

SOME MATHEMATICAL MODELS OF POPULATION GENETICS

SAMUEL KARLIN, The Weizmann Institute of Science, Israel, and
Stanford University, California

Introduction. Theoretical population genetics and mathematical genetics is the study of temporal and spatial changes of frequencies of types (e.g., genes, genotypes, gametes, etc.) in populations subject to various ecological and genetic influences.

Two general opposite tendencies operate on natural population: (i) propensity for adaptability and persistence of specific types favorable to a given environment, and (ii) necessity for populations to maintain potential for variation to cope with situations of changing environments.

The use of mathematics in studying genetic systems is as old as the subject of genetics itself. From the rediscovery of Mendel's work at the beginning of this century it did not take long for the Hardy-Weinberg law (1908)* on the constancy of gene frequency over time to be enunciated. Between 1915 and 1950 mathematical genetics was pioneered and dominated by the names of R. A. Fisher, S. Wright, and J. B. S. Haldane.

The challenge to understand the role of such genetic and ecological factors as mutation and migration rates, the varied manifestations of natural selection, the effects of population behavior and mating patterns, the relevance of recombination, etc., motivated these men to formulate a vast hierarchy of mathematical models describing many facets of population genetic phenomena. Relatively few of these models have as yet yielded to complete analysis.

Haldane, in his famous series of papers in the Proceedings of the Cambridge Philosophical Society in the 1920's, set forth a variety of simple mathematical analyses concerned with the way natural selection might be supposed to act. In particular, he indicated how evolutionary forces such as viability selection, mutation, migration, and sex-linkage could be quantified and brought into these models.

Samuel Karlin received his Princeton Ph.D. under S. Bochner. He has held positions at Cal Tech, Princeton, Stanford, and the Weizmann Institute of Science. At various times he held the Proctor Fellowship, Bateman Fellowship, Wald Memorial Lectureship, Guggenheim Fellowship, and National Science Senior Fellowship. He is a Fellow of the International Statistical Institute, the Institute of Mathematical Statistics, an elected member of the U.S. National Academy of Sciences, and the American Academy of Arts and Sciences.

Professor Karlin has been most productive in a variety of fields. He has supervised 35 Doctoral students, many now recognized scientists, has written over 125 research papers and the following books: *Studies in the Mathematical Theory of Inventory and Production* (with K. Arrow and H. Scarf, Stanford Univ. Press, 1958); *Mathematical Methods and Theory in Games, Programming, Economics*, Volume I: *Matrix Games, Programming and Mathematical Economics*, (Addison-Wesley, 1959); *Mathematical Methods and Theory in Games, Programming, Economics*, Volume II: *The Theory of Infinite Games* (Addison-Wesley, 1959); *A First Course in Stochastic Processes* (Academic Press, 1966); *Chebyshev Systems: With Applications in Analysis and Statistics*, (with W. J. Studden, Interscience, 1966); and *Total Positivity, Volume I*, (Stanford Univ. Press, May 1968). *Editor*.

*This is the G. H. Hardy of mathematical fame.

Fisher and Wright were also involved in the elaboration of these theories. Wright further established that in small populations, evolutionary theory should take account of the sampling effects involved in producing one generation from the previous. He called this effect "random drift". This aspect of population genetics has had significant mathematical consequences especially in stimulating Feller's investigations into boundary theory of diffusion processes on the line.

Again it was Wright and Fisher who pioneered the theory of systems of mating between relatives, such as used by animal and plant breeders. The result was the theory of inbreeding which entails intriguing algebraic and analytic structures much of which is not well understood. Statistical theory probably owes its origin to R. A. Fisher's attempts to design and analyze experiments whose purposes were most often to solve problems in genetics.

The objective of this paper is to acquaint the mathematics student with several classical mathematical genetic models. Attention is mainly given to the formulation of the models accompanied by brief analyses and appropriate references. Some interpretations and implications of the results with reference to evolutionary theory are appended. On occasion relevant unsettled mathematical problems are noted.

It should be underscored that the array of models to be discussed is a very slight representation of the vast number formulated and partly dealt with by geneticists over the past half century and very recently by some mathematicians. We have attempted to highlight several important genetic factors and concepts by presenting models involving different mating patterns, selective forces, migration and mutation pressures, the recombination mechanism, etc. Many types of mathematical genetic models have been omitted in this expository article for lack of space. For example, we avoided entirely the enticing and important excursion into stochastic genetic models. (The interested reader can consult Crow and Kimura [7], Chapters 10–12, for an introduction to this part of mathematical genetics, and references cited therein.) Models based on statistical genetics have also been left out. The general theory of inbreeding systems is given scant attention (see Karlin [16] and [17] for a fuller treatment of this subject). The extensive and important literature of genetic traits determined by several loci is only briefly touched on in Section 8. (For a review on this current very active topic, consult Kojima and Lewontin [27], see also Karlin and Feldman [19], and Karlin [20].)

In closing the introduction, we indicate the organization of the paper. Section I reviews succinctly some of the basic terminology and relevant genetic mechanism. Section II covers a few basic random mating models exhibiting selection balance. Sections III and IV highlight two important situations of non-random mating. Section III is specially devoted to an exposition of some models involving positive assortative mating while Section IV exposes the phenomena of incompatibility mechanisms in mating patterns. These include cases of self-sterility and sex determination. Section V presents briefly the classical model of mutation selection balance

for two alleles (alternative gene forms). Section VI is concerned with the very useful method and concept of identity by descent. Section VII discusses some models of the evolution of a population with an infinite number of possible types. Section VIII introduces the simplest two locus selection model.

I. PERTINENT GENETIC PRELIMINARIES

It is unfortunate but necessary to learn a minimum of the terminology and mechanisms of population genetic systems. **Chromosomes**—usually found in the nucleus—mostly govern the inheritable characteristics of an organism. Chromosomes may occur singly (the **haploid case**) as in some fungi, in pairs (the **diploid case**), as in mammals, or in larger groups (**triploid**, **tetraploid**, in general **polyploid**) as in many plants. The associated pairs, triplets, etc., of chromosomes are called **homologous**. **Locus** is the position at which a gene (a sort of unit of the chromosome) occurs on a chromosome. **Alleles** are alternate gene forms at a given locus. **Genotypes** are the various possible combinations of alleles at corresponding loci on homologous chromosomes. In the diploid case if the alleles are A and a , the genotypes are AA , Aa , and aa .

The populations to be considered here, unless specified otherwise, contain diploid individuals. We concentrate our attention, for the most part, on characters determined by one or two loci, on a given pair of chromosomes. We usually assume that two alternative genes (alleles) may occur at each locus. Consider the case of two loci, where the alleles A and a are possible at the first locus and alleles B and b at the second locus. A typical one of the ten possible genotypes (see listing immediately below) could be written AB/ab . The symbol AB/ab signifies that AB sit on one chromosome A at the first locus, B at the second locus and ab are situated on the second chromosome. The ten genotypes are explicitly

$$\frac{AB}{AB}, \frac{AB}{Ab}, \frac{Ab}{Ab}, \frac{AB}{aB}, \frac{AB}{ab}, \frac{Ab}{aB}, \frac{Ab}{ab}, \frac{aB}{aB}, \frac{aB}{ab}, \frac{ab}{ab}.$$

The physical manifestation of the genotype is called the **phenotype**. If the genotype Aa has the phenotype of the AA individual, then A is said to be a **dominant** gene and a is called **recessive** to A .

We shall assume that an offspring is formed by the donation of a **gamete** (one of each pair of homologous chromosomes) from each of two parents. In the case of one locus, each parent, depending on its genotype, may donate either A or a to form a **zygote** (fertilized egg) having **genotype** AA , Aa or aa . Individuals with genotype AA or aa are **homozygotes**; Aa is a **heterozygote**. For two loci, the donated gametes can be of four kinds, AB , Ab , aB or ab and ten zygotes are possible as listed previously. Generations are taken to be non-overlapping.

Considering the one locus case, we are primarily interested in tracing the frequencies of the three genotypes over time. Assume that the population size is very large,

effectively infinite. Let $u_n, v_n,$ and w_n be the frequencies of AA, Aa and aa , respectively, in the n th generation. In order to follow the vector (u_n, v_n, w_n) as n increases we must describe the mating system, i.e., the way mating pairs are to be selected.

One of the most widely studied systems of mating is **random-mating**. This occurs when any one individual of one sex is equally likely to mate with any one of the opposite sex. Thus, in the one locus case above, the mating $AA \times AA$ would occur with frequency u_n^2 at the n th generation. From this mating only AA offspring result. However, from the mating $Aa \times Aa$, AA, Aa and aa offspring will be produced with probabilities $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ respectively. This equally likely case of segregation is called **Mendelian segregation**.

In an infinite population, not subject to any outside influences, and in which random mating takes place the **Hardy-Weinberg Law** holds. This states that, if in a given generation the frequencies of the A and a gene are p and $q = 1 - p$ respectively, then in all subsequent generations the frequencies remain the same. Verification of this, and the fact that **random mating** is equivalent to **random union of gametes** can be found in most textbooks in population genetics, e.g., Kempthorne [24] Chapter 2.

There are a number of factors (apart from the mating system) which act on populations to influence the path of evolution. Perhaps the three most familiar are **mutation, migration** and **selection**. The first two are self-explanatory. They can be visualized as providing the raw material for selection to mould. We are interested here in three forms of selection. The first is selection through variation in **viability**, i.e. the genotypes differ in their chances of survival to reproduce. The second is through **fertility** variations, i.e., different pairs of parents, on account of the genotype of both parents may produce differing numbers of offspring. **Segregation distortion** from the usual Mendelian ratios is another type of selection. These can be considered particular manifestations of what was called by Darwin (1859) "fitness" in his qualitative description of the different abilities of individuals to survive and contribute to the next generation. Of course, the mating system itself can be another factor affecting evolution. Selection attributable to the mating system is commonly referred to as **sexual selection** to distinguish it from **natural selection**. We shall be partly interested in the mathematical description of the interactions between selection and various mating systems.

Selection is incorporated mathematically in the following ways: If the mating type $AA \times Aa$ is assumed to have fertility f then the offspring are produced in the proportions $\frac{1}{2}f AA, \frac{1}{2}f Aa$. Similar definitions hold for the other matings. The offspring are assumed to have viabilities in the ratio $\sigma_1 : \sigma_2 : \sigma_3$ means that each of the genotypes AA, Aa and aa survives to parenthood with relative chance $\sigma_1 : \sigma_2 : \sigma_3$ respectively.

The frequencies u_n, v_n, w_n of AA, Aa, aa in the n th generation can now be expressed in terms of those in the $(n - 1)$ -th generation using some transformation T which will in general be non-linear.

Another phenomenon of considerable importance to the maintenance of genetic variability will be mentioned before we describe the models in detail. Recombination may occur in the case of two loci when at the first locus we have alleles A and a and at the second B and b , and the two loci are not independent so far as gamete donation is concerned. An individual heterozygous at both loci can produce four types of gametes. For example, an individual of genotype AB/ab can produce gametes of type AB and ab and also gametes of the type Ab , aB . When all four are produced in equal numbers the loci are called **unlinked**. The AB and ab gametes are called **parental** while the Ab and aB are called **recombinant**. If the loci are linked there will be an excess of parental gametes over recombinants. It is found that the parental types AB , ab are produced with equal frequencies $\frac{1}{2}(1-r)$ and the recombinant types with equal frequencies $\frac{1}{2}r$ where the number r , $0 < r \leq 1$, is called the **recombination fraction**. For the physical explanation of the phenomenon and more details on its importance the reader should consult any genetics text book.

This has been a necessarily brief introduction to the terminology we shall use. No attempt has been made to elaborate the biological scope of the terms introduced. For this the reader should consult such texts as Stern [32], Crow and Kimura [7], and Cavalli and Bodmer [6].

II. SOME ONE LOCUS SELECTION MODELS

1. One sex viability model. Consider a population with two possible alleles A , a at a specified locus undergoing random mating and subject to viability selection where the genotypes AA , Aa and aa which survive to maturity (i.e., to reproduce) are in the ratio $\sigma_1 : \sigma_2 : \sigma_3$ respectively.

If the frequencies of A and a in the current generation are p and $q = 1 - p$ respectively, then random union of genes (which is equivalent to random mating) produces the genotypes AA , Aa , aa in the frequencies p^2 , $2pq$, q^2 respectively. The relative frequencies of the three genotypes at maturity taking account of selection effects are then

$$\begin{array}{ccc} AA & Aa & aa \\ \sigma_1 p^2 & \sigma_2 2pq & \sigma_3 q^2. \end{array}$$

With Mendelian segregation (see Section I) the frequency p' and q' of A and a respectively, in the next generation have relative magnitudes $p' \sim p^2 \sigma_1 + \sigma_2 pq$, $q' \sim \sigma_3 q^2 + \sigma_2 pq$. To convert these to *bona fide* frequencies we normalize by dividing by the sum yielding the transformation equation

$$(2.1) \quad p' = \frac{p^2 \sigma_1 + \sigma_2 pq}{p^2 \sigma_1 + 2pq \sigma_2 + q^2 \sigma_3} \stackrel{\text{def}}{=} f(p).$$

The denominator is commonly called the **mean fitness function**, written $W(p)$, and enjoys the remarkable property that $W(f(p)) \geq W(p)$ with equality holding iff $p = f(p)$.

The evolution of the process is obtained by iterating the transformation law (2.1). The following classical results are readily established (cf. Figure 1 below) independent of the initial p ($0 < p < 1$).

$$(2.2) \quad \lim_{n \rightarrow \infty} f_{(n)}(p) = \lim_{n \rightarrow \infty} f(f_{(n-1)}(p)) = 1 \quad (= 0) \quad \text{when } \sigma_1 \geq \sigma_2 > \sigma_3 \quad (\sigma_3 \geq \sigma_2 > \sigma_1),$$

$$(2.3) \quad \lim_{n \rightarrow \infty} p_n = \hat{p} = \frac{\sigma_2 - \sigma_3}{2\sigma_2 - \sigma_1 - \sigma_3} \quad \text{when } \sigma_2 > \max(\sigma_1, \sigma_3).$$

In the case $\min(\sigma_1, \sigma_3) > \sigma_2$ then

$$(2.4) \quad \lim_{n \rightarrow \infty} p_n = 1 \quad \text{for } p > \hat{p}, = 0 \quad \text{for } p < \hat{p}.$$

Figure 1 shows what happens to $f_{(n)}(p)$ in graphical form. The rigorous details are easily supplied.

