METAVARIATION

AND LONG TERM

EVOLUTIONARY PATTERNS

by

ALISTAIR M. BLACHFORD

B.Sc.(Honours) Queen's University, Kingston, 1978

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in
THE FACULTY OF GRADUATE STUDIES
(Department of Zoology)

We accept this thesis as conforming to the
required standard

THE UNIVERSITY OF BRITISH COLUMBIA
November 1984

DE-6 (3/81)

ABSTRACT

By definition "adaptability" is the ability of living systems to cope with change. Genetic adaptability requires the production of genetic variation. The view that variation production is undirected or random, i.e. unconnected with selection, implies that selection does not tailor genetic adaptability. But many genetic elements are known to modify processes of variation production, and secondary selection can act on them, so that view is not justified.

Over the longer term, natural selection 'favors' properties important in <u>maintaining</u> immediate fitness, as well as properties important for persistence in the short term. Genetic adaptability is less important in the short term, and is ignored in models based on short term definitions of fitness (e.g. relative effective rate of increase). If "fitness" is to be "the properties favored by natural selection", then its definition should be time scale dependent. Currently prevalent short term definitions of the action of natural selection should not be allowed to hamper consideration of the role of slow processes in determining long term evolutionary patterns.

A review of patterns in genome size, and the existing explanations for them, reveals that most explanations are based on notions of adaptedness to the <u>state</u> of the environment. An explanation of genome size patterns based on the <u>rate of change</u>

of environments is proposed. It is hypothesized that part of the genome is involved in regulating variation production, and that more DNA means slower production of additive genetic variation. This new hypothesis is simple, general, and testable, but requires more evidence. The question is raised of whether genomes might be organized to facilitate the adjustment of genetic variation production by natural selection.

TABLE OF CONTENTS

<u>LIST</u> <u>OF</u> <u>TABLES</u>

<u>LIST</u> <u>OF</u> <u>FIGURES</u>

ACKNOWLEDGEMENTS

This thesis has been quite an adventure for me. Although I think of myself as an ecologist, the glimmer of some of the ideas explored in this thesis lured me almost completely out of my "standard ecology" background. I would like to acknowledge the support I had while adapting to this new "habitat":

It was decided to fix the proportion of variant offspring arbitrarily (at two thirds) and vary only the offspring distributions. The latter are symmetrical (undirected), bimodal, and vary only in the deviation of the variant offspring from the parental phenotype. The particular deviation characteristic of a given player will be referred to as its "step-size". Figure 8 shows a segment of a population before (a), and after (b)

INTRODUCTION

The Neo-Darwinian tradition regards the generation of new genotypes as a random process, undirected with respect to the adaptive needs of the species. Dobzhansky (1970) states "Mutations...arise regardless of their actual or potential usefulness. It may seem a deplorable imperfection of nature that mutability is not restricted to changes that enhance the adaptedness of their carriers. However, only a vitalist Pangloss could imagine that the genes know how and when it is good for them to mutate". The intent of this thesis is to examine the possibility that part of an organism's genome is involved in adaptively regulating the production of genetic variation -- that genetic control can regulate genetic variation production according to its potential usefulness.

This thesis developed from a search for a framework in which to cast evolutionary hypotheses in terms of patterns of change in, rather than the state of, an environment. Evolution can be viewed as a continual process of creation of new genotypes and alteration of genotype frequencies, or of variant production and variant loss. It is obvious that patterns of evolutionary change could be determined by both patterns in variant production and patterns in variant loss. But the burden of explanation has always been placed on the processes of variant loss, e.g. natural selection and chance. Why? Although we need to understand selection to judge which, if any, of a set of new variants are relatively better, we need to know about variant production to say what types might appear. So unless we can say that all conceivable variants are always present,

variation production, not selection, is limiting. It seems logical, therefore, to have a good look at the potential influence on evolution of patterns in variation production -- even if we end up ruling them unimportant. I will describe how a change in emphasis from variation loss to variation production amounts to increased attention to adaptability and change, rather than adaptedness and state of the environment.

This thesis is written in two parts. In the first part I establish the plausibility of patterned variation production, and describe a general mechanism by which natural selection can adjust variation production patterns. I then expand on the requirements and constraints of a regulation system for variation production.

The merit of a different idea or perspective in science is determined by its success in explaining existing observations, and in generating new questions. Part Two is devoted to exercising the ideas of Part One in this way. I consider patterns in genome size as an example of an area where traditional explanations have not proved fruitful, and where a hypothesis based on variation production might be more successful. After discussing ways in which the hypothesized regulation system might be detected, I conclude by using the ideas to generate new questions for two "bandwagon" topics.

CHAPTER ONE

SOME BASIC CONCEPTS

In this chapter I will define a few terms, and introduce some of the concepts that underly the arguments to be developed later.

"Random" vs. "Directed"

How does one distinguish "random" or "undirected" patterns in variation production from those which are "adaptive" or "directed"? Random variation is uncorrelated with, i.e. not "directed" by, the selection processes to which it will be subjected. It is by reason of the generally assumed lack of connection that "Evolution is (I repeat!) a two step process" (Mayr 1978) of variation production and variation loss.

Randomness and Scale

I will not, in this thesis, be suggesting a level of control that can program mutation at a given locus to a particular end result. I will assume variation production at this level to be uncontrolled, and "undirected" by any influence of selection. I will argue for consideration of possible control of mutation rates across loci, and hence, across traits. Given that the production of genetic variation were not always selected against, the existence of a heritable system for controlling mutation rates across loci would make it possible for selection to influence the pattern of variation production on the level of the genotype, and in the population as a whole.

With this feedback from selection to variation production, evolution is no longer a simple two step process of variation production followed by selection, at the genotypic or organismal level, because step two can influence step one.

These points will be discussed in further detail later. At present it suffices to note that "random variation production" and "directed variation production" are too vague in themselves, because the randomness of a phenomenon can depend on the scale on which it is considered. Secondly, I must emphasize that mutation is but one of many processes of genetic variation production in a population. Above it was discussed alone to simplify explanation.

Kinds of Variation

If we are to deal with the possible connection between selection and variation production, then it will be convenient to distinguish among types of variation according to their differences in interaction with selection. Levins (1964a) describes the different selection regimes favoring non-additive and additive variation respectively:

In a patchy environment in which the patches are sufficiently different, relative to the tolerance of an individual, to make a "patch generalist" strategy inefficient compared with a "patch specialist" strategy, NON-ADDITIVE variation in effect permits a single population to retain genotypes in constant frequencies and become a mosaic of patch specialists.

The function of ADDITIVE variation, on the other hand, is

to _permit_ change in gene frequencies, and evolutionary responsiveness. Additive variation and genetic response are favored if present selection pressures are good predictors of future selection pressures. It is the kind of _change_ in an environment, rather than the _state_ of the environment, which determines the importance of adaptability, and additive variation production.

In this thesis I will be talking about additive variation because I am concerned with genetic adaptability, i.e. evolutionary responsiveness. Furthermore, I will be discussing additive variation production as a process and strategy, rather than as a measure of the amount of additive variation in a population.

Time Scales and Natural Selection

Natural selection is not an agent, but a process, and its components are therefore sub-processes (Ghiselin 1981). The results of selection are the combined results of its many sub-processes, such as competition and predation. Unfortunately the formulation of Neo-Darwinist models, with selection coefficients summarizing the results of all deterministic processes affecting differential reproduction, tends to conjure up the illusion of a selection "force" (agent) acting on "units" (patients, in the terminology Ghiselin 1981 uses).

The "results of selection" will depend on what sub-processes are involved. Longer time scales (longer periods of observation) allow the actions of slower processes to show themselves in the results (i.e. to become "important"). This

implies that the results of selection can differ <u>qualitatively</u> depending on the time scale over which we perceive them. (See Appendices I and II for further discussion of this idea.)

CHAPTER TWO

SELECTION OF VARIATION PRODUCTION

If there exists heritable control of adaptability, then selection, in the form of kinds of environmental change, can mold genetic adaptability, i.e. additive genetic variation production. This would justify the formulation of hypotheses in terms of adaptability and the way environments change (Chapter Four). Hypotheses are more commonly phrased in terms of adaptedness, and the way environments are.

The purpose of this chapter is to establish the plausibility of directed variation production, an alternative explanation for observed patterns of genetic variability. Many known heritable factors influence the production of genetic variation, and can be perceived as being part of a variation regulation system. A selection mechanism which can modify such a regulation system is described by way of analyzing a simple model from the literature.

Patterns in Variation Production

There is no question that there are patterns in the production of genetic variation -- the question is whether the patterns are "random" or not. Mutation is one process of variation production, and patterns in it include the following:

1. Different genes have different mutation rates.

This pattern is easily explained by the assumption that mutation rate is an intrinsic property of the gene. Different genes have different structures, so why not different mutation rates? Ohno (1969), for example, suggests that larger genes probably have higher mutation rates since mutation processes act per base pair and not per locus.

2. Deleterious mutations are more frequent than beneficial ones.

This observation agrees well with the usual model of undirected mutation. An undirected change is unlikely to improve a highly ordered and integrated system.

3. Mutations with small effects are much more frequent than those with large effects (Timofeef-Ressovsky 1935; Kerkis 1938; James 1959; Mukai 1964).

It is less obvious why this should be so, but it could be argued that random mutations would normally affect only one or a few codons and such changes do not usually result in crucial alterations of essential gene products. This refinement of the explanation of pattern (2) should be noted: the organization of the genome is constrained enough that changes to it are very likely to be for the worse, but unconstrained enough that most changes do not matter much. The experimental work referenced above measured the 'severity' of a mutation by its effect on viability.

The tools of molecular biology have revealed patterns in

genetic change among natural populations, between genes (Jeffreys 1981), within genes between introns and exons, and within exons (Holmquist et al. 1983). But these observations are interpreted as being post-selection, cannot be attributed to patterns in variation production and, indeed, are usually hypothesized to be the result of patterns in selection pressures (e.g. less critical parts of a protein cistron are more free to vary (Holmquist et al. 1983)).

An alternative to several of the explanations advanced above is that mutations occur more frequently in less critical parts of the genome or, equivalently, are suppressed to a greater extent in more critical parts of the genome. This hypothesis is uncommon, perhaps because the implied system of mutation control is less parsimonious or because, at first glance, it seems to be teleological or Lamarckian.

Modifiers of variation production processes

Karlin and McGregor (1974) state that

> "...the existence of genes controlling specific
> recombination rates, genes influencing mutation rates at
> particular sites, loci affecting directly or indirectly
> rates of migration, factors controlling outcrossing rates,
> etc., is documented, identified, and studied in the
> genetics literature."

Table 1 lists the many determinants of genetic variation in a population, along with known examples of genetic modifiers. The fact that heritable modifiers of processes of variation production exist implies the potential for adaptive modification

TABLE 1. The determinants of genetic variation in populations,
         along with examples of heritable factors that can modify
         variation production.

MUTATION

     mutator genes and transposons (Ives 1950; McClintock 1965;
     Green 1973; Thompson and Woodruff 1978), genes for repair
     enzymes and polymerases

RECOMBINATION

     -chromosome number

     -frequency of crossing over (inversions, recombinant genes
     (Catchesdide 1968; Stamberg 1969), B chromosomes (Carlson
     1978))

MEIOTIC DRIVE

     segregation distorter locus (SD) in Drosophila (Hiraizumi et
     al. 1960), t-allele in house mouse (Lewontin and Dunn 1960)

POPULATION SIZE

POPULATION STRUCTURE

     -dispersal rates and migration

     "The genetic control of migration rates is exemplified by
     genes determining flagella in protozoa, movements in Hydra,
     bird migration, etc. Bird and fish migration patterns often

Table 1 (cont'd)

display the interesting phenomenon that populations may vary categorically into migrators and non-migrators, and the evidence suggests at times that migration per se may be polymorphic within populations." (Karlin and McGregor 1974)

BREEDING SYSTEM

-frequencies of mating types

-mating patterns (assortativeness) - all of the factors that can contribute to reproductive isolation (see Dobzhansky 1970, p.314), e.g. incompatibility factors in plants, genes controlling seasonal timing of reproduction, recognition (olfaction, vision, sound), etc. "A number of simple Mendelian factors affecting preferences in mating are associated with pigment color or pattern (Mainardi 1968)." (Karlin and McGregor 1974)

SELECTION

-dominance relationships influence the effects of selection. Genetic modification of dominance is often assumed (see Wright 1929; Feldman and Karlin 1971), but I know of no identified genetic elements.

HISTORY

of variation production by selection. How can selection act on these modifiers? Must 'group' selection be invoked? These questions will be answered next.

SELECTION MECHANISMS

Perhaps the greatest barrier to consideration of patterned variation production is the lack of an easily understandable, general mechanism explaining how the patterning of variation production can take place. The literature contains several different mathematical models concerning the modification of production of genetic variation (e.g. Fisher 1930; Kimura 1956,1960,1967; Levins 1967; Leigh 1970,1973; Karlin and MacGregor 1974; Charlesworth 1976; Felsenstein and Yokoyama 1976; Gillespie 1981). But these equilibrium models usually seem more concerned about the dynamics of modifier frequencies than on the relevance of a process of modification of variation production to large scale or long term evolutionary patterns. Layzer (1980) described a more general mechanism of variation modification, and dwelled upon the evolutionary implications of adjustable variation production.

There are two kinds of selection that can act on processes of variation production: group selection, and secondary selection. The conditions under which group selection can take place are rather stringent (Williams 1966). Secondary selection is common, although of uncertain, and variable, importance. In the present section I will explain secondary selection and how it can act on variation production, and criticize Layzer's mechanism for modifying variation production.

Secondary selection

Secondary selection is the process of differential transmission of traits to the next generation, not because selection 'acts' on them, but because they are correlated with traits that are acted on by natural selection. The usual 'primary' selection acts on genes via their correlation with phenotype. Heritability is a measure of this correlation. The higher the heritability, the more effective is a given strength of selection on phenotypes in changing gene frequencies. There is a further level of indirection in secondary selection -- genes are acted upon through their correlation with other genes that are acted on in the usual way.

> N.B. This second correlation is among genes, not traits. Pleiotropy (a gene affecting more than one trait) assures the existence of correlated characters and the dragging along of one character by selection on another. This does not involve the correlation, or linkage disequilibrium of two genes, and is not secondary selection.

In the case of variation production, secondary selection can act on alleles that regulate processes of variation production, if these alleles are correlated in their distribution within a population with alleles producing the traits on which selection acts. For example, let us consider the rate of production of genetic and (given non-zero heritability) phenotypic variation for some all-important, quantitative trait (x). Suppose the population has been at equilibrium in a constant environment. The observation that "factors responsible for increasing variation production will be present in greater

proportion in the phenotypic variants of the population"
establishes the second correlation required for the operation of
secondary selection.

To take a closer look at this important observation, let
there be two alleles (H and L), in an asexual population, whose
only effects are to determine the variance in the distribution
of offspring phenotypes with respect to one trait. One allele
(H) produces a higher variance than the other. The fitness
(survival x fecundity) of a genotype depends only on its
phenotype, which is unaffected by the particular allele H or L .
In time, because of different rates of generating variance in
phenotype, the frequency distributions of the two subpopulations
characterized by the alleles H and L, along the relevant
phenotype dimension (x, a quantitative trait), take on the
following relative shapes:

H

L

μ X→
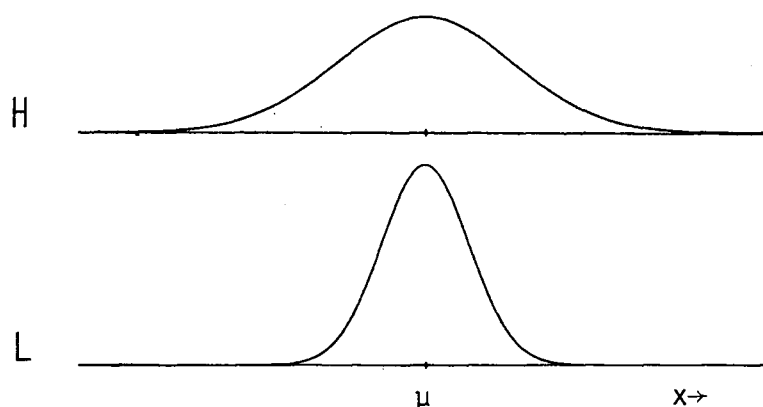
Now we can see that, in a population composed of these two
subpopulations, the genotypes producing greater variation in
their offspring are present in a greater proportion in the
variants of the population. This happens only because of the
relative shapes of the frequency distributions, and regardless
of the relative frequencies of H and L in the population.

With this picture in mind we may follow the action of secondary selection by considering certain patterns of selection on phenotype. It is clear that disruptive selection, favoring <u>both</u> tails of the phenotype distribution, would favor greater variation production, and that stabilizing selection would favor reduced variation production. But it is less easy to see what directional selection (for one tail) would do. The allele for greater variation production, while present in greater proportion in the favored tail, also comprises a greater proportion of the disfavored tail.

f(x)

μ   X→
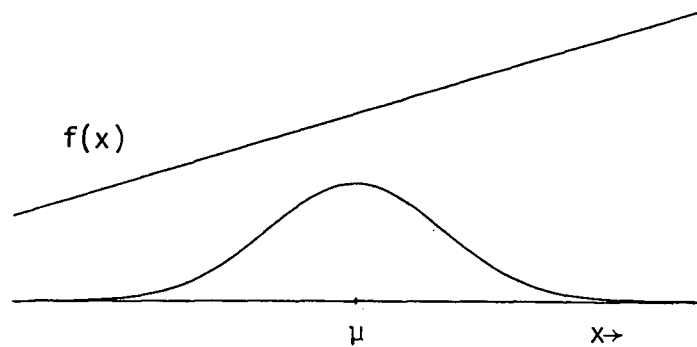
Let's start with symmetric distributions (not necessarily bell-shaped) about the same mean phenotype (μ). Consider a monotonic linear fitness function f(x). A phenotype that deviates to the right of the mean is present in the population as frequently as a phenotype equally deviant to the left of the mean. And the phenotype to the right has a fitness that is greater than $f(\mu)$ by the same amount that the fitness of the left phenotype is <u>less</u> than $f(\mu)$. The differences cancel, both symmetrical subpopulations have the same mean fitness $f(\mu)$ and, with a time horizon of only 1 generation, <u>neither</u> allele has a selective advantage. So no modification of the rate of variation production will occur.

If the fitness function f(x) is <u>not</u> linear, one of the H or L subpopulations will have a higher average fitness. If f(x) is concave up, then the right tail increases the average fitness of a subpopulation more than the left tail decreases it, and allele H has an advantage.

f(x)

μ   X→
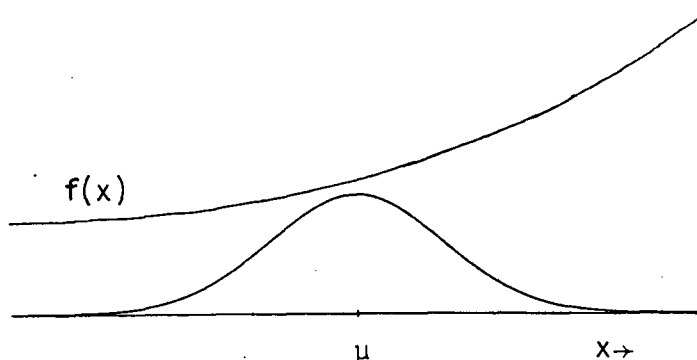
If f(x) is concave down, then members that deviate to the left of the mean phenotype will decrease average fitness more than those members equally deviant to the right will increase it.

f(x)

μ   X→
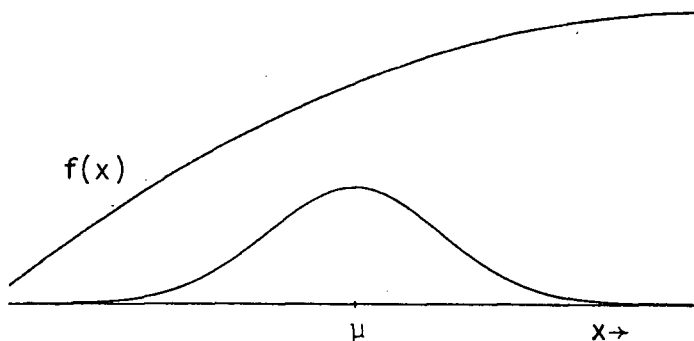
The (symmetrically distributed) subpopulation with the smaller proportion of variants (L) will be favored, and selection will have decreased variation production.

