

# EQUILIBRIA FOR GENETIC SYSTEMS WITH WEAK INTERACTION

SAMUEL KARLIN  
WEIZMANN INSTITUTE OF SCIENCE, ISRAEL  
and  
STANFORD UNIVERSITY  
and  
JAMES MCGREGOR  
STANFORD UNIVERSITY

## 1. Introduction

The following principle (stated here in rough form) bears many applications in the study of ecological and genetic systems.

**PRINCIPLE I.** *If a system of transformations acting on a certain set (in finite dimensional space) has a "stable" fixed point, then a slight perturbation of the system maintains a stable fixed point nearby.*

The theme of this principle is quite intuitive although care in its application and interpretation is vital. Its validity does not require the stability hypothesis to apply in a geometric sense. In fact, for numerous important nonrandom mating genetic models the stability of the relevant equilibrium is manifested only in an algebraic sense. The result is basic in the domain of global analysis and occurs in many other mathematical contexts as well. With the aid of this principle, we are able to establish the existence of equilibria for quite complicated genetic models and these have interesting interpretations for population phenomena.

A converse proposition to Principle I of considerable value in ascertaining all possible equilibria is also now stated in rough form. For a precise mathematical statement the reader should consult Karlin and McGregor [9].

**PRINCIPLE II.** *If  $f(x)$  is a differentiable transformation acting on a certain set  $S$  (in finite dimensional space) having a finite number of fixed points, say  $y_1, y_2, \dots, y_r$ , with the property that the linear approximation to  $f(x)$  in the neighborhood of each fixed point has no eigenvalue of absolute value one, then a slight differentiable perturbation of  $f(x)$  maintains at most a single fixed point  $z_i \in S$  in the neighborhood of each  $y_i$ . Moreover,  $z_i$  is locally stable if and only if  $y_i$  is locally stable.*

It is worth noting that some fixed points of  $f(x)$  (but none of the stable ones) may disappear under small perturbations.

We illustrate the scope of these principles by indicating the application to the investigation of three types of population models subject to a variety of genetic

factors. In Section 2, we treat the effect of small intermigration flow among several genetic subsystems or niches with different selection forces operating in the separate niches. Section 3 highlights the possibilities of mutation selection balance for two locus haploid and diploid populations. The small perturbation in this model arises due to the mutation pressure. The final section reviews results emanating from Principles I and II applied to the study of multilocus viability selection models with small recombination parameters. Some implications for evolutionary theory are also noted.

## 2. Stable equilibrium for multipopulation systems with migration coupling

In this section, we present a general framework for application of Principles I and II for population models involving slight intermigration flow among several separate subpopulations. More specifically, suppose there exists a number (say  $\rho$ ) of ecological or genetic systems  $\mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_\rho$ , for example, separate communities, niches, with a finite number (say  $r$ ) of possible types  $A_1, A_2, \dots, A_r$  that may be represented in each system. We generally denote the frequencies of types  $A_1, A_2, \dots, A_r$  in population (or system)  $\mathcal{O}_\alpha$  by  $\bar{p}_\alpha = (p_{\alpha 1}, \dots, p_{\alpha r})$ , and frequently subscript  $\alpha$  is suppressed when no ambiguity of interpretation is possible. Suppose each system reproduces independently in some fashion such that the frequencies  $\bar{p}' = (p'_{\alpha 1}, \dots, p'_{\alpha r})$  in the next generation are determined by the relations

$$(2.1) \quad p'_{\alpha j} = f_{\alpha j}(p_{\alpha 1}, \dots, p_{\alpha r}), \quad j = 1, 2, \dots, r.$$

We sometimes write (2.1) in vector notation taking the form

$$(2.2) \quad \bar{p}'_\alpha = \bar{f}_\alpha(\bar{p}_\alpha).$$

In most genetic models, the transformation (2.1) is displayed as a ratio of two algebraic polynomials in the frequency variables. These transformations naturally reflect mating and ecological behavior, segregation pattern, selection, migration and mutation pressures, temporal and spatial (cyclical or other) changes, the influence of recombination when more than one locus is involved, and other relevant factors of the process.