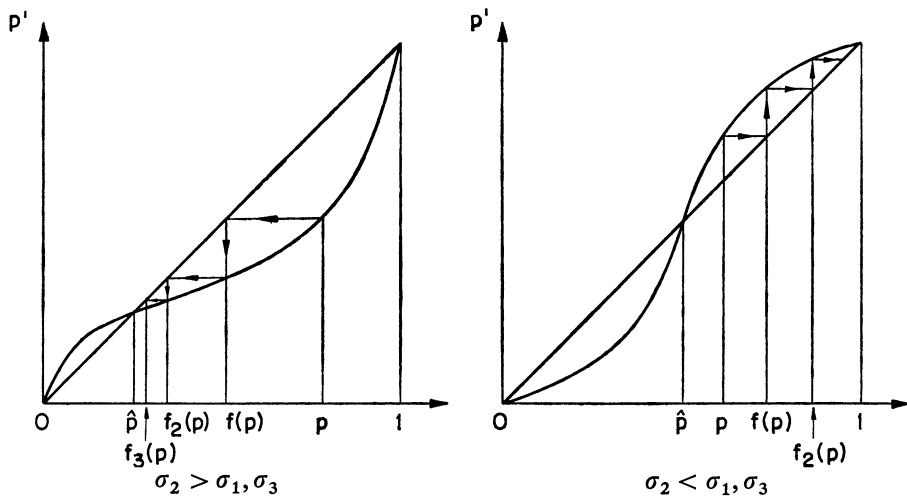


Fig. 1.

The equilibrium \hat{p} is of great importance biologically because it entails the simultaneous existence at an equilibrium involving all genotypes. Thus when the heterozygote is the most fit of the three genotypes a stable **polymorphism** (with all forms) will be maintained. The model of heterozygote advantage (also called the principle of overdominance) has been central to the development of theories on the existence of genetic variability.

2. Two sex viability models with two alleles. (This model was most recently dealt with by Bodmer [2], see also Karlin [20].) Consider next a population divided into males and females, mating randomly subject to viability selection where the fitness coefficients may differ between the sexes. The array in Table 1 describes the process (assuming male and female offspring are produced with equal probability).

Sex	Male			Female		
Gamete		<i>A</i>	<i>a</i>		<i>A</i>	<i>a</i>
Frequency		<i>p</i>	<i>q</i>		<i>P</i>	<i>Q</i>
Genotype	<i>AA</i>	<i>Aa</i>	<i>aa</i>	<i>AA</i>	<i>Aa</i>	<i>aa</i>
Fitness coefficients (viabilities)	σ	1	τ	<i>s</i>	1	<i>t</i>
Relative frequencies after random mating and selection	σpP	$pQ + qP$	τqQ	spP	$pQ + qP$	tqQ

TABLE 1

With Mendelian segregation we obtain for the gene frequencies in the next generation the transformation equations

$$(2.5) \quad p' = \frac{\sigma pP + \frac{1}{2}(pQ + qP)}{\sigma pP + pQ + qP + \tau qQ}, \quad P' = \frac{spP + \frac{1}{2}(pQ + qP)}{spP + pQ + qP + tqQ},$$

where the denominators are the required normalization factors (cf. Model 1).

In the case at hand it is more convenient to express the changes of gene frequencies over successive generations in terms of the equivalent pair of variables $x = p/q, y = P/Q, 0 \leq x, y \leq \infty$. We obtain

$$(2.6) \quad x' = \frac{\sigma xy + \frac{1}{2}(x + y)}{\tau + \frac{1}{2}(x + y)} = f(x, y), \quad y' = \frac{sxy + \frac{1}{2}(x + y)}{t + \frac{1}{2}(x + y)} = g(x, y).$$

Write T for the mapping defined in (2.6). The fixed point $\mathbf{0} = (0, 0)$ corresponds to the pure population of only *aa* genotypes and $\infty = (\infty, \infty)$ represents the pure population of *AA* genotypes.

We wish to ascertain the character of all equilibria of T and their domains of attraction. The analysis of T and its iterates is much facilitated by exploiting the feature that T is monotone, i.e., where $z = (x, y) \leq \tilde{z} = (\tilde{x}, \tilde{y})$ holds (the ordering signifies the inequality for each coordinate). Then we have

$$(2.6a) \quad Tz \leq T\tilde{z} \text{ with strict inequality in each coordinate unless } z = \tilde{z}.$$

The stability nature of any equilibrium is customarily ascertained by analysis of the local linear approximation to the non-linear mapping T in the neighborhood of the fixed point. More specifically, we examine the matrix transformation given by the gradient matrix

$$\| \partial T \| = \left\| \begin{array}{cc} \frac{\partial f}{\partial x} & \frac{\partial g}{\partial x} \\ \frac{\partial f}{\partial y} & \frac{\partial g}{\partial y} \end{array} \right\|$$

evaluated at the fixed point $\hat{z} = (\hat{x}, \hat{y})$.

If both eigenvalues of $\partial T|_{\hat{z}}$ are in magnitude less than 1, then \hat{z} is locally stable. If at least one eigenvalue in magnitude exceeds 1, then *usually* \hat{z} is unstable.

The conditions for local stability of the pure equilibrium $\mathbf{0}$ and ∞ are readily determined by invoking the local linear analysis just described. We get

$$(2.7) \quad \begin{aligned} \mathbf{0} \text{ (fixation in the } a \text{ gene) is stable iff } & \frac{1}{2\tau} + \frac{1}{2t} \leq 1 \\ \infty \text{ (pure } AA \text{ population) is stable iff } & \frac{1}{2\sigma} + \frac{1}{2s} \leq 1. \end{aligned}$$

Algebraic manipulations of the equations (2.6) show that for general positive fitness parameters (σ, τ, s, t) there exist at most 3 fixed points where both coordinates are positive and finite. These are, of course, polymorphic equilibria.

There are five qualitative cases of interest:

(i) *The same homozygote is most fit in both sexes; e.g., $\sigma < 1 < \tau$ and $s < 1 < t$ hold.* Under these conditions adding the relations in (2.6) using obvious inequalities produces

$$(2.8) \quad x' + y' < 2 \frac{xy + \frac{1}{2}(x+y)}{1 + \frac{1}{2}(x+y)}.$$

Since $4xy \leq (x+y)^2$ we see that $x' + y' < x + y$. It follows that $x^{(n)} + y^{(n)}$ decreases in n and its limit is necessarily zero indicating that $\mathbf{0}$ is globally stable.

(ii) *AA is most fit in one sex and aa is most fit in the other sex.* We illustrate with the special symmetric situation $\tau = s$ and $\sigma = t$, $\sigma > 1 > s$. In this case there always exists a unique internal equilibrium $z^* = (\xi_0, 1/\xi_0)$ where ξ_0 is the unique positive solution of the equation

$$\xi^3 + \xi^2(2s - 1) - \xi(2\sigma - 1) - 1 = 0.$$

Analysis reveals that z^* is stable iff the equilibrium point $\mathbf{0}$ (and simultaneously, owing to symmetry, the point ∞) is unstable, i.e., iff $1/2\sigma + 1/2s > 1$.

In the general case of (ii) it can be proved that there can be at most one polymorphic stable equilibrium.

(iii) *Both homozygotes selectively inferior to the heterozygote in one sex but superior in the other sex, i.e.,*

$$(2.9) \quad 1 > \sigma, \tau, \quad 1 < s, t.$$

We illustrate with the symmetric case $\sigma = \tau$ and $s = t$. Then $z^* = \mathbf{1} = (1, 1)$ is a fixed point of the mapping T and is locally stable iff $\sigma s < 1$. If we determine the values of $\sigma = \tau$, $s = t$ satisfying

$$\frac{2}{\frac{1}{s} + \frac{1}{\sigma}} < 1 < \sqrt{\sigma s}$$

which is certainly possible (owing to the harmonic mean, geometric mean inequality) we find that 0 , 1 and ∞ are all unstable. Exploiting the monotonic nature of T , we deduce the existence of two other stable polymorphic equilibrium. Here, then, is a case of the existence of two stable polymorphisms. This phenomenon does not arise in the corresponding one sex model.

(iv) *Heterozygote advantage in each sex* ($1 > \sigma, \tau, s, t$). The expected intuitive result of a unique stable polymorphism is indeed realized.

(v) *Heterozygote advantage in one sex and directed selection in the other sex*, i.e., $1 > \sigma, \tau, s > 1 > t$. In this case, elementary analysis of the transformation (2.6) yields the existence of at most two stable equilibria and when two exist one has to be a boundary equilibrium.

To sum up, the main conclusions are as follows:

There can exist at most two stable equilibria including the possibility that both are polymorphisms. In contrast, the one sex selection model allows at most one stable polymorphism.

3. Two sex multi allele viability model. Suppose there exist $r \geq 3$ alleles A_1, A_2, \dots, A_r possible at the given locus and of course, $r(r+1)/2$ possible genotypes $A_i A_j$. Let the frequencies of the genes in the male population be q_1, q_2, \dots, q_r and p_1, p_2, \dots, p_r for the female population. The viability fitness matrix for females is designated as $F = \|f_{ij}\|_{i,j=1}^r$ where f_{ij} measures the relative average number of the $A_i A_j$ genotype that survive to maturity. The viability fitness matrix for males is denoted by $M = \|m_{ij}\|$.

Stipulating random union of genes and Mendelian segregation quite analogous to (2.5), we obtain for the gene frequencies of the next generation the recursion relations

$$(2.10) \quad p'_i = \frac{\frac{1}{2} \left[p_i \sum_{j=1}^r f_{ij} q_j + q_i \sum_{j=1}^r f_{ij} p_j \right]}{\sum_{i,j=1}^r p_i f_{ij} q_j}$$

$$q'_i = \frac{\frac{1}{2} \left[p_i \sum_{j=1}^r m_{ij} q_j + q_i \sum_{j=1}^r m_{ij} p_j \right]}{\sum_{i,j=1}^r p_i m_{ij} q_j}, \quad i = 1, 2, \dots, r.$$

Call this non-linear transformation of $2r$ variables ($2r - 2$ independent ones) T as before. Results concerning the evolution of this process, i.e., the behavior of the iterates of T and characterizing their limit points, are of primary interest. It would be of much interest to determine precise bounds for the number of stable polymorphisms possible in this r allele selection model. Theorems from algebraic geometry produce upper bounds (but excessive ones) for the number of admissible equilibrium points. We refer to Karlin [20] for a treatment of several non-elementary

cases of (2.10). A rather complete treatment of the special symmetric case $M = F$ is available, e.g., see Kingman [25].

4. Selection model for multi allelic sex linked character. (This model was first formulated by Haldane, see also Cannings [4], [5].)

Consider a character determined by a locus on the sex chromosome with r alleles possible. Suppose the female sex is the homogametic one, the XX chromosome.

The female genotypes assume the form A_iA_j , $i, j = 1, \dots, r$ but the male genotypes take the form A_iY since the Y chromosome carries no complement of the gene.

The fitness coefficients corresponding to females are displayed by the matrix $F = \|f_{ij}\|$ and for males by the vector $\mathbf{m} = (m_1, m_2, \dots, m_r)$. Thus m_i measures the relative fitness of the male genotype A_iY and f_{jk} of the female genotype A_jA_k . Under random mating and selection, the relative number of female offspring of type A_jA_k which survive to maturity is $\frac{1}{2}(p_jq_k + q_jp_k)f_{jk}$ for $j \neq k$ and $p_jq_jf_{jj}$ for $j = k$. For males of genotype A_i , the relative frequency of maturing male offspring is $q_i m_i$, since the male parent always contributes the Y chromosome. With Mendelian segregation, we get the transformation law

$$(2.11) \quad p'_i = \frac{\frac{1}{2} \left[p_i \sum_{j=1}^r f_{ij}q_j + q_i \sum_{j=1}^r f_{ij}p_j \right]}{\sum_{i,j=1}^r p_i f_{ij}q_j}, \quad q'_i = \frac{m_i p_i}{\sum_{i=1}^r m_i p_i}.$$

In general, there exists at most one polymorphic equilibrium \hat{p}, \hat{q} where \hat{p} is calculated by normalizing (so that the sum of components is 1) the positive solution of

$$(2.12) \quad (FI_m + I_m F)\hat{p} = \mathbf{1}.$$

(I_m is the diagonal matrix with m_1, m_2, \dots, m_r down the diagonal and $\mathbf{1}$ is the vector with all components of value 1.) And

$$\hat{q} = \gamma I_m \hat{p} \quad \text{with} \quad \gamma^{-1} = \sum_{i=1}^r m_i \hat{p}_i.$$

Stability conditions of such a polymorphic solution can be determined.

We specialize now to the case $r = 2$. Then it is more convenient to work in terms of the variables

$$x = \frac{p_1}{p_2} \quad \text{and} \quad y = \frac{q_1}{q_2},$$

so that $0 \leq x, y \leq \infty$. The equivalent recursion equations reduce to

$$(2.13) \quad x' = \frac{sx y + \frac{1}{2}(x + y)}{\sigma + \frac{1}{2}(x + y)}, \quad y' = mx,$$

where $s = f_{11}/f_{12}$, $\sigma = f_{22}/f_{12}$ and $m = m_1/m_2$. Designate the transformation (2.13) as $T(x, y) = (x', y')$. It is readily verified that T is a strictly monotonic mapping

(cf. (2.6a)). Exploiting this fact we easily establish by applying a local linear approximation, the existence of a positive pair of numbers (a, b) such that for $\varepsilon > 0$ and sufficiently small $T(\varepsilon a, \varepsilon b) < (\varepsilon a, \varepsilon b)$ iff $m < 2\sigma - 1$. It follows that the fixed point $\mathbf{0} = (0, 0)$ (corresponding to a pure A_2A_2 population) is locally stable iff $m \leq 2\sigma - 1$. In a similar manner, we find that $\infty = (\infty, \infty)$ is locally stable iff $2s - 1 \geq 1/m$. For the case where $2\sigma - 1 < m$ and $2s - 1 < 1/m$ there exists a unique polymorphic globally stable equilibrium (x^*, y^*) with

$$(2.14) \quad x^* = \frac{2\sigma - 1 - m}{(2s - 1)m - 1}, \quad y^* = mx^*.$$

Global stability of (x^*, y^*) results by virtue of the following facts: (i) T is monotone and exactly one interior equilibrium exists, (ii) $T(\varepsilon a, \varepsilon b) > (\varepsilon a, \varepsilon b)$, and (iii) $T(N\tilde{a}, N\tilde{b}) < (N\tilde{a}, N\tilde{b})$ hold for ε small enough and N large enough respectively. (Here a, b are specified to satisfy $m > a/b > 2\sigma - 1$ and \tilde{a}, \tilde{b} to satisfy $1/m > \tilde{a}/\tilde{b} > 2s - 1$.)