The Layzer Model

Layzer's (1980) model makes use of the above observations about the effects of concavity of the fitness function on the average fitness of subpopulations of differing variances (Figure

1). Layzer assumes that the fitness function is bell-shaped, and that

> "...the spread of fitnesses in the population is small compared with the total range of fitnesses associated with the trait in question; see Figure 1. In the neighborhood of the instantaneous population mean $x=\mu$ one may then approximate the fitness function by the first three terms in its Taylor expansion:
>
> $$f(x)= f(\mu) + f'(\mu)(x-\mu) + 0.5f''(\mu)(x-\mu)^2$$
>
> In this approximation, which is adequate if σ is sufficiently small, the mean fitness associated with the trait, obtained by averaging f over the normal distribution of x, is $fav= f(\mu) + 0.5f''(\mu)\sigma^2$. Thus the mean fitness fav is greater than the fitness $f(\mu)$ associated with the mean value of x when $f''(\mu)>0$, and is smaller than $f(\mu)$ when $f''(\mu)<0$. In other words (see fig. 1), $fav>f(\mu)$ during the emergent phase in the evolution of an adaptation, and $fav<f(\mu)$ at or near a fitness peak." (p. 815)

It is obvious that this mechanism works because of the assumed shape of the fitness function. Templeton (1981) severely criticizes the model for this assumption, pointing out that (1), Layzer's conclusions about the evolution of genes that control variation production therefore have little generality, and that (2), Gillespie, who has done much work on the evolution of modifier genes, assumes that adaptive peaks are dome-shaped, i.e. concave down everywhere (Gillespie 1978). If this were the case, Layzer's mechanism would not work except to decrease variation, leaving his evolutionary consequences of adjustable
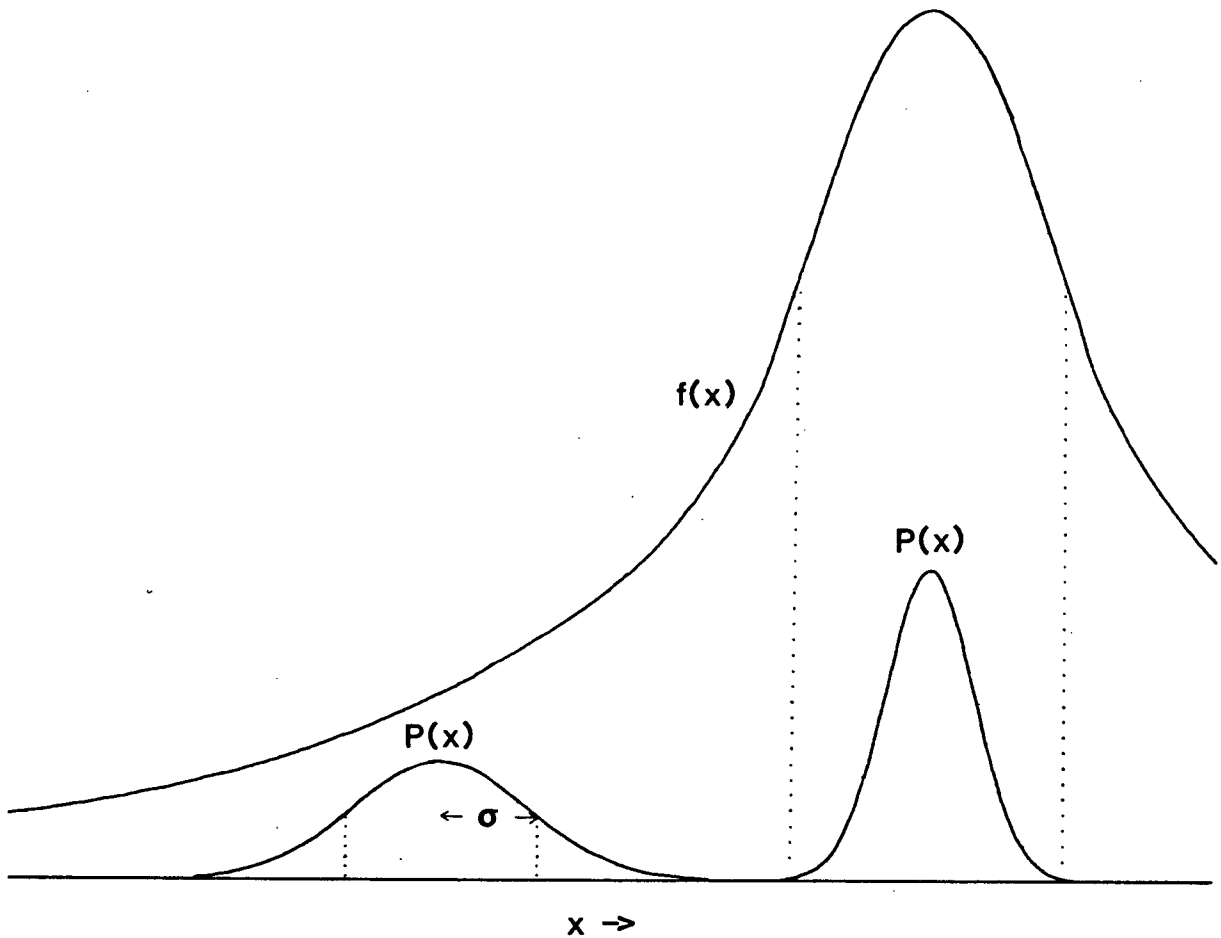
Figure 1.   "Fitness   function   f(x)   and   two   frequency
distributions   P(x)   for   an   adaptation   specified   by
parameter x. During the emergence   of   an   adaptation   the
quantity   c=f"(x) is positive and selection favors a broad
frequency distribution of x. Near the fitness   peak   c   is
negative   and   selection   favors   a   narrow .frequency
distribution." (Layzer 1980)

variation unfounded.

A time scales consideration

A less obvious flaw in Layzer's explanation is that it is based on a one-generation definition of fitness. Yet he discusses implications of his model on the evolutionary (multi-generation) time scale. Since the qualitative results of selection differ on different time scales, as I argue in Appendices I and II, Layzer's conclusions are unsound.

It is convenient to explain this time scales problem in terms of 'adaptedness' and 'adaptability'. Adaptedness is the suitability of a phenotype to the present environment. A one-generation evaluation of fitness would be one way to measure it. Adaptability is the ability to respond so as to maintain adaptedness in the face of environmental change. Levins (1964a) described the function of non-additive genetic variation as being potentially important for adaptedness to a given heterogeneous environment, and additive variation as being important for genetic responsiveness, or adaptability, to a changing environment (see Chapter One). The Fundamental Theorem of natural selection is a statement of the role of additive genetic variation in determining evolutionary responsiveness (Fisher 1958; Price 1972). It says that the rate at which a population (or subpopulation) evolves varies directly with the amount of genetically based (additive) variance in phenotype.

My criticism of Layzer (1980) is that, although he is dealing with additive variation (which is responsive to selection) the function of which is adaptability, his mechanism

and explanation of what sort of variation is advantageous under various circumstances is based entirely on <u>adaptedness</u>. His one-generation evaluation of fitness leaves no room for evolutionary responsiveness, which is the function of the additive variation with which he is dealing.

Given the relationship (Layzer 1980)

$$fav = f(\mu) + 0.5f''(\mu)\sigma^2$$

it is only clear, given a certain concavity (f''(x)), which variance has the advantage if the two subpopulations have the <u>same</u> <u>mean</u>. It could be, for example, that a different subpopulation mean μ more than compensates for any loss in average fitness due to the term $0.5f''(\mu)\sigma^2$. The subpopulation with the greater variance evolves faster and is likely to be nearer an adaptive peak and therefore have a higher average fitness.

The above equation gives the <u>initial</u> relative advantages of subpopulations with the same mean but different variances. Unless the initial relative advantage holds thereafter, it would be wrong to use this initial assessment to evaluate the merits of different variances over longer periods. Over the longer term the subpopulation means μ become functions of time $\mu(t)$, and $\Delta\mu/\Delta t$ is an increasing function of the genetically based variance in phenotype (the direction being toward the nearest peak). Thus comparison of different variances (maintained by differences in variance production) over the long term is more complicated. It is noteworthy that over the long term selection can favor increased variance even under curves that are concave downward.

This last point can be illustrated by computer simulation (details of the model are described in Appendix III). Two subpopulations, A and B, have different rates of production of phenotypic variance. B has the higher rate. They evolve under the parabolic fitness function shown in Figure 2. Note that this curve is concave down so that, given equal means, the subpopulation with greater variance is at an immediate disadvantage in fitness everywhere under the curve. Here both subpopulations begin with the same size and average phenotype. The simulation shows that the subpopulation with the greater variance improves its average phenotype at a greater rate (see Figure 3a), which over the longer term leads to a higher average fitness than that of the subpopulation with lower variance, which lags farther away from the adaptive peak. Short term fitness causes the subpopulation with lower variance to have greater numbers. In generation 7, in this run (see Figure 3b), the ranking of subpopulations reverses due to the cumulative effect of a fitness advantage gained by improvement in average phenotype. Therefore, if we look more than one generation ahead we see that selection can favor greater variance production even under a fitness function that is concave down.

Discussion

Proper evaluation of the relative advantages of different rates of variation production (and therefore, population variances) is dependent on the time scale of comparison. As time scales lengthen, adaptability gains in importance relative to initial adaptedness (Appendices I and II). Variation is
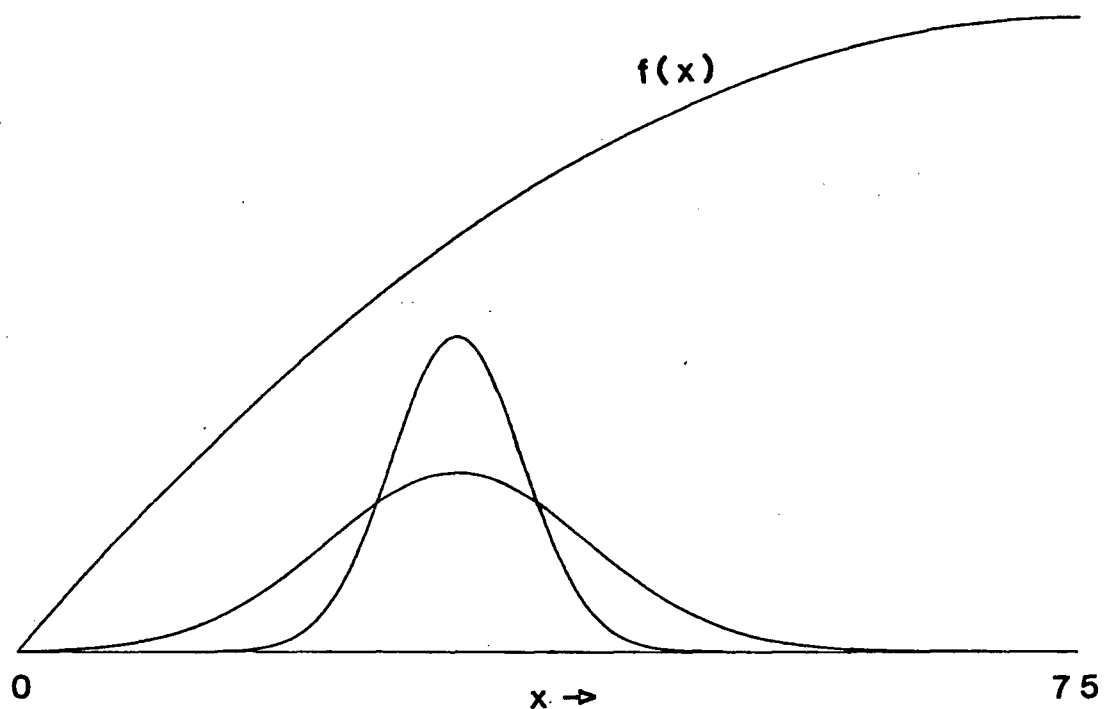
Figure 2. An illustration of the initial conditions of the computer simulation that produced the results in Figure 3. Two subpopulations, initially identical in size and average phenotype, are allowed to evolve with respect to quantitative trait x under a fitness function that is concave down everywhere.
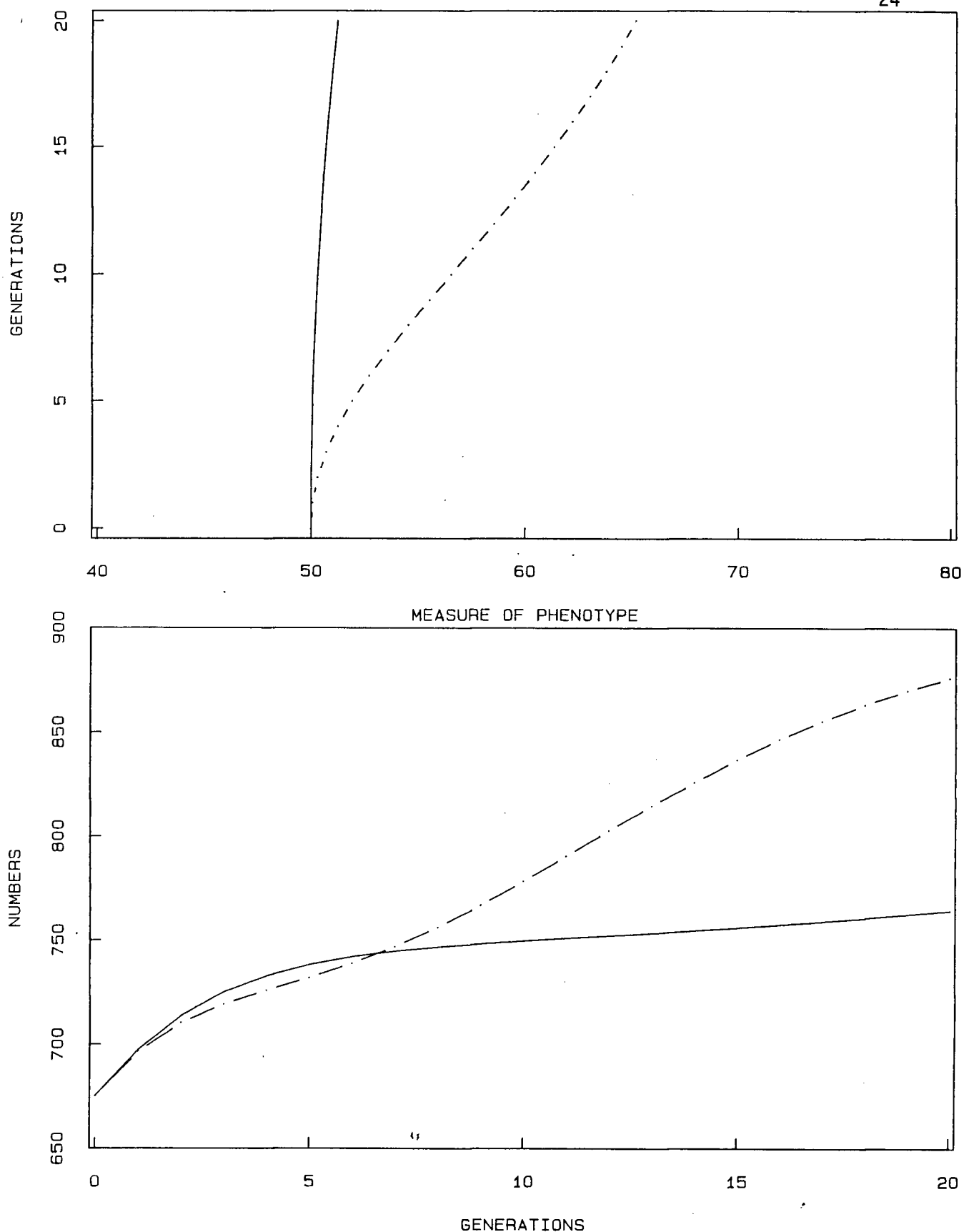
Figure 3. Output from the computer simulation model described in Appendix III. Two subpopulations are evolving towards an adaptive peak at 75 "phenotypic units". Subpopulation B (dotted line) has the higher rate of variation production. (a) Average phenotype over time. (b) Numbers over time.

important for its determination of the rate of phenotypic change over time, not just in its contribution to average population fitness at any one time. The former effect must be taken into account in any realistic analysis of the relative advantages of alleles that confer different rates of variation production.

Population genetics deals with the statistical behavior of populations, and we commonly think of phenotypic variation as a characteristic of a population. I am addressing variation production , which can be treated as a life history strategy, and above I have shown how these two 'variations' are linked. In talking about selection for different variation production, people often think of the population characteristic and deduce that group selection is necessary. In the explanation above, I have shown how selection of more inclusive categories (individuals of many different phenotypes with common rates of variation production) can take place via their respective subcomponents (by secondary selection), rather than by pitting group against group.

It seems obvious that, given heritable variation in variation production, certain patterns of variation will prevail. But this is interesting because different patterns are favored under different conditions -- variation production is not always minimized. The point of this section is to convince the reader that given variation in variation production, selection can produce adaptive patterns of variation production. In other words, it is reasonable to explicitly consider patterns of variation production rather than to assume, as has usually been done in the past, that variation is random, and leave the

explanation of evolutionary patterns to patterns in variation loss, e.g. natural selection and drift.

CHAPTER THREE

METAVARIATION

The last chapter pointed out that many known genetic factors could be perceived as being a part of a system for regulating the production of genetic variation. It also explained how, given variation in these heritable factors, natural selection can act to change the pattern of variation within a population. Chapter One described why we <u>could</u> think about a system for regulating variation production. The present chapter will consider control systems for variation production in the abstract, and take a theoretical look at the conditions under which such a system <u>should</u> exist, if at all.

"Metavariation" is my word for "variation in variation production". Given heritable variation in phenotype, selection can mold the genetic basis of phenotype. Given heritable metavariation, selection can tailor variation production.

<u>An</u> <u>Optimal</u> <u>Rate</u> <u>of</u> <u>Variation</u> <u>Production</u>

The 'function' of additive variation in phenotype is to 'search' phenotype-space for phenotypes better favored by selection. In a changing environment this variation enables the genetic information of a population to track the (changing) optimal phenotype. If, in a changing environment, variation were produced at too slow a rate, the population would go extinct once the changing conditions exceeded the tolerance range of all the phenotypes. Competition with another population of organisms that was better able to track the changing conditions could cause extinction even if all other aspects of the environment

could be tolerated. So a genetic system can produce variation at too slow a rate.

But variation can be produced at too great a rate also. Unless the environment changes to make the original phenotypes less fit, producing experimental phenotypes is a bad gamble rather than a necessity. If the environment has been the same for a long time, the population will likely have evolved to be near an adaptive peak, where the fitness function is concave down. In such a case greater variation production will mean greater production of less fit phenotypes (see Chapter One). If the environment remains static, the best strategy would be to produce no variation at all.

It appears then that there is an optimal rate, or at least a best range of rates, at which to produce variation given a particular rate of change of the environment. In accepting evolution we are accepting the idea that the world is a changing one. But what if the rate of change changes? We have just recognized that there is an optimal rate, or range of rates, at which to produce variation that is determined by the rate of change of the environment. If we live in a world of changes in environmental rates of change it would seem reasonable that genetic systems be capable of tracking the moving compromise between adaptedness and adaptability. Just as variation in phenotype is useful in tracking an optimal phenotype, variation in variation production rates permits the tracking of an optimal variation production rate.

The notion of an optimal rate at which to produce genetic variation is not new. Kimura (1960) quotes Auerbach (1956):

"Thus each species has to strike a balance between the short-term requirement for a low frequency of mutation and the long-term requirement for an ample store of mutant genes. A species in which mutations are too frequent will die out because too many of its individuals are weak, short-lived or sterile. A species in which mutations are too rare may do well for a time, but will not survive when altered conditions demand adaptations for which it does not possess the necessary genes."

Kimura states: "These considerations inevitably suggest that there must be an optimum mutation rate for the survival of a species under a given rate of environmental change."

Kimura (1960) then goes on to describe what he calls the "Principle of Minimum Genetic Load", by which the expected optimal mutation rate can be determined. His essential points are these:

1. Haldane's (1957) substitutional load, or cost of substituting one allele of a gene for another, is independent of selection intensity and depends, instead, on the degree of dominance and initial frequency of the favored mutant.

2. Substitutional load (Le) decreases with greater initial frequency, and therefore decreases with mutation rate.

3. Mutational load (Lm) increases with mutation rate.

4. The genetic system should modify mutation rate (and degree of dominance) so that L= Lm + Le is minimized.

I have presented Kimura's argument only for the sake of interest. Given a familiarity with the concept of genetic load, Kimura's argument is straightforward to explain. Unfortunately, his argument in terms of genetic load is a group selection argument, because a genetic load is "the proportion by which the population fitness is decreased in comparison with an optimum

genotype" (Crow 1958). The reasons I have described for the existence of an optimal rate of variation production are less concise, but they do not involve group selection.

I would like to try to forstall some confusion before I go on. In Chapter One I pointed out that the relative importances of adaptedness and adaptability depend on the time scale over which the results of selection are considered. In the present chapter I have noted that the optimal compromise between adaptedness and adaptability is dependent on the rate of change of the environment. These are separate notions. For a given time horizon the rate of environmental change will determine the optimal compromise. For a given (changing) environment the time horizon of an observer's comparison will determine the best strategy.

To Track or not to Track?

Whether tracking a best phenotype or a best variation production rate, the basic problem is in determining how responsive to be to 'movement of the target'. How adaptable should a system be, or when are the immediate costs of adaptability worth bearing?

I consider Richard Levins (1962,1963,1964a,1964b,1965,1967) to be the pioneer of most of the tracking ideas. The issue of adaptability arises as soon as the world is viewed as one of "changing environments". (He described Layzer's mechanism algebraically in Levins (1964b), p. 638) Levins recognized that variation production was an adaptive system, and that the genome might be tuned to produce the optimal rate of variation

production by secondary selection.

The immediate costs of variation production are only worth bearing if the target is predictable 'enough'. "If the pattern of environmental change is such as to make past environments poor predictors of present environments, populations that have responded adaptively to past environments will be ill-adapted to present ones" (Levins 1965). Predictability, or autocorrelation, is therefore an important environmental parameter in the determination of the adaptedness/adaptability compromise. The determination of what is predictable 'enough' is beyond the scope of this paper. But Levins explored autocorrelation with a simulation model in which phenotypic variance was adjustable via selection on a genetic determinant of the average effect of an allele on phenotype. His analytical calculations (1964) and Monte Carlo experiments (1965) both suggest that phenotypic variance should be greater than zero only if the environmental autocorrelation from one generation to the next is greater than about 0.8.

Metavariation

In my discussion of how genetically responsive a population should be to changes in the environment I have been using a spatial metaphor of tracking, or searching for, a moving target. Levins' problem was to look at the adaptedness/adaptability compromise in the genetic determination of <u>phenotype</u>. My interest here will be in examining the analogous problem of genetic determination of <u>phenotypic variation production</u>. Phenotypic variation has the function of tracking in phenotype-

space; variation in variation production, which I will call 'metavariation', has the function of tracking in variation-production-space. The problems are the same except that they are on different levels.

Earlier work has demonstrated that autocorrelation in the environmentally determined optimal phenotype is an important factor in determining how well to track, i.e. at what rate to produce phenotypic variation. By analogy, autocorrelation is probably important in determining the rate at which to produce metavariation.

Time Scale Dependence

At this time it must be emphasized that the statistic 'autocorrelation' is time scale dependent. "The autocorrelation function ...is a measure of the degree of correlation between a series as observed and that same series if initiated after a specified time lag." (Finerty 1980) Levins was dealing with a one-generation time lag, i.e. how predictable an environment is from one generation to the next. This is because one generation is the minimum time over which differential reproduction can influence gene frequencies; it is the minimum time scale over which the genetic system can respond to a changing environment. Any environmental change on a shorter time scale, i.e. within a generation, must be dealt with by a faster response. Faster responses include developmental adjustment, acclimatization, and shorter term behavioral and physiological responses. Given this time scale consideration, we now make the refinement that a tracking system should respond only if the environment is

predictable enough on the time scale over which the system <u>can</u> respond.

