COO-2472-T1

Three Year Progress Report

December 1, 1974 - November 30, 1977

A Study of Mathematical Models of Mutation
and Selection in Multi-locus Systems

MASTER

R. C. Lewontin

Museum of Comparative Zoology
Harvard University
Cambridge, Massachusetts 02138

9506472

ERDA Contract No. EY-76-S-02-2472.*000

Document No. COO-2472-T1

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

for

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

I. Research Accomplished

Since its inception, this contract has had the rather broad objective of investigating the theory of genetic changes in populations when more than a single gene is taken into account. In the early stages of the contract, there was a very strong emphasis on the primary theory of selection in multi-locus systems, but as this primary work was finished, we have diversified our studies to include environments that fluctuate in time and vary in space, direct selection on phenotypic characters and a large number of problems in the general theory of population genetics and ecological genetics. The research topics covered in the last three years have put primary emphasis on variable environment and on the phenotype in selection. We have also returned to a consideration of the evolution of single loci with multiple alleles where we felt there were serious lacunae in our understanding of these simplest genetic systems. A good deal of the mathematical, as opposed to the more numerical work, has been carried out by Research Associates and graduate students, notably Dr. M. Slatkin, who was late a Co-Principal Investigator, and Dr. Russell Lande who, as will be seen below, has made notable contributions to the theory of phenotypic selection. The Principal Investigator has concentrated more on numerical and simulation studies, but there has been, in general, coordination in our attacks on various problems. A good example is our most recent work (see below) on multiple allelic systems where the Principal Investigator, using chiefly computer methods, found results that were then supported by approximate analytic methods worked out in collaboration with a mathematician, Dr. Ginzburg, and a mathematical physicist now working in ecology, Dr. Tuljapurkar. We believe that the diversity of people associated with the contract over time has added a great deal to its productivity.

A. Spatial variation in environments

1) M. Slatkin studied the effect of clinal variation in selection pressures on the genetic structure of natural populations segregating for alleles at two linked loci. He has shown that linkage between loci could either increase or decrease a population's ability to respond to changes in local selection pressures. If the selection on the two loci changes at the same geographical location, then linkage would produce a steeper cline in frequencies at both of the loci. If the selection changes at different locations, then linkage will interfere with the population's response to the selection and the gene frequencies will not reflect the spatial pattern in selection. According to the general theory of Karlin and McGregor (1974), there would be selection for or against an allele which decreased the recombination fraction between the two loci depending on the spatial pattern of the selection intensities.

Slatkin has also shown that gene flow and selection can produce a significant amount of linkage disequilibrium even in the absence of epistasis. When the recombination fraction between the two loci is of the same order of magnitude as the selection coefficient for either of the loci or less, then there would be expected to be a measurable amount of linkage disequilibrium in the middle of the cline. The degree of epistasis between the loci would be relatively unimportant in determining the general properties of the system. This is in contrast to the results from the model of linkage and epistasis in a single panmictic population (Lewontin and Kojima, 1960) in which there could be linkage disequilibrium maintained only if the recombination fraction were less than the epistatic deviation from additive fitness. When the fitnesses between the loci were multiplicative, this would require that the recombination fraction be less than the square of the selection coefficients, while with gene flow only the

first power of the selection coefficient is important. Thus, Lewontin's (1974) statement that the presence or absence of linkage disequilibrium in a population is quite a sensitive test of epistasis is true only in a panmictic population.

2) An important new result grew out of Slatkin's visit to the National Institute of Genetics in Misima, Japan. While there, he worked with M. Kimura, T. Maruyama, and J. F. Crow (who was visiting from the Department of Genetics at the University of Wisconsin, Madison). Slatkin learned several techniques for analyzing stochastic changes in gene frequencies and applied these techniques to several problems of interest. He analyzed the rate of response of one population to gene flow from a neighboring population and found that there is always a time lag associated with gene flow. Although the two populations would be genetically identical at equilibrium, for a period of several hundred generations or more, they could be genetically distinct. Slatkin argues that when the expected time until extinction of local populations is on the same order of magnitude as the time lag, then the equilibrium might never be reached. If there was a large number of such populations, genetic differences could accumulate between them even though there were sufficient gene flow to prevent divergence at equilibrium.