In the case that $m < 2\sigma - 1$ and $2s - 1 < 1/m$ simultaneously hold then $\mathbf{0}$ and ∞ are both locally stable and possess domains of attraction whose boundary is an algebraic curve containing the point (x^*, y^*) defined in (2.14).

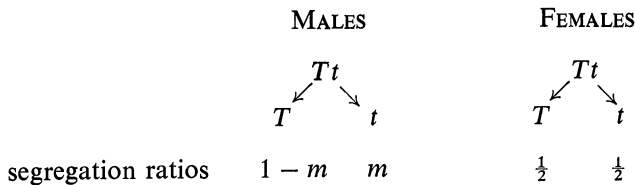
5. Segregation distortion and viability selection balance for the t -locus in house mice. (This model was set up by Lewontin [28].)

The t -locus codes for certain enzyme function essentially involves two alleles labeled T and t . The presence of the t -alleles affects males and females differently. (Morphologically the t allele reveals a shortened tail—hence the name.) With reference to selection, we have

	MALE			FEMALE		
	TT	Tt	tt	TT	Tt	tt
Fitnesses	$1 - s,$	$1,$	0	$1 - s$	1	$1 - \sigma$

($0 \leq s < 1, 0 \leq \sigma \leq 1$). Note that recessive males (tt genotypes) suffer total lethality.

The main difference is revealed in the segregation ratios for the heterozygote in the two sexes. Explicitly



and m is about .90 in the actual example.

Denote by q_1 (q_2) the frequency of T (t) in the males and p_1 (p_2) correspondingly for females. Set $u = q_2/q_1, v = p_2/p_1$. Taking account of the viability selection, segregation bias and assuming random mating, we deduce the recursion relations

$$(2.15) \quad u' = \frac{(1-\sigma)uv + \frac{1}{2}(u+v)}{1-s + \frac{1}{2}(u+v)}, \quad v' = \frac{m(u+v)}{1-s + (1-m)(u+v)}.$$

The transformation (2.15) is strictly *monotonic* as in the earlier two allele models. Direct examination reveals that the transformation Γ in (2.15) satisfies

$$(2.16) \quad \Gamma(\varepsilon a, \varepsilon b) > (\varepsilon a, \varepsilon b)$$

for $\varepsilon > 0$ small enough and appropriate $a, b > 0$ iff $2(1-s)(\frac{1}{2}-m-s) < 0$ or $m+s > \frac{1}{2}$ and the opposite order relation holds in (2.16) when $m+s < \frac{1}{2}$.

It follows that $\mathbf{0} = (0, 0)$ is locally stable iff $m+s \leq \frac{1}{2}$. We now prove global stability for this case. To this end form

$$\begin{aligned} u' + v' &\leq \frac{(1-\sigma)uv + \frac{1}{2}(u+v)}{1-s + \frac{1}{2}(u+v)} + \frac{m(u+v)}{1-s + (1-m)(u+v)} \\ &\leq \frac{(1-\sigma)uv + (m + \frac{1}{2})(u+v)}{1-s + \frac{1}{2}(u+v)}. \end{aligned}$$

But $uv \leq ((u+v)/2)^2$ implies

$$(2.17) \quad z' = u' + v' \leq \frac{(m + \frac{1}{2})z + (1-\sigma)z^2/4}{1-s + \frac{1}{2}z} = h(z).$$

Direct verification shows that h is non-decreasing and $h(z) \leq z$ for $z \geq 0$ with equality iff $z = 0$. Iteration of (2.17) is therefore permissible leading to

$$z^{(n)} \leq h_n(z) = h(h_{n-1}(z)), \quad n = 1, 2, 3, \dots$$

But a simple geometric argument proves $h_n(z) \rightarrow 0$ as $n \rightarrow \infty$ for any initial $z > 0$ and therefore $z^{(n)} \rightarrow 0$. Thus $\mathbf{0} = (0, 0)$ is globally stable as claimed.

The fixed points of (2.15) are obtained as the solutions of the equations

$$(2.18) \quad u = \frac{(1-s-m)v + (1-m)v^2}{m - (1-m)v},$$

where v satisfies $R(v) = A_3v^3 + A_2v^2 + A_1v + A_0 = 0$, where

$$(2.19) \quad \begin{aligned} A_0 &= m(1-s)(s + m - \frac{1}{2}), \\ A_1 &= m(1-\sigma)(1-s-m) + (1-s)[-2m(1-m) + s(m - \frac{1}{2})], \\ A_2 &= (1-m)[(1-\sigma)(2m+s-1) - (1-s)(m - \frac{1}{2})], \\ A_3 &= -(1-\sigma)(1-m)^2. \end{aligned}$$

When $s+m > \frac{1}{2}$ we have $R(0) > 0$ while

$$R\left(\frac{m}{1-m}\right) = \frac{-m}{2} \frac{(1-s)^2}{1-m} < 0.$$

Therefore, in this case there exists v^* ($0 < v^* < m/(1 - m)$) satisfying $R(v^*) = 0$ and u^* determined from (2.18) is > 0 . The point (u^*, v^*) is of course an equilibrium of (2.15). With a little effort, using σ as a parameter ($1 \geq \sigma \geq 0$) it can be proved there exists for $m + s > \frac{1}{2}$ a unique solution v^* of $R(v) = 0$ fulfilling the inequalities $0 < v^* < m/(1 - m)$ and therefore in this case exactly one interior polymorphism occurs. Since $T(Na, Nb) < (Na, Nb)$ prevails for N large enough and appropriate $a > 0, b > 0$, we infer, by virtue of the monotonic nature of T the limit relation $\lim_{n \rightarrow \infty} T^n(u, v) = (u^*, v^*)$ from any initial $(u, v) > 0$.

6. Another model of segregation distortion. We close this section by citing a one-locus two allele segregation distortion model considered by Haldane [14]. There are no fertility differences in the mating types or viability selection differences. There are two alleles A_1 and A_2 where the frequencies of A_1A_1, A_1A_2 and A_2A_2 are x, y and z respectively. The array in Table 2 describes the segregation ratios depending on two parameters.

Mating	Offspring ratios			Mating Frequency
	A_1A_1	A_1A_2	A_2A_2	
$A_1A_1 \times A_1A_1$	1	0	0	x^2
$A_1A_1 \times A_1A_2$	λ	$1 - \lambda$	0	$2xy$
$A_1A_1 \times A_2A_2$	0	1	0	$2xz$
$A_1A_2 \times A_1A_2$	$\frac{\lambda(1 - \mu)}{2 - \lambda - \mu}$	$\frac{2(1 - \lambda)(1 - \mu)}{2 - \lambda - \mu}$	$\frac{\mu(1 - \lambda)}{2 - \lambda - \mu}$	y^2
$A_1A_2 \times A_2A_2$	0	$1 - \mu$	μ	$2yz$
$A_2A_2 \times A_2A_2$	0	0	1	z^2

TABLE 2.

Viability effects only operate in the segregation process. Each mating has output 1. It is straightforward to derive the recursion relations connecting genotype frequencies over two successive generations. We get

$$\begin{aligned}
 x' &= x^2 + 2\lambda xy + \frac{\lambda(1 - \mu)}{2 - \lambda - \mu} y^2 \\
 (2.20) \quad y' &= 2(1 - \lambda)xy + 2xz + \frac{2(1 - \lambda)(1 - \mu)}{2 - \lambda - \mu} y^2 + 2(1 - \mu)yz \\
 z' &= z^2 + 2\mu yz + \frac{\mu(1 - \lambda)}{2 - \lambda - \mu} y^2.
 \end{aligned}$$

All equilibria can be determined in general, and for some special cases, viz., $\lambda = \mu$, $\lambda = 1 - \mu$, $\lambda = 0$ or 1 , the full convergence behavior can be analysed.

Thus, when $\mu = 0$, $x^{(n)} \rightarrow 1$ rapidly.

When $\lambda + \mu = 1$ and $\lambda > \frac{1}{2}$, again we find $x^{(n)} \rightarrow 1$.

For $\lambda = \mu$ and $\lambda < \frac{1}{2}$, then it can be proved that

$$x^{(n)}; z^n \rightarrow \frac{1 - \sqrt{1 - 2\lambda(1 - 2\lambda)}}{2(1 - 2\lambda)}.$$

The following can be readily checked. Assume by symmetry ($0 < \mu \leq \lambda < 1$) then:

- (i) For $0 < \mu \leq \lambda < \frac{1}{2}$, there exists a unique locally stable polymorphism.
- (ii) For $0 < \mu < \frac{1}{2} < \lambda < 1$, there exists no internal equilibrium. It can be proved that fixation in the A_1 allele occurs.
- (iii) If $\frac{1}{2} < \mu \leq \lambda < 1$, there exists a unique internal non-stable equilibrium.

The global convergence behavior of (2.20) for arbitrary parameters λ , μ is in general unsettled.

III. SOME MODELS OF POSITIVE ASSORTATIVE MATING

Consider a two-allele (A and a) single locus population displaying certain preferences in mating behavior. We consider here the case where the preference is exercised by one of the sexes, say the female sex, (this covers most situations of insect and mammal populations). (References and more detailed discussion of the models and related models of this section can be found in Scudo and Karlin [30] and Karlin and Scudo [18].)

1. A model of assortative mating. Assume that A is dominant to a so that phenotypically AA and Aa are alike. The degree of partial assortative mating in the phenotypes is measured by two parameters: α ($0 \leq \alpha \leq 1$) will be the fraction of dominant females preferring to mate with their own kind and β ($0 \leq \beta \leq 1$) that of recessive females preferring their own kind. Thus a fraction, $1 - \alpha$, of A (of AA or Aa) females mate indifferently, i.e., at random. We assume all females are fertilized (i.e., find a suitable mate). This happens if the males are sufficiently abundant and the same male may participate in many matings. Consider the genotypes AA , Aa , aa (A dominant) with the frequencies u , v and w respectively in the female population.

When the prohibitions of assortative mating are operating, it is obligate that each mate of an aa individual is of the same genotype so that the frequency of the $aa \times aa$ mating type is w . Therefore the frequency of the matings of the dominant phenotypes is $1 - w = u + v$. Among the matings of dominants the frequency of occurrence of the $AA \times AA$ mating type is u^2 and its frequency of occurrence considering all admissible matings is then $u^2/(1 - w)$. The frequencies of the mating types are listed in Table 3.

Mating Type	Frequencies	
	Of Assorting Types	Random Mating
$AA \times AA$	$\alpha u^2 / (u + v)$	$(1 - \alpha) u^2$
$AA \times Aa$	$2\alpha uv / (u + v)$	$2(1 - \alpha) uv$
$AA \times aa$		$(2 - \alpha - \beta) uw$
$Aa \times Aa$	$\alpha v^2 / (u + v)$	$(1 - \alpha) v^2$
$Aa \times aa$		$(2 - \alpha - \beta) vw$
$aa \times aa$	βw	$(1 - \beta) w^2$

TABLE 3

The corresponding recurrence relations connecting genotype frequencies over successive generations in accordance with Mendelian segregation laws become

$$\begin{aligned}
 u' &= \left(\frac{\alpha}{u + v} + (1 - \alpha) \right) (u + \frac{1}{2}v)^2 \\
 (3.1) \quad v' &= \frac{\alpha uv}{2(u + v)} + \frac{1}{2}v + (1 - \alpha)u(\frac{1}{2}v + w) + (1 - \beta)w(u + \frac{1}{2}v), \\
 w' &= \beta w + \frac{\alpha v^2}{4(u + v)} + (1 - \alpha)\frac{1}{2}v(\frac{1}{2}v + w) + (1 - \beta)w(\frac{1}{2}v + w).
 \end{aligned}$$

Introducing the A gene frequency, $p = u + \frac{1}{2}v$, and for the next generation, $p' = u' + \frac{1}{2}v'$ and, letting p_n denote the frequency of the gene A in the n th generation, we derive, from (3.1), the relationship

$$(3.2) \quad p' = p[1 + \frac{1}{2}(\alpha - \beta)w].$$

The following inferences can now be made:

(i) For $\alpha > \beta$, p_n increases to 1, the pure homozygous AA state. The rate of convergence is algebraic.

(ii) For $\alpha < \beta$, the population ultimately fixes in the pure homozygous aa state and convergence occurs with an asymptotic factor of decrease per generation $\lambda = 1 + \frac{1}{2}(\alpha - \beta)$.

When $\alpha = \beta$ it is readily checked that $p^{(n)} = p^{(0)}$ for all n . Then v' simplifies to

$$v' = \frac{vp\alpha}{p + \frac{1}{2}v} + (1 - \alpha)2pq = f(v), \quad (q = (1 - p)),$$

where p is the constant gene frequency. Thus $f(v)$ is a linear fractional transformation and therefore the n th generation frequencies $v_n = f_n(v_0) = f(f_{n-1}(v_0))$ can be explicitly evaluated. Indeed, we have

$$\frac{v_n - \gamma_1}{v_n - \gamma_2} = K^n \left(\frac{v_0 - \gamma_1}{v_0 - \gamma_2} \right),$$

where γ_1 and γ_2 are the fixed points of $f(v) = v$ and

$$K = \frac{\gamma_2}{\gamma_1} \left[\frac{2(1 - \alpha)pq - \gamma_1}{2(1 - \alpha)pq - \gamma_2} \right].$$

Because $f(v)$ is concave increasing, we deduce $v_n \rightarrow \gamma_1$. For the case $\alpha = 1$ we obtain $v_n = 2pv_0/(nv_0 + 2p)$ so that $v_n \rightarrow 0$ at an algebraic rate.

2. Model of assortative mating with permanent bonding. In the formulation of the previous model it was tacitly assumed that there was no set order in which the types of mating (random or assortative) took place. The factor of timing of mating for assorting and random mating individuals may be important, and could affect the accessibility and availability of proper mates.

Two simple contrasting assumptions can be made to study the effect of assortment on the timing of pair bonding depending on whether assorting females mate prior to the nonassorting ones, or after. Let u, v, w denote the frequencies of the AA, Aa and aa genotypes respectively.

In the first set up a fraction $\alpha(u + v)$ of the dominant females pair first with an equal number of dominant males; the same occurs for βw of the recessives. The remaining individuals, a proportion $(1 - \alpha)(u + v) + (1 - \beta)w$ of both sexes mate at random. The resulting relative frequencies of the mating types are given in Table 4.

