bulmer (1974) has presented a model of selection in which genetic variance is maintained at equilibrium, not by mutation, but by selection, based on the assumption that genotypes whose phenotypes are similar compete more than those whose phenotypes are more distant. Although Bulmer's specific fitness function is not of the form of equation (3), an analogous function can be developed which is. Fitness as a function of phenotype is taken to be

$$W(P) = [\int e^{-C(Q-P)^2} f(Q)\, dQ\,]^{-D}, \qquad (21)$$

where $f(Q)dQ$ is the density function of the phenotype immediately before this selection acts. There is not sufficient space here to show the derivation, but this leads to a fitness function of form (3), with $\underline{c} = 0$, $\underline{a} = \underline{m}$ (the current mean vector), and

$$\underline{B} = [\,-1/(2/[CD] + \underline{1}'\underline{V}\underline{1}/D - V_E)]\,\underline{J}, \qquad (22)$$

provided that the quantity within the outer brackets is negative.

Note that the presence of $\underline{1}'\underline{V}\underline{1}$, the current diploid genetic variance, in (22) ensures that selection will weaken as the genetic variance increases.

If this selection acts alone, no equilibrium is reached. Not only does $V_G = \underline{1}'\underline{V}\underline{1}$ increase continually, but this increase accelerates, resulting in such a strong selection favoring the population extremes that after a finite number of iterations the distribution of $\underline{x}$ no longer exists, its covariance matrix no longer being positive definite.

We can achieve an equilibrium by assuming that this frequency-dependent selection occurs immediately before the optimizing selection given in equations (19). It is readily demonstrated that this compound selection is of form (3), with $\underline{B}$ the sum of (22) and (19b):

$$\underline{B} = [\, 1/(V_S + V_E) - 1/(2/[CD] + \underline{1}'\underline{V}\underline{1}/D - V_E)]\, \underline{J}. \qquad (23)$$

Equilibrium is reached when

$$\underline{1}'\underline{V}\underline{1} = V_G = D\,(V_S + 2V_E) - 2/C. \qquad (24)$$

When V_G exceeds this value, the net selection reduces V_G, and when it is smaller than this value V_G increases. The result is an equilibrium at this value. The most interesting feature of this equilibrium is that, when it is achieved, $\underline{B} = \underline{0}$. By the same argument as in the case of multiplicative selection, there must then be no linkage disequilibrium at equilibrium. The result is that $\underline{V}$ is then diagonal, its trace being V_G. The individual v_{ii} may be any values which add to V_G. At equilibrium, it will also be true that the mean phenotype equals the optimum phenotype.

Overdominance

The models used so far have had no non-additive variance in the phenotype. We could have nonadditive variance and still be within the framework of (3) if the logarithm of fitness were a quadratic form in the x_i. If we wished to have no epistasis, but to have dominance, then we could take

$$w(\underline{x}) = \exp[\sum_{i=1}^{n} (s_{1i} x_i^2 + 2s_{2i} x_i x_{i+n} + s_{1i} x_{i+n}^2)]$$

$$= \prod_{i=1}^{n} \exp (s_{1i} x_i^2 + 2s_{2i} x_i x_{i+n} + s_{1i} x_{i+n}^2). \quad (25)$$

I have investigated numerically two symmetric special cases of (28). They are the cases (i) $s_{1i} = -s_{2i} = s$, and (ii) $s_{1i} = 0$ and $s_{2i} = -s$. These may be thought of as being cases of over-dominance, in the sense that the more different the two alleles at locus i in an individual are, the higher its fitness. In these cases, $\underline{a} = \underline{0}$, $\underline{c} = \underline{0}$, and the matrix $\underline{B}$ is given by:

in case (i) $b_{ii} = -b_{i,i+n} = -b_{i+n,i} = b_{i+n,i+n} = -s$, (26a)

and in case (ii) $b_{i,i+n} = b_{i+n,i} = s$, (26b)

all other elements b_{ij} being zero in both cases.

Computer iterations of equations (1), (6), (7), and (18) have been carried out for various choices of the r_{ij}, s, and the initial $\underline{V}$. In all cases I have tried, both of these over-dominant models have behaved similarly, so that no further reference will be made to the distinction between them. For simplicity of computation, many runs where made for the case where $r_{ij} = r$ for all i, j such that $i \neq j$. This does not correspond to a realizable genetic map unless $r = 0$ or $r = 1/2$, but it is much easier to compute (if some minor algebra is done) and usually behaves like more realistic cases.