There was a fruitful collaboration on related problems with T. Maruyama. One is on the effect of genetic drift in a cline. Slatkin and Maruyama have shown that genetic drift will always tend to reduce the slope of a cline, sometimes by as much as $1/2$, but that the selection in the cline makes little change in the correlation in gene frequencies predicted in the absence of selection by Kimura and Weiss (1964). Slatkin and Maruyama also analyzed the effect of a small amount of gene flow on measures of genetic distance between the populations. They show that even a small amount of gene flow is sufficient

to significantly change the estimates of divergence times of two populations based on genetic distance. The most important part of their result is that the error in the estimate of divergence times does not depend strongly on the number of steps separating the two populations. They suggest that a comparison of the estimated and actual divergence times of the two populations would provide an estimate of the average level of gene flow between them since divergence.

B. Temporal Variation in Environments

1) Multi-locus simulation

We have written a special 4-locus simulation program, designed to follow the effect of a fluctuating environment on a multi-locus system with linkage. The present model of selection is an optimum deviation model with fitness falling as phenotype deviation from an optimum according to a Gaussian function:

$$\bar{w}(x) = K e^{-(x-0)^2/w^2}$$

The phenotype, x , of any genotype is determined by an additive and a dominance value for each locus and phenotype is additive across loci, although this is easily altered on the program to produce epistasis on the primary scale of phenotype. The optimum, 0 , can be varied from generation to generation in a variety of patterns. If the amplitude of the fluctuation is great enough, then population is being selected directionally in any given generation, while if the amplitude of the fluctuation is within the range of the most extreme phenotypes, then there is optimizing selection in each generation, but with a fluctuating optimum. The period of the fluctuation will determine how far toward the new mean the population will evolve, before being selected for an alternative optimum.

The program was checked with old results by setting the amplitude of

environmental fluctuation to zero, and the result was the same as previous work on the quadratic deviation model had produced. There is fixation of alleles, half the loci fixing at the + allele and half at the - allele to produce an intermediate homozygote. When linkage is very tight, this fixation is extremely slow so that a quasi-equilibrium is obtained at intermediate allelic frequencies. At the same time, the fitness in the tightly linked case goes up very rapidly to unity because of the formation of repulsion linkages, while the fitness in the loosely linked case goes up very slowly as genes go to fixation. All of the phenomena occur in the Gaussian fitness model just as they were previously shown by us for the quadratic deviations model.

We then caused the environment to fluctuate with various amplitudes and periods. Gene frequencies fluctuate but asymmetrically so that 3 of the loci oscillate to fixation at the 1 allele while one oscillates to fixation at the 0 allele, or, in extremely large amplitude environmental fluctuations, all four loci may go to fixation at the 1 allele. The direction of the asymmetry depends, of course, on the initial gene frequencies, so that which of the two alleles become fixed is arbitrary, but asymmetry is always observed when allele frequencies do not take perfectly symmetrical values initially. The second general effect is that tight linkage speeds up the increase in average fitness, as in the case of constant environment, and also retards the fixation of genes. Repulsion linkages are built up during the course of selection. As the amplitude of the fluctuation increases, the quasi-equilibrium of allelic frequencies in the tightly linked cases more nearly approaches the fixation obtained in the loosely linked case. For example, at an amplitude equal to the additive effect of a single gene, after 50 generations, the loosely linked case had allele frequencies of 0, .5, 1 and 1, while the tightly linked case had quasi-

equilibrium frequencies of .33, .59, .74 and .83. When the amplitude of environmental fluctuation was increased to two times the additive effect of a single gene, the loosely linked case had gene frequencies of 0, 1, 1 and 1, while in the tightly linked case these frequencies were .46, .76, .86 and .91. As the amplitude is further increased, the gene frequencies in both tightly and loosely linked areas continue to increase asymmetrically toward unity, so that when fluctuation is strong enough, allele loci fix at 1.0 within 3 cycles of fluctuation.

There are some peculiarities of the small number of loci in the model which result in the equilibrium fitness depending upon whether the amplitude is an integral multiple of gene effects. If it is not, then a stable polymorphism for one locus is established which lowers the mean fitness in half the cycles and raises it in the other half. The net result is a maximization of the geometric mean fitness over time (see below).

2) In a second case, it was assumed that the environmental changes which occurred randomly from generation to generation resulted in symmetric reversals of directional selection (the two fitness sets were mirror images of one another), affecting a single character, determined by four alleles at a single locus. In other words, if the character were size, in some generations the larger members of the population would enjoy a selective advantage, while in other generations the smaller members would. Heterozygotes were intermediate to their respective homozygotes regarding the character, with incomplete dominance. The population size was infinite and there were no mutations, two unrealistic assumptions which tend to cancel one another.