When <u>is</u> <u>Metavariation</u> <u>Important</u>?

The computer model described in Appendix I works well for its intended purpose of comparing different variation production schemes in various environments. The selection regime in a one-dimensional environment is modelled as a fitness function in the form of an adaptive peak defining the optimal phenotype. The form and breadth of the adaptive peak are modifiable to change the immediate cost of variation production (unfit variants). The environmental change from generation to generation can be arbitrarily specified. Up to six asexual populations with "arbitrary" variation production schemes can be run at the same time, independently or in competition with each other. The output consists of the following statistics for each generation: population size, average phenotype and relative average fitnesses for each population, and average rate of variation production for all populations. The data are produced in tabular and plot form.

The above computer model is not adequate for determining the conditions under which metavariation is important. The question is not whether any particular variation scheme is better than another under certain circumstances, but whether <u>adjustable</u> variation, given its costs, is ever at an advantage compared with <u>fixed</u> variation production. The computer model confirms what has been discussed about different optimal rates of variation production for different kinds of environmental

change, but its constraint of asexual reproduction restricts generality, and it is not equipped to compare a population with adjustable variation with one with fixed variation. Furthermore, it was unclear to me how such an exploratory model could be used to confidently delimit all of the conditions under which the costs of a metavariation system would be worth bearing. The costs of metavariation are of two kinds:

1. First, there is the additional metabolic cost incurred to the individual organism in the form of DNA involved only in the regulation of variation production. In the short term a genome comprised solely of DNA involved in development and maintenance of the individual phenotype might be 'cheaper'. It is uncertain what effects "extra" DNA would have on an organism. Some authors are prepared to assume that a significant proportion of eukaryote genomes is "junk", useless to the individual. Given the large amount of DNA in eukaryotes that is of no demonstrated usefulness, it might be safe to assume that the amount of DNA required for a metavariation system would have a very small effect on immediate fitness. See Chapter Four for further discussion of these concerns.

2. A second category of costs arises at the lineage, or population levels. It can be measured by comparing the costs of having a fixed variation production scheme, and not responding to change in environmental change, with the costs of possibly responding inappropriately. As just discussed, such inappropriate response can occur if the environment is not sufficiently autocorrelated on the time

scale over which the regulation system can respond. A system that responds too quickly to immediate conditions is disadvantageous if immediate conditions are anomalous, and therefore bad predictors of future conditions. Given a predictable environment, a system may respond too slowly, and thus always be adapted to the way things were.

A quantitative assessment of the conditions required to favor metavariation over fixed variation requires knowledge of:

1. the tolerance of individuals to the amplitude of environmental variability. This is important in determining the cost of not responding, and the benefit of responding.

2. the potential speed of response of a particular metavariation system

3. the predictability of environmental change, from the viewpoint of the organism, on the time scale of response of its assumed metavariation system.

I concluded that my original intention of quantitatively describing all the sets of conditions under which a metavariation system would be advantageous was too complex to manage within the scope of this thesis. It would have included enough unmeasurable variables to be of little immediate use. Below I offer a qualitative assessment of how fast a variation regulation system might respond, should it exist.

But first, I would like to emphasize that my intended quantitative assessment was to answer just a particular instance of the more general question "What speeds of incorporation of

information about the environment into the genotype are advantageous to living systems?" I can imagine four different models of variation production and selection, each with a different feedback speed of information from the environment to the genetic material:

1. trial and error -- This is the standard assumption of random (undirected) variation. Mutations happen in a way uncorrelated with their potential usefulness. See Warburton (1967) for a model of selection based on the theory of guessing games.

2. patterned trial and error -- Variation production can be focussed in patterns that are more likely to be beneficial. Selection on alpha-genes produces 'heuristics' stored in (longer memory) beta-genes.

3. stress-produced trials -- Lifetime experiences enhance variation production in an individual's germ line with respect to stressed traits (see McDonald 1983).

4. direct programming (Lamarckian) -- No trial and error process moderated by natural selection. Lifetime experience programs an individual's gametes. Characteristics are acquired steadily, if not all at once. Variation production is always in the direction of improvement.

Not all of these feedback speeds have received equal attention. Is that reasonable? A quantitative demonstration that Lamarckian inheritance would seldom or never be advantageous, even if it _were_ possible, would be much more powerful than discrediting it on the basis of what we do not know, i.e. lack

of evidence for the required connections between environment and genetic material. The quantitative theory necessary to specify the potential utility of any of these feedback speeds is, to my knowledge, lacking in the biological literature. This is an area requiring attention.

Consider Figure 4, which pictures a hierarchy of control systems: the phenotype includes several levels of control system (see Bateson 1963), the alpha genes control the phenotype that is developed, and the beta-genes control the generation of new (alpha-) genotypes. (The 'alpha' and 'beta' distinction was made by Layzer 1980.) According to hierarchy theory (see Simon 1973) we expect higher levels in the hierarchy to change more slowly and to determine longer term patterns, and lower levels to be faster and important in the short term.

In the hierarchical organization of Figure 4, changes in each level are selected via the levels below. (This is not group selection.) The 'selection pressure' each level faces is comprised of the differential 'fitnesses' of the level just below. The differential fitnesses of phenotypes determine the change in gene frequencies of alpha-genes within a population. The degree to which the strength of selection on the phenotypes is reflected in a change in gene frequencies depends on the degree of correspondence between phenotype and genotype, i.e. the heritability, h^2 ($0 < h^2 < 1$). This correspondence is never perfect because through dominance and epistasis more than one genotype can code for a given phenotype and, since the development of a phenotype proceeds in interaction with the environment, environmental variation can produce different
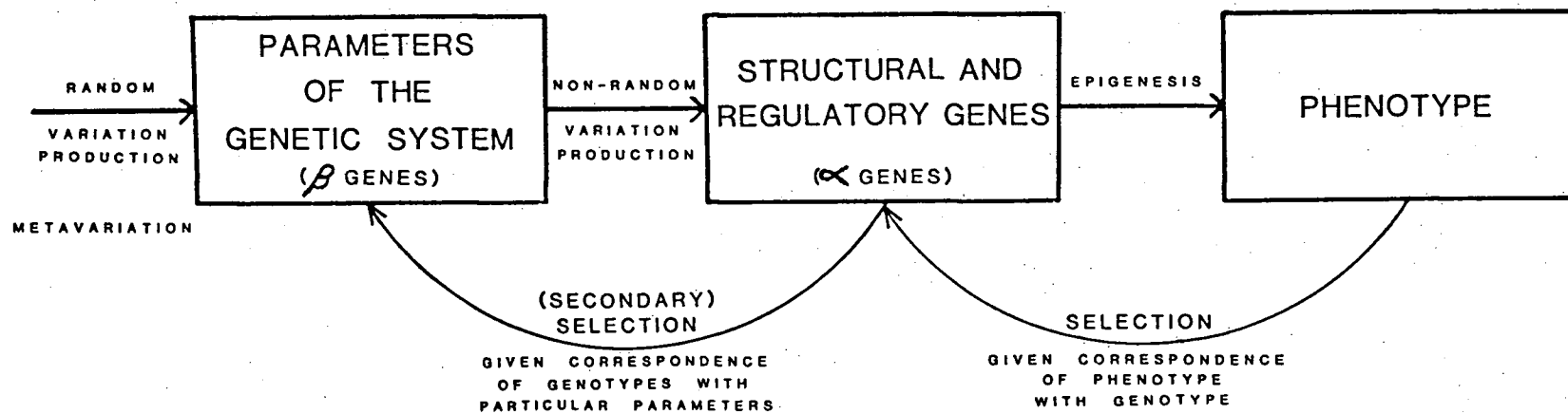
Figure 4. The relationship of alpha and beta genes, and the levels of indirection in secondary selection.

phenotypes from the same genotype.

$$\begin{pmatrix} \text{response to} \\ \text{selection} \end{pmatrix} = h^2 \begin{pmatrix} \text{strength of} \\ \text{selection} \end{pmatrix}$$

(Roughgarden 1979)

So we can say that

$$\begin{pmatrix} \text{change in alpha} \\ \text{gene frequencies} \end{pmatrix} = h_1^2 \begin{pmatrix} \text{strength of selection} \\ \text{on phenotypes} \end{pmatrix}.$$

The differential 'fitnesses' of genotypes (alpha genes) determine the change in frequencies of the 'metagenotypes' (beta genes). The correspondence between these levels is not perfect either ($h_2^2 < 1$). This is because the beta genes (as defined by Layzer(1980) and in Figure 4) just tailor the <u>variability</u> of each alpha locus (rather than prescribe the particular allele, which would involve a Lamarckian mechanism or setting variability=0); a given beta allele may cause the generation of any number of different alpha alleles, and conversely, a particular alpha allele may be associated with many different beta alleles. So we have

$$\begin{pmatrix} \text{change in beta} \\ \text{gene frequencies} \end{pmatrix} = h_2^2 \begin{pmatrix} \text{strength of selection} \\ \text{on alpha-genotypes} \end{pmatrix}$$

Sexual recombination further decreases the correlation, or linkage disequilibrium, between the beta alleles and the alpha alleles that they produce. The reader is referred to Karlin and McGregor (1974) for evidence that the mechanism of secondary selection can work within sexually reproducing populations. They determined that "Linkage between modifier and primary loci appears to affect only the speed of fixation or approach to polymorphism but not the qualitative nature of the outcome" at the modifier (beta) locus.

There are two levels of indirection between selection on phenotypes and changes in beta gene frequencies (see Figure 4). The strength of selection on the phenotype is diluted by two

successive 'heritability' factors, so that the response at the
beta gene level, the variation-patterning level, is very slow.
Levins (1965) calculates that this response would nevertheless
be important on a time scale much shorter than a species
lifetime. But the point is that the patterns of change involving
adjustment of variation production are on an evolutionary time
scale, rather than the generation to generation ecological time
scale usually dealt with in classical population genetics.

The Real World is Multidimensional

For simplicity I have, until now, ignored the fact that the
world is multidimensional. In Chapter One I explained the
modification of variation production only in terms of one, all-
important, trait. That simplification avoided the facts that
different traits can have different variation production rates,
and that there can be covariance in fitness among those traits.

> NOTE: 'Trait' is an awkward term in that it refers to a
> subjectively defined component of phenotype. Almost any
> aspect of phenotype can be called a trait. The term is
> usually used to identify components of phenotype
> particularly relevant to the selection regime being
> discussed in order to facilitate phenotype-phenotype
> comparisons. That is how it is used here.

Just as phenotype- or trait-space is multidimensional, so
is variation-production-space. Adjusting rates of variation
production makes sense only if the rate of environmental change
changes. But the environment may not change in all respects at
once, and only a subset of variants may have potential

usefulness. For example, if the environment only changes in temperature then only temperature variants would be required for adaptability -- variants of any other sort have no potential usefulness.

Consider the problem from another perspective. Let us say that variation production is regulated in populations of a particular type of organism by a mechanism affecting all traits, say by different alleles of a genetic repair enzyme, one of which is more effective than the other. Selection for temperature variants would indirectly favor the less effective allele that produced more variants. But this directional selection would be in opposition to the stabilizing selection acting on all other phenotypic traits. The latter would favor the more effective repair enzyme that caused lower variant production. And if the directional selection were strong enough to select for greater production of variants, because of the general effect of the genetic repair system, variants in all traits would be produced, whereas only temperature variants have a chance of having a higher fitness.

From this kind of argument it is clear that the most responsive variation adjustment system would be one that could independently regulate the production of variants of different traits.

In the population genetics literature, the interference of selection on one locus with selection on other loci that are statistically correlated (in linkage disequilibrium) with it is known as the Hill-Robertson effect (Felsenstein 1974). The process of recombination destroys linkage disequilibrium among

loci in a population, and that function underlies all explanations of the evolutionary advantages of recombination (Felsenstein 1974, Felsenstein and Yokoyama 1976).

Variation Production and Genetic Memory

The spatial metaphor of target-tracking is not the only metaphor useful in discussing evolutionary dynamics. Insights can also be gained by thinking of evolution in terms of acquiring and rejecting information.

The array of gene frequencies in a population contains indirect information about its environment in the present and, since fixation or loss takes time, in the past. Warburton (1967) has, in fact, modelled natural selection as a guessing game where mutations are "guesses" and successful alleles are sufficiently "right" answers.

Levins (1968) points out that the gene pool of a population is a memory, as well as a recipe for phenotypes capable of survival. A very long memory, in which gene frequencies depended on all the environments of the past, with equal weight, would "know" a lot about the environment and its history, but it could not track recent conditions. To track the recent environment, gene frequencies would have to be primarily determined by the recent environment, rather than the history of conditions. Memory is lengthened or shortened by reducing or enhancing genetic variation production, respectively.

"The paradox which now emerges is that only a system with short memory can follow the environment. But the optimum parameters of the tracking system depend on the mean,

variance, and autocorrelation of the environment. These can only be estimated accurately by a system with a long enough memory so that the law of large numbers operates. Since the statistics of the environment are also subject to change, the calibrating system cannot have infinite memory. There is some optimum level of memory for it, which can only be established by systems with longer memory, etc." (Levins 1968)

Levins perceives a problem, or paradox, because he is tacitly assuming that a population can have only one memory, i.e. one gene pool. If organisms contain only alpha genes, that is true. But if they had beta genes as well, that would mean _two_ memories for a population: one short memory of alpha gene frequencies, and another longer, more slowly changing memory of beta genes encoding the strategy for the alpha genes.

So if we suppose that successful organisms are constructed in such a way as to deal with the adaptedness/adaptability tradeoff at the genetic level, since the assessment of the existence of sufficient autocorrelation in the environment requires a long memory, and responsiveness requires a short memory, does it not make sense to expect at least two kinds of genetic memory?

CHAPTER FOUR

GENOME SIZE PATTERNS

I have been asserting the possibility that genetic control determines a pattern of variation production suitable to the rate and direction of change of a population's environment. Since a genetic control system has implications for genome structure, and genome size (the total amount of DNA per cell) is a crude indicator of differences in genome structure, it seems reasonable to examine patterns in genome size among species for evidence of patterns across environments differing in rates of change.

Below I review known patterns in genome size, take a critical look at the existing explanations for them, and propose an explanation based on environmental rates of change and the regulation of variation production.

The Patterns and the C-value Paradox

There is extensive variation in the amount of DNA per genome across animal and plant taxa (see Figure 5). The range in genome size within each group also varies considerably. While there is little species to species variation among reptiles, birds, and mammals, DNA amounts in amphibians span the entire range of these groups. DNA amount in birds varies by less than a factor of two, but it ranges over three orders of magnitude in algae and protozoa.

Genome size has been observed to vary among taxa at all levels. But intraspecific variation in genome size is still the exception; reports of it are probably over-represented in the
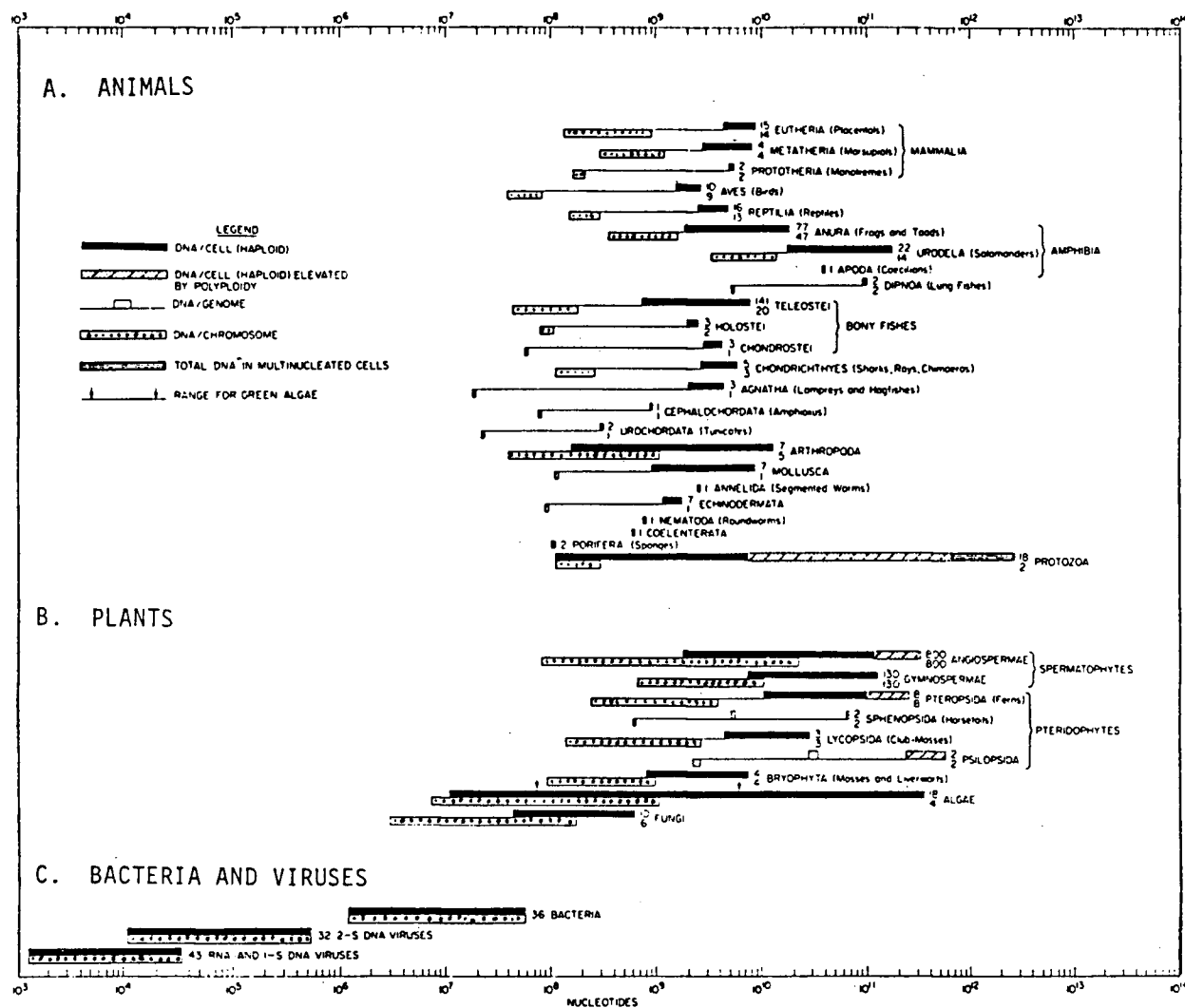
Figure 5. The ranges of DNA (RNA for some viruses) content per cell and per chromosome in major categories of prokaryotic ad eukaryotic organisms. The number of species represented is to the right of each entry. (from Price 1976)

literature, and many of them have not been confirmed by subsequent studies (Bennett and Smith 1976). Current methods cannot reliably detect variation at levels below 3-5% (Bennett and Smith 1976), and it is therefore possible that much intraspecific variation goes undemonstrated.

The notion that the role of DNA is to program an organism and its development leads to the expectation that the amount of DNA should vary with the information requirement of organisms. But from what is known about information requirements, there is no significant correlation (Sparrow et al. 1972); this problem has become known as the 'C-value paradox'. (Whereas the n-value of a cell refers to number of chromosomes, e.g. n for haploid and 2n for diploid, c-value is a relative measure of amount of DNA by weight. Absolute weights are typically in picograms.) Moreover, it is difficult to describe why there is so much DNA in any eukaryote species. Although there are problems in estimation (Bishop 1974), most estimates indicate that only a small proportion of the eukaryote genome codes for proteins (Crick 1971; Ohno 1972). The hypothesis that the remaining DNA is involved in transcriptional control and regulation (Britten and Davidson 1969,1971; Zuckerkandl 1974,1976) seems unable to explain why two closely related species would differ markedly in their DNA contents (Walker 1968).

Although the expected rule of genome size increase with organism complexity does not hold, many patterns in genome size have been observed. It is not known what these patterns mean. Studying them may suggest a function for 'excess' DNA, turn out to be consistent with the dynamics of 'junk' and/or 'selfish'

DNA (see below), or point to some other factor causing particular values of both DNA quantity and the correlated character. At present there seem to be almost as many 'explanations' as there are patterns. In Table 2 are listed patterns in genome size reported in the literature. This list is not intended to be exhaustive, but to illustrate the apparent non-random distribution of DNA amount.

The Explanations

It is obvious that there are two broad possibilities for the mystery DNA. It could have a function, and many authors have cautioned that it might not. It could just be 'junk' and that DNA is, afterall, "... rather an ignorant molecule that frequently gets out of hand and is quite capable of generating a multitude of sequence arrangements" (Dover 1980).

The confounding of correlation and causation appears in the context of genome size patterns in the fact that even function-less DNA might have some phenotypic effect. "Since the range of adaptive stories is as wide as our minds are fertile, new stories can always be postulated" (Gould and Lewontin 1979 in Doolittle and Sapienza 1980). The DNA could be of adaptive value for a certain selection regime (causation), or merely be tolerated by it.

I will now consider published explanations for patterns in genome size. These explanations are most easily treated in groups according to the underlying assumption about what the role is of the DNA that is differentially distributed across species so as to produce the patterns.