When one starts with no linkage disequilibrium, this sort of selection does not generate any (this can also be proven algebraically). This is not surprising. Since we are selecting for ever-more extreme alleles at each locus, we would expect the v_{ii} to increase without limit, which it does. What is less obvious, but not counterintuitive, is that when some critical value of v_{ii} is reached, the next generation's v_{ii} is negative, which actually means that v_{ii} has become infinite in a finite amount of time, and a probability density of $\underline{x}$ no longer exists.

One surprising result does emerge from the numerical runs. If we start with a small amount of covariance (linkage disequilibrium), either positive or negative, between any set of loci, an unusual pattern is seen. At first the covariance declines as expected. But as the v_{ii} approach their point of explosion, the v_{ij} not only begin to increase, but increase faster than the v_{ii}. This means that the correlation between non-alleles is increasing. This phenomenon occurs earlier in the process the smaller the r_{ij}. I believe it to be the result of the Franklin-Lewontin (1970) crystallization effect, found by them in the two allele case. Since, unlike the two-allele case, we are not approaching an equilibrium, this belief is rather hard to test. It does at least merit further investigation.

A Geographic Cline Model

Suppose that, in the absence of mutation, variation is maintained by having optimizing selection, with a cline in the position of the optimum. In particular, we suppose a one-dimensional geographic continuum of infinite length, along which the optimum phenotype changes linearly. Suppose that S measures the slope of the line of optimum phenotype. Let the strength of optimizing selection be V_S everywhere. Suppose that adults migrate after the action of optimizing selection, with the individuals at each point being drawn from a normal distribution of previous positions, with mean displacement zero and variance of displacement V_M. Bulmer (1971c, d) has considered some similar cases.

Let us call some point position 0, and let the optimum phenotype be Sy at position y. We seek an equilibrium, assuming that the mean genotype vector at y after selection is $[y/(2n)]\ \underline{1}$, so that the mean phenotype at each point is at the local optimum. The distribution of genotype vectors at y after migration has some characteristic function $E[\exp(i\ \underline{\theta}'\ \underline{x})]$. Since the migration is equivalent to mixing distributions, we take the expectation

to run over all parental positions z and within each such position over all genotypic vectors $\underline{x}$:

$$E(e^{i\underline{\theta}'\underline{x}}) = E_z E_x (e^{i\underline{\theta}'\underline{x}}) = E_z [\exp(i\underline{\theta}'[z/(2n)] \underline{1} - \tfrac{1}{2}\underline{\theta}'\underline{V}\underline{\theta})] \quad (27)$$

and since z is normally distributed,

$$E(e^{i\underline{\theta}'\underline{x}}) = \exp (i\underline{\theta}'[y/(2n)] \underline{1} - \tfrac{1}{2}\underline{\theta}'(\underline{V} + \underline{V}_B) \underline{\theta}) \quad (28)$$

which shows that after migration, $\underline{x}$ is still multivariate normal, with unchanged mean, but with a between-localities covariance matrix added to $\underline{V}$. This $\underline{V}_B$ is readily computed to be

$$\underline{V}_B = E[S(z-y) \underline{1}\,\underline{1}' S(z-y)] = [S^2 V_M/(2n)^2] \underline{J} \quad (29)$$

The net effect of optimizing selection followed by migration is thus the genotype covariance matrix

$$\underline{V}_2 = (\underline{V}_1^{-1} + [1/(V_S + V_E)] \underline{J})^{-1} + [S^2 V_M/(2n)^2] \underline{J}. \quad (30)$$

Now we try to find a diagonal matrix $\underline{V}$ such that the result of (30) is the same diagonal $\underline{V}_1$. If there is one such, then neither recombination nor mating will further alter it, so that it will be an equilibrium value of $\underline{V}$. We have already established that the mean vector $[Sy/(2n)] \underline{1}$ is unchanged by all of these operations. We will then have found an equilibrium. Since matrices of the form $a\underline{I} + b\underline{J}$ are easily inverted, we can find the required $\underline{V}$ in straightforward fashion (the details are omitted for lack of space). It has

$$v_{ii} = v = \frac{1}{2n} [1/2\, S^2 V_M + 1/2\, (S^4 V_M^2 + 4 [V_S + V_E] S^2 V_M)^{1/2}]. \quad (31)$$

The genetic variance at equilibrium will be $V_G = 2nv$, for which some typical values are:

$(V_S + V_E)/(S^2 V_M)$	$V_G/(S^2 V_M)$
0.01	1.0099
0.1	1.0916
0.2	1.1708
0.5	1.3660
1	1.6180
2	2

5	2.7913
10	3.7016
100	10.5125
1000	32.1267

Note that when selection is strong compared to the amount of migration, $V_G \cong S^2 V_M$. When selection is weak, $V_G \cong S\sqrt{V_M V_S}$.