Mating Types	Frequencies	
	Assorting	Random Mating
$AA \times AA$	$\alpha \frac{u^2}{u + v}$	$(1 - \alpha)^2 u^2 / R$
$AA \times Aa$	$2\alpha \frac{uv}{u + v}$	$2(1 - \alpha)^2 uv / R$
$AA \times aa$		$2(1 - \alpha)(1 - \beta)uw / R$
$Aa \times Aa$	$\alpha \frac{v^2}{u + v}$	$(1 - \alpha)^2 v^2 / R$
$AA \times aa$		$2(1 - \alpha)(1 - \beta)vw / R$
$aa \times aa$	βw	$(1 - \beta)^2 w^2 / R$

TABLE 4.

One can "normalize" back to frequencies (Case A) simply by dividing the proportions in the random mating part by $(1 - \alpha)(u + v) + (1 - \beta)w$. On the other hand, we can assume (Case B) that the delay in pairing causes some decrease in reproduction. One way to express the loss in fertility is to assume that the contribution to the next generation on the part of the population undergoing random mating is

$$[(1 - \alpha)(u + v) + (1 - \beta)w]^2 \text{ instead of } (1 - \alpha)(u + v) + (1 - \beta)w.$$

An alternative formulation in which random mating females pair first can be analyzed (see Scudo and Karlin [30]).

Recurrence relations for genotype frequencies over successive generations are as follows:

$$\text{CASE A. } R = 1 - \alpha + (\alpha - \beta)w$$

$$\begin{aligned} (3.3) \quad u' &= \alpha \frac{(u + \frac{1}{2}v)^2}{u + v} + (1 - \alpha)^2(u + \frac{1}{2}v)^2 / R, \\ v' &= \alpha v \frac{u + \frac{1}{2}v}{u + v} + 2(1 - \alpha)(u + \frac{1}{2}v) \left\{ \frac{1 - \alpha}{2}v + (1 - \beta)w \right\} / R, \\ w' &= \beta w + \alpha \frac{v^2}{4(u + v)} + \left\{ \frac{1 - \alpha}{2}v + (1 - \beta)w \right\}^2 / R. \end{aligned}$$

CASE B.

$$\begin{aligned} (3.4) \quad Nu' &= \alpha \frac{(u + \frac{1}{2}v)^2}{u + v} + (1 - \alpha)^2(u + \frac{1}{2}v)^2, \\ Nv' &= \alpha v \frac{u + \frac{1}{2}v}{u + v} + 2(1 - \alpha)(u + \frac{1}{2}v) \left\{ \frac{1 - \alpha}{2}v + (1 - \beta)w \right\}, \\ Nw' &= \beta w + \alpha \frac{v^2}{4(u + v)} + \left\{ \frac{1 - \alpha}{2}v + (1 - \beta)w \right\}^2, \end{aligned}$$

where $N = 1 - R(1 - R)$.

From (3.3) it follows that the gene frequency $p = u + \frac{1}{2}v$ is invariant over time, i.e., $p' = p$. Using this fact we can rewrite the second equation of (3.3) in the form

$$v' = \frac{\alpha p v}{p + \frac{1}{2}v} + \frac{2p(1 - \alpha)[(1 - \beta)q + \frac{1}{2}v(\beta - \alpha)]}{1 - \alpha p - \beta q + (\beta - \alpha)\frac{1}{2}v} = f(v), \quad (q = 1 - p).$$

The frequency of *Aa* in the *n*th generation is therefore $v_n = f_n(v_0)$ ($f_n(v) = f_{n-1}(f(v))$). By direct verification we find that $f(v)$ is concave and $f(0) > 0$. It follows that $f(v) = v$ admits a unique solution v^* in $(0, 1)$ and, independent of the initial frequency v_n , converges to v^* . The equilibrium v^* depends on p and is computed as the unique root in $(0, 1)$ of the cubic

$$-v^3(\beta - \alpha) - 2(1 - \alpha p - \beta q)v^2 + 4v(1 - \beta)(1 - \alpha)p^2 + 8p^2q(1 - \alpha)(1 - \beta) = 0.$$

We turn to the analysis of case B. Combining appropriately the equations of (3.4) we obtain

$$(3.5) \quad p' = p \left[\frac{1 - (1 - \alpha)(1 - R)}{1 - R(1 - R)} \right],$$

where $R = 1 - \alpha + (\alpha - \beta)w$. Observe that the multiplying factor of p exceeds 1 (is smaller than 1) if and only if $\alpha > \beta$ ($\alpha < \beta$) independent of w ($0 < w < 1$). We deduce easily the following results.

If $\alpha > \beta$, $p_n \uparrow 1$ as $n \rightarrow \infty$, i.e., the population fixes in the homozygote AA state. If $\alpha < \beta$, $p_n \downarrow 0$ as $n \rightarrow \infty$.

3. Assortative mating with no dominance. A general formulation of a model of assortment and random mating would involve 9 parameters. Let α_1, α_2 and α_3 ($0 \leq \alpha_i \leq 1, \alpha_1 + \alpha_2 + \alpha_3 \leq 1$) be measures of the tendency of an AA female to choose an AA, Aa or aa mate respectively. Then $1 - \alpha_1 - \alpha_2 - \alpha_3$ is a measure of ambivalence in the choice of a mate (mates of random). The parameters α_2 and α_3 can be interpreted as propensities of partial disassortment. Similarly, we denote by $\beta_1, \beta_2, \beta_3$ and $1 - \beta_1 - \beta_2 - \beta_3$ the degrees of assortment and random mating respectively for an Aa female. The aa genotype has corresponding assortment parameters γ_1, γ_2 and γ_3 . To illustrate, we discuss the case where all parameters of disassortment are zero, i.e., $\alpha_2 = \alpha_3 = 0, \beta_1 = \beta_3 = 0$, and $\gamma_1 = \gamma_2 = 0$ (for simplicity we drop the subscript and write $\alpha_1 = \alpha, \beta_2 = \beta, \gamma_3 = \gamma$).

Let the frequencies of AA, Aa and aa in the present generation be u, v and w respectively. We assume random mating occurs first, followed by assortative mating. Permanent pairing is assumed and this entails that at the culmination of random mating a total frequency of $\alpha u + \beta v + \gamma w$ males are available to mate with assorting females. Thus the fractions of male and female individuals available for isogenotypic pairings are shown in Table 5.

Genotypes	Proportions of	
	Available Males	Assorting Females
AA	$u(\alpha u + \beta v + \gamma w)$	αu
Aa	$v(\alpha u + \beta v + \gamma w)$	βv
aa	$w(\alpha u + \beta v + \gamma w)$	γw

TABLE 5.

Assorting continues until all possible pairs are formed; the remaining individuals do not contribute to the next generation. Observe that all AA assorting females are

fertilized, if and only if $\alpha u \leq u[\alpha u + \beta v + \gamma w]$ or, what is the same, $\alpha \leq (\beta v + \gamma w)/(v + w)$. If we make the simplifying assumption $\gamma = \alpha$, then if $\gamma = \alpha < \beta$ holds, we find that all AA and aa assorting females can pair. The fraction of unfertilized Aa females is $(\beta - \alpha)v(1 - v)$. Verification of the entries in the Table 6 should now be clear.

Mating Types	Frequencies	
	Random Mating	Assortative Mating
$AA \times AA$	$(1 - \alpha) u^2$	au
$AA \times Aa$	$(1 - \alpha) uv + (1 - \beta) uv$	
$AA \times aa$	$(1 - \alpha) 2uw$	
$Aa \times Aa$	$(1 - \beta) v^2$	$[\alpha(1 - v) + \beta v] v$
$Aa \times aa$	$(1 - \alpha) wv + (1 - \beta) wv$	
$aa \times aa$	$(1 - \alpha) w^2$	αw

TABLE 6

The associated recursion relations connecting genotype frequencies over two successive generations are

$$\begin{aligned}
 Nu' &= \alpha u + f \frac{v}{4} + (1 - \alpha)u \left(u + \frac{1}{2} v \right) + (1 - \beta) \frac{1}{2} v \left(u + \frac{1}{2} v \right), \\
 (3.6) \quad Nv' &= f \frac{v}{2} + (1 - \alpha) \left[\frac{1}{2} v(1 - v) + 2uw \right] + (1 - \beta) \frac{1}{2} v, \\
 Nw' &= \alpha w + f \frac{v}{4} + (1 - \alpha)w \left(w + \frac{1}{2} v \right) + (1 - \beta) \frac{1}{2} v \left(w + \frac{1}{2} v \right),
 \end{aligned}$$

where $N = 1 - (\beta - \alpha)v(1 - v)$ and $f = \alpha(1 - v) + \beta v$.

From (3.6) we have

$$u' - w' = (u - w) \left[\frac{1 - (\beta - \alpha)\frac{1}{2}v}{1 - (\beta - \alpha)v(1 - v)} \right]$$

so that $|u' - w'| > |u - w|$ if and only if $v < \frac{1}{2}$. Moreover, we always have $(u' - w')(u - w) > 0$. Now

$$v' = \frac{\frac{1}{2}v[\alpha + (\beta - \alpha)v] + (1 - \alpha)[\frac{1}{2}v(1 - v) + 2uw] + (1 - \beta)\frac{1}{2}v}{1 - (\beta - \alpha)v(1 - v)}$$

and therefore since $4uw \leq (1 - v)^2$ we have

$$v' \leq \frac{\frac{1}{2}v[\alpha + (\beta - \alpha)v] + \frac{1}{2}(1 - \alpha)(1 - v) + (1 - \beta)\frac{1}{2}v}{1 - (\beta - \alpha)v(1 - v)} = g(v),$$

for all $0 \leq v \leq 1$. Direct computation affirms that $g'(v) \geq 0$ ($0 \leq v \leq 1$). It follows that $v_n \leq g_n(v) = g_{n-1}(g(v))$ where v_n is the frequency of Aa in the n th generation. The theory of iteration of functions tells us that $g_n(v)$ converges as $n \rightarrow \infty$ to the unique fixed point v^* of $g(v) = v$ in $(0, 1)$. We find that v^* satisfies

$$(\beta - \alpha)v^3 - \frac{3}{2}(\beta - \alpha)v^2 + v \left(1 - \alpha + \frac{\beta}{2}\right) - \frac{1 - \alpha}{2} = 0;$$

examination reveals that $v^* < \frac{1}{2}$. Therefore, for n sufficiently large, it follows that $v_n < \frac{1}{2}$ which implies that ultimately $|u_n - w_n|$ continually increases. Its limit is necessarily one. Combining these facts we have established:

- (i) If $u_0 > w_0$ then $u_n \rightarrow 1, v_n \rightarrow 0, w_n \rightarrow 0$.
If $u_0 < w_0$ then $u_n \rightarrow 0, v_n \rightarrow 0, w_n \rightarrow 1$.

The approach of v_n to 0 is geometrically fast at the rate $1 - \beta/2$.

- (ii) When $u_0 = w_0$, then $v_n \rightarrow v^*$ and $u_n = w_n \rightarrow (1 - v^*)/2$ at the geometric rate $|g'(v^*)|$.

The analysis when $\alpha = \gamma > \beta$ paraphrases that above. The conclusions are the same as before, except that now v^* is the solution in $(0, 1)$ of the cubic

$$(\alpha - \beta)v^3 - (\alpha - \beta)v^2 + v(1 - \frac{1}{2}\alpha) - \frac{1}{2}(1 - \alpha) = 0.$$

4. Assortative mating preceding random mating, permanent bonding. Here, assortment is assumed to occur first with permanent pairing. The remaining genotypic proportions of AA, Aa and aa individuals practicing random mating is $(1 - \alpha)u, (1 - \beta)v, (1 - \gamma)w$ respectively. Two cases can be considered according to whether males possess infinite fertility or not. Case B implies a loss of frequency of mating types per generation of magnitude $\alpha u + \beta v + \gamma w$ while Case A assumes no impairment of fertility for females mating randomly. The consequences of the matings are summarized in the recursion relations.

CASE A. $R = 1 - \alpha u - \beta v - \gamma w,$ $N = 1.$	CASE B. $R = 1,$ $N = 1 - R^*(1 - R^*),$ $R^* = 1 - \alpha u - \beta v - \gamma w.$
--	---

$$\begin{aligned}
 Nu' &= \alpha u + \frac{1}{4}\beta v + [(1 - \alpha)u + \frac{1}{2}(1 - \beta)v]^2 / R, \\
 (3.7) \quad Nv' &= \frac{1}{2}\beta v + 2[(1 - \alpha)u + \frac{1}{2}(1 - \beta)v][(1 - \gamma)w + \frac{1}{2}(1 - \beta)v] / R, \\
 Nw' &= \gamma w + \frac{1}{4}\beta v + [(1 - \gamma)w + \frac{1}{2}(1 - \beta)v]^2 / R.
 \end{aligned}$$

We treat only Case B (see Karlin and Scudo [18] for case A).

In the present discussion we restrict attention to the important case where $\alpha = \gamma$. We obtain from (3.7)

$$(3.8) \quad u' - w' = (u - w) \left[\frac{1 - (1 - \alpha)(1 - R^*)}{1 - R^*(1 - R^*)} \right].$$

It follows that $|u' - w'| < |u - w|$ if $\alpha < \beta$ and the opposite inequality holds when $\alpha > \beta$ provided $v > 0$. The recursion relations (3.7) admit a single polymorphic equilibrium $(\hat{u}, \hat{v}, \hat{w})$ where $\hat{w} = \hat{u} = (1 - \hat{v})/2$, and \hat{v} is the unique root in $(0, 1)$ of the equation

$$(3.9) \quad (\alpha - \beta)^2 v^3 + v^2(\alpha - \beta)[1 - 5/2\alpha + \frac{1}{2}\beta] + v[1 - \alpha(1 - \alpha) + \frac{1}{2}\beta - (\alpha - \beta)(1 - \alpha)] - \frac{1}{2}(1 - \alpha)^2 = 0.$$

(i) When $\alpha > \beta$, it can be easily proved that fixation ultimately occurs.

(ii) When $0 < \alpha < \beta$, then for any nontrivial initial values u_0, v_0, w_0 the genotype frequencies at the n th generation u_n, v_n, w_n converge as $n \rightarrow \infty$ to the stable polymorphic equilibrium $(\hat{u}, \hat{v}, \hat{w})$ at a geometric rate. The following is a sketch of the proof.

From (3.8) for the case at hand, we deduce that $u_n - w_n \rightarrow 0$. The second relation of (3.1) can be written in the form

$$(3.10) \quad v_{n+1} = \frac{\frac{1}{2}\beta v_n + (1 - \alpha)(1 - \beta)v_n(1 - v_n) + \frac{1}{2}(1 - \beta)^2 v_n^2 + (1 - \alpha)^2 [\frac{1}{2}(1 - v_n)^2 - \frac{1}{2}(u_n - w_n)^2]}{1 - R_n^*(1 - R_n^*)}$$

where $R_n^* = 1 - \alpha + (\alpha - \beta)v_n$.