TABLE 2

Patterns in Genome size

1. radiosensitivity varies directly with DNA content
        (Sparrow and Miksche 1961; Bowen 1962: Baetcke et al. 1967; Underbrink et
        al. 1968)

2. radiation-induced mutation rates vary directly with DNA content
        (Sparrow et al. 1968; Abrahamson et al. 1973)

3. proportion of genome comprised of repetitive DNA increases with genome size
        (amphibians: Mizuno and MacGregor 1974; Strauss 1971; conifers: Miksche and
        Hotta 1973; angiosperms: Flavell et al. 1974) Contrary evidence: Vicia
        :Chooi 1971b)

4. high DNA content usually implies much heterochromatin
        (Stebbins 1966)

5. DNA/cell varies directly with chromosome size
        (Baetcke et al. 1967; Sparrow et al. 1972)

6. DNA content may or may not correlate with chromosome number
        (see Hinegardner 1976)

7. DNA/cell varies directly with nuclear volume
        (Commoner 1964; Baetcke et al. 1967; Sparrow et al. 1972)

8. DNA/cell varies directly with cell size
        (animals:Commoner 1964; seed and pollen:Jones and Rees 1968, Bennett 1972,
        Bennett 1973, Jones and Brown 1976)

9. adult body size increases with DNA content
        (molluscs: Hinegardner 1974a; Drosophila :Endow and Gall 1975; polyploid
        plants)

10. mitotic cycle time increases with DNA content
        (Van't Hof and Sparrow 1963; Yang and Dodson 1970; Evans and Rees 1971;
        Evans et al. 1972)

11. meiotic cycle time increases with DNA content
        (Bennett 1971; and polyploidy: Bennett and Smith 1972)

12. inbreeding/outbreeding trends are significant in different directions in
    different genera
            (see Rees and Hazarika 1969)

13. minimum generation time increases with DNA content
        (Bennett 1972; Smith and Bennett 1975)

Table 2 (cont'd)

14. annuals have less DNA than perennials
    (Lathyrus : Rees and Hazarika 1969; Vicia : Chooi 1971; many plants:
    Bennett 1972; Ranunculus : Smith and Bennett 1975)

15. annuals have more DNA than perennials
    (Lolium :Jones and Rees 1967; Anthemideae :Nagl 1974; Phalaris :Kadir 1974)

16. temperate plants have larger genomes than tropical plants
    (Avdulov 1931; Stebbins 1966; Levin and Funderberg 1979)

17. within a group DNA increases with latitude
    (Picea sitchensis :Burley 1965, Miksche 1971; Pinus :Mergen and Thielges
    1967)

18. coastal populations have more DNA than inland populations
    (Pseudotsuga menziesii :El-Lakany and Sziklai 1971)

19. deep sea fishes have more DNA than their shallow water relatives
    (Ebeling et al. 1971)

20. primitive species have more DNA than new species
    (Stebbins 1966; Hinegardner 1976; Lathyrus :Rees and Hazarika 1967; Crepis
    :Jones and Brown 1976; Bachmann et al. 1972)

21. generalists have more DNA than specialists
    (plants: Stebbins 1966; teleosts: Hinegardner 1968; amphibians: Bachmann et
    al. 1972; insects: Bier and Muller 1969; mammals: Bachmann 1972; molluscs:
    Hinegardner 1974a; echinoderms: Hinegardner 1974b)

22. the distribution of DNA content within major groups is usually asymmetrically
    distributed, skewed toward the high end
    (Hinegardner 1976, and refs. there)

23. fish families with smaller average genome size also have less variation in
    genome size
    (Hinegardner and Rosen 1972).

24. average genome size of a taxon is inversely related to the number of subtaxa
    (Bachmann et al. 1972; Mirsky and Ris 1950; Hinegardner 1968; Goin and Goin
    1968)

Extra DNA plays no role in phenotype

If 'excess' DNA has no effect on phenotype, then patterns in DNA amount could be explained by processes acting internally, within the genome. Stochastic processes could generate random patterns in genome size. This would not explain the many non-random patterns reported. On the other hand, it might be that increases and decreases in amount of DNA are not equally probable, in which case an orthogenetic trend would result. Hinegardner (1976), for example, labels groups as capable or incapable of DNA increase, thus presuming a difference in genome organization.

Means by which changes in genome size are brought about (polyploidization, unequal crossing over, insertion, deletion, tandem duplication, changes in chromosome number) are documented in Ohno (1970). It is obvious that these processes cannot proceed without at least occasionally affecting the viability of an organism and thereby being subject to selection at one level (internally). So trends in genome size are affected not only by the probable net direction of change, but also by probable viability. It does not seem unreasonable, as an example, that insertions always have a lower probability (than deletions) of disrupting essential translation. This would favor genome increase over time.

Hinegardner and Rosen (1972) consider the 'time hypothesis' that genomes accumulate DNA over evolutionary time. This could potentially explain certain patterns such as primitive species having more DNA than new species (#20) and generalists having larger genomes than specialists (#21, given that the direction

of evolution is toward specialization) but Hinegardner and Rosen (1972) found no significant correlation between the age of a taxon and its genome size. Regardless of the existence or non-existence of a correlation this hypothesis has other serious problems. How is a genome reset to 'small' at the 'start' of a species? It is difficult to cast this hypothesis in terms of mechanism rather than mere correlation.

It is a very different and more limited idea, as Dover (1980) emphasizes, that a significant proportion of 'junk' DNA be 'selfish' in that its sequences promote the accumulation of like sequences, again with little affect on the phenotype (Doolittle and Sapienza 1980; Orgel and Crick 1980). Selfish replication is another reason why DNA might increase over time.

At this point it seems impossible to proceed without appealing to natural selection as a factor in generating genome size patterns. Since the pioneering work of Mirsky and Ris (1951) it has been apparent that genome size has both increased and decreased during the course of evolution. How can one explain reversal in an orthogenetic trend? By a change in genome organization? For what reason?

And what determines the limit to the accumulation of selfish DNA in different species? Obviously there would have to be logistical limits to the total amount of DNA at some point. The proponents of selfish DNA invoke "metabolic disadvantage relative to organisms with less selfish DNA" (Orgel and Crick 1980) and "energetic burden" and the destruction of needed sequences by selfish elements (Doolittle and Sapienza 1980) -- they appeal to selection forces on the phenotype. No doubt

metabolic costs would exist for junk as it would for selfish
DNA. In the end the concepts of junk or selfish DNA can only
justify the existence of 'excess' DNA in the genome, and
contribute nothing to the explanation of patterns in genome size
across species. The explanation of patterns in the particular
amount of functionless DNA, or the ratio of useless to useful
DNA (Orgel and Crick 1980) requires the explanation of the
relative tolerances of various selection regimes to the
consequences of its presence. "Intragenomic selection is clearly
important for understanding evolution within genomes and for
understanding the significance of certain DNA sequences, but it
does not in itself clarify the evolutionary forces determining
genome size..." (Cavalier-Smith 1980b).

Extra DNA has nucleotypic effects

DNA could play a role intermediate between those of mere
junk and of an alphabetic code -- its bulk alone could determine
aspects of phenotype, independent of sequence. If this were the
case, selection could act not only in favour of smaller genomes
(against the accumulation of junk or selfish DNA), but in favour
of larger genomes under conditions where a greater quantity of
DNA produces a phenotype with a selective advantage. Bennett
(1971) called this a 'nucleotypic' effect, and the set of traits
resulting from such sequence-independent effects the
'nucleotype'. "The nucleotype is therefore a genotypic, but not
a genic character" (Bennett 1981).

The idea of nucleotype appealed to Cavalier-Smith
(1978, 1980a, 1980b) who expounds a 'nucleotypic theory'. He

(1978) proposes two functions of DNA other than those directly or indirectly involved in coding for proteins:

1. control of cell volume, and

2. determination of nuclear volume.

Function (1) is determined by the number of replicon origins in the genome. (A replicon is a unit of DNA replication.) Cavalier-Smith (1978) refers to models (Sompyrac and Maaloe 1973; Donachie 1974) that postulate the accumulation of an initiator, or the dilution by cell growth of a repressor, specific for replicon origins. When the concentration of the initiator/repressor attains a particular concentration replication is initiated (or ceased to be repressed) and the cell divides. Thus maximum cell size is determined.

Cavalier-Smith (1978) assumes that the volume of the nucleus is determined by the bulk of its contents. He feels that this may be important because it thereby determines nuclear surface area (and pore number) with a consequent effect on nucleocytoplasmic transport of RNA and, therefore, 'growth rate'.

Cavalier-Smith attempts to apply these non-genic functions to the explanation of genome size patterns (and the resolution of the C-value paradox) by linking up with some ideas about r- and K-selection theory:

" The great diversity of cell volumes and growth rates, and therefore of DNA contents, among eukaryotes results from a varying balance in different species between r-selection, which favours small cells and rapid growth rates and therefore low DNA C-values, and K-selection which favours

large cells and slow growth rates and therefore high DNA C-
values." (Cavalier-Smith 1978).

The link between K-selection and large cells and slow growth is
never explained by him. But in any case, given selection for
larger cell size, "Though there are undoubtedly several ways of
evolving larger cells, a simple and direct way would be by
increasing the number of replicon origins involved in the
volume-dependent control of DNA replication." (Cavalier-Smith
1980b). But in the next sentence he points out that "This cannot
provide the fundamental explanation of the C-value paradox,
since it does not necessitate a larger genome (as it involves
only a small fraction of the genome)..." So where is the link
between selection and C-value?

"On this theory it is the extra replicon origins not the
larger nucleus, or the larger genome as such, which increases
cell size: what the C-value controls more directly is nuclear
volume." (Cavalier-Smith 1980b) So let us look at the causal
chain between selection and the second proposed function of DNA,
that of determination of nuclear volume.

"... larger cells require more rRNA transport to the
cytoplasm per cell cycle than do smaller cells. One would
therefore expect selection to increase the rate of RNA
transport in larger cells, lest it become rate limiting to
cell growth and unduly lengthen the cell cycle. This could
be done by increasing the amount of skeletal DNA (S-DNA) so
as to increase the nuclear surface area and the number of
nuclear pores. The suggestion is therefore that the excess
DNA is used, not, as has been incorrectly stated, to slow

development but rather to prevent the excessive slowing of development which would otherwise be caused by large increases in cell size." (Cavalier-Smith 1980b)

In other words it is selection for <u>shorter</u> cell cycle time that favours increased C-value. At this point it is apparent that "K-selection for large cells and slow growth" has no effect on C-value. Cell volume can be increased by increasing the number of replicon origins with little effect on C-value, and large C-values in large cells are really due to r-selection to prevent excessive lengthening of cell cycle time.

"... what should never occur on my theory -- and has not been observed -- is that high C-value organisms have small cells and short life cycles." Why not? A large genome (and high rate of RNA transport) with few replicon origins would fulfill the requirements given the two non-protein coding functions given at the outset.

A final caution is that the DNA functions proposed by Cavalier-Smith are not entirely sequence-independent. Replicon origins are presumably particular sequences, and one role of structural DNA is to code for structural RNA.

Cavalier-Smith presents us with several new ideas but, as I have shown, they are not sufficient to form a 'nucleotypic theory'. At present all that exists is the suggestion that DNA quantity may have a sequence-independent (nucleotypic) role (Bennett 1971).

DNA differences are in amount of primary DNA

Another selectionist hypothesis of great potential generality in explaining genome size patterns is what I shall call the 'loss-of-parts' hypothesis advanced by Hinegardner (1968,1976; Hinegardner and Rosen 1972). The main components of this hypothesis are as follows:

1. Some groups of organisms can increase their genome size, while others cannot, or do not.

2. In groups that do not increase their DNA "The consequence of DNA change and loss is evolution toward specialization and eventual extinction." (Hinegardner 1976)

3. Specialization involves loss of parts and/or functions. "A specialized fish would be expected to have fewer parts, since the derivative condition of specialization is the adaptation to a restricted mode of life not requiring the use of all structures present in the generalized form. Though it would be possible to quantitate fish parts, it is hardly necessary; even the gross picture one gets from examining fish anatomy shows that the generalized fishes have more parts. They tend to have more separate elements in the hyoid apparatus and gill arches, more vertebrae, intermuscular bones, skull bones, and fin rays. Certainly the trend is not in the opposite direction." (Hinegardner and Rosen 1972)

4. DNA coding for the lost traits can be lost, and is.

There are several difficulties with this explanation of genome size patterns. One is the postulation of orthogenetic trends, that some groups can and others cannot increase their

DNA. No reason for these trends is given. Another difficulty is the interesting question of whether the direction of evolution is toward specialization. The definitions of generalized and specialized were as follows:

> "Generalized will be used to describe organisms that share numerous features with other members of their taxon. In contrast, a specialized organism shares fewer features with the members of its taxon and will differ from them in presence or absence of certain features." (Hinegardner 1976)

These definitions imply that evolution will be in the direction of specialization, making the stated trend tautological. Evolution is change, and if change means gain or loss of characteristics it will imply specialization by the above definition (unless all subtaxa change in the same way so that there is no diversification).

A third problem is the connection between specialization and the loss of DNA. Some modes of specialization would seem not to require fewer parts and functions. Viviparity, for example, is viewed (Hinegardner and Rosen 1972) as a particular form of specialization within the Atherinomorpha (Teleostei), although it is not an obvious simplification of oviparity. Also, some authors (e.g. Valentine 1976) assume that the more precise metabolic requirements of specialists require more copies of an enzyme locus. At any rate, the mechanism by which unused DNA is lost is never explained. And why are DNA increases not lost for the same reason that DNA that has fallen into disuse is lost?

Hinegardner(1976) distinguishes two types of DNA in

organisms. "There are the selectively constrained sequences. These are the ones that affect events in the organisms, and include genes and their control. This will be called primary DNA. Then there are the much less constrained sequences produced by duplication; this is the secondary DNA. A fuzzy area undoubtedly lies between the two; however, the two types are probably bigger than the overlap and can be examined as two populations of nucleotide sequences." (Hinegardner 1976)

Why is secondary DNA retained if there is selection capable of removing a piece of primary DNA fallen into disuse? The most serious problem with the loss-of-parts hypothesis is that it is only relevant to primary DNA, "the selectively constrained sequences", and for a few reasons it is likely that primary DNA is a relatively unimportant part of genome size differences. These reasons are:

1. Primary DNA, according to most estimates, comprises only a small proportion of the genome. Hinegardner (1976) considers this and concludes that "At the maximum then, in our average organism primary DNA accounts for 0.6 to 24 percent of the haploid DNA." If primary DNA is a small proportion then selection-determined changes in it are likely to be swamped by changes in the amount of secondary DNA, unless the quantity of secondary DNA is very stable. But in secondary DNA "Changes or losses would be more rapid than in the primary DNA." (Hinegardner 1976, p.195)

Table 3. The range in genome size in teleost taxa as a proportion of the minimum and maximum genome measured in each taxon. The minimum (or maximum) genome size can be viewed as a generous over-estimate of the primary DNA for the group. Calculated from the data of Hinegardner and Rosen (1972).

Teleost Taxon	min.	max.	% of min.	% of max.	# of spp.
Osteoglossomorpha	.77	1.3	69	41	8
Osteoglossiformes	.77	1.3	69	41	4
Osteoglossoidei	.77	1.0	30	23	3
Mormyriformes	1.0	1.2	20	17	4
Elopomorpha	1.2	2.5	108	52	4
Anguilliformes	1.4	2.5	79	44	3
Clupeomorpha	.77	1.9	147	59	6
Protacanthopterygii	2.7	3.3	22	18	3
Ostariophysi	.65	4.4	577	85	75
Cypriniformes	.65	2.2	238	70	43
Characoidei	.71	2.1	196	66	22
Cyprinoidei	.65	2.2	238	70	21
Siluriformes	.88	4.4	400	80	32
Callichthyidae	1.7	4.4	159	61	8
Scolpelomorpha	1.2	1.2	0	0	2
Paracanthopterygii	.68	3.0	341	77	12
Gadiformes	.68	.98	44	31	5
Batrachoidiformes	1.7	3.0	76	43	4
Lophiiformes	.74	1	35	26	3
Acanthopterygii	.48	2.1	338	77	168
Atherinomorpha	.72	1.6	122	55	19
Exocoetoidei	.74	1.2	62	38	5
Cyprinodontoidei	.72	1.6	122	55	11
Atherinoidei	1.1	1.3	18	15	3
Percomorpha	.48	2.1	338	77	149
Gasterosteiformes	.58	.70	21	17	6
Gasterosteoidae	.58	.70	21	17	3
Syngnathoidei	.64	.66	3	3	3
Scorpaeniformes	.76	1.4	84	46	15
Scorpaenoidei	.96	1.4	46	31	4
Hexagrammoidei	.79	.99	25	20	5
Cottoidei	.76	1.1	45	31	6
Perciformes	.59	2.1	256	72	104
Percoidei	.72	1.4	94	49	68
Sphyraenoidei	.83	1.2	45	31	2
Labroidei	.91	2.1	131	57	5
Blennioidei	.81	1	23	19	6
Gobioidei	1.2	1.4	17	14	3
Scombroidei	.88	1.1	25	20	7
Stromateoidei	.80	.81	1	1	2
Anabantoidei	.59	.88	49	33	9
Pleuronectiformes	.65	1.1	69	41	12
Pleuronectoidei	.65	1	54	35	10
Soleoidei	.65	1.1	69	41	2
Tetraodontiformes	.48	1.1	129	56	10
Balistoidei	.64	1.1	72	42	5
Tetraodontoidei	.48	.90	131	57	5

2. If the differences in genome size lie in primary DNA then differences in genome size between related species must be less than the entire amount (most likely a small fraction) of primary DNA in the generalized species. Table 3 is derived from the data of Hinegardner and Rosen 1972 and shows the range in genome size within teleost taxa as a percentage of both the minimum and maximum genome within a given taxon. The ranges are generally much greater than a fraction of a very generous estimate of the amount of primary DNA (e.g. Hinegardner's estimate was 24% and the range is often greater than 24%). This is reason to believe that at least some, if not all, of the differences in genome size are due to varying amounts of secondary DNA. The loss-of-parts hypothesis seems unimportant, if not implausible.

DNA variation is differential redundancy for gene dosage effects

It is often speculated that the causal factor in the positive correlation between genome size and latitude (#16,#17) is temperature, and that has led to the suggestion that more DNA is adaptive in cooler environments.

"... One way to increase protein production is to increase the number of gene copies. Since enzyme activity has a strong dependence on temperature (approximately doubling with each 10°C. rise up to an optimum) it could be one limiting factor with respect to efficient growth at low temperatures in plants and cold-blooded animals. Then, any

organisms having duplications of rate-limiting genes would

have an immediate selective advantage..." (Sparrow et al.

1972).

Stebbins(1966) and others have cautioned that high DNA content

is probably not directly adaptive to low temperatures since a

few strictly tropical (plant) groups have a high DNA content.

Stebbins suggested that gene redundancy governs the rate of

development, which generally has to be slower at cooler

latitudes. He (1966) proposed a model in which gene multiples

are organized in a time-chain, one copy active at a time, so

that the duration of a stage in development is set by the length

of the chain of copies.

Grime and Mowforth (1982) proposed a different model to

relate DNA content and temperature. The idea is that there can

be a strategy of separation in time of cell expansion and cell

division. They cite evidence that mitosis is inhibited at (low)

temperatures that still permit high rates of cell expansion, and

show that shoot expansion at lower temperatures means higher DNA

content (in British flora). Plants which grow in continually

warm climates, or only in the warm season of variable climates,

have small genomes. The association of fast cell expansion with

large genome could be due to dosage repetition requirements, or

to the need for a large nuclear envelope (see nucleotypic

effects).

Grime and Mowforth (1982) suggest that temperature could

influence genome size in animals, too.

"As might be expected, small genomes are characteristic of

warm-blooded animals and it is particularly interesting

that reptiles which flourish in dry hot conditions have uniformly small genomes whereas amphibia include species with exceptionally large genomes and cells. The respiratory system of amphibians depends on the maintenance of a moist permeable skin, so the ecology and behavior of most species involve the avoidance of insolation and maintenance of low body temperature."

Extra DNA influences adaptability

All of the proposed functions for the 'extra' DNA up to this point have centered around 'adaptedness', or suiting the properties of organisms to conditions existing in their environments. It is a different idea that the DNA be important for 'adaptability'. There are several suggestions of this type.

Ohno (1970) stresses the importance of extra, or redundant DNA as experimental material in which to test new genes without suffering the loss of previously existing, and needed, genes. And there is evidence to show that several genes have evolved by gene duplication. Whereas Ohno was primarily concerned with structural genes, Britten and Davidson (1971) stress the importance of having redundant DNA in order to evolve new gene regulation systems, and tried to explain the existence of repetitive DNA in this manner.

Introns (intragenic regions) are portions of the genome that are transcribed, but spliced out of RNA before translation. Gilbert (1978) suggested that introns serve to increase the rate of evolution:

1. large scale changes could be caused by a mutation that altered a splicing pattern

2. they could increase the frequency of recombination between parts of a single gene.

Repetitive DNA could also be important through the spacing out of genes in complex organisms to facilitate recombination.

On the other hand, extra DNA might retard evolution.

"... the failure of the polyploids to evolve new characteristics can best be ascribed to the retardation of evolutionary progress which results from the presence of many duplicated gene loci." (Stebbins 1966)

DISCUSSION

The existence of the C-value paradox is evidence that much is to be learned about the structure and function of the genome. There is much more DNA in cells than is required to code for proteins, and if the remainder plays a role in regulation, why can congeneric species, with apparently similar requirements, differ so much in DNA content? What else could the DNA be doing there?

There are many non-random correlations of genome size with physiological, ecological, and taxonomic parameters (see Table 2). These are clues to the role of the differentially distributed DNA. Since many of the patterns are difficult to explain with existing models of the genome, the differentially distributed DNA could be the same as the 'extra' DNA, linking the two aspects of the paradox.

I have reviewed the explanations of patterns in genome size, treating them according to the underlying assumption of what the extra DNA is. Many of these explanations 'make sense', and this is at least a reminder, if not an indication, that there could be multiple causes underlying the patterns observed. Furthermore, many of the explanations are not mutually exclusive.

As an illustration consider the number of explanations that can be brought to bear on the annual/perennial pattern (#14, Table 2). Annuals generally have less DNA than perennials. One could assume that the difference lies in regulatory DNA, and argue that perennials have a greater regulatory requirement because they are longer lived and therefore must contend with a more variable environment during their life cycles. Or one could assume that the DNA difference was due to 'junk' or 'selfish' DNA, and that the metabolic cost of it were greater in a species with fast development and a short life cycle, so there is stronger selection against useless DNA in annual species. Another alternative is that the DNA effects are nucleotypic (sequence-independent) and annuals have less DNA because they need a shorter minimum generation time (faster mitosis and meiosis), or (for some reason) smaller cells. Hinegardner might suggest that annuals are specialized, usually simpler and with fewer 'parts', and therefore require less DNA. Grime and Mowforth (1982) would look to see if the growth period for annuals was confined to the warm part of the season, when extra DNA would not be required for fast cell expansion temporally separated from mitosis. Someone else might point out that

adaptability might be more important for perennials, and that the additional DNA they contain could be to that end. Careful experiments could distinguish among some of these alternate hypotheses, but several would be quite difficult to falsify.