In each system  $\mathcal{O}_\alpha$  there usually exist certain equilibria (invariant points under the transformation (2.1)) which are locally stable. The collection of equilibria include polymorphic (all types represented) and peripheral (that is, boundary equilibrium) points, where in the latter case some types are not represented. Local stability is to be understood in the following generalized sense. A frequency vector  $p^*$  is said to be locally stable if for any prescribed neighborhood  $U$  of  $p^*$  there exists another neighborhood  $V$ ,  $p^* \in V \subset U$  such that  $f(\bar{V}) \subset V$  ( $\bar{V}$  denotes the closure of  $V$ ), and therefore the iterates of  $\bar{f}_\alpha^{(n)}(\bar{p}) = \bar{f}_\alpha^{(n-1)}(\bar{f}_\alpha(\bar{p}))$  for any starting point  $\bar{p} \in \bar{V}$  never depart from  $V$ . In most cases, local stability of an equilibrium  $p^*$  actually entails that if the initial frequency vector  $\bar{p}$  is sufficiently close to  $p^*$ , the iterates  $\bar{f}_\alpha^n(\bar{p})$  indeed converge to  $p^*$ .

The notion of stability prescribed above makes no stipulations on the rate of convergence to the equilibrium. However, in most genetic and ecological systems

when the equilibrium expresses a stable polymorphic balance, then convergence takes place at a geometric rate. But on the other hand when the equilibrium is of the boundary kind (mostly a population of a single type), then convergence not uncommonly occurs at an algebraic rate.

Suppose now that the system  $(\mathcal{O}_1, \dots, \mathcal{O}_\rho)$  is coupled by some form of interaction. To fix the ideas, we consider here the example when there are migration coefficients  $m_{\alpha\beta}$  with the interpretation that after reproduction a proportion  $m_{\alpha\beta}$  of individuals from  $\mathcal{O}_\beta$  migrate to  $\mathcal{O}_\alpha$ . The recursion relations describing the evolution of type frequencies in the migration coupled system are then

$$(2.3) \quad p'_{\alpha i} = \sum_{\beta=1}^{\rho} m_{\alpha\beta} f_{\beta i}(p_{\beta 1}, \dots, p_{\beta r}), \quad \alpha = 1, \dots, \rho; i = 1, \dots, r.$$

It is naturally assumed that  $m_{\alpha\beta} \geq 0$  and  $\sum_{\beta=1}^{\rho} m_{\alpha\beta} = 1$ .

The migration specified by the  $\rho$  square matrix  $M = ||m_{\alpha\beta}||$  would be called weak if  $M$  is sufficiently close to the identity matrix, that is, the  $m_{\alpha\beta}$  with  $\alpha \neq \beta$  are all sufficiently small. In this event, the flow between systems is slight. (All subsystems are presumed in the present theory to be of large size.) *If the uncoupled system has a locally stable equilibrium point  $p^*$  (that is, set of equilibrium vectors  $\bar{p}_\alpha^*$ ), then we expect the coupled system to have a locally stable equilibrium point  $q^*$  near  $p^*$  provided the migration coupling is sufficiently weak.*

The above statement is a special case of the general Principle I.

We say a full polymorphism is attainable in the uncoupled system  $(\mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_\rho)$  if there is a set of equilibrium frequency vectors  $p_\alpha^*$ ,  $\alpha = 1, 2, \dots, \rho$ , where  $p_\alpha^*$  is a locally stable solution of

$$(2.4) \quad p_\alpha^* = f_\alpha(p_\alpha^*), \quad \alpha = 1, 2, \dots, \rho,$$

such that for each  $i$ ,  $1 \leq i \leq r$ ,  $p_{\alpha i}^*$  is positive for at least one  $\alpha$  ( $\alpha$  may depend on  $i$ ). The set of equilibrium frequency vectors  $p^*$  comprise a fixed point  $p_\alpha^*$  of the uncoupled system in which every possible type is represented in at least one subsystem. Now if  $q^*$  is the nearby equilibrium point of the system under weak coupling, then every type will still be represented in at least one subsystem provided the coupling is sufficiently weak. In the migration coupling example, there is a simple condition which will guarantee that at the equilibrium state  $q^*$  each possible type is actually represented in every subsystem. The matrix  $M = ||m_{\alpha\beta}||$  is called irreducible if there is a power  $M^k$  whose elements are all strictly positive. This means that a  $k$ th generation descendent of an individual from any subsystem has a positive probability to be in any other subsystem.