We regard $\frac{1}{2}(u_n - w_n)^2 = \epsilon_n$ as a parameter, and the transformation then achieves the form

$$(3.11) \quad v_{n+1} = f_\epsilon(v_n),$$

where $f_\epsilon(v)$ is the function of (3.10) with ϵ_n replaced by ϵ . Simple analysis shows that $f_0(v)$ on $(0, 1)$ is monotone increasing and crosses the 45° line at the unique root of the cubic (3.9). Furthermore, $f_\epsilon(v)$ is monotone increasing for $\eta(\epsilon) < v < 1 - \eta(\epsilon)$ with $\eta(\epsilon)$ tending to zero as $\epsilon \rightarrow 0$.

Inspection of (3.10) reveals that v_n is bounded away from 0 and 1 provided $0 < v_0 < 1$. We infer from (3.11) that

$$f_0^{(m)}(v_{n_0}) \geq v_{n_0+m} \geq f^{(m)}(v_{n_0}),$$

where $f_\epsilon^{(m)}$ denotes the m th composed function f_ϵ with itself. Letting $m \rightarrow \infty$ and exploiting the cited monotonicity properties of f_ϵ we find that

$$\lim_{m \rightarrow \infty} v_m \geq \hat{v}^{(\epsilon)} \quad \text{and} \quad \overline{\lim}_{m \rightarrow \infty} v_m \leq \hat{v},$$

where \hat{v}_ϵ is the unique fixed point of $f_\epsilon(v) = v$ in $(\eta(\epsilon), 1 - \eta(\epsilon))$. Obviously $\hat{v}^{(\epsilon)} \rightarrow \hat{v}$ as $\epsilon \rightarrow 0$ and thus the convergence of v_n to \hat{v} is established. The convergence $u_n \rightarrow \frac{1}{2}(1 - \hat{v})$ and $w_n \rightarrow \frac{1}{2}(1 - \hat{v})$ readily ensue.

For the case of general parameters α, β, γ a complete analysis as above appears difficult; however, investigation of local stability of the fixations provides a good

qualitative picture of the properties of the system (3.7). (See Karlin and Scudo [18] for details.)

5. Partial assortative mating with no priorities. We now consider the case of mixed assortative and random mating where the two mating patterns occur in no predetermined order. Enough males are assumed to be present so that all females contribute to the next generation with no reduction in fertility. The recursion relations connecting genotype frequencies over successive generations are

$$(3.12) \quad \begin{aligned} u' &= \alpha u + \frac{1}{4} \beta v + (1 - \alpha)u(u + \frac{1}{2}v) + (1 - \beta)\frac{1}{2}v(u + \frac{1}{2}v), \\ v' &= \frac{1}{2} \beta v + (1 - \alpha)u(w + \frac{1}{2}v) + (1 - \beta)\frac{1}{2}v + (1 - \gamma)w(u + \frac{1}{2}v), \\ w' &= \gamma w + \frac{1}{4} \beta v + (1 - \gamma)w(\frac{1}{2}v + w) + (1 - \beta)\frac{1}{2}v(w + \frac{1}{2}v). \end{aligned}$$

Some algebraic manipulations reveal that there exists at most one nontrivial equilibrium given by

$$(3.13) \quad \begin{aligned} \hat{u} &= \frac{(L + \gamma - \alpha)(\gamma - \beta)(2 - \alpha - \beta)}{L[L(4 - \alpha - \gamma) - (\gamma - \alpha)^2]}, \quad \hat{v} = \frac{(L + \gamma - \alpha)(L + \alpha - \gamma)(2 - \alpha - \gamma)}{L[L(4 - \alpha - \gamma) - (\gamma - \alpha)^2]}, \\ \hat{w} &= \frac{(L + \alpha - \gamma)(\alpha - \beta)(2 - \gamma - \beta)}{L[L(4 - \alpha - \gamma) - (\gamma - \alpha)^2]}, \end{aligned}$$

where $L = (1 - \alpha)(\gamma - \beta) + (1 - \gamma)(\alpha - \beta)$.

The equilibrium (3.13) exists and is globally stable if $L + \gamma - \alpha < 0$ and $L + \alpha - \gamma < 0$ hold.

The symmetrical case $\alpha = \gamma$ is especially interesting. For $\alpha = \gamma < \beta$ the equilibrium simplifies to

$$\hat{u} = \hat{w} = \frac{1}{2(2 - \alpha)}, \quad \hat{v} = \frac{1 - \alpha}{2 - \alpha}.$$

which is independent of the parameter β and is stable. The symmetric multi allele version of this model can also be analyzed.

IV. INCOMPATIBILITY SYSTEMS AND SELF STERILITY

When not all possible matings can take place, incompatibility mechanisms usually operate for the prohibition of certain matings. An example which springs to mind is the human population where male-male and female-female incompatibility are in force and only male-female matings can occur. There are many other subtle incompatibilities in nature, especially involving plant populations (e.g., see East [8]), and we now study some simple mathematics of this phenomena.

1. A pollen elimination model. Consider a plant species in which the phenotype

in question is controlled by a single diploid locus at which there are three possible alleles A , B and C . Each plant produces both pollen and ova, but we prohibit the mating between a given pollen grain and an ovule of a plant whose genotype contains the same allele as the pollen. The model decrees that an ovule of a plant of type AB may be fertilized by only pollen of type C so that the offspring will be $\frac{1}{2}AC$ and $\frac{1}{2}BC$.

Suppose now that at the n th generation we have x_n , y_n and z_n as the proportions of AB , AC and BC respectively and suppose further that *all ova are fertilized*. It is trivial to verify that

$$(4.1) \quad x_{n+1} = \frac{y_n}{2} + \frac{z_n}{2} = \frac{1 - x_n}{2} = -\frac{1}{2}x_n + \frac{1}{2}.$$

Iterating and by symmetry we obtain

$$(4.2) \quad \begin{aligned} x_n &= \frac{1}{3} + (x_0 - \frac{1}{3})(-\frac{1}{2})^n, & y_n &= \frac{1}{3} + (y_0 - \frac{1}{3})(-\frac{1}{2})^n, \\ z_n &= \frac{1}{3} + (z_0 - \frac{1}{3})(-\frac{1}{2})^n. \end{aligned}$$

Hence x_n , y_n and z_n all converge to $\frac{1}{3}$, at an oscillating geometric rate.

So far the incompatibility we have discussed arises as an incompatibility between the diploid genotype of the ovule and the haploid genotype of the pollen. Thus pollen of the incompatible type, although it contacts the female organ of the plant, dies leaving the ova intact and still available for a compatible fertilization. The incompatibility is determined by the genotype of the diploid ovule. The type of incompatibility system described above occurs in the tobacco plant (*nicotiana*).

2. A zygote elimination model. We next examine the case in which the chance of a mating is proportional to the product of the relative frequencies of both parents subject to the same incompatibility as before. In this case the chance that an AB female mates with the male genotypes AC or BC is proportional to $x(y+z) = x(1-x)$. Table 7 is relevant at the n th generation.

Females		Frequencies of mating	Offspring	
x_n	AB	$x_n(1-x_n)$	$\frac{AC}{2}$,	$\frac{BC}{2}$
y_n	AC	$y_n(1-y_n)$	$\frac{AB}{2}$,	$\frac{BC}{2}$
z_n	BC	$z_n(1-z_n)$	$\frac{AB}{2}$,	$\frac{AC}{2}$

TABLE 7

From Table 7 we find the frequencies in the next generation:

$$(4.3) \quad \begin{aligned} Nx' &= \frac{1}{2}y(1-y) + \frac{1}{2}z(1-z), & Ny' &= \frac{1}{2}x(1-x) + \frac{1}{2}z(1-z), \\ Nz' &= \frac{1}{2}x(1-x) + \frac{1}{2}y(1-y), \end{aligned}$$

where N is the normalizing constant $1 - x^2 - y^2 - z^2$ measuring the loss in fertility due to the diploid-diploid incompatibility.

Subtracting the pairs of equations readily shows that if $y > x$ then $x' > y'$ in the next generation and similarly if $z > x$ then $x' > z'$, etc.

Suppose, for definiteness that $z_0 < y_0 < x_0$ in the initial generation and so, $\min(x_n, z_n) < y_n < \max(x_n, z_n)$ in every succeeding generation. Clearly $y_0 \leq \frac{1}{2}$ and therefore

$$(4.4) \quad \frac{y_0}{2(1-x_0^2-y_0^2-z_0^2)} \leq \frac{1}{4(1-y_0)} \leq \frac{1}{2}$$

we deduce that $|x_1 - z_1| \leq \frac{1}{2}|x_0 - z_0|$, and so

$$|x_{n+1} - z_{n+1}| \leq \frac{1}{2}(x_n - z_n) \leq \frac{1}{2^n}|x_0 - z_0|$$

which implies that $x_n \rightarrow \frac{1}{3}$, $z_n \rightarrow \frac{1}{3}$ and $y_n \rightarrow \frac{1}{3}$ at a geometric rate.

Model 1 is an example of what is called **pollen elimination** since unsuitable pollen is not accepted while the ova remains intact until compatible pollen arrives. Model 2 corresponds to that called **zygote elimination** as pollen derived from an incompatible parent destroys the contacted ova.

3. A multi allelic self sterility model. In practice the number of alleles in a self sterility system of the kind discussed in IV §1 is much larger than 3. In fact as many as 35 alleles have been identified in a sample of 500 plants of *Oxalis Rosa*. We now consider a multi-allele version of IV §1 where once again it is assumed that all ova are fertilized.

Let the r alleles be denoted by A_1, A_2, \dots, A_r . Then our model postulates that an $A_1 A_2$ ovule may be fertilized by A_3, A_4, \dots, A_r pollen only, etc. Let s_{ij} be the frequency of the $A_i A_j$ genotype and we distinguish between $A_i A_j$ and $A_j A_i$. Hence $s_{ii} = 0$, $\sum_{ij} s_{ij} = 2$. Then, at a given generation, the frequency of the pollen containing A_i is $q_i = \frac{1}{2} \{ \frac{1}{2} \sum_j s_{ij} + \frac{1}{2} \sum_j s_{ji} \}$. Now noting that $\sum_j s_{ij} = \sum_j s_{ji}$ we have

$$(4.5) \quad q_i = \frac{1}{2} \sum_j s_{ij}.$$

We next calculate the frequency s_{ij} of the $A_i A_j$ genotype in the next generation. The frequency of a particular ovule, say $A_i A_k$, in the present generation is s_{ik} . This ovule will produce one half A_i gametes and one half A_k gametes. The proportion of A_j pollen which is available to the ovule is taken to be the probability of its being

fertilized by A_j pollen. Since the proportion of compatible pollen is $1 - q_i - q_k$ we have $q_j/(1 - q_i - q_k)$ for the frequency of compatible pollen which will produce the desired A_iA_j zygote. Thus, from the A_iA_k ovule we expect a frequency $\{s_{ik}q_j/(1 - q_i - q_k)\}^{1/2}$ of A_iA_j zygotes. Note that A_jA_k, A_kA_i, A_kA_j ovules also produce A_iA_j zygotes. Combining and simplifying we obtain (when $i \neq j$) the recursion relations

$$(4.6) \quad s'_{ij} = \frac{1}{2} \sum_{k \neq i,j} s_{ik} \frac{q_j}{1 - q_i - q_k} + \frac{1}{2} \sum_{k \neq i,j} s_{jk} \frac{q_i}{1 - q_j - q_k}, \quad i, j \in 1, \dots, r.$$

The following facts can be checked directly. For any ($l \leq r$),

$$s_{ij} = \frac{2}{l(l-1)} \quad \text{for } i \neq j, s_{ii} = 0$$

is a fixed point of (4.6) where the indices i, j vary over a subset l of the original indices and the other frequencies are zero.

It can be shown that the gene frequency $q_i = 1/r$ ($i = 1, 2, \dots, r$), $r \geq 3$ is a locally stable equilibrium. The problem of global stability has not been settled as yet.

4. Sex Determination Models. The first mathematical analysis of diploid-diploid incompatibility systems were concerned with certain naturally occurring plant genetic systems (see Fisher [13], Finney [12], Bodmer [1]). Subsequent investigators treated such models as special cases of the more general phenomenon of negative assortative matings, the most prominent being that of the XX, XY determination of sex in humans; although this is undoubtedly the most familiar diploid-diploid incompatibility many organisms exhibit other forms of sex determination and associated incompatibility mechanisms. The genotypes can be considered to be partitioned into two sets, with matings possible only between individuals in different sets although at random within this restriction. In the terminology set previously the models are of the zygote elimination type. For a biological justification of this formulation, see Scudo [29].

The first model treated here is extremely simple. We allow three genotypes AA, AB and BB , but the only matings producing viable offspring are those between a homozygote and heterozygote.

MODEL Γ

Set 1		Set 2
AA	BB	AB

TABLE 8

Matings are possible only between members of different sets. If the frequencies of the AA , AB and BB in the n th generation are respectively u_n , v_n , w_n , we obtain the recursion relations

$$(4.7) \quad \begin{aligned} T_{n-1} u_n &= u_{n-1} v_{n-1}, \\ T_{n-1} v_n &= u_{n-1} v_{n-1} + w_{n-1} v_{n-1}, \\ T_{n-1} w_n &= w_{n-1} v_{n-1}, \end{aligned}$$

where T_{n-1} is a normalizing constant inserted to keep everything in terms of frequencies.

Obviously $u_n/w_n = u_0/w_0 = \alpha$ and $v_n = \frac{1}{2}$ for $n \geq 1$. It follows that

$$u_n = \frac{\alpha}{2(1 + \alpha)}, \quad w_n = \frac{1}{2(1 + \alpha)} \text{ for } n \geq 1.$$

Significant changes occur when a third allele is incorporated into the above model. We consider two cases according to whether the third allele C is introduced into set 1 (model Γ_1) or set 2 (model Γ_2). In the model Γ_1 the incompatibility is specified by Table 9.