There are two conceivable ways in which this paradox might be solved. One is that a function be found for the extra DNA, and that this new understanding make us realize why the various trends exist -- what their 'common denominator' is. The second is the possibility of noticing such a 'common denominator', a grand correlation that accommodates all of the patterns found, and exploring the reality of mechanisms (with assumed roles of DNA) that could plausibly generate the patterns.

Genome Size and Rates of Evolution

The grand correlation 'the faster the rate of evolution, the smaller the genome size' could be argued to fit all patterns. Ancient or 'living fossil' species are recognized as such because their rate of change has been very slow for a long time -- and they have larger genomes. Most people would agree that specialists are in more of a 'Red Queen' situation than generalists. Biotic (evolving) factors play a greater role in defining specialists's niches, and they more often find themselves in coevolutionary 'arms races'. Specialists have smaller genomes than generalists (#21). The environment of dispersing annuals could generally differ more from generation to generation than it does for perennials, exposing them more often to directional selection. That the average genome size of a taxon is inversely related to the number of subtaxa (#24),

temperate genomes be bigger than tropical ones (#16), and deep sea fishes have more DNA than their shallow water relatives (#19), are all fairly easily cast in 'rate of evolution' terms.

A New Explanation for Genome Size Patterns

Correlation does not imply causation. What mechanism might explain this grand correlation? Evolution is change in the genetic constitution of a population. From a Neo-Darwinian perspective this implies the action of either directional or disruptive selection. So the problem becomes how these modes of selection can influence genome size. I hypothesize that the answer lies in the observation that both of these evolution-causing modes of selection favor phenotypic variants in the population. This means that genome size or any other characteristic that is non-randomly distributed among phenotypic conservatives and variants in the population will be indirectly influenced by directional or disruptive selection acting on phenotype. This non-random distribution can come about by association of the characteristic with rate of production of phenotypic variants. If, for example, smaller genomes produce variants at a greater rate, then on average phenotypic variants will have smaller genomes and selection for variants will indirectly select for smaller genomes. Stabilizing selection would favor larger genomes. In this manner rate of evolution can be correlated with genome size.

Stated more precisely, my new hypothesis is:

1. The function of the 'mystery' DNA responsible for most genome size patterns is regulation of variation production; more DNA means slower production of additive genetic variation.

2. Secondary selection (Chapter Two) can act via phenotypic variation to influence DNA amount, given (1).

3. Directional selection indirectly favors greater production of genetic variation, and smaller genome size.

Support for the Genome Size/Variation Production Hypothesis

Unfortunately the hypothesized link between genome size and rates of additive genetic variation production has yet to be established. There is only circumstantial evidence in the literature.

Pierce and Mitton (1980) reported a strong negative relationship, in the species they examined, between genome size and genetic variation as measured by average heterozygosity (H) and percent of polymorphic loci (P). Their work does not constitute evidence supporting the hypothesis for these reasons:

1. The statistics H and P cannot be equated with either additive or non-additive variation. Only additive variation influences evolutionary responsiveness.

2. H and P are not measurements of variation production per se.

3. Larson (1981) severely criticized Pierce and Mitton's statistical analyses and concluded that the reported relationship was undemonstrated.

There is, however, considerable literature that is

explicitly or implicitly suggestive of a role of DNA in either promoting or suppressing rates of variation production. The question that needs to be answered is, "Does the kind of DNA observed to be differentially distributed across organisms generally increase or decrease genetic variation production?"

Some "nucleotypic effects" of DNA could influence variation production as well as other phenotypic traits. Selection for greater evolutionary rate would select for shorter generation times as well as more additive variation production. DNA bulk might <u>cause</u> slower cell divisions and longer generations and, therefore, decreased evolutionary responsiveness.

Another nucleotypic effect could occur if the speed of cell divisions (patterns #10, #11), or of certain phases of them (Hotta and Stern 1965), had any influence on the fidelity of DNA replication, either by facilitating error-checking processes or by allowing more time for the proper organization and alignment of genetic material.

The organizing function of heterochromatin may lead to higher replication fidelity. "... heterochromatin may help maintain the proper spatial relationships necessary for the efficient operation of the cell through the stages of meiosis and mitosis", for example it may aid in the initial alignment of chromosomes prior to synapsis (Yunis and Yasmineh 1971). Asynapsis has long been causally linked to mutability (Thompson 1962). But heterochromatin may also increase variation production, for example by allowing chromosomal rearrangement (Yunis and Yasmineh 1971). Supernumerary chromosome segments and B chromosomes are both heterochromatic. They have been observed

to both increase and decrease recombination rates (refs. in Rees 1972; Carlson 1978).

It is relevant to ask what fraction of DNA is responsible for the differences in genome size. On the one hand, there have been numerous reports that genome size differences are largely differences in amount of intermediately repetitive DNA (pattern #3). Evidence suggests (Yunis and Yasmineh 1970) that most intermediately repetitive DNA resides in heterochromatin, and Stebbins (1966) has noted that plants with larger genomes generally have a greater proportion of heterochromatin. On the other hand, there is some evidence of the proportion of repetitive DNA not changing with genome size (Chooi 1971), and evidence that other fractions, e.g. single-copy DNA, may vary significantly with genome size (refs. in Larson 1981).

Probably the best supported of all relationships between DNA amount and variation production is Hsu's (1975) "bodyguard hypothesis". He marshalls considerable circumstantial evidence in favor of his proposal that constitutive heterochromatin functions as a shield against "Mutagens, clastogens or even viruses attacking the nucleus..." (Hsu 1975). A preliminary experiment supported his hypothesis.

It is a common idea that genetic redundancy leads to the reduction of expressed variation. Bachman et al. (1972) describe the suggestion of Bier and Muller (1969) that

> "... the repetitiveness inherent in larger genomes results in a genetic inertia; mutations in single copies of repetitive DNA are not quantitatively important enough to provoke natural selection, even if the genes are

functional."

Stebbins (1966) ascribes the retarded evolution of polyploid plants to the presence of duplicated gene loci. The idea is that mutations, as long as they are not dominant, will have less of an impact on phenotype if they are "covered for" by redundant copies. One problem in applying this coverup notion to species which are not polyploid is that the assumed existence of extensive tandem duplication has not been supported (Price 1976). Some evidence against the "coverup" effect of redundancy is provided by Paquin and Adams (1983), who found the rate of variation production in populations of diploid yeast to be almost twice that of coisogenic haploids.

It might be a general truth that, at levels common in organisms, variation suppression requires more organization, and more DNA, than does variation promotion. The hypothesis that amount of DNA and rate of variation production are negatively correlated is testable (Hsu 1975), but the critical experiments have yet to be performed.

CONCLUSION

I have critically reviewed the existing explanations for genome size patterns. Some are "special-case" and narrow in scope. The more general explanations are internally inconsistent, and untestable.

The correlation "the faster the rate of evolution, the smaller the genome size" is consistent with most of the reported patterns. It is hypothesized that greater amounts of DNA are causally connected with lower rates of additive variation

production. Although it is testable, there is at present only circumstantial evidence for this hypothesis. Given the hypothesized relationship, directional selection for phenotypic variants could indirectly favor smaller genomes, thus producing the observed correlation.

Testing this explanation would require the quantification of "rate of evolution" for many groups of organisms. This problem is akin to that of comparing "niches". It will probably only be possible to make relative statements about rates of evolution, under specially controlled circumstances.

CHAPTER FIVE

TESTING THE IDEAS

I have suggested (Chapter Four) that a significant
proportion of the genome can be considered a genetic memory for
variation production patterns, and I have described a selection
mechanism (Chapter One) by which it can be modified.

How does one go about testing the hypothesized existence of
a regulation system for variation production? The available
options are to seek:

1. genetic evidence of the system itself, and/or

2. evidence of the action of such a system.

Each sort of evidence could be sought in a "natural experiment",
or a purposefully created experiment. After briefly noting which
organisms are most likely to have metavariation systems, I will
speculate on the structure of a metavariation system. An
expected structure is a requirement for seeking evidence of the
system itself. I will then discuss problems in measuring the
expected patterns in genetic variation. A suggested experiment
is outlined, and the outcomes of some related previous
experiments are discussed. Throughout this chapter I exercise
the "metavariation perspective" by generating questions. The
chapter ends with a discussion of the relevance of the ideas to
two "bandwagon" topics.

Where to Look

Organisms most likely to have a metavariation system would live in an environment

1. which changed over time, over a greater range than that which could be accommodated by phenotypic flexibility, necessitating a genetic response.

2. where the change in rates and/or direction of change of the environmental change has to be predictable "enough" to make modification of variation production patterns worthwhile.

The difficulties of choosing a likely organism are those of determining "environmental grain" (i.e. how the environment is perceived by the organism) and the lack of a quantitative theory to properly describe "enough".

Could it be more essential that complex organisms, of lower fecundity, find a way of reducing their losses (unfit variants) in variant production? Making the assumption that the differentially distributed DNA among taxa has its function in the regulation of variation production, it would be interesting to see if there is a significantly better correlation of DNA content with estimated information requirement for simpler organisms as compared with complex ones.

Speculation on Metavariation System Structure

In the last chapter I considered known patterns in genome size. I noted at the outset that differences in genome size could be interpreted as rough indicators of differences in genome structure. Although, as predicted from the metavariation

model, the patterns seemed to vary according to the "rate of evolution" of a group, the linking of greater amount of DNA with less variation production was strictly post hoc. There was no prior expectation about how such variation regulation might be organized. It would be useful to speculate about the genetic structure of a system for variation production adjustment. A hypothetical model system could yield new insight into present knowledge. If some necessary characteristics could be established, the way would be open to falsification of the idea that variation control exists in an organized system. We know that:

1. To maximize the effectiveness of secondary selection on a set of variation production schemes, the variation production modifiers should be closely linked to the genes that influence the phenotype upon which (primary) selection acts. In that way the 'heritability' or correlation between genotypes created from a certain variation production scheme, and the beta genes that program that scheme, is as high as possible.

I intentionally used asexual populations to describe the action of secondary selection (Chapter One). In a panmictic sexual population, if beta and alpha gene were not linked, then they would remain associated for an average of only two generations (Leigh 1970). In that situation there is little association between phenotype and the hypothetical variation control genes responsible for producing it, and secondary selection is least effective.

2. Since evolutionary responsiveness will not necessarily be required of all traits at the same time, the independent control of variation for different traits is advantageous.

3. There must be a certain amount of recombination between the genes involved in producing the traits undergoing selected change, and those genes of traits under different selection regimes, in order to avoid the Hill-Robertson effect (Felsenstein 1974).

One way to have independent control of traits would be to have separate modifiers controlling each locus in the genome (closely linked to the locus for reason (1)). But selection is more effective on modifiers if one modifier affects many loci (Karlin and McGregor 1974). So the most responsive variation regulation system would have only one modifier for each independent trait. It appears that the ways to satisfy the dual requirements of tight linkage to each locus, and one modifier for several loci, are:

1. combine all loci for a trait, and the modifier, in one linkage group

2. have a regulator-controller system isomorphic to the gene regulation system proposed by Britten and Davidson (1969) for gene regulation.

The pleiotropic relationship that can exist between genes and traits makes solution (1) less effective than (2). In a regulator-controller system (2), one or more 'controllers' ('receptors') would be linked to each primary locus, and they would receive signals from particular monomorphic 'regulator' loci (the modifier genes) to, for example, allow or disallow

error-checking by a certain enzyme.

The fact that the structural predictions of the ideal gene regulation system and the ideal variation regulation system coincide is important in itself, since it gives two different interpretations to the same structural evidence. It might affect confidence in the hypothesized gene regulation system. McClintock (1965) has described a regulator-controller system for mutation in maize. And several authors, including Nagl (1979), have voiced suspicion that variation production and differentiation may be related processes.

Two important points are:

1. The most responsive system is not always the most appropriate one, the latter being determined by the predictability of the rate of change of environmental change. It should be apparent that responsiveness would be slowed by the increase of recombination rates between loci and their variation 'controllers', and the decrease of recombination between loci involved in different traits, to produce the Hill-Robertson effect.

2. It may require a less highly formalized variation regulation system than the one just suggested to pattern variation to a significant extent. Gene duplication, or changes in the bulk of DNA present (see other influences in previous chapter), for example, may be enough.

Epigenesis and Canalization

So far I have been making the standard Neo-Darwinian simplification of ignoring the process of epigenesis. The process of epigenesis lies between the genetic mechanisms I have been discussing, and the phenotype (and phenotypic variation). It is the relationships of genes and traits (one-to-one, epistatic, pleiotropic), an epigenetic consideration, that determine the "additivity" of genetic variation.

```
----------> GENOTYPE ----------> PHENOTYPE

     mutation              epigenesis
```
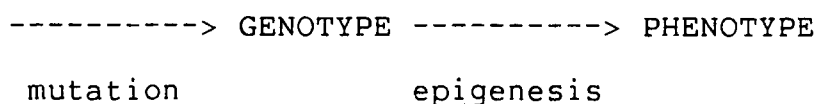
Earlier I mentioned a possible relatedness between differentiation and variation production in the context of the similarity of the genetic organization that each require. Now I would like to note the great similarity of the changing patterns of variation production expected from a metavariation system, and the phenomenon of "canalization" causally attributed to the epigenetic system.

Canalization is the "deepening" of a developmental response so that the same norm is produced despite considerable variation in the environment (Waddington 1957,1975). It refers, not to the decrease in phenotypic variation per se, but to the decreasing sensitivity of the developmental pathway to variation in the environment. At the phenotypic level this implies less variation in the trait, a change that could be mimicked by stabilizing selection on metavariation. Modifier frequencies change slowly. Canalization "...has been observed in situations in which selection pressure continues for many generations." (Waddington

1974) Stable environments and stabilizing selection favor canalization, whereas changing environments and disruptive selection favor its breakdown (Waddington and Robertson 1966, and references therein). Waddington gives credit to the epigenetic system, rather than to a category of (beta) genes, for focussing variation production.

On Measuring Genetic Variation Production

One prediction based on the assumed existence of a metavariation system is that genetic variation in populations be adaptively patterned to suit the environment. (This is the "niche variation hypothesis" about genetic variation.) This is a weak prediction, because finding such patterns would not imply the existence of the hypothesized metavariation system. Stronger predictions would be in terms of changes in variation production patterns. In either case a test involves the measurement of genetic variation production.

Genetic variation (generally measured via phenotype) is usually quantified in terms of average heterozygosity (H) and percent of polymorphic loci (P). These convenient statistics bear no clear relationship to the categories "additive" and "non-additive" genetic variation. Levins (1964a) was clear in his explanation that additive and non-additive variation have different functions in adaptation, and are favored by natural selection under different circumstances. For example, an environment constant in time would not favor the maintenance of additive variation. But that same stability could enable populations to perceive the environment as spatially patchy, a

situation favoring the maintenance of non-additive variation in the population. All predictions of the metavariation model are in terms of additive variation only. Any pattern comparing genetic variation measured by H and P with some other factor is potentially influenced by differing amounts of non-additive variation.

The rest of the problem is in how to estimate variation production. Variation in a population is influenced by several factors such as population size, population structure, and history (time since last bottlenecking). And, should the variation be interpreted as pre-selection, or post-selection? Is low variation indicative of low variation production, or high selection?

Perhaps the best way to measure the additive variation relevant to these hypotheses would be indirectly via measurements of evolutionary responsiveness. This would best be done through selection experiments using the methods of analysis of quantitative genetics.

Suggested Experiment

Ayala (1966,1967,1969) performed experiments in which he compared the evolutionary response of irradiated fruit flies with that of unexposed flies. The increasing population densities in his cages provided the directional selection pressure. Adaptation was assessed in terms of population size, and productivity. After an initial delay, the irradiated populations increased in fitness and became better adapted than the controls.

I propose a similar experiment in which directional selection, instead of ionizing radiation, is used to trigger increased variation production. The hypothesis is that a metavariation system exists. This implies that stabilizing selection will cause variation production to decrease in a population, and directional selection will cause it to increase. The rate of variation production will be compared in two groups of populations by comparing their evolutionary responsiveness. The latter will be measured as ability to maintain productivity in a changing environment. The experimental trick is to stimulate variation production in one group by directional selection in a way that does not confer it with an advantage in appropriate stored variation. This is to be accomplished as follows:

1. Choose a variable that can be conveniently manipulated to produce directional selection in two (opposite) directions, and strong stabilizing selection. A variable such as developmental time, that permits catastrophic cutoff beyond a predetermined range, is superior to a variable like temperature tolerance.

2. Start two groups of populations from the same stock. Each population should be large enough to minimize the importance of drift, yet small relative to the carrying capacity of its container.

3. Let us call the starting value of the selection variable "relative zero". Relative zero should be as close as possible to the conditions under which the founder stock has been maintained for a long time, so that "stabilizing"

selection around relative zero does not amount to directional selection. Subject group A to stabilizing selection at relative zero.

4. Subject group B to directional selection in the "negative" direction (arbitrary). For lack of a quantitative theory I cannot specify the duration and intensity of directional selection required to significantly change the frequencies of modifier alleles. (Remember that the speeds of selection at modifier loci will be much slower than the speed of selection on alleles determining phenotype.) Group B is to be directionally selected to half that undetermined extent in the negative direction, and then back again to relative zero. This treatment is intended to "turn on" variation production without storing alleles in the population appropriate for the next genetic response.

The important environmental change for producing directional selection is <u>between</u> generations, not <u>within</u> generations. The classical experiments testing the effectiveness of heterogeneous environments in maintaining polymorphisms (Powell 1971; Ayala and McDonald 1974) used within-generation temporal heterogeneity, which would not be expected to favor the maintenance of <u>additive</u> variation.

5. The productivity of groups A and B is to be measured at the same times, over regular intervals. Productivity is to be measured in terms of the "surplus" over a population size "quota". The quota is set large enough to avoid the

influence of genetic drift. The populations are reset to this quota at each census. The rate of change of the selection variable, which determines the intensity of directional selection, is to be controlled so that the adult population size never falls below the quota.

6. As group B is returned to conditions at relative zero, begin directional selection of <u>both</u> groups, together, in the "positive" direction. Monitor their relative productivities.

My prediction is that group A will initially be more responsive to the selection because its stored variation is likely to be greater, and more appropriate to conditions slightly "positive". Group B, if it did indeed produce variation at a greater rate by this time, should gain and surpass group A in adaptedness as the latter's stored variation is depleted.

A possible criticism of this experiment is that the altered environmental parameter itself (e.g. temperature), rather than directional selection, may be responsible for changing variation production. The relative merits of stored versus produced variation are an important consideration. Slow variation production might require that very large populations be used in order to obtain appreciable genetic response. And if phenotypic variation production were moderated at the epigenetic level, then group A, under stabilizing selection, could maintain a sizable store of potentially useful alleles while producing little phenotypic variation.

Previous Experiments

Several experiments have been done comparing the competitive ability of strains differing in rates of variation production. These have compared, for example, mutator and wild-type strains of Escherichia coli (Gibson et al. 1970; Cox and Gibson 1974; Chao and Cox 1983), and isogenic haploid and diploid strains of the yeast Saccharomyces cerevisiae (Paquin and Adams 1983). In both of these cases the competing strains were asexual, and considered to be facing the problem of adaptation to a new environment. The mutator E. coli and the diploid S. cerevisiae proved to be competitively superior by reason of a faster rate of fixation of advantageous new mutations -- they evolved faster. The mutator E. coli had a higher mutation rate. The diploid yeast was assumed to have the same per-locus mutation rate as the isogenic haploid, and therefore twice the overall mutation rate of the haploid. Chao and Cox (1983) concluded that "...it is surprising that mutator genes are rarely found in nature." Evolution did not stop over the course of their experiment -- an equilibrium, or "adaptive peak", was never reached. I would expect that stabilizing selection, if it could be produced in their system, would be to the advantage of the non-mutator.

These are comparisons of the suitability of different variation schemes to a given environment, rather than a test for the existence of metavariation, i.e. the ability of a strain to change its variation production scheme.

Metavariation as Explanation in Two Bandwagon Topics

SEX. Sex permits a myriad of different ways in which gene flow, and consequently genetic variation production, can be regulated. (See the factors in Table 1.) The concept and cost of adaptability have been discussed more in the literature on the value of sex than anywhere else. I think it is important to realize that, in being a between-individual phenomenon, sex opens the way for using spatial (between individual) heterogeneity for temporal cues, damping genetic response to local, short-term fluctuations. (This kind of response is wasteful. See Levins (1964) on the adaptive significance of gene flow.) More than just recombination, a process of genetic variation production, sex is a mechanism by which variation production can be adjusted and regulated in many ways, ranging from assortativeness of matings, to dispersal rates, to the extreme of reproductive isolation. I prefer to view sex as a very flexible system for producing metavariation.

HYBRID DYSGENESIS. From the standpoint of metavariation, the phenomenon of hybrid dysgenesis is very interesting:

"Hybrid dysgenesis arises in one of the two reciprocal male-female crosses, usually males from recently wild-caught strains crossed with females from long-established laboratory strains ... It is characterized by various associated germ line dysfunctions including high mutability, frequent chromosomal rearrangements, male recombination (normally absent in Drosophila), failure of early embryonic development, and sterility due to germ line extinction in both males and females... The only puzzle is

the ostensible normality of the somatic tissue. " (Rose and
Doolittle 1983)

So far two interaction systems (P-M and I-R) have been found to
underly hybrid dysgenesis, and they have been shown to involve
transposable P and I elements (Rose and Doolittle 1983). The
distributions of P and I strains show striking temporal trends:
laboratory age is inversely correlated with the presence of
active P and I elements. Two hypotheses have been advanced to
explain why older laboratory populations of Drosophila have
fewer P and I elements (Bregliano and Kidwell 1983):

1. The "stochastic loss hypothesis" (Bucheton et al. 1976;
   Engels 1981) suggests that R and M populations result from
   the loss of I and P factors in small isolates, mainly
   those kept under laboratory conditions.