Our first application of Principle I is the following. *If a full polymorphism (defined in paragraph above) is attainable in the uncoupled system and if the migration matrix is irreducible and sufficiently weak (that is,  $M$  close to  $I =$  identity matrix), then the coupled system has a locally stable equilibrium state in which actually every possible type is represented in every subsystem.*

As a simple application of the assertion stated above, consider the case of a simple genetic system involving two alleles  $A_1$  and  $A_2$  at a single locus with two niches with selection coefficients  $1, 1 - \sigma, 1$  for the genotypes  $A_1A_1, A_1A_2$ , and

$A_2A_2$ , respectively, or more generally, we can suppose that the fitness coefficients of the genotypes are  $s_1, s_2, s_3$  in niche 1;  $\sigma_1, \sigma_2, \sigma_3$  in niche 2, where  $s_2 < \min(s_1, s_3)$  and  $\sigma_2 < \min(\sigma_1, \sigma_3)$ . Thus, in each separate niche disruptive selection operates and ordinarily the population would be fixed. However, invoking the above result, we find that if a small fraction  $m$  (necessarily small) of the population of each niche migrates to the other niche then there are possible sets of stable polymorphisms with both alleles represented in each population. Of course, the possibility of stable global fixations also exists.

The polymorphic equilibria of this example have the property that a preponderance of one homozygote occurs in one niche while a preponderance of the alternative homozygote is maintained in the second niche. It is reasonable to speculate that some forms of habitat selection confer an advantage, say on  $A_2A_2$  in niche 2, on  $A_1A_1$  in niche 1, while the heterozygote (or hybrid type) bears marked disadvantage to both homozygotes in each of the niches, yet a global balance is preserved.

What evolves depends crucially on the initial composition of all the subpopulations. Thus, whether fixation transpires or polymorphism is attained could be a function of founders and random fluctuation effects determining the initial conditions. Small colonies of different homozygotes could inhabit neighboring localities with selection favoring both homozygotes over the heterozygotes in each locality. Subsequently, population size grows and presumably some slight gene flow binds the two localities. A suitable application of the principle then points to a polymorphism with most existing types being homozygotes (see Karlin and McGregor [10] for a more detailed quantitative analysis of this two niche model).

The extension to the case of three alleles is as follows. Consider a three allele model involving alleles  $A_1, A_2, A_3$  with viabilities of  $A_iA_j$  specified by the matrix with the obvious interpretation

$$(2.5) \quad \begin{pmatrix} 1 + \alpha_1 & 1 - \epsilon_1 & 1 - \epsilon_2 \\ 1 - \epsilon_1 & 1 + \alpha_2 & 1 - \epsilon_3 \\ 1 - \epsilon_2 & 1 - \epsilon_3 & 1 + \alpha_3 \end{pmatrix}, \quad \epsilon_i, \alpha_i > 0; \quad i = 1, 2, 3,$$

so that each homozygote is favored. Consider two replicate systems of the above structure with slight migration between them. It can be proved that there exists no stable polymorphism with all genotypes represented. (The proof involves the converse version of Principle I, that is, Principle II.) However, if the above three allele genetic population is replicated in *three* systems with slight gene flow between them, then a stable polymorphism is possible involving all types. (The proof is accomplished by application of Principle I.)

### 3. On mutation selection balance for two locus models with small mutation rate

The haploid model considered is the traditional one. The parameters are listed in Table I. The recombination fraction is denoted by  $r$ . The mutation rate of

TABLE I

Gamete	$AB$	$Ab$	$aB$	$ab$
Fitness coefficient	$\sigma_1$	$\sigma_2$	$\sigma_3$	1
Frequencies in a given generation of haploid individuals	$x_1$	$x_2$	$x_3$	$x_4$

$a \rightarrow A$  or  $b \rightarrow B$  is  $\mu$  in both cases with mutations occurring independently at each locus. (All our results carry over as well for different mutation rates at each locus; specifically, rate  $\mu_1$  for  $a \rightarrow A$  and  $\mu_2$  for  $b \rightarrow B$ .) Thus, we are assuming unidirectional mutation to the  $AB$  gamete.

