

# SOME NATURAL VIABILITY SYSTEMS FOR A MULTIALLELIC LOCUS: A THEORETICAL STUDY\*

SAMUEL KARLIN

*Department of Mathematics, Stanford University, Stanford, California 94305*

Manuscript received August 8, 1980  
Revised copy received December 9, 1980

## ABSTRACT

The maintenance of genetic polymorphism under various natural structured viability regimes *vs.* general unrestricted fitness assignments are compared. The selection models considered include a generalized dominance fitness system, a generalized viability model based on allelic activity values, viability matrices based on multilocus activity levels, viability matrices defined by partitioned "resource" or "substrate" variables, and circulant-type viability matrices. A number of examples that support these formulations are discussed. Detailed results on the nature of the genotype frequency equilibrium configurations for the specified viability models are presented. An increased likelihood for a globally stable equilibrium is predicted for the more structured viability models.

**M**ULTIPLE alleles have been documented for many morphological, physiological, serological and electrophoretic markers (*e.g.*, see reviews of FORD 1975; HARRIS 1975; NEVO 1978; BROWN 1979). These include genes responsible for color and shape patterns (*e.g.*, *Cepaea*, *Papilio*), a host of red and white blood typings, (*e.g.*, primates), inversion regions (*Drosophila*), incompatibility and sex determinants (*e.g.*, *Primula*, *Hymenoptera*), segregation distorter genes (*Drosophila*), protein variants (*e.g.*, esterase  $\alpha$ , *Xdh*) and representations of supergenes. These conditions frequently manifest a spectrum of dominance orderings and variable expressivity that may be influenced by environmental effects and specific and/or nonspecific genetic modifiers.

The continued refinements and developments in biochemical separation techniques and sequencing methodology have led to a wider recognition of the existence of multigene families (blocks of very tightly linked genes). These include chorion genes in the silkworm (GOLDSMITH and BASEHOAR 1978), immunoglobulin series (*e.g.*, SEIDMAN *et al.* 1978; TONEGAWA *et al.* 1978), androgen-regulated major urinary proteins in mice (HASTIE, HELD and TOOLE 1979), copies of  $\alpha$ -hemoglobin chains, other repeating units, etc. It is surmised that gene duplication and subsequent partial differentiation tend not to be deleterious and, thereby, they can greatly augment allelic diversity (OHNO 1970; BODMER 1979). Moreover, the discovery of the elaborate gene domain incorporating mRNA transcription of flanking regions, intervening (intron) segments with subsequent

\* Supported in part by Public Health Service Grant 5R01 GM10452-17 and NSF Grant MCS79-24310.

processing of deletions and splicing further supplement the multiplicity and complexity of the allelic spectrum (GILBERT 1978).

When tight linkage and fitness interactions occur together among several loci, the study of these multilocus models effectively reduces to one-locus multiallele models. Then, by invoking the small perturbation principle (KARLIN and MCGREGOR 1972a,b, 1973), the results available for multiallele one-locus selection models can be translated into a number of properties pertinent to multilocus multiallele systems associated with small recombination rates (see also KARLIN 1979).

To further our theoretical understanding of one-locus multiallelic systems, we developed and investigated a series of multiallele selection models founded on dominance relationships among phenotype classes, interactions accommodating biochemical allelic activity levels and symmetries or invariants endowed to heterozygote and homozygote genotype groups. Several results are presented on the nature of polymorphisms for these models.

## Section 2. *The selection model*

The standard one-locus multiallele selection model has the following structure. Consider a population characterized by  $r$  possible alleles  $A_1, A_2, \dots, A_r$  at a given locus. The population frequency vector of the  $r$  alleles for the current generation is  $(x_1, x_2, \dots, x_r)$ . Random mating is assumed with discrete nonoverlapping generations and differential viability selection, where the fitness (viability) matrix is given by  $W = \|w_{ij}\|_{i,j=1}^r$ . Under these conditions, the resulting frequency state in the next generation is given by

$$x'_i = \frac{x_i \sum_{j=1}^r w_{ij} x_j}{\sum_{i,j=1}^r w_{ij} x_i x_j}, \quad i = 1, 2, \dots, r. \quad (1)$$

The denominator  $W(\mathbf{x}) = \sum_{i,j=1}^r w_{ij} x_i x_j$  is the *mean fitness function* at population state  $\mathbf{x}$ , which is proportional to the mean number of progeny persisting to the next generation.

This paper delineates three classes of one-locus multiallele viability selection structures that may possess biological relevance. These are:

- (i) Selection expressing partial or complete dominance relations among the allele types and associated heterozygote classes;
- (ii) Selection as a function of allelic activity level;
- (iii) Selection induced by patterns of multilocus associations.

We first highlight a number of concrete classes of viability matrices before describing results on their stable equilibrium forms.

I. *Generalized dominance fitness model* (fitness correlated with a dominance hierarchy among the possible alleles): The alleles (or haplotypes) are assumed to possess an intrinsic dominance ordering, say,

$$A_1 < A_2 < A_3 < \dots < A_{r-1} < A_r \quad (3)$$

such that allele  $A_j$  is dominant to allele  $A_i$  for  $i < j$ , where

$$\begin{aligned} w_{ij} &= \text{fitness } (A_i A_j) = \alpha_j \text{ for all } i < j \\ w_{jj} &= \text{fitness } (A_j A_j) = a_j, j = 1, 2, \dots, r \end{aligned} \quad (4)$$

so that only the dominant allele determines the fitness expression in a heterozygous state. Note that the homozygote fitness  $a_j$  need not agree with  $\alpha_j$ , thus also allowing a distinction in the dominance relation between heterozygotes and homozygotes, possibly due to pleiotropic effects.

This model of multiple alleles subject to complete or partial dominance relationships is motivated by numerous natural examples. A classical case is the melanic trait in the moth *Biston betularia* involving three main color morphs with a dominance ordering black > grey > white (KETTLEWELL 1973). In the ladybird beetle, *Adalia bipunctata* (CREED 1971a,b), a complete dominance series of at least 12 morphs exists. Shell color for *Cepaea nemoralis* (CAIN and SHEPPARD 1952; FORD 1975) involves a dominance hierarchy, including phenotypes white < pale yellow < yellow < dark yellow < pale pink < pink < dark pink < brown, further exhibiting strong (epistatic) linkage of the brown phenotype with banding pattern. There are relevant dominance structures in melanic feral pigeons (MURTON, WESTWOOD and THEARLE 1973), in melanic Skua (O'DONALD 1979) in colors of the male ruff, *Philomachus pugnax*, and a polar dominance series polymorphism in the guppy, *Poecilia reticulata*. The classic example of *Papilio dardenus* (SHEPPARD 1958; FORD 1975) exhibits a hierarchy of (not complete) dominant alleles in a supergene structure. Only a few combinations are present, mostly those that produce mimics.