It turns out that a good summarizing statistic for the behavior of the population was the geometric mean of the two selective sets, which were imposed

on the population. This geometric mean was based on the frequencies of the sets for a given run, which was usually 50 to 100 generations. The population always moved to increase the frequency of the genotype with the highest geometric mean. Due to historical effects, namely the added importance of the last generations on the final gene frequencies, the geometric mean would not predict the final frequencies exactly, especially when the autocorrelations between generations was high.

It soon became apparent that two factors were especially important in their influence on the system. The first is the strength of selection. If the selective differences between genotypes were small, then an intermediate genotype would turn out to be optimal, even though the selective sets by themselves would push the population to one end or other of the character spectrum. Nonetheless, even if this intermediate were a heterozygote, there would still be some loss of genetic variability, with the other two genes of the four tending to be lost. The weaker the selective differences between genotypes, the more slowly the loss of variability. However, if we increase the selective differences, there comes a point where we obtain two optima separated by a valley, and in such cases the possibility of maintaining the polymorphism arises. Now this brings us to the second important consideration, which is the actual pattern of the fluctuations in time. It has just been mentioned that polymorphism seemed possible if the selective differences were strong enough, as can best be seen by plotting the geometric mean for the two selective sets, assuming equal frequency for the two. However, the geometric mean depends strongly on the frequencies involved, and in the case of the strong selective differences, it doesn't take much of a departure from 50:50 frequencies to lose the bimodality upon which the polymorphism depends. In a finite number of runs, an increase in the autocorrelation in the otherwise random choice of selective sets may shift

the final frequency of the selection sets away from the expected equality. When this occurs, the fact that the selection is strong will quickly shift the population to a single optimum, with a concurrent loss of variability.

Thus the conditions for genetic polymorphism with such a scheme seem to be either selective neutrality between genotypes, or strong selective differences, but, in the latter case, only when the environmental fluctuations are actually alternations, back and forth every generation.

3) Fluctuating environments and continuous characters

R. Lande has investigated a model of continuously varying traits in which mutation constantly replenishes the variation lost from stabilizing selection. Selection is, again, based on a Gaussian deviations model. It was previously shown with the model by Lande and Slatkin that in a constant environment, any decrease in mutation or recombination, which then decreases the genetic variance, gives a higher fitness. Lande had now examined two forms of varying environment for this model. In the first, the environment changes continuously in one direction. In this case, no matter how slowly the environment changes, there is an intermediate value of genetic variance that is stable, this value increasing as the optimum moves more quickly. There is, however, no threshold value for the rate of change of the environment, below which genetic variance is always selected against. This contrasts with the finding of Lande and Slatkin that in a random, uncorrelated environment, there is a critical value of the environmental variance in time, below which genetic variance is always disadvantageous.

The other form of varying environment was a regular oscillation, both square wave and sine wave with various amplitudes and periods. There is a minimum (threshold) amplitude of fluctuation necessary to select for genetic variance to be maintained in the population. Thus modifiers that increase recombination

or mutation would be selected if the amplitude exceeds this value. The longer the period of fluctuation, the lower is this threshold, approaching zero as the period grows extremely long, converging to the directional case. This is analogous to the stochastic environmental model (with correlation time analogous to period). Thus, long-term changes appear to be most important in selecting for non-zero genetic variance to be maintained in a population.

4) Mutation and phenotypic variation

Slatkin and Lande studied the determination of the variance of phenotypic characters subject to temporally varying selection. The evolutionarily stable equilibrium variance of a character subject to temporally varying, density independent, optimizing selection was found with different degrees of autocorrelation in the environment. They discovered that only if the variance in the optimal phenotype is above a certain threshold will selection favor a non-zero variance of the character. Also, they found that autocorrelation in the changes in the optimal phenotype tends to reduce the variance, indicating that a species could afford to specialize more in a more predictable environment.

Slatkin, in collaboration with Jonathan Roughgarden of Stanford University, has used the same type of analysis on density dependent optimizing selection to find the evolutionarily stable variance when intraspecific competition is a factor. Slatkin and Roughgarden have shown that selection on a modifier would tend not to maximize the total population size but rather to equalize the fitnesses in each of the phenotypic classes. They show that, in general, the total population size is not maximized at equilibrium.