For definiteness, we postulate as in most of the literature cited (although irrelevant to the qualitative conclusions deduced in this paper) that the effects occur in the order mutation  $\rightarrow$  random union of gametes  $\rightarrow$  segregation  $\rightarrow$  selection. Thus, the population can be envisioned as consisting of mature haploids who produce gametes to be fertilized at which stage mutations occur. After segregation, selection operates.

The recursion relations connecting the gamete frequencies in two successive generations  $(x_1, x_2, x_3, x_4) \rightarrow (x'_1, x'_2, x'_3, x'_4)$  are derived in the standard way yielding

$$(3.1) \quad \begin{aligned} Wx'_1 &= \sigma_1\{x_1 + \mu(x_2 + x_3) + \mu^2x_4 - r(1 - \mu)^2D\}, \\ Wx'_2 &= \sigma_2\{(1 - \mu)x_2 + \mu(1 - \mu)x_4 + r(1 - \mu)^2D\}, \\ Wx'_3 &= \sigma_3\{(1 - \mu)x_3 + \mu(1 - \mu)x_4 + r(1 - \mu)^2D\}, \\ Wx'_4 &= (1 - \mu)^2x_4 - r(1 - \mu)^2D, \end{aligned}$$

where  $D = x_1x_4 - x_2x_3$  (the disequilibrium expression) and  $W$ , as usual, stands for the sum of the right members of the four equations and is a quadratic function of the variables  $x_1, x_2, x_3, x_4$ .

We shall deal mostly with the situation

$$(3.2) \quad 1 > \sigma_2, \quad 1 > \sigma_3, \quad 1 < \sigma_1,$$

that is, a single mutation has a deleterious effect while a double mutant  $AB$  is endowed with a selective advantage relative to the wild type  $ab$ . The assumption (3.2) is a major case of interest in Crow and Kimura [2], Eshel and Feldman [3], especially relevant to their discussions concerning the advantages of recombination in evolution.

Let the conditions (3.2) prevail; that is,  $\sigma_2 < 1$ ,  $\sigma_3 < 1$ ,  $\sigma_1 > 1$ . If  $(1 - r)\sigma_1 < 1$ , then there exists a positive  $\mu_0$  such that for  $\mu$  satisfying  $0 < \mu < \mu_0$  there exists a stable polymorphism  $\bar{x}^* = (x_1^*, x_2^*, x_3^*, x_4^*)$  of the system (that is,  $x_i^* > 0$ ,  $i = 1, \dots, 4$ ) satisfying  $x_1^* + x_2^* + x_3^* \leq \epsilon_0$ , where  $\epsilon_0(\mu_0)$  tends to zero as  $\mu_0$  tends to zero. When  $(1 - r)\sigma_1 > 1$  holds, fixation of the  $AB$  gamete occurs independent of the rate of mutation.

We can write the transformations (3.1) in the form

$$(3.3) \quad x'_i = f_i(x_1, x_2, x_3) + \mu g_i(x_1, x_2, x_3, \mu), \quad i = 1, 2, 3,$$

where  $0 \leq g_i(x_1, x_2, x_3, \mu) \leq C$  and  $C$  is independent of  $\mu$  and  $x_i$ .

The proof then involves an appropriate application of Principles I and II. We refer to Karlin and McGregor [8] for details and further discussion of the implications of the results. There are versions of the above result valid also in the corresponding diploid model.

The result of our Theorem 1 also has relevance to the finding of Feldman [4] that in a haploid two locus population, selection pressures alone cannot maintain a stable polymorphism. Thus, to achieve polymorphism some other influences apart from selection pressures should be operating.

Along these lines, it is implicit in the work of Raper and others that random mating is not applicable to a number of haploid models of fungi populations. In these cases, certain incompatibility mechanisms are in force. For such two locus haploid populations, a stable polymorphism can be maintained by the balance of the force of selection in conjunction with the incompatibility mating behavior.