Fitness with dominance ordering may relate to the period of maturation in plant populations, mating pattern, competition, predation and other ecological correlates. Coat color in a variety of animals and plants, spot formation and degree of smoothness often fall under simple gene control (or involve a few genes that are not always closely linked) that exhibit a hierarchical dominance structure (e.g., SHEPPARD 1958; SEARLE 1968; O'DONALD 1980; FORD 1975; JONES, LEITH and RAWLINGS 1977).

Some traits relevant to sexual selection, preferential and assortative mating propensities can also be mediated in terms of multiallelic models subject to forms and levels of dominance relations (O'DONALD 1980; KARLIN and O'DONALD 1981). The frequency distribution of phenotypes in relation to their dominance hierarchy can, in some cases, be directly interpreted in terms of a behavioral model for the mechanism of mating choice.

Further, there is a wide scope of dominance relations for biochemical activities, e.g., a single autosomal locus governs lactase production and involves at least three alleles,  $L$ ,  $l_1$  and  $l_2$  with  $L$  dominant to  $l_1$  and  $l_2$ , and  $l_1$  dominant to  $l_2$  (EHRMANN and PARSONS 1976, p. 50). The mode of inheritance of many genetic diseases that are classified as dominant or recessive also frequently entails multiple alleles (McKUSICK 1978). Partial dominance structures abound in serological identifications (e.g., ABO, Duffy and the Rh blood types, Gm series).

II. *A generalized viability model based on allelic activity values*: A determi-

nation of fitness induced by an appropriate (*e.g.*, biochemical or physiological) performance level has an activity  $v_i$  associated with allele  $A_i$ , with the genotype fitness expression

$$w_{ij} = f(v_i, v_j) = f(v_j, v_i) . \tag{5}$$

Generally,  $w_{ij}$  values of the above form may relate to an enzyme kinetic mechanism that is translated into some kind of genetic selection. Without loss of generality, we will stipulate different activity levels,  $v_1, v_2, \dots, v_r$  for each allele, so that allele  $A_i$  and its activity level  $v_i$  are equivalent descriptions.

An important special case of (5) has

$$f(v_i, v_j) = f(|v_i - v_j|) = c_{|i-j|} , \tag{6}$$

so that the fitness of  $A_i A_j$  depends only on the difference of the indices involved. We refer to a fitness matrix of the form (6) as the *distance index viability model*.

One possible situation for a sequence of alleles whose fitnesses depend on the difference of the allele indices is the set of temperature-sensitive monomers indexed by the temperature of peak activity. Suppose: (i) The organism containing the alleles passes through a range of temperatures on a regular basis. (ii) At optimal temperature, the monomer provides sufficient activity to saturate the system. Then, if the horizontal axis is temperature-scaled to reflect the extent of time that the organism spends at each temperature, and the vertical axis describes the activity level for an allele functioning at the indicated temperature, a possible schematization is shown in Figure 1.

In the spirit of model (6), the fitness conferred by this activity performance is specified such that the amount of activity below the saturation level increases with the distance between the optimal temperatures.

As a second example, some kinetic considerations of enzymes suggest the function  $f(\xi, \eta) = \frac{\alpha(\xi + \eta)}{\delta + \xi + \eta}$ ,  $\alpha, \delta > 0$  is appropriate to (5) (*cf.*, GILLESPIE 1978). This is plainly monotone increasing in  $\xi$  and  $\eta$ . When  $f(\xi, \eta)$  is monotone in the same

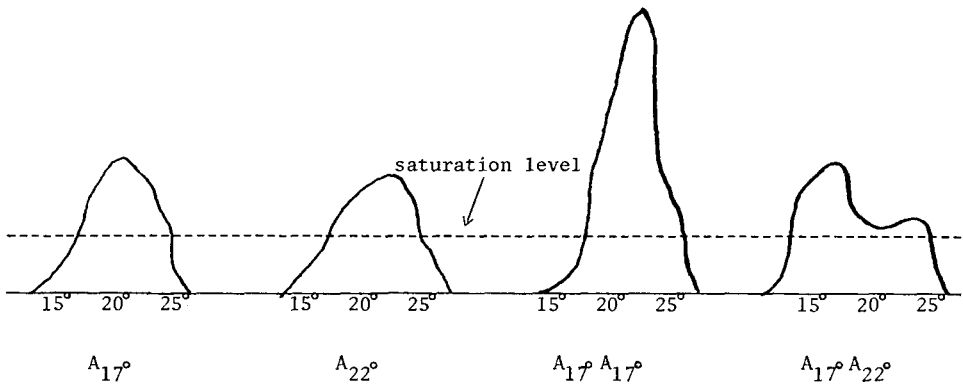


FIGURE 1.—The activity level for alternative alleles or genotypes functioning over certain temperature ranges.

direction for both variables, it is easy to prove that a unique stable fixation evolves under (1) to the allele that maximizes  $f(v_i, v_i)$ .

Other prescriptions of  $f(v_i, v_j)$  can be based on crystallographic, structural, specificity, type of reaction, allosteric properties and inhibitor factors. If  $f(\xi, \eta)$  is not monotone, perhaps with an intermediate optimum or when interaction from different activity levels enhances fitness (e.g., over a range of temperatures, or chemical environments), then the expression of (5) can lead to a stable polymorphism. In this connection, see Results 3 and 5 of Section 5.

III. *Viability models based on multilocus activity values:* A few examples of viability regimes for multilocus systems involving aggregate or compounded allelic effects have been investigated by numerical means. These are proposed as mechanisms relevant to quantitative and polygenic inheritance. We describe a general construction of viability values determined by the activity values of the allele components and the conglomerate genotype.