R. Lande has investigated the effect of selecting for an intermediate phenotype value (stabilizing selection) when there is a constant input of new variation from mutation. In this and subsequent work, Lande has been able to

show that, to a first order of approximation, most of the behavior of the distribution of phenotypes can be predicted without recourse to a detailed genetic model. It is assumed that a character under stabilizing selection is determined genetically by n linked, mutable loci with additive effects and a range of many possible allelic effects at each locus. A general qualitative feature of such systems is that the genetic variance for the character is independent of the linkage map of the loci, provided linkage is not very tight. A particular detailed model shows that certain aspects of the genetic system are molded by stabilizing selection, while others are selectively neutral. With reference to experimental data on characters of Drosophila, maize and mice, it is concluded that large amounts of genetic variation can be maintained by mutation in polygenic characters even when there is strong stabilizing selection.

Using a similar approach involving the moments of the distribution of phenotypes, Lande has made an analysis of the deterministic and stochastic forces affecting the mean value of the phenotypic characters. Many evolutionary arguments have been weakened by the finding of Moran that deterministic changes in gene frequencies caused by natural selection do not necessarily increase the level of adaption or the average fitness of a population. Lande derived an alternative principle showing that under certain conditions with respect to changes in the average phenotypes of a set of correlated characters, evolution does maximize the average fitness of individuals in the population. These conditions are: 1) the phenotype variance must be independent of the mean and 2) no density or frequency dependence of the fitness values. These conditions also apply to Fisher's Fundamental Theorem. He also has developed a simple model of random genetic drift which applies to phenotypic characters under stabilizing selection. He has shown that considerable change in the mean value

of a character can be produced by genetic drift in populations with effective size in the tens or hundreds, within a timespan typical for speciation. Random genetic drift may thus play an important role in the differentiation of populations in the process of geographic speciation. Unlike in Wright's theory of random genetic drift in gene frequencies, this conclusion is valid even for completely isolated populations.

C. Patterns of Phenotypic Variation

R. Lande and R. Lewontin have investigated the behavior of the coefficient of variation statistic in two cases: 1) the coefficient of variation of measurements of different dimensionality, as, for example, length, area, and volume and 2) coefficient of variation of parts and wholes. Their finding is that the coefficient of variation in the first case has an upper limit of kV , where k is the dimensionality of the measurement and V is average coefficient of variation of the linear measurement. If there are less than perfect correlations between the linear measurements, the coefficient of variation of the higher dimensional measurement may be much smaller than the upper limit. In the second case, the coefficient of variation of the whole is less than the average coefficient of variation of the parts, weighted by their means. The lower the correlation between the parts the lower the coefficient of variation of the whole. For the case of a countable character such as bristle or cell number, an expression for the relationship between the variance of the parts and the whole has been derived showing that an apparent positive correlation between parts is generated by variance between individuals.

D. Game-theoretical Formulations in Population Genetics

R. Lewontin and I. Eshel have investigated the nature of game-theoretic solutions to population genetic and evolutionary problems. Starting with Maynard Smith's suggestion of the so-called "evolutionary stable strategy" (ESS), they have shown that this is a weak Nash solution and cannot give rise to a mixed strategy solution. They then derive these other solution criteria including a strong Nash solution and various behavioral solutions designated as "strongly evolutionarily stable" (SESS). In particular, they have shown that in many cases, none of these game theoretic solutions correspond to actual stable equilibria in real genetic systems.

E. Dynamics of Gene Frequency Dependent Fitness Systems

G. Oster and R. Lewontin, in a long distance collaboration with R. May of Princeton University, have investigated the equilibrium structure of one gene systems with frequency dependent selection. Lewontin had already shown that a single stable fixed point could be changed into an oscillatory solution with increasing amplitude by simple frequency dependence. They have investigated whether such frequency dependent selection could result in "chaotic behavior" such that the oscillation would be aperiodic and give the appearance of a random variation in gene frequency. They were able to show that fitness functions of the exponential kind,

$$w_{AA} = e^{-a(p-1/2)}$$

will never produce chaotic behavior and that the fitness function must be of the kind that fitness decreases with increasing frequency of the genotype, but only up to a certain frequency p^* , whereupon it decreases again. An

example that does exhibit chaotic behavior is

$$w = \frac{ax^2 - ax + 1}{ax^2 + ax + 1} \quad \text{where } x = 2p - 1$$

Thus, frequency dependent selection can produce almost any kind of gene frequency dynamic from stable equilibrium to oscillations to chaotic behavior.