We can also prove that selection coupled with certain assortative (and not only disassortative or incompatibility) patterns of mating can produce stable polymorphisms for a two locus haploid population. Also, *multiniche* two locus haploid populations subject only to selection forces can exhibit stable polymorphism (see Karlin and McGregor [9]).

#### 4. Applications of the basic principle to the study of multilocus models with small recombination parameters

We record some results on multilocus genetic models deduced by appropriate application of Principle I. The rigorous proofs and further developments on multilocus phenomena will be set forth in a separate publication.

Consider first a two locus diploid population. The notation adopted is the traditional one (Bodmer and Felsenstein [1]). The selection parameters of the ten genotypes are listed below

$$(4.1) \quad \begin{array}{ccc} & BB & Bb & bb \\ AA & \left( \begin{array}{ccc} w_{11} & w_{12} & w_{22} \\ w_{13} & w_{14} & w_{24} \\ & w_{23} & \\ w_{33} & w_{34} & w_{44} \end{array} \right) & , & w_{14} = w_{23}, \\ Aa & & & \\ aa & & & \end{array}$$

such that the subscripts 1, 2, 3, 4 refer to the gametes  $AB$ ,  $Ab$ ,  $aB$ , and  $ab$ , respectively, and  $w_{ij}$  is the fitness parameter of the genotype composed from gametes  $i$  and  $j$ . It is convenient to write the parameters in the form

$$(4.2) \quad \begin{pmatrix} 1 - \alpha_1 & 1 - \beta_1 & 1 - \alpha_2 \\ 1 - \beta_2 & 1 & 1 - \beta_3 \\ 1 - \alpha_3 & 1 - \beta_4 & 1 - \alpha_4 \end{pmatrix}.$$

We will concentrate on the situation where  $0 < \alpha_i < 1$ ,  $0 < \beta_i < 1$ ,  $i = 1, 2, 3, 4$ , so that the two double heterozygotes have equal viability and are most fit. The unperturbed system is that where recombination is zero (complete linkage). The model then reduces to a standard four allele viability set up. Principle I can be applied where the small mixing parameter is the recombination fraction  $r$ . The following result can be achieved.

**THEOREM 1.** *If the viability matrix (4.2) satisfies one of the following:*

(i) *symmetric viability pattern* (that is,  $\alpha_1 = \alpha_4$ ,  $\alpha_2 = \alpha_3$ ,  $\beta_1 = \beta_4$ ,  $\beta_2 = \beta_3$ , so that the viabilities are symmetric with respect to the two loci and the labeling of the alleles);

(ii) *multiplicative viabilities* (that is, the selection coefficients are determined as multiplicative effects of the viabilities at each locus);

(iii) *additive viabilities* (that is, the selection coefficients result as the additive effects of the viabilities at each locus), *then for  $r$  sufficiently small there exists at least one and at most two locally stable polymorphic equilibria; and*

(iv) *more generally, for  $r$  small enough and any viability pattern there exists at most two stable polymorphic equilibria.*

It is conjectured that the bound 2 on the number of stable polymorphic equilibria persists for all  $r$ .

Some additional special information applies in the situations of (i) and (ii).

(a) In the additive viability model with overdominance (meaning here that the double heterozygotes are most fit), there exists a unique interior stable polymorphism exhibiting linkage equilibrium with global convergence to this equilibrium occurring from any initial composition involving all gametes. This occurs for all recombination values  $r > 0$  (see Karlin and Feldman [7]).

(b) By appropriate choices of the selection parameters for the case of multiplicative viabilities, we always have two stable internal equilibria provided  $r$  is near 0. On the other hand, when  $r$  is near  $\frac{1}{2}$  Moran [12] established global convergence to an internal equilibrium which is in linkage equilibrium.