Suppose that at locus  $k$  the feasible alleles consist of  $\{A_1^{(k)}, \dots, A_{m_k}^{(k)}\}$ . Consider an  $n$ -locus model based on interactions accruing from the allele-haplotype-genotype composition. For each allele, there exists an intrinsic activity value  $v_i^{(k)}$ , e.g., activity level, performance index, reaction rate. Therefore, each gamete  $\xi = (A_{i_1}^{(1)}, A_{i_2}^{(2)}, \dots, A_{i_n}^{(n)})$  has an associated activity vector  $[v_{i_1}^{(1)}, v_{i_2}^{(2)}, \dots, v_{i_n}^{(n)}]$ . We postulate that the fitness of the genotype

$$\mathbf{g} = \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} A_{i_1}^{(1)}, A_{i_2}^{(2)}, \dots, A_{i_n}^{(n)} \\ A_{j_1}^{(1)}, A_{j_2}^{(2)}, \dots, A_{j_n}^{(n)} \end{pmatrix} \quad (7)$$

is determined by an appropriate function of the constituent allele variables denoted by  $f(\xi_1, \xi_2, \dots, \xi_n; \eta_1, \eta_2, \dots, \eta_n)$ . In particular, the fitness of  $\mathbf{g}$  is  $w(\mathbf{g}) = f(v_{i_1}^{(1)}, \dots, v_{i_n}^{(n)}; v_{j_1}^{(1)}, \dots, v_{j_n}^{(n)})$ .

The simplest prescription for  $f$  has

$$w(\mathbf{g}) = f(\xi_1, \dots, \xi_n; \eta_1, \dots, \eta_n) = \phi \left[ \sum_{i=1}^n (\xi_i + \eta_i) \right], \quad (8)$$

where the fitness depends on the aggregate allelic value. The selection is directional when  $\phi$  is monotone increasing in its argument. If  $\phi$  is *unimodal*, the expression of (8) is sometimes called the *optimum* (or intermediate) selection model.

There is a qualitative divergence in the evolutionary (stable equilibrium) realizations for viability regimes of the structure (8) depending on whether  $\phi$  is concave or convex, as depicted in Figure 2.

Other versions can reflect loci effects, dominance relations among gametes or more complex intra- and interlocus interactions. In this perspective, an assignment of interest takes

$$w(\mathbf{g}) = \phi \left[ \sum_{i=1}^n |\xi_i - \eta_i| \right] \quad (9)$$

such that the viability value is a function of the aggregate number of heterozygous loci (KARLIN 1977).

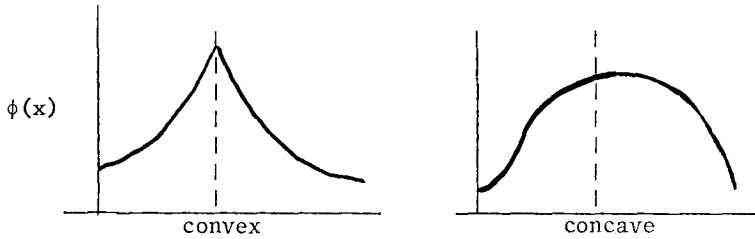


FIGURE 2.—Examples of optimum fitness models.

IV. *Viability matrices defined by partitioned "resource" or "substrate" variables:* An important parameterization used by McARTHUR (e.g., 1971) for the study of ecological competitive systems pertains to the concept of species-specific resource utilization and production functions. In a parallel manner, it is of interest to investigate the equilibrium structure for a viability matrix  $W = \|w_{ij}\|_1^r$  with  $w_{ij}$  represented in the form

$$w_{ij} = [(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})], \quad i, j = 1, \dots, r \quad (10)$$

where  $\{\mathbf{x}^{(i)}\}_1^r$  constitutes a sequence of linearly independent vectors and  $[(\mathbf{z}, \mathbf{w})]$  stands for an appropriate bilinear function in the vectors  $\mathbf{z}$  and  $\mathbf{w}$ . Following (10), the vector  $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_p^{(i)})$  associated with the allele  $A_i$  is specified such that each component contributes to fitness in a prescribed manner. For allele  $A_i$ , the components in  $\mathbf{x}^{(i)}$  may correlate with age, an activity parameter and/or performance indices with respect to chemical, electrical or other environmental conditions.

More specifically, we consider the fitness of the genotype  $A_i A_j$  determined in the manner

$$w_{ij} = w(A_i A_j) = [(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})] = \lambda_1 x_1^{(i)} \cdot x_1^{(j)} - \sum_{\nu=2}^p \lambda_\nu x_\nu^{(i)} x_\nu^{(j)}, \quad (11)$$

where the coefficients  $\lambda_i > 0$  are fixed weights (scalings to common measurement units) independent of the genotype. The first component,  $x_1^{(i)}$ , can be interpreted as providing the optimum contribution to viability for allele  $A_i$ , while the other variables  $x_2^{(i)}, x_3^{(i)}, \dots$  inhibit the natural functioning of allele  $A_i$ . In this context, the variable  $x_1^{(i)}$  corresponds to the natural substrate for the gene (enzyme) activity at appropriate temperature and pH conditions; whereas, the variables  $x_2^{(i)}, \dots, x_p^{(i)}$  can be construed as competitor substrates that detract from the effectiveness of allele  $A_i$ . For example, for a hemoglobin variant  $A_i$ ,  $x_1^{(i)}$  may refer to a proper oxygen pressure, while  $x_2^{(i)}, x_3^{(i)}, \dots$  refer to the inhibitor molecules  $\text{CO}_2$ , DPG and nonoptimal pH ambience. In the model of (11), the allele vectors  $\mathbf{x}^{(i)}$  are measured on a multiplicative scale, but other standardizations can be handled by similar means.

Consider the viability matrix

$$W = \| \| [(\mathbf{x}^{(\nu)}, \mathbf{x}^{(\mu)})] \|_{\nu, \mu=1}^r \quad (12)$$

with the generalized inner product  $[(,)]$  calculated as in (11).

*Result 1. Let  $W$  be defined as in (12). There exists a unique stable equilibrium (not necessarily an internal polymorphism with all possible alleles segregating).*

In contrast, consider for  $[(, )]$  the (usual) inner product  $\langle , \rangle$ , namely

$$[(\mathbf{z}, \mathbf{w})] = \langle \mathbf{z}, \mathbf{w} \rangle = \sum_{i=1}^p z_i w_i a_i, \quad (13)$$

where  $a_i$  are fixed positive weights. The viability matrix (12) now evaluated according to (13) has all its stable equilibrium states of the corresponding recursion (1) among the fixation outcomes and there are at least two such stable states.