MODEL Γ_1

	set 1					set 2
genotype	AA	BB	BC	AC	CC	AB
n th generation frequency	u_n	w_n	x_n	y_n	z_n	v_n

TABLE 9

Again matings are considered to take place only between individuals in different sets. The relations connecting genotype frequencies over successive generations are

$$(4.8) \quad \begin{aligned} T_{n-1} v_n &= v_{n-1} u_{n-1} + v_{n-1} w_{n-1} + \frac{v_{n-1}(x_{n-1} + y_{n-1})}{2}, \\ T_{n-1} u_n &= v_{n-1} u_{n-1} + \frac{v_{n-1} y_{n-1}}{2}, \quad T_{n-1} w_n = v_{n-1} w_{n-1} + \frac{v_{n-1} x_{n-1}}{2}, \\ T_{n-1} x_n &= \frac{v_{n-1} x_{n-1} + v_{n-1} y_{n-1}}{2}, \quad T_{n-1} y_n = \frac{v_{n-1} x_{n-1} + v_{n-1} y_{n-1}}{2}, \end{aligned}$$

$z_n = 0$ for $n > 1$, where $T_{n-1} = 2v_{n-1}(1 - v_{n-1})$ is the normalizing factor. Notice that $u_n + w_n = v_n$ and $x_n = y_n$ for $n \geq 1$. Hence

$$\frac{u_n}{x_n} = \frac{u_{n-1}}{x_{n-1}} + \frac{1}{2} = \frac{u_{n-2}}{x_{n-2}} + 1 = \dots = \frac{n}{2} + \frac{u_0}{x_0} \text{ and } \frac{w_n}{x_n} = \frac{n}{2} + \frac{w_0}{x_0}.$$

Therefore $x_n \rightarrow 0$ and then $y_n \rightarrow 0$, so that $u_n + v_n + w_n = 2v_n \rightarrow 1$ or $v_n \rightarrow \frac{1}{2}$. Since

$$\frac{u_n}{w_n} = \frac{\frac{u_0}{x_0} + \frac{n}{2}}{\frac{w_0}{x_0} + \frac{n}{2}} \rightarrow 1 \text{ as } n \rightarrow \infty$$

we have $u_n \rightarrow \frac{1}{4}$, $w_n \rightarrow \frac{1}{4}$. The ultimate configuration of the population is therefore $u_e = w_e = \frac{1}{4}$, $v_e = \frac{1}{2}$ and is independent of the initial makeup of the population.

Note that the previous continuum of fixed points of model Γ is reduced to the single point $u_e = w_e = \frac{1}{4}$, $v_e = \frac{1}{2}$. In model Γ the equilibrium point (which depends on the initial conditions) is achieved in one generation. In model Γ_1 the third allele disappears quite slowly at an algebraic rate.

The incorporation of the third allele into model Γ to form model Γ_1 profoundly alters the equilibrium behavior, as shown above. The only change from model Γ_1 in constructing model Γ_2 is the set to which C has been added. There are three families of equilibrium points, and the initial conditions determine which is reached. The equilibrium behavior differs markedly from that of the previous model. The following is a brief discussion of the results obtained.

MODEL Γ_2

genotype	Set 1		Set 2			
	<i>AA</i>	<i>BB</i>	<i>AB</i>	<i>AC</i>	<i>BC</i>	<i>CC</i>
<i>n</i> th generation frequencies	u_n	w_n	v_n	y_n	x_n	z_n

TABLE 10

There exist exactly three families of equilibria

$$F_1 : \hat{w} = \frac{1}{2}, \hat{v} + \hat{x} = \frac{1}{2}, \quad F_2 : \hat{u} = \frac{1}{2}, \hat{v} + \hat{y} = \frac{1}{2},$$

$$F_3 : \hat{u} + \hat{w} = \frac{1}{2}, \hat{v} = \frac{1}{2}$$

and the vector $(u_n, w_n, v_n, x_n, y_n)$ always converges as $n \rightarrow \infty$. It is possible to determine precise domains of attraction to the respective equilibria.

In fact if $u_0/w_0 \leq 1$, $y_0/x_0 \leq 1$ and $u_0y_0/w_0x_0 < 1$, then the limiting equilibrium is in F_1 . Symmetrically, if $w_0/u_0 \leq 1$, $x_0/y_0 \leq 1$ and $w_0x_0/u_0y_0 < 1$ then the limiting equilibrium belongs to F_2 . If $(u_{n-1}/w_{n-1}) - 1$ and $(u_n/w_n) - 1$ alternate continually as $n \rightarrow \infty$ or are zero, then the limit equilibrium belongs to F_3 (see Karlin [16] for proofs). The domain of attraction to F_3 is usually a hypersurface.

One final example is where each sex is characterized by three genotypes as follows. (A nine allele expression of this model arises in a strain of wasp. For certain fungi, including yeast, sex determination appears to be controlled at a single locus.)

	Set 1			Set 2		
	<i>AA</i>	<i>BB</i>	<i>CC</i>	<i>AB</i>	<i>AC</i>	<i>BC</i>
frequency	<i>x</i>	<i>y</i>	<i>z</i>	<i>u</i>	<i>v</i>	<i>w</i>

The recurrence relations are as follows:

$$\begin{aligned}
 Tx' &= x(u + v), & Tu' &= (x + y)u + vy + wx, \\
 Ty' &= y(u + w), & Tv' &= (x + z)v + uz + xw, \\
 Tz' &= z(v + w), & Tw' &= (y + z)w + vy + uz, \\
 T &= 2(x + y + z)(u + v + w).
 \end{aligned}$$

The stable equilibria are precisely the fixed points

$$\begin{aligned}
 x + y &= \frac{1}{2} & u &= \frac{1}{2}; & x &= \frac{1}{2} & v + w &= \frac{1}{2} \\
 x + z &= \frac{1}{2} & v &= \frac{1}{2}; & y &= \frac{1}{2} & u + w &= \frac{1}{2} \\
 y + z &= \frac{1}{2} & w &= \frac{1}{2}; & z &= \frac{1}{2} & v + w &= \frac{1}{2}.
 \end{aligned}$$

An interior unstable fixed point $x = y = z = 1/9$, $u = v = w = 2/9$, also exists.

It can be proved generally in the case of three alleles at a single locus that, any grouping for sex determination exhibits only the $\frac{1}{2}$ sex ratio in a stable configuration. When sex is determined involving at least two loci, then a stable sex ratio may be different from $\frac{1}{2}$.

V. MUTATION SELECTION BALANCE

1. Mutation balance. We assume that each *A* allele has a probability μ of mutating to *B* (and hence $1 - \mu$ of not mutating), that ν similarly is the mutation rate of *B* to *A* and that no other forces are acting to change gene frequencies. It is easily seen that

$$(5.1) \quad p_n = (1 - \mu) p_{n-1} + \nu(1 - p_{n-1}),$$

where p_n is the gene frequency of *A* in the n th generation. This equation can be rewritten in the form

$$\left(p_n - \frac{\nu}{\mu + \nu} \right) = (1 - \mu - \nu) \left(p_{n-1} - \frac{\nu}{\mu + \nu} \right) = (1 - \mu - \nu)^n \left(p_0 - \frac{\nu}{\mu + \nu} \right).$$

Thus $p_n \rightarrow \nu/(\mu + \nu)$ as $n \rightarrow \infty$ at a rate $(1 - (\mu + \nu))^n$ i.e., $p_n - \nu/(\mu + \nu)$ is of order $(1 - \mu - \nu)^n$. There is thus a stable intermediate equilibrium point, whose position depends on the ratios of the two mutation rates. However, since mutation rates are generally less than 10^{-5} , the rate of convergence to the equilibrium is exceedingly slow. As we shall see below, it seems likely that selection differentials are nearly always large enough to mask these balancing effects of opposing mutation rates.

2. Immigration balance. We assume that a proportion m of the population is replaced in each generation by individuals from another population with constant A and B gene frequencies P and Q respectively. The change in gene frequency is then given by

$$(5.2) \quad p_n = (1 - m) p_{n-1} + mP.$$

As $n \rightarrow \infty$, then $p_n \rightarrow P$, the frequency of the immigrant population, at a rate $(1 - m)^n$. If we put $v = mP$ and $\mu = m(1 - P)$ then equation (5.2) is identical to equation (5.1), so that this situation is exactly analogous to the mutation balance. Both factors cause linear changes in the gene frequencies.

3. Mutation-selection balance for disadvantageous genes. Assume genotypes AA , AB and BB have relative fitnesses 1, $1 - hs$, and $1 - s$ where $s, h \geq 0$, and that p and q are the gene frequencies in fertilized zygotes. Gene frequencies are measured in the gametes which combine at random to form the fertilized zygote, before selection has acted, and mutation is assumed to occur after selection during the formation of the next generation's gametes.

As in Section 2, the gene frequencies of A and B after selection, before mutation, are

$$\frac{p^2 + (1 - hs)pq}{1 - 2hspq - sq^2} \text{ and } \frac{(1 - s)q^2 + (1 - hs)pq}{1 - 2hspq - sq^2},$$

respectively. Allowing only one way mutation $A \rightarrow B$, the new frequency of B will be

$$(5.2) \quad q' = \frac{(1 - s)q^2 + (1 - hs)pq}{1 - 2hspq - sq^2} + \mu \frac{[p^2 + (1 - hs)pq]}{1 - 2hspq - sq^2}.$$

Equilibria are obtained as the solutions of

$$(5.3) \quad sq^3(2h - 1) + sq^2[1 - 3h - h\mu] + q[\mu + hs(1 + \mu)] - \mu = 0.$$

The mutation rate μ is always very small. One stable equilibrium is approximately

$$(5.4) \quad q \sim \mu/hs$$

provided μ is small compared with hs . The n th generation frequency q_n approaches its equilibrium value of μ/sh at a geometric rate of approximate order $1 - sh$. It is noteworthy that this solution depends only on the product sh and not on s alone, indicating that the fitness of the heterozygote dominates the situation. Given q and hs for any particular gene, assumed to have reached its equilibrium frequency, we can estimate from (5.4) the magnitude of the mutation rate μ . This was, in fact, the way that mutation rates in man were originally derived by Danforth in 1920 and later by Haldane.

When $h = 0$ the allele B is recessive with respect to its effect on fitness and (5.4) reduces to $(q - 1)(sq^2 - \mu) = 0$. The solution $q = \sqrt{\mu/s}$ is the only stable equilibrium, of course provided $\mu \leq s$.

The results of this model have been frequently applied in estimating the mutation rate for recessive human diseases.

Criteria for selection mutation balance for a character controlled at two loci are given in Karlin and McGregor [21]. In Section 8 we present a model for mutation selection balance involving an infinite number of types. Those considerations are also relevant to an understanding of polygenic inheritance (characters determined by many loci).

VI. THE CONCEPT OF IDENTITY BY DESCENT AND APPLICATIONS

The **inbreeding coefficient** of an individual (introduced first by Wright) is defined to be the probability that two genes at a single locus are **identical by descent** by which we mean that the genes can be traced back to copies of the same gene in a particular individual of a previous generation. Certain finite size population genetic problems can be solved relatively easily using calculations for probabilities of descent. We expose a series of important models exemplifying the method. (This method has been exploited by many including Malécot, Kimura, Kempthorne and others. See Karlin [16] and [17] for further applications and references on this subject.)

1. Monoecious diploid finite population. A **monoecious individual** is one that can contribute both male and female gametes (e.g., as occurs commonly in plants).

Consider a population of N monoecious individuals diploid at an autosomal locus, reproducing randomly but maintaining constant population size. More specifically we may stipulate that each individual produces an infinite number of copies of each of his genes to form a pool from which the next generation is formed by choosing N pairs at random where each parental gene is represented to the extent of $\frac{1}{2}N^{-1}$ -th of the complete gene pool.

Let I_t denote the probability that two homologous chromosomes at a given locus in an individual in the t th generation carry genes identical by descent. Let J_t be the probability that two homologous chromosomes of the t th generation, chosen at random one from each of two different individuals, carry genes identical by descent.

Under random mating two genes are derived from the same parental individual with probability $1/N$ or from different individuals with probability $1 - 1/N$. In the former event either they are copies of the same gene or they are copies of the homologous pair, each occurring with probability $\frac{1}{2}$. We may evidently compute I_t and J_t according to the same recursion relations

$$(6.1) \quad J_t \text{ and } I_t = \frac{1}{N} \left(\frac{1}{2} + \frac{1}{2} I_{t-1} \right) + \left(1 - \frac{1}{N} \right) J_{t-1}, \quad t \geq 1.$$

Thus $I_t = J_t$ for $t \geq 1$ and (6.1) reduces to

$$(6.2) \quad I_t = \frac{1}{2} N^{-1} + \left(1 - \frac{1}{2} N^{-1} \right) I_{t-1}, \quad t \geq 2.$$

We introduce the quantity $H_t = 1 - I_t$ and then (6.2) is converted into

$$(6.3) \quad H_t = (1 - \frac{1}{2}N^{-1})H_{t-1} = (1 - \frac{1}{2}N^{-1})^{t-1}H_1, \quad t \geq 1,$$

where $H_1 = 1 - I_1$ and $I_1 = \frac{1}{2}N^{-1}(1 + I_0) + (1 - N^{-1})J_0$. Equation (6.3) shows that H_t tends to zero at a geometric rate $(1 - \frac{1}{2}N^{-1})$.

The above analysis implies two interesting conclusions. Firstly, the ultimate population is composed exclusively of inbred individuals, i.e., individuals with inbreeding coefficient 1. Secondly, even for the process of random mating, limitation of population size imposes a certain degree of inbreeding which eliminates, at an exponential rate, the heterozygote types.

2. Dioecious finite diploid population. We consider a two sex population consisting of N_1 males and N_2 females. Let I_t be the probability that two homologous genes from the same male or female of the t th generation are identical by descent. Let J_t be the probability that two genes chosen at random one from each of two different males or females in the t th generation are identical by descent. Let K_t be the probability that two genes chosen at random, one from a male, the other from a female of the t th generation, are identical by descent. Finally, let \tilde{J}_t denote the probability that two genes chosen at random in the t th generation, one each from different individuals (with no reference to sex), are identical by descent. Symmetry suggests and indeed it can be easily proved that I_t and $J_t = \tilde{J}_t$ are well defined.

We now develop recursion formulas for the quantities introduced above by examining the source of the two genes in a given individual traced two generations back. Consider two genes in a given individual. Conditional that they both come from males, two generations back, the probability they derive from the same male (say A) is $(N_1/N_1^2) = N_1^{-1}$.

The probability is $\frac{1}{4}$ that two children B and C of A transmit to their offspring D the genes received from A . Now the genes given B and C by A are copies of the same gene or correspond to distinct homologous genes with probability $\frac{1}{2}$ each. In the latter event the genes are identical by descent with probability I_{t-2} . This accounts for the first term of the recursion relation

$$(6.4) \quad I_t = \frac{1}{4}N_1^{-1}(\frac{1}{2} + \frac{1}{2}I_{t-2}) + \frac{1}{4}N_2^{-1}(\frac{1}{2} + \frac{1}{2}I_{t-2}) + (1 - \frac{1}{4}N_1^{-1} - \frac{1}{4}N_2^{-1})J_{t-2}.$$

The second term reflects the circumstance when both genes derive from the same female parent. The probability is $(1 - \frac{1}{4}N_1^{-1} - \frac{1}{4}N_2^{-1})$ that the two genes of D derive from distinct individuals of the $(t-2)$ -th generation, in which case the probability is J_{t-2} that they are identical by descent.