2. The "recent invasion hypothesis" (Kidwell 1979,1982)
   suggests that there has recently been a rapid spread of I
   and P factors through natural populations. Old laboratory
   strains record the original state of wild populations.

Again, there is a missing hypothesis dealing with the
constancy, or rate of change, rather than the particular state,
of the lab environment. Maybe the P and I elements are related
to variation production and are less beneficial in the constant
lab environment than they are in a perhaps more temporally
heterogeneous natural environment.

From my perspective, particularly given the "puzzling"
normality of the soma, the phenomenon of hybrid dysgenesis looks
like a runaway variation production system -- too much, or
uncontrolled, variation is produced in the germ line

(offspring). Notice that, like sex, it is a between-individual phenomenon, and could therefore be playing a function in using spatial heterogeneity as a means of predicting and regulating variation useful in temporal responsiveness. Bregliano and Kidwell (1983) note that data on the timing of appearance of I and P factors in wild populations "... roughly coincides with the appearance of strong new selective pressures in populations of insects". The dates of appearance of I and P strains correspond with intensive use of DDT, and first use of organophosphates, respectively (Bregliano and Kidwell 1983).

I would like to emphasize hybrid dysgenesis as an important place to start looking for a system involved in regulating variation production.

DISCUSSION

"There has grown up, within the Neo-Darwinist paradigm of evolutionary theory, a dogma that the character of any new elements that may appear in a population as potential raw materials for evolution are quite unconnected with the nature of the selection process to which they will be subjected...This seems to be not only tenaciously believed, but deeply felt..." (C.H. Waddington 1974).

Population-level Genetic Memory

Waddington and his followers have stressed the importance of the "epigenetic system", a phenomenological model, in adaptively focussing the production of phenotypic variation. The epigenetic system is presumably the result of a long period of evolution.

In the course of this thesis I have described not a phenomenological model, derived from experimental observations, but rather a mechanism that might be considered a first approximation to an epigenetic system. It is premised on a conceptual division of the genome into two, not necessarily mutually exclusive, sets of genes (Layzer 1980):

1. those (alpha genes) which govern the development and maintenance of the individual organism, and

2. those (beta genes) which govern genetic adaptability, i.e. genetic variation production.

To use Levins' (1968, and see end of my Chapter Three) memory analogy, the alpha memory stores what selection has favored in developmental programs, and the beta memory stores what selection has favored in genotypic variation production. Each memory has a characteristic lag time due to the time it takes to fix new information. For example, the "alpha memory" is last influenced by the environment of the <u>past</u> generation. Unless the last generation was a good predictor of the present generation, it would not have been an advantage to respond genetically, i.e. to update alpha memory, at all.

The same rationale holds for beta memory, only its lag time is much longer because it is more slowly changed (by secondary selection) via changes in alpha memory. Because adaptability (beta memory) changes more slowly than adaptedness, this means that evolutionary patterns determined by the way <u>adaptability</u> changes are necessarily of a longer time scale than those patterns based on changes in adaptedness.

It is interesting to realize that the beta genes, storing favorable patterns of genetic variation production, are really a population-level memory. Edstrom (1975) has described an ingenious model of evolution in which he, too, postulates the existence of a population-level memory in the genome. In his model the second genetic memory is a sample of allele frequencies in the population, obtained by a censusing process

made possible by sexual reproduction (see Edstrom 1975). This memory is used in a different way: through gene conversion, segregation ratios are distorted slightly in favor of rare alleles. This would cause rare alleles to increase much faster, and at much lower cost, than would be possible by natural selection. Edstrom (1975) calculates that it would take 9240 generations, at a positive selection coefficient of 0.01, to increase the frequency of a recessive mutation from 0.01 to 0.1. A segregation ratio of 1.01 in favor of the recessive could produce the same result in only 233 generations, without the reproductive excess required by natural selection.

Variation Production and Evolutionary Patterns

Despite explicit recognition, by people such as Levins and Layzer, of variation production as an adaptive system subject to selection, there still seems to be a reluctance to openly discuss the idea in the literature. I have yet to see a discussion about Layzer's adjustable variation ideas, although he made them clear. Templeton (1981) criticized Layzer's selection mechanism and made no comment on the central drive of Layzer's (1980) paper.

Although considerable attention is now being paid to adaptability and adjustment of variation in the life history and optimal foraging literature (Kaplan and Cooper 1984; Lacey et al. 1983; Crandall and Stearns 1982; Caraco 1980), the importance of patterns of variation production in evolution has received direct attention only by the epigeneticists (eg. Waddington 1957; Ho and Saunders 1979). But processes of

variation production are starting to come under greater scrutiny for their potential role in determining evolutionary patterns.

The 'effect hypothesis' of Vrba (1980,1983) is one explanation for 'trends', which are long-term directional tendencies in evolution (Vrba 1983). It states that trends may be unselected 'effects' of characters and processes within species, rather than the result of selection and adaptation. This is based on a notion of secondary selection, that adaptations driven by individual selection may have incidental effects, and that these effects might influence net speciation rate (R) in a monophyletic group.

My stance is intermediate between the effect hypothesis and species selection (see Stanley 1975,1979). Although Vrba refers to second order selection, she is emphatic in allowing adaptation only at the 'individual' level, and does not include the possibility of a more inclusive adaptive system. My point is precisely that such an adaptive system might exist. On the other hand, Stanley's species selection is a form of group selection, rather than a process of direct upward causation by selection on individuals, as I advocate.

Dover's (1982) concept of "molecular drive" is unusual in that it suggests patterns based on variation production rather than on variation loss, eg. selection and drift. "Molecular drive" refers to the action of a set of processes including unequal crossing over, transposition, and gene conversion, that are capable of moving a variant repeat sequence intrachromosomally, to homologous chromosomes, and to nonhomologous chromosomes (Lewin 1982). In a panmictic sexual

population this diffusion of a variant throughout the gene pool is faster than the fixation of a variant within a family of ·repeats. This implies that "there is in each individual the same average ratio of old and new variants for a particular family" (Dover 1982), and that selection will not discriminate among individuals because they are all more or less similar. This is an explanation of the observation that variation in members of a repeat family is mostly between species, and not within species. It also led Dover to speculate on the possibility of "accidental speciation" due·to the cohesive divergence of subpopulations to the point where the viability of hybrids might be affected.

Molecular drive is independent of selection and drift, and seems to operate on a longer time scale than those two processes.

Old Ideas

The conceptual division between types of genes is not a recent idea. Dobzhansky (1970) writes

"Lamprecht argued in a series of papers (summary in Lamprecht 1964) that there are two categories of genes and mutations, some distinguishing species and others only varieties. Bocher (1951) believed that there are two kinds of mutations, some responsible for adaptation to the environment and others for progressive evolution."

Bocher's dichotomy, in particular, is remarkably similar to that of alpha and beta genes (Layzer 1980), which I described. Dobzhansky continues, "These views have very few adherents at present." Perhaps these ideas should be reconsidered!

SUMMARY

1. Undirected variation production by definition has no connection with the selection forces to which the new variants will be subjected, and usually lowers the immediate fitness of parents.

2. Genetic adaptability is important for maintaining immediate fitness in a changing world, and it requires the production of genetic variants, so there is an "adaptedness/adaptability" tradeoff.

3. The best adaptedness/adaptability compromise depends on

    a. environmental parameters, including rates of change and predictability, and

    b. the time scale over which the compromise strategy is observed. A long time scale of observation favors more investment in adaptability.

4. Given the existence, in a population, of heritable "variation in variation production patterns" (i.e. "metavariation"), selection can tailor genetic adaptability. Many genetic elements are known that modify processes of genetic variation production. Random variation in these elements produces metavariation. Secondary selection, in particular, can act on these modifiers. Thus variation production at the level of the genotype is not necessarily undirected, or random. The question is raised of whether genomes might be organized to facilitate modification of genetic variation production.

5. We can expect evolutionary patterns to be caused by patterns in variation production, as well as the traditionally emphasized patterns in variant loss (e.g. selection and drift). Hypotheses can be formulated on the basis of adaptability and change, as well as adaptedness and state of the environment.

6. The mechanism of secondary selection, and the notion that a part of the genome is involved in regulating genetic variation production, are used to build a possible explanation for the observed patterns in genome size.

7. Selection changes allele frequencies at modifier loci much more slowly than at primary loci. So changes in adaptability occur more slowly than changes in adaptedness. Slow processes might explain long term evolutionary patterns better than fast processes and special contingencies.

Patterned production of genetic variation:

consider it,

"...the main weakness of modern evolutionary theory is its
lack of a fully worked out theory of variation, that is, of
canditure for evolution..."

(Medawar 1967)

or dismiss it.

"Mutations...arise regardless of their actual or potential
usefulness. It may seem a deplorable imperfection of nature
that mutability is not restricted to changes that enhance
the adaptedness of their carriers. However, only a vitalist
Pangloss could imagine that the genes know how and when it
is good for them to mutate."

(Dobzhansky 1970)

REFERENCES CITED

Abrahamson, L., M.A. Bender, A.D. Conger and S. Wolff. 1973. Uniformity of radiation-induced mutation rates among different species. Nature 245:460-462.

Auerbach, C. 1956. Genetics of the atomic age. Oliver and Boyd, Edinburgh.

Avdulov, N.P. 1931. Karyo-systematische Untersuchungen der Familie Gramineen. Bull.Appl.Bot.Genet.Plant Breed. 44(Suppl.):1-428.

Ayala, F.J. 1968. Biology as an autonomous science. American Scientist 56:207-221.

Bachmann,K. 1972. Genome size in mammals. Chromosoma 37:85-93.

Bachmann,K., B.A. Harrington, and J.P. Craig. 1972. Genome size in birds. Chromosoma 37:405-416.

Baetcke,K.P., A.H. Sparrow, C.H.Nauman, and S.S. Schwemmer. 1967. The relationship of DNA content to nuclear and chromosomal volume and to radiosensitivity (LD50). Proc.Natl.Acad.Sci. USA 58:533-540.

Bateson,Gregory. 1963. The role of somatic change in evolution. Evolution 17:529-539

Bennett,M.D. 1971. The duration of meiosis. Proc.R.Soc.(Lond.) B 178:277-299

Bennett,M.D. 1972. Nuclear DNA content and minimum generation time in herbaceous plants. Proc.R.Soc.(Lond.) B 181:109-135

Bennett, M.D. 1973. Nuclear characters in plants. Brookhaven Symp.Biol. 25:344-366.

Bennett,M.D. And J.B.Smith. 1972. The effects of polyploidy on meiotic duration and pollen development in cereal anthers. Proc.R.Soc.(Lond.) B 181:81-107

Bennett,M.D. And J.B.Smith. 1976. Nuclear DNA amounts in angiosperms. Phil.Trans.R.Soc. B 274:227-274

Bier,V.K. And W. Muller. 1969. DNA-Messungen bei Instekten und eine Hypothese uber retardierte Evolution und besonderen DNS-Reichtum im Tierreich. Biol.Zentralbl. 88:425-449.

Bishop,J.O. 1974. The gene numbers game. Cell 2:81-86

Bocher, T.W. 1951. Studies on morphological progression and evolution in the vegetable kingdom. Dan.Biol.Medd. 18(13):1-51.

Bowen, H.J.M. 1962. Radiosensitivity of higher plants and correlations with cell weight and DNA content. Radiat.Bot. 1:223-228.

Britten,R.J. And E.H.Davidson. 1969. Gene regulation for higher cells: a theory. Science 165:349-357

Britten,R.J. And E.H.Davidson. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Quart.Rev.Biol. 46:111-133

Burley, J. 1965. Karyotype analysis of Sitka spruce, Picea sitchensis (Bong.) Carr. Silvae Genet. 14:127-132.

Carlson, W.R. 1978. The B chromosome of corn. Ann.Rev.Genet. 16:5-23.

Cavalier-Smith,T. 1978. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. J.Cell Sci. 34:247-278.

Cavalier-Smith,T. 1980a. r- and K- tactics in the evolution of protist developmental systems: cell and genome size, phenotype diversifying selection, and cell cycle patterns. Biosystems 12:43-59.

Cavalier-Smith,T. 1980b. How selfish is DNA? Nature 285:617-618.

Chao,L. And E.C.Cox. 1983. Competition between high and low mutating strains of E. Coli. Evolution 37(1):125-134.

Charlesworth, B. 1976. Recombination modification in a fluctuating environment. Genetics 83:181-195.

Chooi,W.Y. 1971. Variation in nuclear DNA content in the genus Vicia . Genetics 68:195-211.

Commoner,B. 1964. Roles of deoxyribonucleic acid in inheritance. Nature 202:960-968.

Cox, E.C. And T.C. Gibson. 1974. Selection for high mutation rates in chemostats. Genetics 77:169-184.

Crick, F. 1971. General model for the chromosomes of higher organisms. Nature 234:25-27.

Crow, J.F. 1958. Some possibilities for measuring selection intensities in man. Hum.Biol. 30:1-13.

Dawkins, R. 1976. The Selfish Gene. Oxford Univ. Press.

Dawkins, R. 1978. Replicator selection and the extended phenotype. Z.Tierpsychol. 47:61-76.

Dawkins,R. 1982. The extended phenotype: the gene as the unit of selection. Freeman, San Fransisco.

Dobzhansky, T. 1970. The genetics of the evolutionary process. Columbia Univ. Press, N.Y.

Donachie, W.D. 1974. Cell division in bacteria. IN Mechanism and Regulation of DNA Replication, ed. A.R. Kolber and M. Kohijama. Plenum Press, N.Y.

Doolittle,W.F. And C.Sapienza. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601-603.

Dover,G. 1980. Ignorant DNA? Nature 285:618-620.

Dover,Gabriel. 1982. Molecular drive: a cohesive mode of species evolution. Nature 299:111-117.

Ebeling,A.W., N.B. Atkin, and P.Y. Setzer. 1971. Genome sizes of teleostean fishes: increases in some deep-sea fishes. Amer.Nat. 105:549-561.

Edstrom, J.-E. 1975. Eukaryotic evolution based on information in chromosomes on allele frequencies. J.Theor.Biol. 52:163-174.

El-Lakany, M.H., and O. Sziklai. 1971. Intraspecific variation in nuclear characteristics of Douglas fir. Adv.Front.Plant Sci. 28:363-378.

Endow.S. And J.G.Gall. 1975. Differential replication of satellite DNA in polyploid tissues of Drosophila virilis. Chromosoma 50:175-192.

Evans, G.M., and H. Rees. 1971. Mitotic cycles in dicotyledons and monocotyledons. Nature 233:350-351.

Evans, G.M., H. Rees, C.L. Snell and S. Sun. 1972. The relation between nuclear DNA amount and the duration of the mitotic cycle. Chromosomes Today 3:24-31.

Felsenstein, J. 1974. The evolutionary advantage of recombination. Genetics 78:737-756.

Felsenstein, J. And S. Yokoyama. 1976. The evolutionary advantage of recombination. Genetics 83:845-859.

Finerty, J.P. 1980. The population ecology of cycles in small mammals: mathematical theory and biological fact. Yale Univ. Press. 234 pp.

Fisher, R.A. 1958. The genetical theory of natural selection (2nd ed.). Dover, N.Y. 291 pp.

Flavell, R.B., M. Bennett, J. Smith and D.B. Smith. 1974. Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem.Genet. 12:257-269.

Ghiselin,M.T. 1974. A radical solution to the species problem. Syst.Zool. 23:536-544.

Ghiselin,M.T. 1981. Categories, life, and thinking. Behav.Brain 4:269-313.

Gibson, T.C., M.L. Scheppe and E.C. Cox. 1970. Fitness of an E. coli mutator gene. Science 169:686-688.

Gilbert,W. 1978. Why genes in pieces? Nature 271:501.

Gillespie, J.H. 1978. A general model to account for enzyme variation in natural populations.V. The SAS-CFF model. Theor.Popul.Biol. 14:1-45.

Gillespie,J.H. 1981. Mutation modification in a random environment. Evolution 35(3):468-476.

Grime,J.P. And M.A.Mowforth. 1982. Variation in genome size -- an ecological interpretation. Nature 299:151-153.

Haldane, J.B.S. 1957. The cost of natural selection. J.Genet. 55:511-524.

Hinegardner,R.T. 1968. Evolution of cellular DNA content in Teleost fishes. Amer.Nat. 102:517-523.

Hinegardner,R. 1974a. Cellular DNA content of the mollusca. Comp.Biochem.Physiol. 47A:447-460.

Hinegardner,R. 1974b. Cellular DNA content of the echinodermata. Comp.Biochem.Physiol. 49B:219-226.

Hinegardner,R. 1976. Evolution of genome size. Pp.179-199 IN Molecular Evolution ,F.J.Ayala(ed.).Sinauer Assoc.Inc.,Sunderland,Mass.

Hinegardner,R., and D.E. Rosen. 1972. Cellular DNA content and the evolution of teleostean fishes. Amer.Nat. 106:621-644.

Ho,M.W. And P.T.Saunders. 1979. Beyond neo-Darwinism -- an epigenetic approach to evolution. J.Theor.Biol. 78:573-591.

Holmquist, R., M. Goodman, T. Conroy and J. Czelusniak. 1983. The spatial distribution of fixed mutations within genes coding for proteins. J.Molec.Evol. 19:437-448.

Hsu, T.C. 1975. A possible function of constitutive heterochromatin: the bodyguard hypothesis. Genetics 79:137-150.

Hull,David L. 1980. Individuality and selection. Ann.Rev.Ecol.Syst. 11:311-332.

James, A.P. 1959. The spectrum of severity of mutant effect. I. Haploid effects in yeast. Genetics 43:1309-1324.

James, A.P. 1960. The spectrum of severity of mutant effects. II. Heterozygous effects in yeast. Genetics 45:1627-1647.

Jeffreys, A.J. 1981. Recent studies of gene evolution using recombinant DNA. Pp1-48 IN Genetic Engineering Vol 2, R.Williamson (ed.), Acad. Press, N.Y.

Jones, R.N., and L.N. Brown. 1976. Chromosome evolution and DNA variation in Crepis. Heredity 36:91-104.

Jones, G.H., and H. Rees. 1967. Chromosome evolution in Lolium. Heredity 20:1-18.

Jones, R.N., and H. Rees. 1968. Nuclear DNA variation in Allium. Heredity 23:591-605.

Kadir, Z.B.A. 1974. DNA values in the genus Phalaris (Gramineae). Chromosoma 45:379-386.

Kaplan,R.H. And W.S. Cooper. 1984. The evolution of developmental plasticity in reproductive characteristics: an application of the "adaptive coin-flipping" principle. Amer.Nat. 123(3): 393-410.

Karlin, S., and J. McGregor. 1974. Towards a theory of the evolution of modifier genes. Theor.Popul.Biol. 5:59-103.

Kerkis, J. 1938. The frequency of mutations affecting viability. Bull.Acad.Sci. USSR (Biol.) 1938:75-96.

Kimura, M. 1956. A model of a genetic system which lends to closer linkage by natural selection. Evolution 10:278-287.

Kimura, M. 1960. Optimum mutation rate and degree of dominance as determined by the principle of minimum genetic load. J.Genetics 57:21-34.

Kimura, M. 1967. On the evolutionary adjustment of spontaneous mutation rates. <u>Genet</u>.<u>Res</u>. 9:23-34.

Lamprecht, H. 1964. Species concepts and the origin of species. The two categories of genes -- intra- and interspecific ones. <u>Agi</u>.<u>Hortique Genetica</u> 22:<u>272-280</u>.

Larson,Allan. 1981. A reevaluation of the relationship between genome size and genetic variation. <u>Amer</u>.<u>Nat</u>. 118:119-125.

Layzer,D. 1980. Genetic variation and progressive evolution. <u>Amer</u>.<u>Nat</u>. 115(6):809-826.

Leigh,E. 1970. Natural selection and mutability. <u>Amer</u>.<u>Nat</u>. 104:301-305.

Leigh,Egbert. 1973. The evolution of mutation rates. <u>Genetics</u> Supp. 73:1-18.

Levin,D.A. And S.W. Funderburg. 1979. Genome size in angiosperms: temperate versus tropical species. <u>Amer</u>.<u>Nat</u>. 114:784-795.

Levins, R. 1962. Theory of fitness in a heterogeneous environment I. The fitness set and adaptive function. <u>Amer</u>.<u>Nat</u>. 96:361-373.

Levins, R. 1963. Theory of fitness in a heterogeneous environment II. Developmental flexibility and niche selection. <u>Amer</u>.<u>Nat</u>. 97:75-90.

Levins, R. 1964a. Theory of fitness in a heterogeneous environment III. The response to selection. <u>J</u>.<u>Theor</u>.<u>Biol</u>. 7:224-240.

Levins, R. 1964b. Theory of fitness in a heterogeneous environment IV. The adaptive significance of gene flow. <u>Evolution</u> 18:635-638.

Levins, R. 1965. Theory of fitness in a heterogeneous environment V. Optimal genetic sustems. <u>Genetics</u> 52:891-904.

Levins, R. 1967. Theory of fitness in a heterogeneous environment VI. The adaptive significance of mutation. <u>Genetics</u> 56:163-178.

Levins, R. 1968. <u>Evolution in Changing Environments</u>. Princeton Univ. Press, Princeton.

Lewin,R. 1982. Molecular drive: how real, how important? <u>Science</u> 218:552-553.

Lewontin, R.C. 1965. Selection in and of populations. IN
      J.A.Moore (ed.) Ideas in modern biology Proc. Vol.6,
      Nat.Hist.Press,N.Y.

Maynard Smith, J. 1971. The origin and maintenance of sex. IN
      Group Selection, ed. G.C. Williams. Aldine-Atherton,
      Chicago.

Mayr, E. 1978. Evolution. Sci.Amer. 239:46-55.

Medawar, P.B. 1967. The Art of the Soluble. Methuen, London.

Mergen, F., and B.A. Thielges. 1967. Intraspecific variation in
      nuclear volume in four conifers. Evolution 21:720-724.