V. *Circulant type viability matrices:* Consider

$$W = ||w_{ij}|| = ||w_{i+j}||,$$

where the subscript is reduced modulo  $r$ , that is if  $i + j > r$ , then  $w_{i+j}$  is taken to be  $w_{i+j-r}$ . Now for any set of  $w_i$ , there always exists a central internal equilibrium  $\mathbf{x} = (\frac{1}{r}, \frac{1}{r}, \dots, \frac{1}{r})$  that, except for  $r = 2$ , is always unstable. We can represent the above fitness scheme in the form  $w_{ij} = f(v_i, v_j) = f(v_i + v_j)$  for a periodic function  $f$ . The following set of scenarios may help motivate the existence of periodic optimal phenotypes.

Whenever a quantitative trait is governed by additive contributions from the relevant alleles (*e.g.*, net charge of a multimer enzyme, pigmentation, lethality of venom) and the optimum value is intermediate, there will be an optimum phenotype, but no unique optimal genotype. Moreover, the optimal phenotype may depend on the environment. For example, habitat selection in the sense of TEMPLETON and ROTHMAN (1978) ("seek out best environment") may imply a varied environment consisting of a few distinct microenvironments. Also, there may be several optimal phenotypes with the fitness of other phenotypes decreasing with distance from the optima.

The side groups (*i.e.*, amino acids in a polypeptide chain) may significantly affect the pH of the medium. It may be appropriate to raise or lower the medium pH due to external causes (sources) in order to provide the optimal pH for an enzyme system. Thus, an organism should choose the "external" pH (*e.g.*, food source) that its own proteins will bring closest to the optimal pH.

If it is necessary to stun, but not kill, prey in order to capture it, yet keep it suitable for consumption, the ideal potency of the venom will depend on the hardness of the prey (large prey may require more venom; too little venom permits escape, too much kills the prey). Assuming that a few prey sizes are available, individuals should attack those prey for which their venom potency is most appropriate. There will be as many optimal potencies as prey sizes; lower viabilities accompany those individuals whose venom potencies are farthest from all optima.

Section 3. *Levels of multiallele overdominance.*

It is useful to distinguish different assessments of the strength of a *stable complete* (= internal) *polymorphism*. KARLIN (1981) singled out four levels of

multiallelic polymorphisms and discussed their interrelationships. For our present purposes, we recall two principal forms. The first is, in essence, a tautology.

*Definition 1.* An  $r \times r$  viability matrix  $W = \|w_{ij}\|_r^r$  corresponding to  $r$  alleles at a single locus is said to be simply overdominant if there exists a stable complete polymorphism (i.e., all alleles are segregating).

A stable interior equilibrium is globally attracting (KINGMAN 1961).

*Definition 2.* The viability matrix  $W$  is said to be totally overdominant if every principal submatrix of  $W$  is simply overdominant in the sense of Definition 1.

Thus, with total overdominance in force, any subcollection of alleles can maintain a stable equilibrium restricted to these alleles that is, however, unstable with the introduction of any new allele.

*Robust properties of totally overdominant viability matrices:* Consider an overdominant viability matrix  $W$  in the sense of Definitions 1 or 2. Suppose  $W$  is modified to the new viability matrix

$$W^* = W - D, \quad (14)$$

where  $D$  is a positive diagonal matrix with  $(d_1, d_2, \dots, d_r)$  down the diagonal and  $0 \leq d_i \leq w_{ii}$ . After this perturbation, the fitness array of  $W^*$  retains the identical heterozygote fitness values as for  $W$ ; whereas, the homozygote fitness values may be diminished. We may inquire whether  $W^*$  is overdominant, at least to the same extent as  $W$ . There is the intuitive sense that overdominance is promulgated owing to some combination of "heterozygote superior fitness," so that with a reduction of the homozygote fitness values we may expect that the contribution of the heterozygotes and their interactions will be accented, enhancing the establishment of a stable polymorphism.

A specialization of the foregoing problem is to compare the overdominance endowments for  $W$  and  $\tilde{W} = W - \beta I$ ,  $I =$  identity matrix, where  $\tilde{W}$  is obtained from  $W$  in reducing the homozygote fitnesses by the same amount,  $\beta$ .

We know of classes of examples for which  $W$  is simply overdominant, while  $W^*$  constructed as in (14) and even  $\tilde{W} = W - \beta I$  for appropriate  $\beta$  is not simply overdominant. Indeed, the following three-allele viability matrix is simply overdominant, but ceases to be overdominant after some decrease of all its homozygote fitnesses. Consider

$$W = \begin{pmatrix} \alpha - \varepsilon^3 & \alpha & \alpha + \varepsilon^3 \\ \alpha & \gamma + \varepsilon & \gamma \\ \alpha + \varepsilon^3 & \gamma & \varepsilon \end{pmatrix}, \quad (15)$$

where  $\gamma < \alpha < 2\gamma$  and  $\varepsilon > 0$  is sufficiently small. It can be directly verified that  $W$  is simply overdominant, while for  $\tilde{W}_\lambda = W - \lambda I$ ,  $\lambda = \varepsilon/2$ , no polymorphism exists. For all  $0 \leq \lambda \leq \varepsilon$ , a unique globally attracting equilibrium prevails. But, for  $\lambda = \varepsilon/2$ , the unique stable equilibrium of the viability matrix  $\tilde{W}_{\varepsilon/2}$  involves only alleles  $A_1$  and  $A_2$ , while for  $\lambda = 0$  and  $\lambda = \varepsilon$ , a complete polymorphism is realized.

It can be proven that, if  $W$  is totally overdominant, then  $W^*$  of (14) is again totally overdominant. (A hierarchy of levels of overdominance is compared in KARLIN 1981).



Section 4. *Equilibrium structure for the generalized dominance fitness model.*

The relevance of a fitness regime of the structure (4) was discussed in Section 2. In this section, we inquire as to the nature of its stable equilibrium configurations. The necessary and sufficient conditions for a polymorphism for the general case of the selection regime (4) are not known, but a workable sufficient condition is described in the next result.

*Result 2. For the general dominance fitness model (4), a sufficient condition for a stable polymorphism is the fulfillment of the inequalities*

$$\alpha_k > \max \left[ a_k, \frac{\alpha_{k-1}}{2} + \frac{\alpha_{k-2}}{2^2} + \dots + \frac{\alpha_2}{2^{k-2}} + \max \left( \frac{a_1}{2^{k-2}}, \frac{a_2}{2^{k-2}}, \frac{a_3}{2^{k-3}}, \dots, \frac{a_{k-1}}{2} \right) \right],$$

$$k = 3, \dots, r . \tag{16}$$

*In the presence of these inequalities, the fitness matrix (4) is totally overdominant in the sense of Definition 2.*

Observe that when  $\alpha_i$  are nondecreasing, and  $\alpha_i > a_i, i=1, \dots, r$ , then (16) holds. The relations of (16) are satisfied even where  $\alpha_k$  are decreasing, but not too rapidly.