The feature of the equilibrium (31) which is of interest is that it applies irrespective of the amount or pattern of linkage, and it involves no linkage disequilibrium. This is perhaps surprising, since both migration and optimizing selection tend to produce linkage disequilibrium. But they produce disequilibrium of opposite sign, and apparently in this model these exactly cancel. Slatkin (1977) has made a general analysis of clines in optimizing selection for a model with infinitely many loci. The present result indicates that his results may generalize to cases with finitely many linked loci.

While the above anlysis neither proves uniqueness nor stability of the equilibrium, I have not observed convergence to any other equilibrium in numerical runs of this model.

ACKNOWLEDGMENTS

This work has been supported by ERDA contract AT(45-1)2225 TA 5 with the University of Washington.

BIBLIOGRAPHY

[1] Bodmer, W. F. & Cavalli-Sforza, L. L. (1972). Variation in fitness and molecular evolution. *Proc. Sixth Berkeley Symp. Math. Statist. & Prob., Vol. V,* 225-275.

[2] Bulmer, M. G. (1971a). The effect of selection on genetic variability. *Amer. Naturalist 105,* 201-211.

[3] Bulmer, M. G. (1971b). The stability of equilibria under selection. *Heredity 27,* 157-162.

[4] Bulmer, M. G. (1971c). Stable equilibria under the two-island model. *Heredity 27,* 321-330.

[5] Bulmer, M. G. (1971d). Stable equilibria under the migration matrix model. *Heredity 27,* 419-430.

[6] Bulmer, M. G. (1972). The genetic variability of polygenic characters under optimizing selection, mutation, and drift. *Genet. Res. 19*, 17-25.

[7] Bulmer, M. G. (1973). The maintenance of the genetic variability of polygenic characters by heterozygous advantage. *Genet. Res. 22*, 9-12.

[8] Bulmer, M. G. (1974a). Density-dependent selection and character displacement. *Amer. Naturalist 108*, 45-58.

[9] Bulmer, M. G. (1974b). Linkage disequilibrium and genetic variability. *Genet. Res. 23*, 281-289.

[10] Cavalli-Sforza, L. L., and Feldman, M. W. (1976). Evolution of continuous variation: direct approach through joint distribution of genotypes and phenotypes. *Proc. Natl. Acad. Sci. USA 73*, 1689-1692.

[11] Cochran, W. G. (1951). Improvement by means of selection. *Proc. Second Berkeley Symp. Math. Statist. & Prob.*, 449-470.

[12] Eshel, I. (1971). On evolution in a population with an infinite number of types. *Theoret. Pop. Biol. 2*, 209-236.

[13] Eshel, I. (1972). Evolution processes with continuity of types. *Adv. Appl. Prob. 4*, 475-507.

[14] Eshel, I. (1973). Evolution in diploid populations with continuity of types. *Adv. Appl. Prob. 5*, 55-65.

[15] Franklin, I. and Lewontin, R. C. (1970). Is the gene the unit of selection? *Genetics 65*, 707-734.

[16] Guess, H. A. (1974). Evolution in finite populations with infinitely many types. *Theoret. Pop. Biol. 5*, 417-430.

[17] Kimura, M. (1965). A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proc. Natl. Acad. Sci. USA 54*, 731-736.

[18] Lande, R. (1975). The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genet. Res. 26*, 221-235.

[19] Lande, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution 30*, 314-334.

RP. 395

[20] Latter, B. D. H. (1970). Selection in finite populations with multiple alleles. II. Centripetal selection, mutation and isoallelic variation. *Genetics 66*, 165-186.

[21] Latter, B. D. H. (1972). Selection in finite populations with multiple alleles. III. Genetic divergence with centripetal selection and mutation. *Genetics 70*, 475-490.

[22] Roughgarden, J. (1972). Evolution of niche width. *Amer. Naturalist 106*, 683-718.

[23] Roughgarden, J. (1974a). The fundamental and realized niche of a solitary population. *Amer. Naturalist 108*, 232-235.

[24] Roughgarden, J. (1974b). Niche width: biogeographic patterns among *Anolis* lizard populations. *Amer. Naturalist 108*, 429-441.

[25] Slatkin, M. (1970). Selection and polygenic characters. *Proc. Natl. Acad. Sci. USA 66*, 87-93.

[26] Slatkin, M. and Lande, R. (1976). Niche width in a fluctuating environment - density independent model. *Amer. Naturalist 110*, 31-55.

[27] Slatkin, M.* (1977). Spatial patterns in the distribution of polygenic characters. *Submitted for publication.*

[28] Wiorkowski, J. (1972). Some statistical methods for the study of quantitative genetic traits. *Unpublished Ph.D. Dissertation, University of Chicago.*