A similar kind of reasoning establishes the relation

$$(6.5) \quad J_t = \frac{1}{4}N_1^{-1}(\frac{1}{2} + \frac{1}{2}I_{t-1}) + \frac{1}{4}N_2^{-1}(\frac{1}{2} + \frac{1}{2}I_{t-1}) + (1 - \frac{1}{4}N_1^{-1} - \frac{1}{4}N_2^{-1})J_{t-1}.$$

Notice that the subscript on the right now involves the $(t-1)$ -th generation rather than the $(t-2)$ -th.

The identical formula as in (6.5) obtains with the left side replaced by K_t . It follows that $J_t = K_t$. Comparing (6.5) and (6.4) we may conclude that $J_{t-1} = I_t$ and then we rewrite (6.4) in the form

$$(6.6) \quad I_t = N_e^{-1}(\frac{1}{2} + \frac{1}{2}I_{t-2}) + (1 - N_e^{-1})I_{t-1},$$

where $N_e^{-1} = \frac{1}{4}N_1^{-1} + \frac{1}{4}N_2^{-1}$, a quantity commonly called the **effective population number**. Let $H_t = 1 - I_t$ and then (6.6) becomes

$$(6.7) \quad H_t = (1 - N_e^{-1})H_{t-1} + \frac{1}{2}N_e^{-1}H_{t-2}, \quad t \geq 2.$$

The solution of this second order difference equation has the form $H_t = a\lambda_1^t + b\lambda_2^t$, $t \geq 2$, where λ_i ($i = 1, 2$) are roots of the quadratic equation $\lambda^2 - (1 - N_e^{-1})\lambda - \frac{1}{2}N_e^{-1} = 0$. Hence as $t \rightarrow \infty$, H_t behaves asymptotically as

$$(6.8) \quad H_t \sim \frac{1}{2}a[1 - N_e^{-1} + (1 + N_e^{-2})]^t.$$

The special case of sib mating arises when $N_1 = N_2 = 1$ and so $N_e = 2$. Then $H_t \sim a(\frac{1}{4}(1 + \sqrt{5}))^t$.

3. Loss of k alleles out of p in a haploid model. Consider a finite constant size (say N individuals) **haploid** population (each individual carries one dose of an allele) undergoing some general pattern of reproduction where the number of alternative alleles represented in the population is at least $p > 2$. We investigate the problem of determining the rate at which k of the p alleles are lost from the population.

The reproduction mechanism is as follows. Each individual replicates his type in some general fashion but with no selection differences operating among the types. The next generation is formed by choosing at random N progeny from the output of the previous generation. The parameters of the reproduction mechanism are the numbers g_{ij} = to the probability that i randomly chosen progeny derive from j distinct parents ($i, j = 1, 2, \dots, N$). Obviously $g_{ij} = 0$ for $j > i$ so the matrix $G = \|g_{ij}\|_1^N$ is lower triangular. Clearly $g_{11} = 1$ and we postulate that

$$(6.9) \quad g_{kk} > g_{k+1,k+1} > 0 \quad (k = 1, 2, \dots, N - 1) \text{ and } g_{k,k-1} > 0$$

in order to avoid pathological algebraic annoyances. These conditions are satisfied in almost all examples. In the special case where each parent contributes exactly r replicas of his own type then an elementary combinatorial analysis shows that

$$(6.10) \quad g_{ij} = \begin{cases} \sum^* \frac{\binom{r}{i_1} \binom{r}{i_2} \dots \binom{r}{i_j} \binom{N}{j}}{\binom{Nr}{i}} & i \geq j \\ 0 & i < j, \end{cases}$$

where \sum^* indicates summation over all $i_1, i_2, \dots, i_j \geq 1$ subject to $i_1 + i_2 + \dots + i_j = i$.

The conditions of (6.9) are obviously satisfied in this circumstance. Notice that here $g_{ii} \rightarrow (N(N-1)\cdots(N-i+1))/N^i$ as $r \rightarrow \infty$.

Let $P_{ij}^{(t)}$ be the probability that i randomly chosen different individuals of the t th generation consist of j different types (alleles). Our objective is to ascertain the asymptotic properties of $P_{Nj}^{(t)}$ as $t \rightarrow \infty$ for $j = 1, 2, \dots, p$. Since the population size is kept constant we expect ultimate fixation in one type, i.e., $P_{Nj}^{(t)} \rightarrow 0$ as $t \rightarrow \infty$ for $j = 2, \dots, N$. We wish to determine the rate of this approach to zero. The key to the analysis is the recursion relation

$$(6.11) \quad P_{ij}^{(t+1)} = \sum_{k=1}^N g_{ik} P_{kj}^{(t)} \quad (i, j, \dots, N).$$

The derivation is simple and follows by considering the various possibilities describing the parental genes that can produce the given sampled genes.

If we introduce the matrices $P^{(t)} = \|P_{ij}^{(t)}\|$, then (6.11) can be written concisely as the matrix product $P^{(t+1)} = GP^{(t)}$, and iteration produces

$$(6.12) \quad P^{(t)} = G^t P^{(0)},$$

where G^t is the t th power of the matrix G and $P^{(0)}$ provides the information of the initial frequencies of types. Since G is lower triangular and the diagonal elements are distinct by assumption, we may conclude that the eigenvalues of G are $\lambda_1 = g_{11} = 1, \lambda_2 = g_{22}, \dots, \lambda_k = g_{kk}, \dots, \lambda_N = g_{NN}$.

A system of left eigenvectors of G can be constructed of the form

$$v^k = (v_1^{(k)}, \dots, v_k^{(k)}, 0, \dots, 0), \quad k = 1, 2, \dots, N$$

with the property $v_k^{(k)} \neq 0$. This last fact derives from the condition $g_{kk} > g_{k-1, k-1}$. Let V be the matrix with row vectors $v^{(1)}, v^{(2)}, \dots, v^{(N)}$ and $U = V^{-1}$. Since V is lower triangular, so is U . Of course, $G = U\Omega V$, where Ω is the diagonal matrix of eigenvalues of G whose values are $g_{11}, g_{22}, \dots, g_{NN}$. It is not difficult to prove inductively that $u_i^{(k)} > 0$ for all $i \geq k$. Consider now

$$P'_{Nj} = \sum_{k=1}^N G_{Nk}^{(t)} P_{kj}^0 = \sum_{k=j}^N G_{Nk}^{(t)} P_{kj}^0.$$

Expanding

$$\begin{aligned} G_{Nj}^t &= \sum_{k=1}^N u_N^k g_{kk}^t v_k^{(j)} = \sum_{k=j}^N u_N^{(k)} [g_{kk}]^t v_k^{(j)} \\ &= [g_{jj}^t] u_N^{(j)} + O[g_{j+1, j+1}^t] \end{aligned}$$

where the last reduction is valid since $v_k^{(j)} = 0$ for $k < j$. Since $u_N^{(j)} > 0$ we have proved the following theorem.

THEOREM. Suppose (6.9) holds. If $P_{jj}^0 > 0$ then the probability that a population

of N haploid individuals contains at least j types in the t -th generation is of the order of magnitude $c_j [g_{jj}]^t$ where c_j is a positive constant depending on the initial set of frequencies.

The condition $P_{jj}^0 > 0$ for $j \leq p$ is very weak and would ordinarily be satisfied. For further discussion of this model and ramifications we refer to Karlin [17] Section 6, and Felsenstein [11].

5. Identity by descent and mutation effects. Consider a population of N diploid individuals or $2N$ genes with an infinite series A_1, A_2, \dots of possible alleles at a locus with no selective differences among the allelic types. The population is randomly reproducing as in Model I, i.e., the $2N$ genes of the next generation are formed by repeated sampling with replacement from the $2N$ genes of the present generation. Suppose moreover that as each gene is drawn there is a probability u that a mutation occurs and any new mutant allele is of a not previously existing type.

Let I_t be the probability in generation t that two genes sampled at random are identical by descent. A recursion formula analogous to (6.1) with due account of mutation is

$$I_t = \left[\frac{1}{N} \left(\frac{1}{2} + \frac{1}{2} I_{t-1} \right) + \left(1 - \frac{1}{N} \right) I_{t-1} \right] (1 - u)^2.$$

Letting $t \rightarrow \infty$, we get the equilibrium value $\lim_{t \rightarrow \infty} I_t = I$, where

$$I = \frac{(1 - u)^2}{1 + 4Nu - 2Nu^2}$$

and for u small and N large such that $4Nu = \theta$ we have the approximate formula $I = 1/(1 + \theta)$.

Of considerable interest for discussions relevant to non-Darwinian evolution (Neutral mutation theory) is the evaluation of the probability

$$(6.13) \quad P\{2N, u, n_1, n_2, \dots, n_k\}$$

that a sample of r genes, chosen from the population, contains just k different allelic types with n_1 of one kind, n_2 of a second kind and so on, n_k of a k th kind where the n_i are positive integers with sum r . For the significance of the computation of (6.13) and its utility in evaluating the relevance of neutral mutation theory, we refer to Ewens [10]. The quantity (6.13) is a complicated function of $2N$ and u . However, if we let $N \rightarrow \infty$ and $u \rightarrow 0$ in such a way that $4Nu$ converges to a finite non-zero limit θ , then (6.13) converges to a relatively simple limit formula

$$(6.14) \quad P(\theta; n_1, n_2, \dots, n_k) = \frac{r!}{n_1 n_2 \dots n_k \alpha_1! \alpha_2! \dots \alpha_p! L_r(\theta)},$$

where p is the number of distinct integers in the set $\{n_1, n_2, \dots, n_k\}$ of which there are

exactly α_1 indices equal to an integer, α_2 indices equal to a different integer, and so on, and exactly α_p indices equal to the p th distinct value among the numbers n_1, n_2, \dots, n_k . Here

$$L_r(\theta) = \theta(\theta + 1)(\theta + 2)\cdots(\theta + r - 1).$$

The formula was suggested by Ewens [10] and rigorously established in Karlin and McGregor [22]. The method relies heavily on the concept of identity by descent.

VII. EVOLUTION OF A POPULATION WITH POLYGENIC CHARACTERS

1. A model of a polygenic trait. Consider a population with an infinite number of possible types. Assume that the different types are identified with points of the real line \mathbb{R} . One example is where the type x can be associated with the "fitness" of the given individual. A second case is where x corresponds to a measurable numerical trait whose value is determined by the combined effects of many loci.

Consider the frequency distribution of the types in the population. More precisely, let A be any interval (or Borel measurable set) in \mathbb{R} and let $m_t(A)$ be the proportion of the population (population size is for our purposes, regarded of large-infinite size) of types corresponding to A at generation t . Selection and mutation affect changes in m_t over successive generations in the following manner:

(i) The relative viability of an offspring of type x compared to that of type y is in the ratio $\gamma(x)/\gamma(y)$ which we stipulated as a first approximation to be independent of time. Assuming each parental type replicates its identical type, the change of frequency distribution due to this selection is to be calculated by the formula

$$\tilde{m}_{t+1}(A) = \frac{\int_A \gamma(x)m_t(dx)}{\int_{\mathbb{R}} \gamma(x)m_t(dx)}$$

for all intervals (and sets) A .

(ii) Mutation acts after selection as follows: let $p_t(B, x)$ be the conditional probability that an offspring of an x -type parent of generation t alter its form to that of type in B . Then a parent of type x , affected by selection and mutation will produce offspring of type in A is calculated modulo a proportionality constant by the expression $\gamma(x)p_t(A, x)$. It follows that the total number of A -type offspring in generation $t + 1$ is proportional to $\int_{\mathbb{R}} \gamma(x)p_t(A, x)m_t(dx)$ which after converting to frequencies, becomes

$$(7.1) \quad m_{t+1}(A) = \frac{\int_{\mathbb{R}} \gamma(x)p_t(A, x)m_t(dx)}{\int_{\mathbb{R}} \gamma(x)m_t(dx)}.$$

The evolution of the frequency distributions m_t over time is the primary object under investigation. To achieve qualitative results and deeper insights into the behavior of m_t as t increases we now specialize to the situation where

$$(7.2) \quad p_t(B, u) = \int_B dG(x - u) \text{ and } \gamma(x) = \lambda^x, \quad \lambda > 1$$

so that the difference between a parent and offspring has the same distribution $G(u)$ (called the **distribution of the mutation**) over the whole population. The reproduction rate of an x -type parent is λ^x so that a type is more advantageous with larger values. For the case of $\gamma(x) = \lambda^x$ the meaning of x is strongly correlated with the actual fitness of the x -individual.

Let $F_t(x)$ be the proportion of types $\leq x$ in the population at time t . Manifestly, $F_t(x)$ is a distribution function of the variable x . Define $E_t = \int_{-\infty}^{\infty} x dF_t(x)$ as the average fitness and $V_t = \int_{-\infty}^{\infty} [x - E_t]^2 dF_t(x)$ as the fitness variance. Define for any distribution $H(x)$ the quantity $\bar{H} = \inf\{x \mid H(x) = 1\}$ as the largest point in the spectrum of $H(x)$. The following results were proved by Eshel [9], (see Karlin [23] for improvements and extensions).

THEOREM I. *Assume $\bar{F}_0 < \infty$ (i.e., the initial fitness distribution in the population is bounded). Suppose that $\bar{G} < \infty$, that is the maximal possible mutation change is bounded. Then*

$$(7.3) \quad \lim_{t \rightarrow \infty} (E_{t+1} - E_t) = \bar{G}.$$

The rate of evolution (= the rate of change of the average fitness in the population) approaches \bar{G} .

A more refined result pertains to the changes in the centered fitness distribution $F_t(x - E_t)$ as $t \rightarrow \infty$.

THEOREM II. *Under the assumptions of Theorem I $F_t(x - E_t)$ tends to a limit distribution $F(x)$ whose variance is finite.*

In particular the proportion of types compared to the mean fitness in any given region tends to a positive value. We state as a consequence of Theorem II: *If $\bar{G} = 0$ (i.e., all mutations are deleterious or neutral), then it follows that $F_t(x)$ approaches a limiting mutation selection balance with distribution of types $F(x)$ iff $G(x)$ has a positive jump at 0.*

The results cited above hinge strongly on the assumptions of (7.2). To what extent are corresponding conclusions valid for other choices of the selection functions $\gamma(x)$ not of exponential growth λ^x ? Cases where $\gamma(x)$ is bell-shaped (e.g., $\gamma(x) = e^{-x^2}$ or $1/(1+x^2)$) would be of interest in treating the evolution of quantitative traits where the optimum type has an intermediate value.