A stable polymorphism for a dominance fitness regime is preserved when reducing the fitness of any homozygote, *i.e.*, if  $W$  of the form (4) admits a stable polymorphism, then  $W^* = W - D$  as in (14) maintains a stable polymorphism for all diagonal  $D$  [compare to example (15)].

The generalized ordered dominance selection regime can be construed as a multiallelic version of disruptive selection to the extent that no two co-existing stable equilibria can involve  $A_l$  as a common allele. More generally, no two stable equilibria co-exist segregating  $A_k$  and  $A_l, l > k$ .

*Random dominance viability matrices:* We inquire as to the chance for a polymorphism when the parameters

$$\{a_1, a_2, \dots, a_r, \alpha_1, \alpha_2, \dots, \alpha_r\} \tag{17}$$

for  $r$  alleles, inducing the fitness matrix (4), are all independently uniformly distributed on  $[0,1]$ . The viability matrix constructed in this way can be construed as a random dominance viability matrix. For this class of matrices, we can establish the theoretical bounds:

$$\left(\frac{1}{2}\right)^r < \text{Probability \{existence of a polymorphism\}} \leq \left(\frac{3}{4}\right)^{r-1} \tag{18}$$

and

$$\left(\frac{1}{4}\right)^r < \text{Probability \{stable polymorphism\}} \leq \left(\frac{1}{2}\right)^{r-1} . \tag{19}$$

In specializing  $\alpha_1 = \alpha_2 = \dots = \alpha_r = \alpha$ , ( $\alpha$  uniformly distributed on  $[0,1]$ ) we obtain the asymptotic relation: Probability of a stable polymorphism  $\sim \frac{c}{r}$  for some positive constant  $c$ .

The orders of magnitude of (18) and (19) contrast sharply with the chance for a stable polymorphism in a *general* random viability matrix, *i.e.*, where *all*

fitnesses of  $W = \|w_{ij}\|$  with  $r$  alleles are *independent* uniformly distributed random values, in which case the probability of a stable polymorphism is miniscule (as  $r \rightarrow \infty$ ) of the order  $\exp[-\frac{r^2}{2} \log r]$ .

*Some numerical results on the existence of a complete polymorphism for the dominance viability model:* We considered four general classes of dominance viability matrices. Class I (referred to as random dominance-diagonals and columns unordered) matrices of structure (4) are constructed with  $a_i$  and  $\alpha_i$  selected independently uniformly distributed over the unit interval  $[0,1]$ . Class II matrices (diagonals ordered and columns unordered) first select the  $\{a_i\}$  at random as in Class I, but these values are then arranged in increasing order (we refer to this determination "random increasing" on the diagonal), while  $\{\alpha_i\}$  the non-diagonal column elements are ascertained at random and unordered as before. Class III has  $\{\alpha_i\}$  generated random monotone increasing and  $\{a_i\}$  random unordered. Class IV matrices are randomly generated with both  $\{a_i\}$  and  $\{\alpha_i\}$  ordered (random-ordered). We generated 10,000 random matrices in each class.

The tabulations in Table 1 show that when the dominance property across the allele combination is concordant with enhanced fitness such that  $\{a_i\}$  and  $\{\alpha_i\}$  are "randomly increasing," the opportunities for a stable internal polymorphism are considerably increased compared to the general random dominance viability matrices, *i.e.*, where the  $\{a_i\}$  and  $\{\alpha_i\}$  are random and unordered.

TABLE 1

*The nature of polymorphism of random dominance viability matrices\**

Size of matrix $r$	(Class I) nothing ordered			(Class III) columns ordered		
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
3	0.4731	0.2246	0.1018	0.4747	0.2483	0.1385
4	0.3344	0.0831	0.0269	0.3276	0.1146	0.0566
5	0.2381	0.0274	0.0062	0.2235	0.0546	0.0250
6	0.1728	0.0091	0.0017	0.1455	0.0290	0.0131
7	0.1176	0.0025	0.0003	0.1008	0.0132	0.0059
8	0.0882	0.0010	0.0001	0.0675	0.0062	0.0024
9	0.0647	0.0003	0.0000	0.0453	0.0027	0.0012
	(Class II) diagonal ordered			(Class IV) diagonal and column ordered		
3	0.5267	0.2341	0.1437	0.3697	0.3447	0.2056
4	0.4349	0.1004	0.0562	0.2203	0.2537	0.1386
5	0.3686	0.0405	0.0222	0.1653	0.1994	0.1113
6	0.3009	0.0160	0.0068	0.1276	0.1691	0.0896
7	0.2537	0.0064	0.0030	0.1096	0.1462	0.0783
8	0.2085	0.0022	0.0010	0.0936	0.1293	0.0689
9	0.1647	0.0011	0.0003	0.0793	0.1127	0.0581

\* We ascertained for each fitness matrix the extent to which the following three properties hold: (i) When  $W\mathbf{x} = \mathbf{I}$  admits a positive solution  $\mathbf{x} > 0$ , *i.e.*, whether  $W$  possesses an internal equilibrium (a polymorphism), stable or unstable. (ii) Whether the eigenvalues of  $W$  consist of one positive and  $n-1$  negative values. (iii) Where properties (i) and (ii) hold together, implicating a stable polymorphism.

Section 5. *Overdominance properties of the distance index viability model (6).*

We discuss various aspects of the overdominance properties for the “distance” viability regime, where the allelic array  $\{A_1, A_2, \dots, A_r\}$  possess an intrinsic ordering  $A_1 \sim A_2 \sim A_3 \sim \dots \sim A_r$  such that the genotype fitness of  $(A_i A_j) = c_{|i-j|}$  depends only on the index distance defining its constituent alleles. In Section 2, we suggested some models of biochemical activity implicating a fitness scheme of the form (6). In ascertaining the existence and degree of overdominance for some natural cases of (6), we commence with the following result.

*Result 3. (i) Let  $c_k$  be a strictly concave sequence connoting*

$$c_{k+1} + c_{k-1} < 2c_k \text{ for every } k = 1, 2, \dots, r-2 . \tag{20}$$

*Then the only stable allelic configurations involve a contiguous group of alleles*

$$(A_k, A_{k+1}, \dots, A_l) \text{ for some } k \text{ and some } l \text{ (} 1 \leq k \leq l \leq r \text{)} .$$

*(ii) If  $c_k$  is also increasing, then a globally stable polymorphism exists and the fitness matrix  $w_{ij} = c_{|i-j|}$  is totally overdominant (see Definition 2).*

A schematization of the  $\{c_k\}$  conforming to the hypothesis (20), is shown in Figure 3.