F. Multiple Allelic Equilibria

The mathematical statement of the solution to a multiple allelic selective equilibria and an algorithm for testing the stability of such an equilibrium has been known for a long while due to the work of Kimura and Mandel for constant fitness models, and Lewontin for frequency dependent models. In any event, these results are a straightforward application of the basic methods of linear algebra and the standard conditions for the neighborhood stability of difference and differential equations. Unfortunately, these mathematical formulations give no direct answer to the question of how easy or difficult it is to maintain multiple allelic equilibria, because the conditions on the fitnesses are so complex as not to have any intuitive use. For example, it is neither a necessary nor a sufficient condition for stable multiple allelic equilibria, that the heterozygotes be more fit than the homozygotes, although this condition is both necessary and sufficient for a two-allele equilibrium.

We have used a combination of numerical and mathematical methods to ask the question: "What kinds of fitness relations will lead to stable equilibria of n alleles at a locus?" With n alleles, there are $n(n + 1)/2$ different genotypes. If we scale all fitnesses between 0 and 1, then there

is an $n(n+1)/2$ dimensional hypercube which is the fitness space and we have investigated the density and location in this fitness space of equilibria that are not only stable but feasible, that is that have all alleles segregating and therefore have solutions inside the allele frequency simplex. This simplex is, in general, an $(n-1)$ dimensional tetrahedron. Our method was to generate fitness arrays at random in the fitness space, subject to various restrictions corresponding to different regions of the space, then to test each solution to see if the equilibrium was inside the allele frequency simplex and if it was stable. If a stable feasible equilibrium is found, the vector of allelic frequencies is recorded. Tens of thousands up to hundreds of thousands of fitness matrices for each region of the fitness space were generated.

a) Measure of stability and feasibility

The first determination was the measure of stability and feasibility in the parameter space as a whole. This fell off extremely rapidly as the number of alleles in the system was increased. The proportion stable and feasible falls off approximately as e^{-n^2} . Thus, there were no stable feasible equilibria found out of 100,000 fitness matrices generated for $n = 6$ and only a few for $n = 5$. This is perhaps not so startling since the entire fitness space includes large regions of underdominance, partial dominance, etc. The next step was to restrict fitnesses to pairwise heterosis. This means that each heterozygote is more fit than its constituent homozygotes. The problem of efficiently generating random fitnesses subject to such a restriction is not a trivial one, but a computer algorithm was worked out which had the correct properties. Again, the measure of stability fell off

faster than exponentially with n and no stable feasible equilibria were found for $n = 7$ and only a few for $n = 6$. This result was more surprising. We then generated fitnesses at random subject to the restriction of total heterosis, that is all heterozygotes more fit than all homozygotes. While there was some increase in the measure of stability and feasibility, it still fell off more than exponentially with n and the results were of the same order of magnitude as the previous case.

b) Small variations in fitness

We have discovered one region of the fitness space with high feasibility and stability. This corresponds to heterozygotes more fit than homozygotes and very low variation in fitness from heterozygote to heterozygote and from homozygote to homozygote, but especially the former. The likelihood of stable equilibrium depends upon d/σ where d is the average superiority of heterozygotes and σ the standard deviation of fitnesses among heterozygotes. For a value of 10 or greater, very large numbers of alleles can be kept in equilibrium. For a value of 5 or less of the ratio, very few alleles can be maintained.

Using a model of random variation from homozygote to homozygote and heterozygote to heterozygote, we were able to find analytic results on the measure of stability, based on the theory of the eigenvalues of random matrices developed by Wigner. In addition, we have proved several inequalities on fitnesses that are necessary conditions for stability including the "triangle inequality" for heterozygotes, namely that for every heterozygote fitness w_{ij} , it must be possible to find an allele k such that $w_{ik} + w_{jk} < w_{ij}$.

c) Distribution of equilibrium vectors

We have studied the equilibrium allelic frequency vectors resulting from stable feasible equilibria by comparing the volumes of successive shells of the frequency simplex around the centroid, to the vectors actually found for stable feasible equilibria. The result is that there is a concentration of vectors near the centroid, especially for smaller numbers of alleles. That is, the alleles tend to be equal in frequency, near the frequency point $p_i = \frac{1}{n}$. This is especially true for the models of loci fitness variance among heterozygotes. The prediction, then, is that naturally occurring polymorphisms with large numbers of alleles should have the alleles in nearly equal frequency. This is not what is observed. Usually in nature one or two alleles are in intermediate to high frequency and all others are in low frequency. Thus, it is unlikely that heterosis accounts for these observed polymorphisms.