Miksche, J.P. 1971. Intraspecific variation of DNA per cell
      between Picea sitchensis (Bong.) Carr. Provenances.
      Chromosoma 32:343-352.

Miksche, J.P., and Y. Hotta. 1973. DNA base composition and
      repetitious DNA in several conifers. Chromosoma 41:29-36.

Mirsky,A.E. And H. Ris. 1951. The deoxyribonucleic acid content
      of animal cells and its evolutionary significance.
      J.Gen.Physiol. 34:451-462.

Mukai, T. 1964. The genetic structure of populations of
      Drosophila melanogaster I. Spontaneous mutation rate of
      polygenes controlling viability. Genetics 50:1-19.

Nagl, W. 1974. Role of heterochromatin in the control of cell
      cycle duration. Nature 249:53-54.

Ohno, S. 1969. The spontaneous mutation rate revisited and the
      possible principle of polymorphism generating more
      polymorphism. Can.J.Genet.Cytol. 11:457-467.

Ohno,S. 1970. Evolution by gene duplication. Springer-Verlag,
      N.Y.

Ohno,S. 1972. So much "junk" DNA in our genome. Brookhaven
      Symp.Biol. 23:366-370.

Orgel,L.E., and F.H.C.Crick. 1980. Selfish DNA: the ultimate
      parasite. Nature 284:604-607.

Pattee,H.H.(ed.) 1973. Hierarchy Theory: the challenge of
      complex systems. George Braziller, N.Y. 156pp.

Pierce,B.A. And J.B. Mitton. 1980. The relationship between
      genome size and genetic variation. Amer.Nat. 116:850-861.

Powell,J. 1971. Genetic polymorphisms in varied environments. Science 174:1035-1036.

Price,G.R. 1972. Fisher's fundamental theorem made clear. Annals Hum.Genet. 36:129-140.

Price,H.J. 1976. Evolution of DNA content in higher plants. Botan.Rev. 42:27-57.

Rees,H., and M.H.Hazarika. 1969. Chromosome evolution in Lathyrus . Chrom.Today 2:158-165.

Roughgarden, J. 1979. Theory of population genetics and evolutionary ecology: and introduction. MacMillan, N.Y.

Salthe,S.N. 1983. An extensional definition of functional individuals. Amer.Nat. 121:139-144.

Simon,H.A. 1962. The architecture of complexity. Proc.Amer.Philosophical Soc. 106:467-482.

Simon,H.A. 1973. The organization of complex systems. IN Hierarchy Theory, George Braziller, N.Y. 156pp.

Slobodkin,L.B. 1964. The strategy of evolution. American Scientist 52:342-357.

Slobodkin,L.B. 1968. Toward a predictive theory of evolution. IN R.C.Lewontin (ed.), Population biology and evolution, Syracuse Univ. Press.

Smith,J.B. And M.D.Bennett. 1975. DNA variation in the genes of Ranunculus. Heredity 35(2):231-239.

Sompyrac, L., and C. Maaloe. 1973. Autorepressor model for control of DNA replication. Ature,New Biol. 241:133-135.

Sparrow, A.H., K.P. Baetcke, D.L. Shaver and V. Pond. 1968. The relationship of mutation rate per roentgen to DNA content per chromosome and to interphase chromosome volume. Genetics 59:65-78.

Sparrow, A.H., and J.P. Miksche. 1961. Correlation of nuclear volume and DNA content with higher plant tolerance to chronic radiation. Science 134:282-283.

Sparrow,A.H., H.J.Price and A.G.Underbrink. 1972. A survey of DNA content per cell and per chromosome of prokaryotic and eukaryotic organisms: some evolutionary considerations. Brookhaven Symp.Biol. 23:451-494.

Stanley, S.M. 1975. Clades versus clones in evolution: why we have sex. Science 190(4212):382-383.

Stanley, S.M. 1979. Macroevolution: pattern and process. Freeman, San Francisco.

Stebbins, G.L. 1966. Chromosomal variation and evolution. Science 152:1463-1469.

Strauss, N.A. 1971. Comparative DNA renaturation kinetics in Amphibians. Proc.Natl.Acad.Sci. 68:799-802.

Templeton, A.R. 1981. Some comments on "Genetic variation and progressive evolution" by D. Layzer. Amer.Nat. 117:1049-1051.

Thoday,J.M. 1953. Components of fitness. Symp.Soc.Exp.Biol. 7:96-113.

Thompson, P.E. 1962. Asynapsis and mutability in D. melanogaster. Genetics 47:337-349.

Underbrink, A.G., A.H. Sparrow and V. Pond. 1968. Chrosomes and cellular sensitivity II. Use of interrelationships among chromosome volume, nucleotide content and Do of 120 diverse organisms in predicting radiosensitivity. Radiat.Bot. 8:205-238.

Valentine, J.W. 1976. Genetic strategies of adaptation. IN Molecular Evolution, F.J. Ayala (ed.), Sinauer Assoc. Inc., Mass.

Van't Hoff,J. And A.H. Sparrow. 1963. The relationship between DNA content, nuclear volume, and mitotic cycle time. Proc.Natl.Acad.Sci. U.S. 49:897-902.

Vrba,E. 1980. Evolution, species and fossils: how does life evolve? South African J.Sci. 76(2):61-84.

Vrba,E. 1983. Macroevolutionary trends: new perspectives on the roles of adaptation and incidental effect. Science 221:387-389.

Waddington, C.H. 1957. The strategy of the genes. The MacMillan Co., N.Y.,ix+262pp.

Waddington, C.H. 1974. A catastrophe theory of evolution. Ann.N.Y.Acad.Sci. 231:32-41.

Waddington, C.H. And E. Robertson. 1966. Selection for developmental canalization. Genet.Res. 7:303-312.

Walker,P.M.B. 1968. How different are the DNAs from related animals? Nature 219:228-232.

Warburton,F.E.   1967.   A model of natural selection based on the
     theory of guessing games. J.Theor.Biol. 16:78-96.

Williams, G.C. (ed.) 1971. Group Selection Aldine-Atherton, N.Y.
     210pp.

Yang, D.P., and E.O. Dodson. 1970. The amounts of nuclear DNA
     and the duration of DNA synthesis period (S) in related
     diploid and autotetraploid species of oats. Chromosoma
     31:309-320.

Yunis,J.J.   And W.B.Yasmineh.   1971. Heterochromatin, satellite
     DNA, and cell function. Science 174:1200-1209.

Zuckerkandl,E. 1976. Gene control in eukaryotes and the c-value
     paradox: "excess" DNA as an impediment to transcription of
     coding sequences. J.Molec.Evol. 9:73-104.

APPENDIX I

FITNESS, PERSISTENCE, AND TIME SCALES

I am convinced that not enough attention has been paid to the time scale dependence of the results of selection, and that this oversight is the source of an enormous amount of confusion. I will try to illustrate the qualitative change in the results of selection with time scale, and follow it with an example of confusion arising out of failure to distinguish among different time scales.

Fitness is defined to mean "that or those properties relevant to selection", to deliberately make "natural selection favors the fittest" a truism (Dawkins 1982). Looking at a population in the short term, say from one generation to the next, the only phenomena we see are relative changes in proportions of various characteristics, and fitness gets defined as the relative effective rate of reproduction.

If one assumes that in our world the passing of time implies change, then over the longer term the ability to cope with change, i.e. adaptability, becomes more important compared with adaptedness to the state of the environment at any one time (e.g. relative rate of increase). Unless anything that contributes to adaptedness increases adaptability in the same proportion (plainly not so, e.g. sex and no sex) we will need a different definition of fitness to satisfy the truism. Different time scales produce different results of selection, requiring a different definition of fitness.

Instead of having the word "fitness" be so malleable, and

therefore vague, as to take on the meaning of whatever properties are convenient, it is probably better to leave it as the short time scale notion of differential reproduction and adaptedness. Dawkins (1982) shows that even with this constraint the word "fitness" has at least five different definitions. For properties relevant to selection on longer time scales the words "persistence" or "persistability" could perhaps be used.

Crandall, Stearns and Dudman (in prep) recently constructed two computer models, and tested the worth of various definitions of fitness as predictors of the outcome. The two models differed in the spatial scale in which the world was viewed. The microscopic model viewed the world as consisting of a finite number of patches, whereas the macroscopic world view was one of an infinite number of patches. In the latter model "productivity" was the best predictor, whereas in the microscopic model the best predictor was "something like a persistence" -- and the two measures were found not to be the same even as the number of finite patches was increased without bound. This indicated a qualitative difference in the two models, which to my mind was embodied in the possibility of extinction in the finite model. And, just as a spatial scale that made extinction a factor made "fitness" a "persistence", I am claiming that a temporal scale that makes extinction a factor will make "fitness" a "persistence".

The results of natural selection and, in consequence, the appropriate properties to be embodied in "fitness", are dependent on the time scale over which the results are viewed.

As an example of the confusion possible if time scale is

not taken into account, consider Lewontin's (1965) statement:

> "...the course of evolution is determined by a similar maximizing principle both within and between populations. Within populations the result of natural selection, by and large, is to change the frequency of genotypes to maximize the intrapopulation fitness. Between populations it is the probability of survival that is maximized and the net result is a compromise between these forces, the degree to which one or the other is important depending upon the autecology of the species."

The different "maximizing principles" that Lewontin perceives at different levels of organization are due not to different goals of natural selection, but to the fact that in switching from one level to another Lewontin has changed the time scale over which he views the results of natural selection. The two maximizing principles are really just the two sides of the compromise on any level of selection: "short term gain versus long term persistence". The "gain" in the short term for a whole population can be in terms of average fitness (W), absolute size, or size relative to other populations. But over a longer term what matters is not the history of the population, its record of short term gains (W,size), but that it manages to persist.

Similarly, on a lower level, the short term gains of a lineage within a population are in absolute numbers or proportion of the population. But in the long run what matters is not the number or relative number of offspring, but that the number of offspring exceeds the number of deaths -- that the lineage persists.

When looking at selection within populations Lewontin considered only short term gain, embodied in the notion of fitness. He overlooked the fact that in the long term the

genotypes and traits favored by selection within a population are those enhancing persistence or probability of survival. On the population level Lewontin saw only the long term "maximization" of probability of survival and not the short term population increases and other features that would be important in a population-level definition of "fitness" analogous to the "individual", or within-population concept (W).

The results of natural selection differ depending on the time scale over which they are examined. They do not depend on the level of organization at which selection is being considered. "Fitness" belongs on a shorter time scale, "persistence" on the evolutionary one.

APPENDIX II

THE FUNCTIONAL HIERARCHY

The word 'fitness' is intended to reflect the property(ies) of phenotypes that natural selection favors (Dawkins 1982). Natural selection is the net result of many processes at work in a population. Some processes are slower than others. A quality advantageous with respect to a slow process becomes important over the longer term as that process becomes an important cause of change. Thus, time scale determines what processes can be important, and what properties will be favored by natural selection, so 'fitness' should be defined time scale dependently (see Appendix I).

What are all the properties that fitness potentially could embody? Can we somehow make a list of properties in order of their importance over longer and longer time scales? Or, equivalently, can we describe how the results of natural selection change with time scale? This will involve breaking out of the traditional habit of confining natural selection to a time scale of one generation, when the most important quality is relative effective reproductive rate (the usual definition of fitness).

The Selfish Individual

Let us first consider a time scale shorter than one generation. An organism has certain qualities that enable it to survive to maturity as it must do to reproduce. Note that a phenotype could arise that invests highly in these qualities,

spending all of its energies on self-maintenance and leaving nothing for reproductive investment. Reproduction always costs an individual organism in energy or risk, probably lowering its individual survival rate.

We generally recognize that such a strategy would not be favored by natural selection on the generation to generation time scale -- we expect a balance in investment between self-maintenance and reproduction such that a zygote contributes the greatest number of zygotes to the next generation.

Note that if we consider increasing proportional representation in a population to be 'success', i.e. to indicate being 'selected', then on a within-generation time scale, the strategy of putting all energies into self-preservation would appear to be favored, since these strategists would comprise a greater proportion of the adult population than of the zygote population.

'Individual' selection is a misnomer. It is not individuals that are being selected, or sorted out, but the genes they carry from one generation to the next (consider Hamilton's inclusive fitness). The individual organisms do not even persist, so how can they be said to be selected? The entity that persists, and can be viewed as being selected on that time scale is the genotype, or semi-conserved genotype. That is Dawkins' (1976, 1982) basic point. If the individual is all-important, why should it take the risk and spend the energy to reproduce? Selection of individuals is within a generation, and favors strategies (e.g. self-maintenance with no reproduction) that are not fit according to our usual trans-generation view of

natural selection and definition of fitness.

I could be criticized for playing a semantic game here, or paying too much attention to words. Let me emphasize that the problem _is_ one of semantics. If we are going to talk about the qualities that make some phenotypes more or less successful, then our definition of success is critical.

The Selfish Lineage

On a longer-than-usual time scale, in a changing world, natural selection will favor properties involved in maintaining (one-generation) fitness. Genetic adaptability becomes important, and it requires the production and maintenance of genetic variation in a population. Undirected production of genetic variants implies the production of less fit genotypes, and genotypes that may be fitter in the future (important for adaptability) may be less fit at present. So adaptability is at odds with immediate fitness, or "adaptedness", and by changing the time scale over which we view the results of natural selection we have changed the qualities that natural selection "favors". Adaptability is now worth some cost in terms of immediate fitness.

Note the concurrent change in 'selected unit' with time scale. Within a generation it is the individual. From generation to generation it is the genotype. Over many generations genetic variation production becomes a virtue, and the selected unit is something more inclusive than the genotype -- the latter is sacrificed for variation production. I have called this unit the lineage, and over the long term _it_ looks selfish -- not the

genotype, and not the individual.

Another Selfish Level?

We now have three different properties that a definition of fitness could embody. In order of increasing importance with longer time scale they are: self-maintenance, effective reproduction rate, and variation production. The effect of increasing time scale in changing the properties that natural selection favors is not due to time itself. It is the change in an environment that can take place over a longer time period that exerts the selection pressure favoring the ability to cope with change (adaptability), that requires variation production. The 'function' of variation production is to permit genetic adaptation in a changing environment.

There is one last strategy that I would like to consider. If the environment is one of varying rates and/or direction of change, it might be "useful" to vary the frequency and kind of variant production (i.e. to produce metavariation). That way, if the variants produced by one lineage were rarely suited to the way the environment was changing, a lineage with a more successful scheme would begin to predominate. The more inclusive unit consisting of these lineages would thereby become more adaptable to the current pattern of environmental change. I think that the species corresponds to this more inclusive unit, since the member populations and lineages of a species are generally considered to share the many mechanisms by which variant production can be modified. These include breeding structure, assortativeness of mating, and dispersal tendencies.

Table 4. Properties required for persistence under various environmental conditions.

Environmental Conditions	Persistence-related Properties	Persisting Entity	Comment
CONSTANT over time and space	durability and self-maintenance (phenotypic flexibility)	individual	The individual is vulnerable to certain destructive forces and lack of supplies for metabolism.
CONSTANT over time and heterogenous in space	reproduction	genotype	Since individuals occupy different locations in space, reproduction creates "spatial refuges" for the genotype.
CHANGING over time and homogeneous in space	variation	lineage	Without spatial refuges, a genotype will go extinct. Within a lineage, only the variant genotypes that are tolerant to the change will persist.
CHANGING in rate and/or direction of change	metavariation	species	Metavariation in effect explores different variation production schemes for the one best suited to the particular kind of environmental change.

The environments and strategies described above are summarized in Table 4. Note that higher level strategies are required for persistence in successively more complicated environments, which roughly correspond to longer and longer time scales. In more complicated environments, where the persistence of units at one level of organization becomes impossible, a kind of persistence is only possible by moving one level of organization higher, to the system that generates the ephemeral units at the level below.

The Existential Game

I find it conceptually useful to look at evolution as a problem of observation. We observe changes in the living forms existent from one point in time to another. Some previously existing forms might be absent, or changed. Some might be new. (Note that these categories are highly dependent on our definitions of these "forms".) The things we see at any time

1. could be constantly re-created by the abiotic environment, or

2. could have come into being in the more distant past and have some quality permitting PERSISTENCE.

In the case of living things, because of their complexity and high degree of organization, we can rule out (1). Persistence-oriented properties imbue living systems with internal teleology (see Ayala 1968). We do not see those forms with less or no 'persistability'. Of the things we do see, their purpose seems to be to persist. This justifies Slobodkin's metaphorical description of evolution as an existential game among "players"

with various persistence-oriented properties, or "strategies".
Unfortunately, Slobodkin was not clear about who the players
are, and he mixed examples of individual and population level
considerations. The present approach makes it clear that who the
players are is determined by what strategies are allowed,
i.e. what the rules of the existential game are. Each level of
strategy is a way of keeping a more inclusive, longer-lived
"player" in the "existential game". Table 4 gives the "players"
corresponding to various basic strategies.

The Functional Hierarchy

It is clear that the persistence-oriented strategies of
Table 4 are 'nested'. Effective reproduction rate is dependent
on the 'lower' strategy, self-maintenance (survival of the
individual) as well as fecundity. But self-maintenance does not
require reproduction. Likewise, variation production and
adaptability are superfluous if they cost extinction in the
short term. The persistence-oriented strategies of Table 4 can
be rearranged into the hierarchical control system or functional
hierarchy portrayed in Figure 6. Although it is common to
recognize a hierarchical organization of biology, it is usually
a structural hierarchy that is described (molecules, cells,
tissues, organs, organisms, populations, ecosystems), not a
functional one. This is awkward in view of the fact that
evolutionary explanations are in terms of function.

Others have suggested reorganizing biology around
categories based on function rather than structure. Dawkins
(1978) suggested "replicator" as a general term for a "unit of

slower response →

PHENOTYPES

NUMBERS[1] → REPRODUCTION → VARIATION → METAVARIATION → etc. ?

NUMBERS
NEXT TIME

REPRODUCTION
NEXT TIME

VARIATION
NEXT TIME

METAVARIATION
NEXT TIME

Persisting
'Unit'

SPECIES

LINEAGE

GENOTYPE

INDIVIDUAL
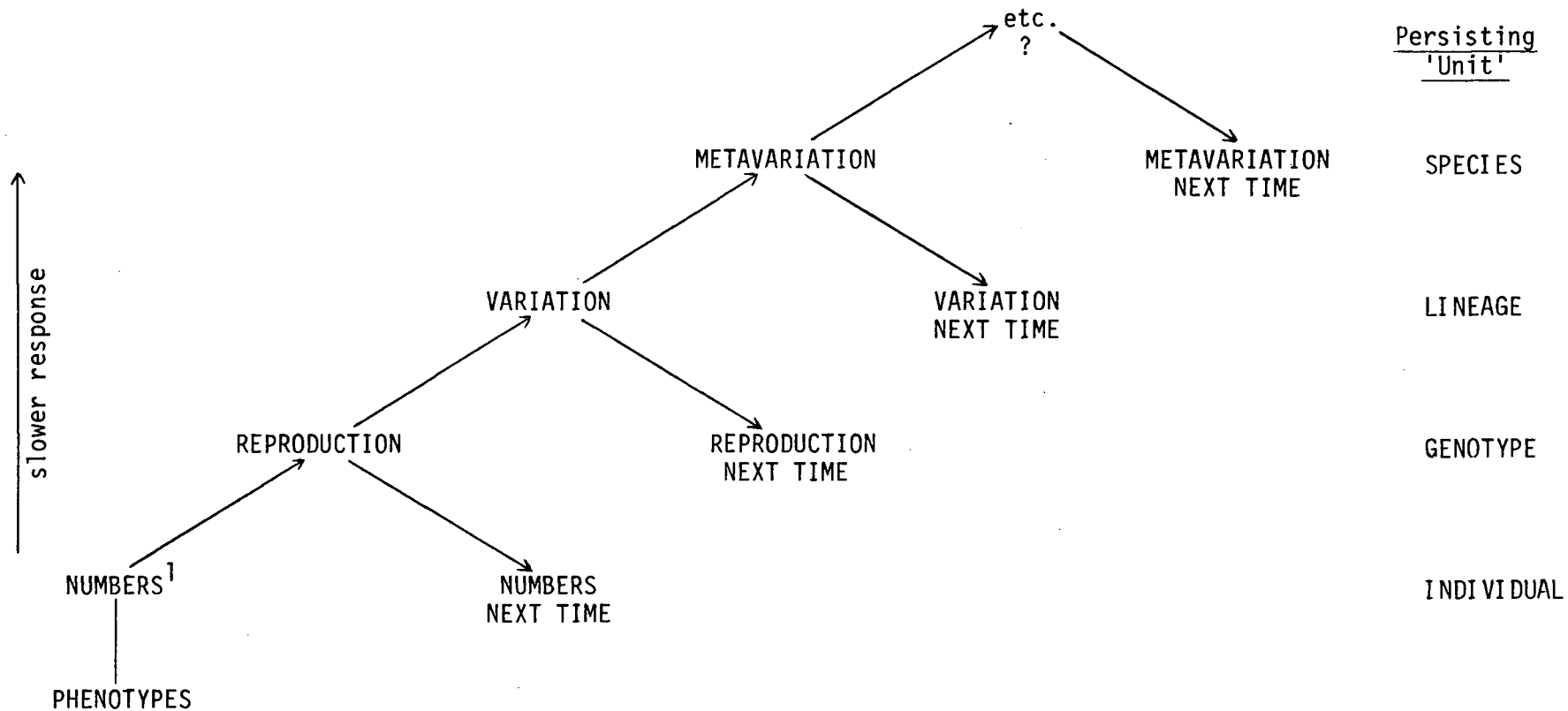
[1]Maintenance of numbers (short term) requires maintenance of reproductive fitness (longer term).

Figure 6. A hierarchical control system of biological persistence-oriented strategies.

selection". Hull (1980) thinks that the pair of terms "replicator" and "interactor" are more appropriate. These terms stem directly from the usual one-generation time scale description of the mechanism of natural selection. They over-emphasize reproduction, and variability is considered little more than slop from less-than-perfect replication. Ghiselin (1974) also suggests that biological categories be defined in terms of the causes of evolution, and suggests that "This is one of the main reasons why the usual formulation of the biological species definition is so attractive. Gene flow and reproductive isolation obviously profoundly influence the properties of organisms."