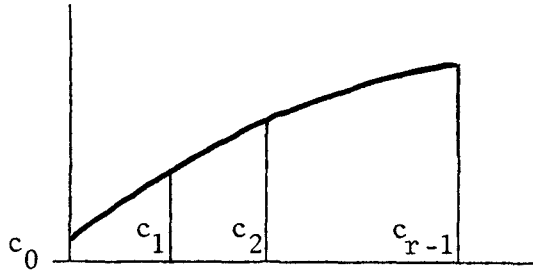


FIGURE 3.—A concave increasing fitness function for the distance index viability model.

In contrast to the concave case, the stable configurations markedly differ under the stipulation that  $\{c_k\}$  is convex.

$$c_{k-1} + c_{k+1} > 2c_k \text{ for all } k = 1, 2, \dots, r-2 . \tag{21}$$

The shape of  $\{c_k\}$  is shown in Figure 4.

The following result describes the possible stable realizations in the convex case.

*Result 4. (i) If  $c_k$  constitutes a strictly convex sequence, then any stable equilibrium is either a fixation state or involves only two alleles.*

*(ii) Where the  $\{c_k\}$  are strictly convex and strictly increasing, then the allele array  $\{A_1, A_n\}$  is the unique stable equilibrium.*

A far reaching extension of Results 3 and 4 is as follows.

*Result 5. Suppose that  $f(\xi, \eta)$  is increasing and concave in each variable on each side of the diagonal, i.e.,*

$$\frac{\partial f}{\partial \eta}(\xi, \eta) > 0, \frac{\partial^2 f}{\partial \eta^2}(\xi, \eta) < 0 \text{ for } \eta > \xi \text{ and for } \xi > \eta , \tag{22}$$

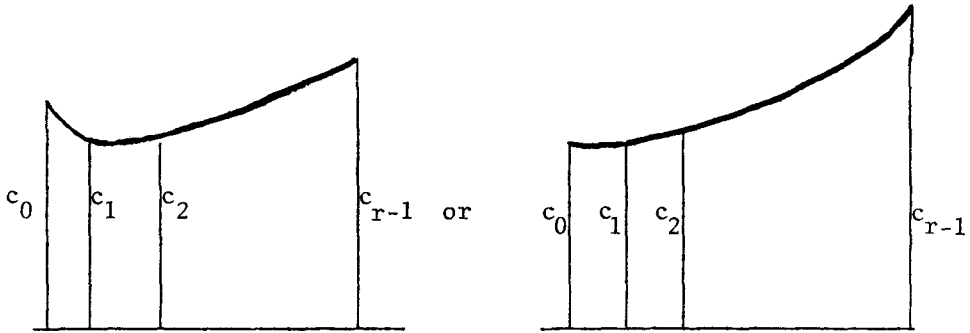


FIGURE 4.—A convex fitness function for the distance index viability model.

then the viability matrix (5) in this case admits a complete stable polymorphism. On the other hand, where

$$\frac{\partial f}{\partial \eta}(\xi, \eta) > 0, \text{ but } \frac{\partial^2 f}{\partial \eta^2}(\xi, \eta) > 0 \text{ for } \eta > \xi, \tag{23}$$

then the extreme allele pairing  $\{A_1, A_n\}$  is uniquely stable.

The implications of Result 5 for biochemically based selection regimes will be presented elsewhere.

Some numerical results on the existence of a complete polymorphism for the distance index viability model: Table 2 indicates the relevant stable realizations when the fitness parameters  $\{c_i\}$  of (6) are irregular or “random” for the corresponding  $r$  allele models,  $r \leq 10$ . It is of interest to assess the degree of polymorphism when the  $\{c_k\}$  of (6) vary in an irregular manner. To this end, for the allelic activity model and more particularly for the index distance fitness model

$$W = \begin{pmatrix} c_0 & c_1 & c_2 & \dots & c_{r-1} \\ c_1 & c_0 & c_1 & \dots & c_{r-2} \\ \vdots & & & \ddots & \vdots \\ \vdots & & & & \vdots \\ c_{r-1} & c_{r-2} & \dots & \dots & c_0 \end{pmatrix} \tag{24}$$

we generated 10,000 random matrices by the format:  $c_i$  independently chosen uniformly in  $(0,1)$ . The matrix obtained in this way is referred to as a random distance viability matrix  $W$ . Another 10,000 random selections of  $\{c_i\}$  were generated and the resulting  $c_i$  arranged in increasing order to give  $\{c_i^*\}$ . The corresponding fitness matrices are then referred to as “random monotone” index distance viability matrices.

Table 2 summarizes the simulation results with respect to properties (i) to (iii) delimited in the legend of Table 1.

The possibilities for the existence of a polymorphism for the random unordered distance index viability matrix contrast sharply with those for the random monotone distance index viability matrix. The probability of an unstable internal

TABLE 2

For random viability matrices of type (24), the matrix order is  $r$ . 10,000 random matrices  $M$  and 10,000 random monotone  $W$  were generated (as described in Table 1)

Matrix size $r$	$W$			$M$		
	Proportion of cases of internal equilibrium property (i)	Proportion of cases with property (ii)	Proportion of cases of stable polymorphism	Proportion of cases of property (i)	Proportion of cases with property (ii)	Proportion of cases of stable polymorphism
4	0.5152	0.5746	0.3464	0.7726	0.1904	0.1556
5	0.2088	0.1568	0.0796	0.6530	0.0694	0.0322
6	0.1914	0.1206	0.0636	0.6212	0.0424	0.0204
7	0.0816	0.0554	0.0192	0.5696	0.0194	0.0080
8	0.0784	0.0448	0.0156	0.5532	0.0140	0.0074
9	0.0370	0.0240	0.0064	0.5076	0.0036	0.0016
10	0.0308	0.0206	0.0056	0.4772	0.0030	0.0012

equilibrium (and, therefore, the existence of multiple stable boundary equilibria) is surprisingly high, realized with a probability close to 0.5. A theoretical lower bound for stable fixation states is of order exceeding  $1/r$ . The chance for complete stable polymorphism is of smaller order, but nontrivial even with 8 to 10 alleles of the order 0.5%. For the random monotone model, the chance of a stable polymorphism exceeds by four-fold that for the unordered model.