G. Time Dependent Behavior of Gene Frequencies in Stochastic Models

Nearly all the work on the stochastic theory of gene frequencies gives the form of the stationary distribution of gene frequencies, but does not give the time-dependent solutions. That is because the process is described as a partial differential equation with $\phi(x,t)$ changing in time as a function of differentials of $\phi(x,t)$ with respect to x , and these equations are not, in general, soluble explicitly. The steady state can be solved for, however, since the derivative with respect to time is zero and a simple algebraic equation can be produced by integration over x .

A few cases have been solved explicitly for arbitrary time and several attempts at approximate methods for time-dependent solutions have been made,

but not with great success. Dr. Welch has now developed a technique for time-dependent solutions which gives excellent results when checked against actual numerical evaluation of the differential equations. This technique depends upon the fact that the solution $\phi(x,t)$ is a separable function. This then allows it to be expressed in terms of Chebyshev polynomials. This in turn allows the use of the very powerful property of these polynomials, namely that their derivatives can be themselves expressed in terms of the polynomials. Thus, a differential equation becomes an algebraic equation, and a solution is possible. The algebraic equation, of course, is in terms of the sum of a series of polynomials which must be evaluated numerically, but this is easier than the numerical solution of a differential equation from some arbitrary starting value.

The solution obtained by the method gives extremely good results for t of the order of $.1N$ or larger. For small numbers of generations, the results are not nearly as good and for some cases may be quite poor. Dr. Welch has derived results for the probability that two alleles are still segregating at various generation times from various initial conditions of gene frequency.

H. The Effect of Selfing on Linkage and Selection

I. Eshel has studied the interaction of selection and linkage in populations with mixed selfing and random mating, as, for example, many agricultural plants such as corn. When there is partial selfing, then even in the absence of epistatic fitness interaction, and with free recombination, the behavior of different loci are not independent. More specifically, he has

shown that polymorphism at one locus will facilitate the maintenance of polymorphism at a second locus. Moreover, a protected polymorphism at two loci may be maintained under conditions of variable environment that are not sufficient to maintain a polymorphism at each locus in isolation. He has also shown that recombination between non-epistatic loci weakens the influence of selfing on overdominant equilibrium. That is, selfing makes the requirements on heterozygous fitness more extreme for the maintenance of polymorphism. If there are two such loci linked, however, the necessary value of the heterosis is less. On the other hand, linkage between loci increases the effect of selfing on the fixation of deleterious genes. Selfing increases the speed of fixation, and linkage between loci enhances this effect of selfing.

II. Plans for the Future

1. Multi-locus Models

In our previous three year summary, we projected two major problems for the period just passed: variable environment and multiple alleles. As our progress report shows, we have made a good deal of progress on the problem of environments that vary in space or time, but we are just beginning a thorough study of multiple allelic systems. Some small analytic progress has been made on this latter problem both by the contract and by other workers, but much remains to be done.

For the problem of variable environment we intend to study the effect of linkage on damping or exaggerating the effect of low level environmental variation in time. This will be done almost entirely by simulation techniques using our 5 locus and possibly our 36 locus simulation program. The question to be asked is what relationship exists, if any, between the variance and autocorrelation of the environment, the degree of linkage, and the variance around a steady state or rate of decay of genetic variation. This is the logical next step following our exploration during the past 3-year period of large oscillations of a regular nature.

The problem of multiple alleles cannot be separated from the problem of asymmetry of selection coefficients. Most of what is known about selection in linked systems concerns highly symmetric selection models. Some recent work by Feldman and Franklin and by Karlin shows that asymmetric models have a much more complex behavior and we have already found (bibliography item 8) that if there are 3 alleles at one locus and 2 alleles at a second that asymmetrical frequency solutions occur. We must put the two together. First, we will make an extensive study of asymmetric fitness models, generating

very large numbers of such models. Rather than simply generating them at random, we will begin with symmetric models of known properties and perturb these by making small changes in the parameters. We will then observe the equilibria that are produced. We will then go back to the symmetric models, add more alleles and follow the equilibria. There will be a great variety of these and we do not expect to exhaust this problem during the next contract period.

2. Single-locus Models with Multiple Alleles

We will pursue our study of multiple allelic models in two directions. First, we will study the behavior of frequency dependent selection, in part because this is a realistic kind of fitness assumption for a number of cases in nature and in part because the multiple locus model with recombination is formally equivalent to a multiple allele model with frequency dependent selection. We have already begun a study of linear frequency dependence and this will be expanded to include fitnesses that are convex and concave functions of frequency with various assumptions about heterozygotes. Up to now, our work on the frequency dependent question has been entirely analytic, but we will soon outrun the analytic possibilities and we will use numerical (not simulation) techniques to find equilibria and test their stability, much as we did for the case of constant fitnesses.