*Extensions to multilocus model.* Consider the corresponding model of  $k$  loci and assume all complete multiple heterozygotes have equal viability and are of superior fitness to all other genotypes. The following general result prevails. *There exists a prescription of viability parameters leading to  $2^{k-1}$  distinct locally stable polymorphic equilibria. Moreover, for the multiplicative viability model if all the pairwise recombination parameters are sufficiently small then there exists at most  $2^{k-1}$  distinct stable polymorphic equilibria. Any even number  $\leq 2^{k-1}$  can occur.*

These results suggest that there are generally more cases of stable polymorphism for multilocus selection models involving tight linkage than for weak linkage. The relevant polymorphisms for small recombination parameters are usually in substantial linkage disequilibrium, where a preponderance of a few special chromosome types are abundantly present. These facts argue for the conclusion that if variability (latent or actual) is desired for a population to cope with a multitude of environments, then recombination reducing mechanism may be favored in order to produce increased possibilities of polymorphism. Actually,

tight linkage is a vehicle for maintaining diverse forms of polymorphism and, of course, also serve for easy transcription of a series of actions (biochemical or otherwise) controlled by a sequence of closely linked loci. On the other hand, loose linkage or large recombinations parallel more the effects of mutation and sexuality. Moderate recombination appears to serve less well both the objectives of maintaining polymorphism and/or simple regular gene transcription.

The analysis of multilocus models provides an important application of the general principle of Section 2. Principle I can be interpreted as a perturbation or continuity theorem. Starting with a given genetic system for which the nature of the equilibria can be fully delineated (for example, the classical multiallelic viability model), it is desired to investigate a perturbed version of the model. The perturbing factors can be in the form of small mutation and/or migration pressures, weak selection effects, slight seed load or some other genetic carry over from previous generations or small recombination effects superimposed on a multiallelic selection model corresponding to a multilocus situation with no recombination. In this last case, we have established the remarkable fact that in a  $k$  locus genetic model with multiplicative viability selection coefficients, provided all pairwise recombination parameters are sufficiently small, there exist at most  $2^{k-1}$  stable polymorphic equilibria and for certain specifications of the selection coefficients this upper bound is achieved. This result underscores the increased potentialities for polymorphism corresponding to large cistrons or supergene complexes with slight intragenic recombination present.

The validation of these assertions for multilocus selection genetic model is not accomplished by determining explicitly the actual equilibria in the perturbed system and testing their stability properties (undoubtedly a prohibitive task), but rather the procedure exploits continuity methods coupled with the implicit function theorem and certain fixed point theorems.

The assertions of this section, their proofs, and implications for evolutionary theory will be elaborated elsewhere.

#### REFERENCES

- [1] W. F. BODMER and J. FELSENSTEIN, "Linkage and selection: theoretical analysis of the deterministic two locus random mating model," *Genetics*, Vol. 57 (1967), pp. 237-265.
- [2] J. F. CROW and M. KIMURA, "Evolution in sexual and asexual populations," *Amer. Natur.*, Vol. 99 (1965), pp. 439-450.
- [3] I. ESHEL and M. W. FELDMAN, "On the evolutionary effect of recombination," *Theor. Pop. Biol.*, Vol. 1 (1970), pp. 88-101.
- [4] M. W. FELDMAN, "Equilibrium studies of two locus haploid populations with recombination," *Theor. Pop. Biol.*, Vol. 2 (1971), pp. 299-318.
- [5] I. FRANKLIN and R. C. LEWONTIN, "Is the gene the unit of selection?," *Genetics*, Vol. 65 (1970), pp. 707-734.
- [6] S. KARLIN and M. W. FELDMAN, "Linkage and selection: the two locus symmetric viability model," *Theor. Pop. Biol.*, Vol. 1 (1970), pp. 39-72.
- [7] ———, "Convergence to equilibrium of the two locus additive viability model," *J. Appl. Probability*, Vol. 7 (1970), pp. 262-271.



- [8] S. KARLIN and J. MCGREGOR, "On mutation selection balance for two-locus haploid and diploid populations," *Theor. Pop. Biol.*, Vol. 2 (1971), pp. 60-70.
- [9] ———, "Polymorphisms for genetic and ecological systems with weak coupling," *Theor. Pop. Biol.*, Vol. 3 (1972), pp. 210-238.
- [10] ———, "Application of methods of small parameters to multi-niche population genetic models," *Theor. Pop. Biol.*, Vol. 3 (1972), pp. 186-209.
- [11] ———, "Polymorphisms for multi-locus selection system with tight linkage," to appear, 1973.
- [12] P. A. P. MORAN, "On the theory of selection dependent on two loci," *Ann. Hum. Genet.*, Vol. 32 (1968), pp. 183-190.