It is useful for comparative purposes to record numerical results on the existence of polymorphism for general random viability matrices, *i.e.*, each  $w_{ij}$ ,  $i \leq j$  is selected independently uniformly distributed on  $[0,1]$ .

By theoretical considerations we know that asymptotically for columns (ii) and (iii) of Table 3 the rate of approach to zero is

$$\exp\left[-\frac{r^2}{2} \log r\right],$$

which is incredibly small for large  $n$  (even  $n \geq 6$ ). The chance for a polymor-

TABLE 3

Random viability matrices based on 10,000 samples

Matrix order $r$	Frequency of internal equilibrium; property (i)	Frequency of sample satisfying property (ii)	Frequency of stable polymorphism; property (iii)
2	0.6731	0.571	0.338
3	0.4148	0.175	0.041
4	0.2160	0.020	0.000
5	0.1341	0.002	0.000
6	0.0660	0.006	0.000
7	0.0324	0.000	0.000
8	0.0189	0.000	0.000
9	0.0086	0.000	0.000
10	0.0034	0.000	0.000

phism goes to zero at a rate slower than  $(1/2)^n$  for the dominance fitness model (4) and the activity allele fitness forms (4) and (5).

A comparison of Tables 1–3 underscores the miniscule opportunities for polymorphism under viability selection governed by general random viability matrices with  $n \geq 5$  possible alleles. The results are poignantly of a different order for the dominance viability model and the activity allele selection models.

### Section 6. Discussion

Several simulation studies for one-locus multiallele “random” viability systems conducted by LEWONTIN, GINZBERG and TULJAPURKAR (1978) predicted miniscule probabilities (*cf.*, Table 3) for the existence of a stable polymorphism involving a large number of alleles maintained by selection interactions. However, real selection systems are not likely to be random, but are expected to possess some intrinsic structure. The heterozygotes generally fall into natural classes, and it is in this perspective that one might better assess the likelihood and nature of polymorphisms.

In appraising the contribution of heterozygotes in maintaining polymorphism, it is essential to classify types of heterozygosity. This paper highlights three natural classes of viability regimes that are based on allelic dominance hierarchies and functional allele activity levels. In the context of a multiallele, multilocus model, we can order the extent of heterozygosity through criteria depending on the numbers, locations and composition at the various loci. The aggregate heterozygosity determination of KARLIN (1977, see also 1979) is a prototype case.

The objectives of this study have been two-fold: (i) To describe natural classes of viability expression connected with biochemical activities or morphological and behavioral patterns; and (ii) To ascertain the nature of the equilibrium realizations for these highly structured selection matrices.

The dominance selection models based on a series of alleles of increasing order of dominance may reflect degrees of fitness mediated by ecological conditions (*e.g.*, relations to predation, dessication). The relevance and scope of this kind of viability expression is discussed in Section 2. Dominance hierarchies occur in connection with preferential mating patterns (O'DONALD 1980), in sex determination mechanisms, in biochemical pathways (*e.g.*, EHRMANN and PARSONS 1976, Chap. 5) and elsewhere.

Result 2 shows that the opportunities for polymorphism are greatly facilitated when a heterozygote carrying a larger dominance component implies greater fitness. On the other hand, when there is no concordance between fitness and the degree of dominance, the resulting stable allelic array generally involves only a few alleles (*cf.*, KARLIN and FELDMAN 1981).

For irregular patterns of the dominance selection matrix, *i.e.*, where  $a_i$  and  $\alpha_i$  are random unordered or random monotone, the possibilities are more varied, although not as unlikely as with general random viability matrices (compare Tables 1 and 3).

A general tenet of polymorphism concerns the the influence of recognized patterns *vs.* random fitness assignments. For mechanisms of fitness reflecting bio-

chemical, behavioral or physiological structure (*e.g.*, dominance ordering, allelic activity influence), a specified association (monotonicity, convexity, concavity) underlying the biological relationship tends to predict more regularity in the stable equilibrium possibilities, while random fitness values create more multiplicity in the evolutionary outcomes, even when superimposed on a structured model.

The allelic activity selection model (5) tends to exhibit more polymorphism, at least when fitness increases for genotypes involving more distant allelic combinations, provided the marginal increase in fitness is reduced for alleles of more diverse optimal activity (Result 5). This result is especially germane for the distance index selection form (6).

On the other hand, if the marginal increase in fitness is strongest for alleles corresponding to indices furthest apart (*i.e.*, their optimum activity occurs at the extremes of the environmental range), then the unique stable equilibrium segregates only the extreme allele types  $A_l$  and  $A_r$  (Result 4).

For a viability matrix structured *via* "partitioned" variables (see paragraph IV of Section 2) where a single resource variable predominates in its contribution to fitness, as in the case of (11)–(12), there exists a unique globally stable equilibrium (not necessarily a polymorphism).

To sum up, an increased likelihood for a globally stable (unique) equilibrium is predicted for the structured fitness models (as exemplified by the dominance ordering, distance index allelic and resource partitioning fitness models) considered in this study, and also for the multilocus forms (aggregate heterozygosity model KARLIN 1977) and generalized nonepistasis (KARLIN 1979; KARLIN and LIBERMAN 1979). Accordingly, if the observed allele frequency data exhibit a reasonably consistent common set over different population areas and epochs, the contingencies for a structured selection mechanism may be relevant. On the other hand, where the allele frequency observations vary significantly in space or time with few segregating alleles in any particular sample, an explanation of the observed variability based on fitness interactions is unlikely. Other forces, such as migration, population structure, mating pattern, genetic frequency and/or ecological density factors and strong randomizing recombination interactions, may be important.

The emphasis of this study is the involvement of an inherent structure of biologically relevant parameters in evaluating selection regimes and in classifying and comparing genotypes. The study of random viability matrices is arbitrary, independent of clear biological mechanisms and is likely to lead (as it does) to "pathological" results. An important step in advancing ecological theory was taken by McARTHUR when he introduced concepts of resource utilization functions as a way to parameterize interspecies community competition. The dominance viability structure, activity allele form, distance index model, "resource" partitioning selection expression proposed in this paper and other well-conceived models derived from multilocus interactions (KARLIN and AVNI 1981) may serve as a point of departure for more insightful genotype fitness assessments.