Second, we will explore the interaction of random drift processes with selection for multiple alleles. This is a subject that was first treated by Crow and Kimura analytically and has recently been the subject of investigation by Ohta. The models that can be treated analytically are very simple selection schemes, usually with partial dominance of deleterious genes. We

will explore the effect of drift on the polymorphisms that are stable in deterministic models. Since the equilibrium vectors in frequency space tend to be close to the centroid for stable equilibria, drift will have less effect of destroying genetic variation than if many alleles had very small equilibrium values. Nevertheless, for large numbers of alleles, drift will play an important role. It may be that the skewed distributions actually observed for multiple allelic polymorphisms in nature, are, in fact, the result of drift acting on the deterministic equilibria predict near the center of the frequency simplex.

III. Graduate Students and Postdoctoral Associates

As in the previous periods of the contract, we have had very few graduate students and many more postdoctoral associates. There seems to be something in the style of the work of the Principal Investigator that intimidates or otherwise discourages students from theoretical work while at the same time encouraging more advanced professionals. There have been fewer people altogether associated with the contract in the past 3 years than there were in previous years in Chicago. This is because the Principal Investigator had a very large training grant in population biology in Chicago which brought in large numbers of postdoctoral fellows. These associated themselves with the work of the contract with no cost to it. The termination of that training program after the move to Harvard meant that the contract itself and the occasional individual fellowship would have to support all the people who could contribute to the work. Since there is insufficient money in the contract for even one postdoctoral appointment, the variety of workers has been seriously reduced. The following have worked on the contract investigations during the last three years:

Professor Dan Cohen of the Hebrew University, Jerusalem, spent several months working on optimality solutions to problems in determining sex ratio in insects. He has since returned to Israel.

Dr. Ilan Eshel of the Department of Statistics at Tel Aviv University in Israel spent one semester working on a great variety of problems in linkage, inbreeding, optimality theory and ecological genetics. His mathematical expertise was particularly valuable to the contract. He has returned to his position in Israel.

Professor Marc Feldman of the Biology Department at Stanford University spends several weeks each year with the contract, working with the Principal Investigator on multiple-locus problems and on phenotypic variance theory. This is

an ongoing collaboration.

Dr. Peter Haas, now of the U. S. National Museum, was a graduate student with the contract. His thesis work was concerned with the oscillating environmental models with multiple loci and linkage.

Dr. Russell Lande was the other graduate student on the contract. He was a prolific and imaginative worker whose contributions to the theory of phenotypic selection (see the bibliography) were one of the most important elements in the contract work. He is currently a postdoctoral fellow at the Department of Genetics, University of Wisconsin, and will become Assistant Professor of Biology at the University of Chicago at the end of his fellowship.

Professor George Oster of the University of California, Berkeley, spent several months at Harvard and during that time collaborated with several people including the Principal Investigator, Dr. Cohen and Dr. Eshel on several problems in optimization theory and frequency dependent selection.

Professor Montgomery Slatkin began as a Research Associate and then became a Co-Principal Investigator on the contract at Chicago. He has been a major contributor to the work of the contract, especially by his investigations of selection clines, as well as his early work, reported in a previous 3-year summary, on the continuous model of selection. Professor Slatkin is now Professor of Biology at the University of Utah.

Dr. Joel Sohn has worked continually with the contract over the last two years on the population genetics of sex determination. While his work has been mostly experimental, he has made a number of analytic and numerical investigations of the stability of sex ratios in collaboration with Mr. Stephen Orzack, an undergraduate visiting investigator from the University of Rochester.

Dr. S. D. Tuljapurkar is a 1/4 time Research Associate with the contract.

He was originally trained as a mathematical physicist but has worked for several years in mathematical ecology. Although he is primarily engaged in a study of the epidemiology of schistosomiasis, he has already made several important contributions to our study of the stability of multiple-allelic systems.

Dr. Diane Wagener Welch came as a Postdoctoral Research Associate from Stanford University. Her main work on the contract was the study of the time dependent solutions for random genetic drift with selection. She also carried out studies, not completed, of the "hitch-hiking" problem in multi-locus theory. She is now in the Department of Biostatistics at the University of Pittsburgh.