2. Another model of a polygenic trait. Another model of a polygenic trait involving a selection balance and the mating process proposed by Haldane has the following structure.

The set of all possible phenotypes are again identified with the real line. Let the proportion of the population exhibiting phenotype in an interval A in generation t be

$$(7.4) \quad m_t(A) = \int_A p_t(x) dx.$$

(For ease of exposition we have assumed the existence of a density p_t for the frequency measure $m_t(dx$.) The basic assumption for this model is that the distribution of the type of the offspring depends on the type of each parent, through the conditional probability (**segregation function**) $L(x; x_1, x_2) dx$ equal to the probability that the offspring is of type x to $x + dx$ given the parental types are x_1 and x_2 . Clearly,

$$\int_{-\infty}^{\infty} L(x; x_1, x_2) dx = 1.$$

In theory, L could be determined from careful analysis of breeding experiments. Assuming random union of types the density of phenotypes in the next generation would ordinarily be calculated by the formula

$$(7.5) \quad \check{p}_{t+1}(x) = \iint L(x; x_1, x_2) p_t(x_1) p_t(x_2) dx_1 dx_2$$

before selection has acted. The action of selection is determined as in Model 1 by a function $\gamma(x)$ which is the relative survival probability for individuals of type x . Taking account of selection, the density $p_t(x)$ is altered to

$$\check{p}_t(x) = \frac{\gamma(x) p_t(x)}{\int_{-\infty}^{\infty} p_t(x) \gamma(x) dx}.$$

Subject to random mating, segregation (described by $L(x; x_1, x_2)$) and selection (measured in relative terms by $\gamma(x)$) we obtain the non-linear transformation law

$$p_{t+1}(x) = \frac{\iint p_t(x_1) p_t(x_2) \gamma(x_1) \gamma(x_2) L(x; x_1, x_2) dx_1 dx_2}{\left(\int_{-\infty}^{\infty} p_t(\xi) \gamma(\xi) d\xi \right)^2}.$$

For certain choices of $L(x; x_1, x_2)$ for a large class of bell-shaped functions $\gamma(x)$ we can deduce the fact that $m_t(x)$ converges to a limiting stable frequency distribution.

Other models for polygenic traits were studied by Kimura (see Crow and Kimura [7], pages 294–296, Slatkin [31], Haldane [14], among others).

VIII. SOME SELECTION MODELS FOR TWO LOCUS MARKERS

Consider a diploid population of a character determined by two loci with possible alleles A, a and B, b at the first and second locus respectively. There are therefore four types of chromosomes (or referred to as gametes):

$$(8.1) \quad AB \quad Ab \quad aB \quad ab$$

and 10 genotypes

$$\frac{AB}{AB}, \frac{AB}{Ab}, \frac{AB}{aB}, \frac{AB}{ab}, \frac{Ab}{Ab}, \frac{Ab}{aB}, \frac{Ab}{ab}, \frac{aB}{aB}, \frac{aB}{ab}, \frac{ab}{ab}$$

where the symbol AB/aB , for example, means that the alleles A and B sit on one of the chromosomes while the alleles a and B are found on the other. Let $M = \| m_{ij} \|_{i,j=1}^4$ denote the fitness matrix, where m_{ij} is the fitness of the genotype composed from the i and j type chromosomes.

Let x_1, x_2, x_3 and x_4 be the frequencies of the four gamete types in the order of (8.1). Assuming random union of gametes (= random mating) and recalling the nature of Mendelian segregation involving recombination frequency r (refer here back to Section I), it is easy to check Table 11.

Reading off from the table we find that the frequency x'_1 of AB in the next generation is proportional to

$$\begin{aligned} x'_1 &\sim x_1^2 m_{11} + 2x_1 x_2 m_{12} \frac{1}{2} + 2x_1 x_3 m_{13} \frac{1}{2} + 2x_1 x_4 m_{14} (1-r) \frac{1}{2} + 2x_2 x_3 r \frac{1}{2} \\ &= x_1 m_1 - rD, \end{aligned}$$

where $m_1 = \sum_{j=1}^4 m_{1j} x_j$, $D = x_1 x_4 m_{14} - x_2 x_3 m_{23}$. Similar expressions result for x'_2, x'_3 and x'_4 . The recursion relations connecting frequencies over successive generations become

$$(8.2) \quad x'_i = \frac{x_i m_i + \varepsilon_i r D}{W}, \quad i = 1, 2, 3, 4,$$

where $\varepsilon_2 = \varepsilon_3 = -\varepsilon_1 = -\varepsilon_4 = 1$, $m_i = \sum_{j=1}^4 m_{ij} x_j$, $W = \sum_{i,j=1}^4 m_{ij} x_i x_j$.

1. No selection differences. The special case where $m_{ij} \equiv 1$ (no selection differences) is the most classical case treated. Then (8.2) reduces to

$$(8.3) \quad x'_i = x_i + \varepsilon_i r D, \quad i = 1, 2, 3, 4.$$

It is convenient to introduce the **gene frequency variables**

$$(8.4) \quad \begin{aligned} p_1 &= x_1 + x_2 = (\text{frequency of } A), \quad p_2 = x_1 + x_3 = (\text{frequency of } B), \\ D &= x_1 x_4 - x_2 x_3 \text{ (linkage disequilibrium function)}. \end{aligned}$$

We can obviously recapture the gamete frequency according to

Mating type	Frequency	Viability	Segregation
$\frac{AB}{AB}$	x_1^2	m_{11}	AB
$\frac{AB}{Ab}$	$2x_1 x_2$	m_{12}	$\frac{1}{2}AB + \frac{1}{2}ab$
$\frac{AB}{aB}$	$2x_1 x_3$	m_{13}	$\frac{1}{4}AB + \frac{1}{2}aB$
$\frac{AB}{ab}$	$2x_1 x_4$	m_{14}	$(1 - r)(\frac{1}{2}AB + \frac{1}{2}ab) + r(\frac{1}{2}Ab + \frac{1}{2}ab)$
$\frac{Ab}{Ab}$	x_2^2	m_{22}	Ab
$\frac{Ab}{aB}$	$2x_2 x_3$	m_{23}	$(1 - r)(\frac{1}{2}Ab + \frac{1}{2}aB) + r(\frac{1}{2}AB + \frac{1}{2}ab)$
$\frac{Ab}{ab}$	$2x_2 x_4$	m_{24}	$\frac{1}{2}Ab + \frac{1}{2}ab$
$\frac{aB}{aB}$	x_3^2	m_{33}	aB
$\frac{aB}{ab}$	$2x_3 x_4$	m_{34}	$\frac{1}{2}aB + \frac{1}{2}ab$
$\frac{ab}{ab}$	x_4^2	m_{44}	ab

TABLE 11

$$(8.5) \quad \begin{aligned} x_1 &= p_1 p_2 + D, & x_2 &= p_1(1 - p_2) - D, \\ x_3 &= (1 - p_1)p_2 - D, & x_4 &= (1 - p_1)(1 - p_2) + D. \end{aligned}$$

On the basis of (8.3) and (8.4) we obtain

$$(8.6) \quad p'_1 = p_1, \quad p'_2 = p_2, \quad D' = (1 - r)D$$

and therefore $D^{(n)} = (1 - r)^n D^{(0)} \rightarrow 0$ provided $r > 0$. Combining (8.6) with (8.5) we see that

$$\begin{aligned} x_1^{(n)} &= p_1 p_2 + D^{(n)} \rightarrow p_1 p_2 \text{ as } n \rightarrow \infty \\ x_2^{(n)} &\rightarrow p_1(1 - p_2), \text{ etc.} \end{aligned}$$

Letting $p^0(A)$ ($p^0(B)$) denote the initial frequency of the A gene (B gene) etc. we can express the limiting frequencies in the form limit frequency of

$$\begin{aligned}
 f^\infty(AB) &= x_1^{(\infty)} = p_1^{(0)}p_2^0 = p^{(0)}(A)p^{(0)}(B) \\
 (8.7) \quad f^\infty(Ab) &= p^0(A)p^0(b), \quad f^\infty(aB) = p^0(a)p^0(B) \\
 f^\infty(ab) &= p^0(a)p^0(b)
 \end{aligned}$$

so that the two loci act in the limit independently provided, recombination is positive.

2. Additive viabilities. This is the case where the fitness of a genotype is determined as the additive effects of the fitness contributed by each locus separately. Specifically, suppose $\sigma_1, \sigma_2, \sigma_3$ denote the relative fitnesses of AA, Aa, aa respectively and s_1, s_2, s_3 represent the relative fitnesses of BB, Bb and bb respectively. Then m_{11} the fitness of AB/AB is $\sigma_1 + s_1$, the sum of the fitnesses of AA and BB . Similarly, m_{14} for AB/ab is $\sigma_2 + s_2$ and m_{24} of Ab/ab is $\sigma_2 + s_3$, etc.

In the case of additive fitnesses and heterozygote advantage at each locus, i.e., $\sigma_2 > \max(\sigma_1, \sigma_3)$ and $s_2 > \max(s_1, s_3)$, it can be proved that the limiting gamete frequencies are

$$(8.8) \quad \lim_{n \rightarrow \infty} x_1^{(n)} = \hat{p}_1 \hat{p}_2, \quad \lim_{n \rightarrow \infty} x_2^{(n)} = \hat{p}_1(1 - \hat{p}_2) \text{ etc.,}$$

where

$$\hat{p}_1 = \frac{\sigma_2 - \sigma_3}{2\sigma_2 - \sigma_1 - \sigma_3}, \quad \hat{p}_2 = \frac{s_2 - s_3}{2s_2 - s_1 - s_3}$$

valid for any initial frequency vector $(x_1^0, x_2^0, x_3^0, x_4^0)$ provided $x_1^0 x_2^0 x_3^0 x_4^0 > 0$.

Other examples of viability arrays that can be mostly analyzed include the cases of multiplicative viabilities and the symmetric viability model (e.g., see Bodmer and Felsenstein [3], Kojima and Lewontin [27] and Karlin and Feldman [19]).

Work supported in part by the National Institute of Health Grant USPRS 10452-09.

References

1. W. F. Bodmer, (1960), Discrete stochastic processes in population genetics, *J. Roy. Stat. Soc. B* 22: 218-244.
2. W. F. Bodmer, (1965), Differential fertility in population genetic models, *Genetics* 51: 411-424.
3. W. F. Bodmer and J. Felsenstein, (1967), Linkage and selection: Theoretical analysis of the deterministic two locus random mating model, *Genetics* 57: 237-265.
4. C. Cannings, (1967), Equilibrium, convergence and stability at a sex-linked locus under natural selection, *Genetics* 56: 613-617.
5. C. Cannings, (1969), Unisexual selection at an autosomal locus, *Genetics* 62: 225-229.

6. L. Cavalli Sforza and W. Bodmer, (1971), *Genetics of Human Population*, Freeman, San Francisco.
7. J. F. Crow and M. Kimura, (1970), *An introduction to population genetics theory*, Harper & Row, New York.
8. E. M. East, (1940), The distribution of self sterility in flowering plants, *Proc. Amer. Phil. Soc.*
9. I. Eshel, (1971), On evolution in a population with an infinite number of types, *Theor. Pop. Biol.* 2: 209-236.
10. W. J. Ewens, (1972), The sampling theory of selectively neutral genes, *Theor. Pop. Biol.*, March 1972.
11. J. Felsenstein, (1971), On the loss of alleles from a Haploid population, *Theor. Pop. Biol.*, December 1971.
12. D. J. Finney, (1952), The equilibrium of a self-incompatible polymorphic species, *Genetica* 26: 33-64.
13. R. A. Fisher, (1941), The theoretical consequence of polyploid inheritance for the mid style form of *lythrum salicaria*, *Ann. Eugen.* 11: 31-38.
14. J. B. S. Haldane, (1924), A mathematical theory of natural and artificial selection, *Biol. Proc. Camb. Phil. Soc.*, *Biol. Sci.* 1: 158-163.
15. J. B. S. Haldane, (1932), *The causes of evolution*, Harper & Row, New York.
16. S. Karlin, (1968), Equilibrium behavior of population genetic models with non-random mating, Part I: Preliminaries and special mating systems, *J. Appl. Prob.* 5: 231-313.
17. S. Karlin, (1968), Equilibrium behavior of population genetic models with non-random mating, Part II: Pedigrees, homozygosity, and stochastic models, *J. App. Prob.* 5: 487-566.
18. S. Karlin and F. M. Scudo, (1969), Assortative mating based on phenotype, II. Two autosomal alleles without dominance, *Genetics* 63: 499-510.
19. S. Karlin and M. W. Feldman, (1970), Linkage and selection: The two locus symmetric viability model, *Theor. Pop. Biol.* 1: 39-72.
20. S. Karlin, (1970), Lecture notes on mathematical genetics, Notes at Weizmann Institute.
21. S. Karlin and J. McGregor, (1971), On mutation selection balance for two-locus haploid and diploid populations, *Theor. Pop. Biol.* 2: 60-70.
22. S. Karlin and J. McGregor, (1972), Addendum to a paper of Ewen, *Theor. Pop. Biol.*, March 1972.
23. S. Karlin, On some genetic models with an infinite number of types involving mutation selection balance, (to appear).
24. O. Kempthorne, *An Introduction to Genetic Statistics*, Wiley, New York, 1957.
25. J. F. C. Kingman, (1961), A mathematical problem in population genetics, *Proc. Camb. Phil. Soc.* 57: 574-582.
26. K. Kojima, (1970), Volume Edited "Mathematical Topics in Population Genetics". *Bio-mathematics Vol. 1*. Springer-Verlag, New York.
27. K. Kojima, and R. C. Lewontin, (1970), Evolutionary significance of linkage and epistasis, *Mathematical Topics in Population Genetics*, Vol. edited K. Kojima, *Biomathematics Vol. 1*. Springer-Verlag, New York.
28. R. C. Lewontin, (1968), The effect of differential viability on the population dynamics of t alleles in the house mouse, *Evolution* 22: 262-273.
29. F. M. Scudo, (1964), Sex population genetics, *La Ricerca Scient.* 34: 93-146.
30. F. M. Scudo and S. Karlin, (1969), Assortative mating based on phenotype I, Two alleles with dominance, *Genetics* 63:479-498.
31. M. Slatkin, (1970), Selection and polygenic characters, *Proc. Nat. Acad. Sci.* 87-93.
32. C. Stern, (1960), *An introduction to human population genetics*, Freeman, San Francisco.