## LITERATURE CITED

- BODMER, W., 1979 Gene clusters and the HLA system in human genetics: Possibilities and realities. *CIBA Foundation Series 66*. Elsevier, North Holland: 205-229.
- BROWN, A. H. D., 1979 Enzyme polymorphism in plant populations. *Theoret. Pop. Biol.* **15**: 1-44.
- CAIN, A. S. and P. R. SHEPPARD, 1952 The effects of natural selection on body colour on the land snail, *Cepaea nemoralis*. *Heredity* **6**: 217-223.
- CREED, E. R., 1971a Melanism in the two-spot ladybird, *Adalia bipunctata*. In Great Britain, in *Ecological Genetics and Evolution* (ed. CREED). Blackwell, Oxford: 134-151. —, 1971b Industrial melanism in the two-spot ladybird and smoke abatement. *Evolution* **25**: 290-293.
- EHRMANN, L. and P. PARSONS, 1976 *The Genetics of Behavior*. Sinauer Assoc. Sunderland, Mass.
- FORD, E. B. 1975 *Ecological Genetics*. John Wiley & Sons, New York.
- GILBERT, W. 1978 Why genes in pieces? *Nature* **271**: 507.
- GILLESPIE, J. 1978 A general model to account for enzyme variation in natural populations. V. The SAS-CFF model. *Theoret. Pop. Biol.* **14**: 1-46.
- GINZBERG, L. R., 1979 Why are heterozygotes often superior in fitness? *Theoret. Pop. Biol.* **15**: 264-267.
- GOLDSMITH, M. R. and C. BASEHOAR, 1978 Organization of the chorion genes of *Bombyx mori*, a multigene family. I. Evidence for linkage to the second chromosome. *Genetics* **90**: 291-310.
- HARRIS, H., 1975 *The Principles of Human Biochemical Genetics*. American Elsevier, North Holland.
- HASTIE, N. D., W. A. HELD and J. J. TOOLE, 1979 Multiple genes coding for the androgen-regulated major urinary proteins of mouse. *Cell* **17**: 449-456.
- JONES, J. S., B. H. LEITH and P. RAWLINGS, 1977 Polymorphism in *Cepaea*. *Ann. Rev. Ecol. and System.* **8**: 109-143.
- KARLIN, S., 1977 Selection with many loci and possible relations to quantitative genetics. pp. 207-226. In: *Proceedings of the International Conference on Quantitative Genetics*. Edited by E. POLLAK, O. KEMPTHORNE and T. B. BAILEY. Iowa State University Press, Ames.
- , 1978 Comparison of positive assortative mating and sexual selection models. *Theoret. Pop. Biol.* **14**: 281-312. —, 1979 Principles of polymorphism and epistasis for multilocus systems. *Proc. Natl. Acad. Sci. U. S.* **75**: 541-546. —, 1981 Some theoretical concepts and results for one-locus multiallele selection models. *Theoret. Pop. Biol.* (in press).
- KARLIN, S. and H. AVNI, 1981 Analysis of central equilibria in multilocus systems: A generalized symmetric viability regimes. *Theoret. Pop. Biol.* (in press).
- KARLIN, S. and M. W. FELDMAN, 1981 A theoretical and numerical assessment of genetic variability. *Genetics* **97**: 475-493.
- KARLIN, S. and U. LIBERMAN, 1979 Central equilibria in multilocus systems. I. Nonepistatic selection. *Genetics* **91**: 777-798.
- KARLIN, S. and J. L. MCGREGOR, 1972a Application of method of small parameters to multi-niche population genetic models. *Theoret. Pop. Biol.* **3**: 186-209. —, 1972b Polymorphisms for genetics and ecological systems with weak coupling. *Theoret. Pop. Biol.* **3**: 210-238. —, 1973 Equilibria for genetic systems with weak interactions. pp. 79-87. In: *Proc. Sixth Berkeley Symp. Math. Stat. & Prob.* Edited by L. M. LECAM, J. NEYMAN and E. L. SCOTT, Univ. Calif. Press, Berkeley.
- KARLIN, S. and P. O'DONALD, 1981 Sexual selection at a multiallele locus with complete or partial dominance. *Heredity* (in press).



- KETTLEWELL, H. B. D., 1973 *The Evolution of Melanism*. Clarendon Press, Oxford.
- KINGMAN, J. F. C., 1961 A mathematical problem in population genetics. *Proc. Cambridge Phil. Soc.* **57**: 547-582.
- LEWONTIN, R. C., L. R. GINZBERG and S. D. TULJAPURKAR, 1978 Heterosis as an explanation for large amounts of genic polymorphism. *Genetics* **88**: 149-169.
- McARTHUR, R. H., 1971 Species packing and competitive equilibrium for many species. *Theoret. Pop. Biol.* **1**: 1-11.
- McKUSICK, V. A., 1978 *Mendelian Inheritance in Man*. Johns Hopkins Press, Baltimore, Md.
- MURTON, R. K., N. J. WESTWOOD and R. J. P. THEARLE, 1973 Polymorphism and the evolution of a continuous breeding season in the pigeon *Columbia livia*. *J. Reprod. Fert. Suppl.* **19**: 563-577.
- NEVO, E., 1978 Genetic variation in natural populations: Patterns and theory. *Theoret. Pop. Biol.* **13**: 121-177.
- O'DONALD, P., 1980 *Genetic Models of Sexual Selection*. Cambridge University Press, London.
- OHNO, S., 1970 *Evolution by Gene Duplication*. Springer-Verlag, Berlin.
- SEARLE, A. G., 1968 *Comparative Genetics of Coat Color in Mammals*. Academic Press, New York.
- SEIDMAN, J. G., A. LEDER, M. N. B. NORMAN and P. LEDER, 1978 Antibody diversity. *Science* **202**: 11-17.
- SHEPPARD, P. M., 1958 *Natural Selection and Heredity*. Hutchinson and Co. ———, 1961 Some contributions to population genetics resulting from the study of the *Lepidoptera*. *Adv. Genet.* **10**: 165-216.
- TEMPLETON, A. R. and R. D. ROTHMAN, 1978 Evolution in fine grained environments: I. Environmental runs and the evolution of homeostasis. *Theoret. Pop. Biol.* **13**: 340-355.
- TONEGAWA, S., A. M. MAXAM, R. TIZARD, O. BERNARD and W. GILBERT, 1978 Sequence of a mouse germ line gene for a variable region of an immunoglobulin light chain. *Proc. Natl. Acad. Sci. U. S.* **75**: 1485-1489.

Corresponding editor: B. S. WEIR