Bibliography

The following publications have appeared or are in press since our last three-year report. Items 1 and 2 were reported at that time as "in press". Book reviews and general articles are not listed.

1. Slatkin, M. 1974. Gene flow and selection in a cline. *Genetics* 75:733-756.
2. Slatkin, M. 1974. Competition and regional coexistence. *Ecology* 55:128-134.
3. Lewontin, R. C. 1974. The analysis of variance and the analysis of causes. *Amer. J. of Human Genetics* 26:400-411.
4. Slatkin, M. 1975. Gene flow and selection in a two locus system. *Genetics* 81:787-802.
5. Slatkin, M. and T. Maruyama. 1975. Genetic drift in a cline. *Genetics* 81:209-222.
6. Lewontin, R. C. 1975. Genetic aspects of intelligence. *Ann. Rev. of Genetics* 9:387-405.
7. Feldman, M. and R. C. Lewontin. 1975. The heritability hang-up. *Science* 190:1163-1168.
8. Feldman, M., R. C. Lewontin, I. Franklin and F. B. Christiansen. 1976. Selection in complex genetic systems. III. An effect of allele multiplicity with two loci. *Genetics* 79:333-347.
9. Slatkin, M. and R. Lande. 1976. Niche width in a fluctuating environment. *Amer. Nat.* 110:39-55.
10. Lande, R. 1976. The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genet. Research* 21:1-15.
11. Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30:314-344.

12. Slatkin, M. and R. Lande. 1976. Niche width in a fluctuating environment - density independent model. *Amer. Nat.* 110:31-55.
13. Lande, R. 1977. The influence of mating system on the maintenance of genetic variability in polygenic characters. *Genetics* 86:485-498.
14. Lande, R. 1977. On comparing coefficients of variation. *Systematic Zoology* 25:214-217.
15. Lande, R. 1977. Statistical tests for natural selection on quantitative characters in experimental populations. *Evolution* 31:442-444.
16. Lewontin, R. C. 1977. Single and multiple locus measures of genetic distance between groups. *Amer. Nat.* vol. 111, in press.
17. Lewontin, R. C., L. Ginzburg and S. D. Tuljapurkar. 1977. Heterosis as an explanation for large amounts of genic polymorphism. *Genetics*, in press.
18. Lewontin, R. C. 1977. Fitness, survival and optimality. *Ohio State Symposium on Ecology*. Ohio State Press, in press.
19. Lewontin, R. C. 1977. Population genetics. *Proc. 5th Internat. Congress of Human Genetics*. Elsevier, in press.
20. Lewontin, R. C. 1977. The relevance of molecular biology to plant and animal breeding. *Proc. International Workshop on Quantitative Genetics*. Iowa University Press, in press.
21. Tuljapurkar, S. D. 1977. A note on Gillespie's model of enzyme variation in natural populations. *Evolution*, in press.

V. Present State and Future Prospects

As we pointed out in our previous overviews, the main difficulty with population genetic theory has been its distance from the actual biological objects of study. The theory is highly developed, yet it fails in three critical ways. First, many of the quantities in theory are not measurable, especially effective population sizes and selection coefficients of genotypes. Second, despite its richness the theory is badly oversimplified at critical points, especially a proper description of the variation of environment in space and time and a proper genetic dimensionality for phenotype. If population genetic theory is finally to make contact with the observations of genetic diversity in populations and its evolution over time, it must deal in phenotypes in a varying environment. Third, the theory is largely a steady state theory or equilibrium theory that has great difficulty in dealing with historical processes in their transient states.

The problem is, as ever, to close the gap between a highly developed theory and a rich set of observations which are in minimal contact. This is to be done, in our view, by

- 1) a concentration on multiple locus models of phenotype
- 2) much greater attention to multi-allelic models, in correspondence with the known experimental facts
- 3) an attempt to make the model of the environment more in accord with nature. Specifically strictly cyclic models of environment with no variation in period or amplitude or "white noise" models of stochastic environments are so far from reality as to be useless, unless it can be shown that these simplified models nevertheless contain the relevant

information for understanding genetic changes in real environmental sequences. Much more attention should be given to autocorrelated temporal variation and distance correlated spatial variation.

4) Much more attention to transient states of populations, and to relaxation times from perturbation, rather than an exclusive emphasis on equilibrium and steady state. Unfortunately, much of this can only be done numerically and the results may be so cumbersome as to be uninterpretable. Very careful strategic planning of simulation and numerical studies is absolutely essential. Otherwise, there will be so much information produced that none of it can be understood.