A hierarchy of physiological response systems was described by Bateson (1963), and extended into the ecological realm of populations and species by Slobodkin (1964, 1968). I have added metavariation as a logical extension. Thoday (1953) defined fitness as survival (persistence) probability, and in his discussion of the "components of fitness" he made several of the same points I have made. To my knowledge, Figure 6 is the first graphical portrayal of such a functional hierarchy.

Figure 6 puts the process of reproduction in perspective. It also shows us a scale of processes of decreasing speeds as one moves up the hierarchy. Hierarchy theory (Simon 1962, Pattee 1973) tells us to expect the faster processes at the bottom to be more important in short term patterns, and the slower processes at the top to be more important in the long term, evolutionary patterns. What this implies is that the logical place to look for explanation of very long term patterns like

the mode of evolution (e.g. punctuated vs. gradual) is at the metavariation level. Instead, most current explanations construe special situations in which the fast processes of drift and selection can be more or less effective.

Figure 6 also illustrates a duality: the slower processes important in the longer terms correspond to successively more inclusive "units of selection". So talking about evolution over longer time scales is the same as talking about "selection" of more inclusive "units", and vice-versa. (See my discussion of Lewontin's statement in Appendix I.)

Fitness

I think it is correct to assume that in evolutionary biology the desired meaning of fitness is "those qualities favored by natural selection". It is when this notion is translated into something more empirically tractable, e.g. physiological vigor or effective reproductive rate, that the problems and ambiguities arise. In my opinion at least some of this confusion is explained by the fact that different investigators have different perspectives and unconsciously adopt different time scales. That fact would assure that the same concept "fitness" give rise to different empirical definitions.

The definitions of fitness are so numerous that in any work it must be specified which definition pertains. I recommend a greater conciousness of the time scale being considered. I have shown above how the "favoritism" of natural selection changes with the time scale of observation, and have given a range of

strategies or properties that are differentially emphasized by different time scales. On the longest of time scales, metavariation becomes a valuable strategy.

APPENDIX III

A COMPUTER MODEL FOR COMPARING
VARIATION PRODUCTION STRATEGIES
IN VARIOUS ENVIRONMENTS

This computer program is written as an exploratory evolutionary "game". The "players" are asexually reproducing populations with different variation production "strategies", living in arbitrarily specified, one-dimensional environments. The strategy of each player is fixed for a given run -- this program cannot compare the virtues and costs of adjustable and fixed variation. It can compare the costs and benefits of different fixed variation production strategies. In the following description, variable names are printed in capital letters. (In the actual program code they are lower case.)

Variation Production

There are two possible components to a variation production strategy. One is the proportion of offspring that are variant. The other is the distribution of variant offspring phenotypes. It was decided to fix the proportion of variant offspring arbitrarily (at two thirds) and vary only the offspring distributions. The latter are symmetrical (undirected), bimodal, and vary only in the deviation of the variant offspring from the parental phenotype. The particular deviation characteristic of a given player will be referred to as its "step-size". Figure 8 shows a segment of a population before (a), and after (b)
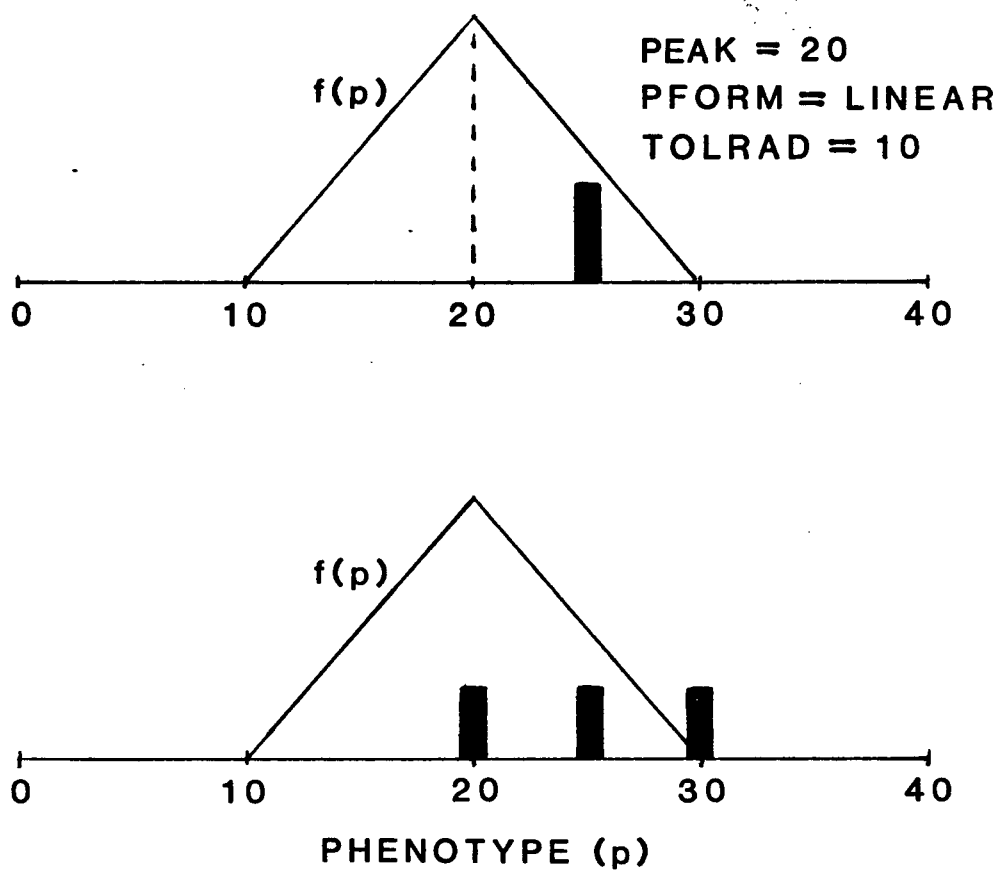
Figure 7.   Reproduction of the portion of the population that has a phenotype of '25'. This population has a step-size of 6.   The offspring with phenotype '31' will die without reproducing if the environment remains the same.

reproduction. Variation production strategies were modelled in this manner because of the potential for studying what the ideal offspring distribution should be. It should be a composite of several step-sizes dependent on the probabilities of occurrence of the environmental conditions for which each is optimal. The work of James (1959) and others showed that the distribution of lethality of effects of mutations is bimodal. One mode is that of slightly deleterious, neutral, and beneficial mutations. The other mode is of very deleterious and lethal mutations. The modelled variation production strategy gives a reasonable approximation to this distribution.

The Environment

The environment was modelled as a symmetrical fitness function. The form of the adaptive peak (PFORM) can be set as triangular (LINEAR) or PARABOLIC. The breadth of the peak is determined by the variable TOLRAD ("tolerance radius"), which is the distance in "phenotypic units" from the optimal phenotype to the nearest phenotype with zero fitness. Narrowing the breadth of the peak by diminishing TOLRAD increases the immediate costs of variation production. The adaptive PEAK can be moved anywhere over the permissible range of 0 to 100 phenotypic units. Any relationship between PEAK and TIME can be specified. The environment has a carrying capacity (KK) which may (ONEPOP:=1) or may not be shared by the populations in a given run.

Dynamics

Any initial frequency distribution may be specified for a

population, either by entering it anew, or by referring to a previously stored distribution. Reproduction occurs at the start of every (discrete) generation. As the adaptive peak moves, individuals of a given phenotype may experience a variety of different fitnesses during a generation. For simplicity, these cumulative effects were ignored. The contribution of a phenotype to the next generation (survival * fecundity) was determined only by the value of the fitness function at the time of reproduction. Fitness is a function only of phenotype, and not of variation production scheme. Phenotypes outside the adaptive peak die without leaving progeny. Phenotypes underneath the peak reproduce as prescribed by the fitness function and the size of the population(s) relative to the carrying capacity. The number of generations to be run can be specified.

Output

The program requests a run description which becomes part of the run summary printed with the output. The available output data are:

1. population sizes over time,

2. mean phenotypes of the populations, and the optimal phenotype, over time,

3. relative average fitnesses (i.e. relative rates of increase) of the populations over time, and

4. mean step-size for all populations over time.

The output facilities are modularized so that each of the four kinds of data may be requested separately. Furthermore, the model is constructed for easy extension of runs through

additional generations. After such a continuation has been run, the original output and the additional output can be requested separately, or together. All output consists of a paired data table and plot.

Program Language

This model is a system of programs running on a PDP 11/45 at the Biosciences Data Centre, University of British Columbia, under the Berkeley UNIX V7 operating system. It is written in Berkeley Pascal, the Bourne shell, and Awk. See Figure 13 for the program listings.

Example Run

Figure 8 shows an example run of the computer model. Lines entered by the user are prefixed with an angle bracket ">". PEAKGAME was called in order to alter the specifications for PEAK, PFORM, TOLRAD or ONEPOP. Otherwise, the model could have been started by calling GAME. The example run modelled the effects of three different variation production strategies in a constantly changing environment. All populations were started with the same size and phenotype distribution. The output of the run is pictured in Figures 9 through 12.

At this rate of environmental change, a greater proportion of the offspring of populations with greater variation production land outside the adaptive peak. For the first four generations (Figure 11) relative fitness was negatively correlated with variation production. The populations with greater variation production tracked the environment more

closely (Figure 10). After the fourth generation, population "X" fared worse, as measured by population size (Figure 9) and relative mean fitness (Figure 11), than the populations with greater variation production. By the end of the run it is apparent that population "+", with an intermediate rate of variation production, has the best "adaptedness/adaptability" compromise for this environment.

Figure 8. Example run of the computer model.

```
>peakgame

 enter PEAK function of TIME, tolrad, {pform, onepop} in Pascal:
>peak:=30+2*time; tolrad:=10; onepop:=1;


 off to work...
   (The peak dynamics are now written into the program,
    and the program is compiled and run. The program begins:)

 This run
                                             is a new run -- 1
                      is a continuation of the last (data in file pd) -- 2
        is a continuation of some other run (data in file other than pd) -- 3
                              uses a stored player distribution -- 4
 Enter 1,2,3 or 4:
>4
 Which file has the stored player distribution?
>dist7
 Enter number of players (<=6)
>3
 Enter the step-size of
 player #1:
>2
 player #2:
>4
 player #3:
>7


 Magic Line:
  peak:=30+2*time; tolrad:=10; onepop:=1;
 Write description of this run:
>constant rate of change
 Number of generations to be run?
>20

        players    step-size
            1          2
            2          4
            3          7
 generations: 0 to 20 = 20
 description: constant rate of change
 Magic Line:
  peak:=30+2*time; tolrad:=10; onepop:=1;

 Does everything look O.K.? (y/n)
>y
```

```
     .
    .
   .
  (population sizes are printed at the end of each generation)
   .
   .
   .
>steplot 1 2
    (produces Figures 9 and 10, respectively)
>fitnesses
    (produces Figure 11)
>meanstep
    (produces Figure 12)
END OF EXAMPLE RUN
```

```
players    step-size
   1          2
   2          4
   3          7
generations: 0 to 20 = 20
description: constant rate of change
Magic Line:
 peak:=30+2*time; tolrad:=10; onepop:=1;

:DATA ECHO==>
point x:step-size of 2      point rt:step-size of 7
      1.    50.98                 1.    50.98
      2.   170.469                2.   119.126
      3.   351.718                3.   181.851
      4.   407.523                4.   169.067
      5.   396.209                5.   141.474
      6.   363.758                6.   126.307
      7.   318.968                7.   119.704
      8.   270.765                8.   120.448
      9.   222.921                9.   125.323
     10.   177.914               10.   131.2
     11.   136.999               11.   134.552
     12.   101.896               12.   136.963
     13.    72.994               13.   136.863
     14.    50.594               14.   133.451
     15.    33.968               15.   128.77
     16.    22.209               16.   124.233
     17.    14.128               17.   118.47
     18.     8.802               18.   111.934
     19.     5.369               19.   105.85
     20.     3.222               20.   100.546
point +:step-size of 4
      1.    50.98         point x:20 xy pairs:step-size of 2
      2.   153.557        point +:20 xy pairs:step-size of 4
      3.   290.326        point rt:20 xy pairs:step-size of 7
      4.   315.935
      5.   296.898
      6.   289.886
      7.   298.499
      8.   326.026
      9.   365.152
     10.   414.908
     11.   463.65
     12.   513.459
     13.   552.504
     14.   588.925
     15.   614.211
     16.   639.855
     17.   653.794
     18.   670.905
     19.   679.037
     20.   692.164
```
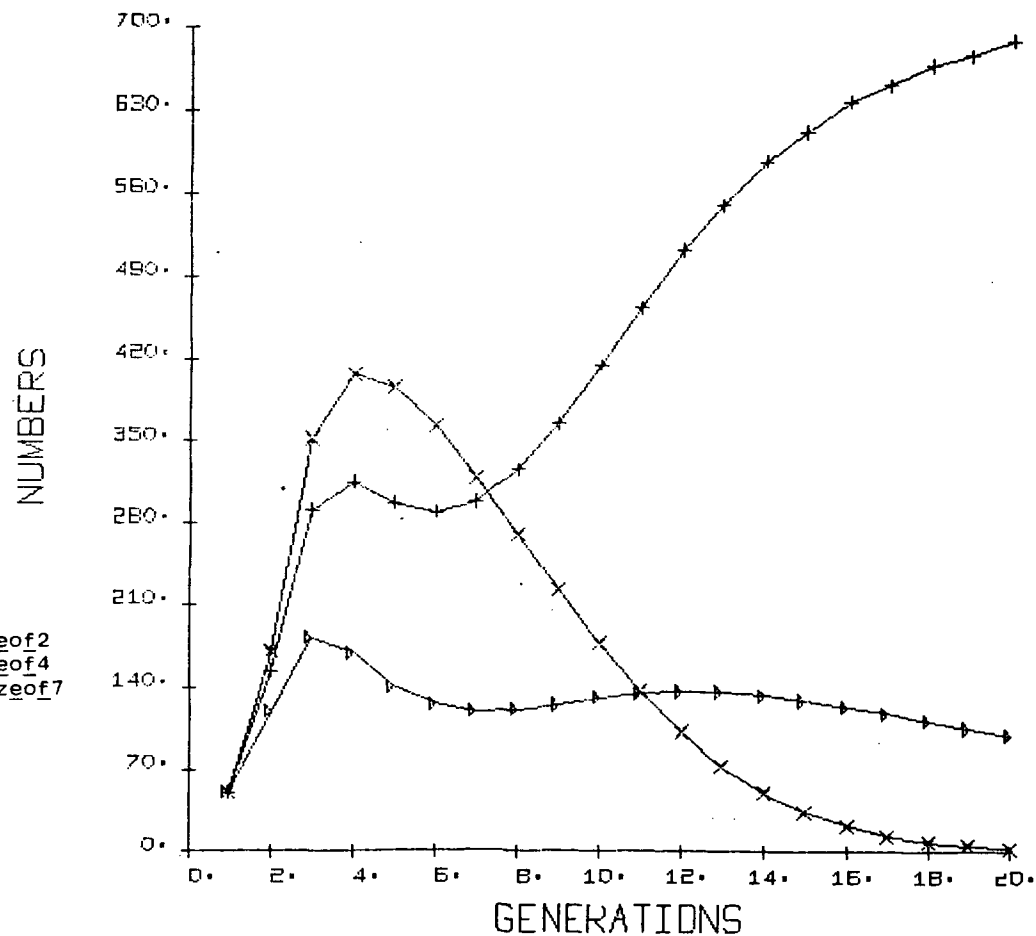

Figure 9. Simulation model output: population sizes over time. Command: steplot 1.

```
      players    step-size
         1           2
         2           4
         3           7
generations: 0 to 20 = 20
description: constant rate of change
Magic Line:
 peak:=30+2*time: tolrad:=10; onepop:=1;

:DATA ECHO==>
point np:                    point +:step-sizeof4
        30.   1.                    30.   1.
        32.   2.                    31.064  2.
        34.   3.                    32.396  3.
        36.   4.                    33.702  4.
        38.   5.                    35.339  5.
        40.   6.                    37.216  6.
        42.   7.                    39.224  7.
        44.   8.                    41.233  8.
        46.   9.                    43.276  9.
        48.  10.                    45.263  10.
        50.  11.                    47.27   11.
        52.  12.                    49.228  12.
        54.  13.                    51.219  13.
        56.  14.                    53.169  14.
        58.  15.                    55.166  15.
        60.  16.                    57.126  16.
        62.  17.                    59.129  17.
        64.  18.                    61.094  18.
        66.  19.                    63.106  19.
        68.  20.                    65.078  20.
point x:step-sizeof2          point rt:step-sizeof7
        30.   1.                    30.   1.
        30.796  2.                  31.626  2.
        31.687  3.                  33.441  3.
        32.345  4.                  35.244  4.
        33.268  5.                  37.14   5.
        34.603  6.                  39.086  6.
        36.205  7.                  41.128  7.
        37.961  8.                  43.192  8.
        39.792  9.                  45.211  9.
        41.659  10.                 47.186  10.
        43.535  11.                 49.201  11.
        45.421  12.                 51.201  12.
        47.308  13.                 53.149  13.
        49.205  14.                 55.158  14.
        51.109  15.                 57.186  15.
        53.026  16.                 59.182  16.
        54.948  17.                 61.14   17.
        56.883  18.                 63.156  18.
        58.823  19.                 65.16   19.
        60.773  20.                 67.116  20.

point np:20 xy pairs:
point x:20 xy pairs:step-sizeof2
point +:20 xy pairs:step-sizeof4
point rt:20 xy pairs:step-sizeof7
```


Figure 10.   Simulation model output: average
phenotypes over time.   Command: steplot 2.

point x:
    1. 1.
    2. 1.
    3. 1.
    4. 1.
    5. .940304
    6. .851567
    7. .777206
    8. .735084
    9. .702394
   10. .689079
   11. .671621
   12. .665733
   13. .65026
   14. .643744
   15. .627618
   16. .622576
   17. .607128
   18. ..60267
   19. .588731

point +:
    1. .900791
    2. .916362
    3. .939192
    4. .966579
    5. 1.
    6. 1.
    7. 1.
    8. 1.
    9. 1.
   10. 1.
   11. 1.
   12. 1.
   13. 1.
   14. 1.
   15. 1.
   16. 1.
   17. 1.
   18. 1.
   19. 1.

point rt:
    1. .698813
    2. .739878
    3. .80239
    4. .860688
    5. .914389
    6. .920377
    7. .921259
    8. .928987
    9. .921351
   10. .917736
   11. .919174
   12. .928652
   13. .914769
   14. .925199
   15. .926101
   16. .93328
   17. .920733
   18. .934322
   19. .931877

point x:19 xy pairs:
point +:19 xy pairs:
point rt:19 xy pairs:
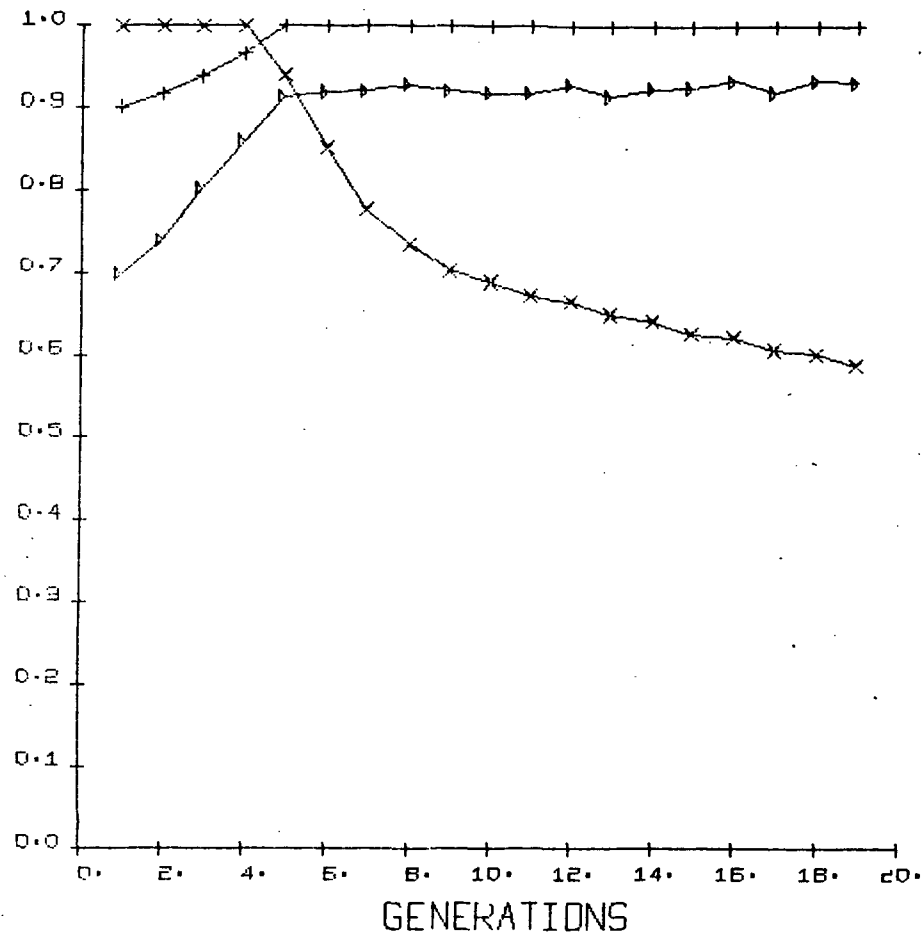
Figure 11. Simulation model output: relative average
fitnesses over time. Command: fitnesses.

DATA ECHO==>

point x:
       1.   4.33333
       2.   4.0371
       3.   3.80837
       4.   3.65509
       5.   3.55906
       6.   3.55306
       7.   3.62176
       8.   3.74878
       9.   3.90206
      10.   4.05217
      11.   4.17636
      12.   4.27528
      13.   4.34708
      14.   4.38703
      15.   4.40977
      16.   4.4175
      17.   4.41602
      18.   4.40195
      19.   4.38824
      20.   4.37088

point x:20 xy pairs:
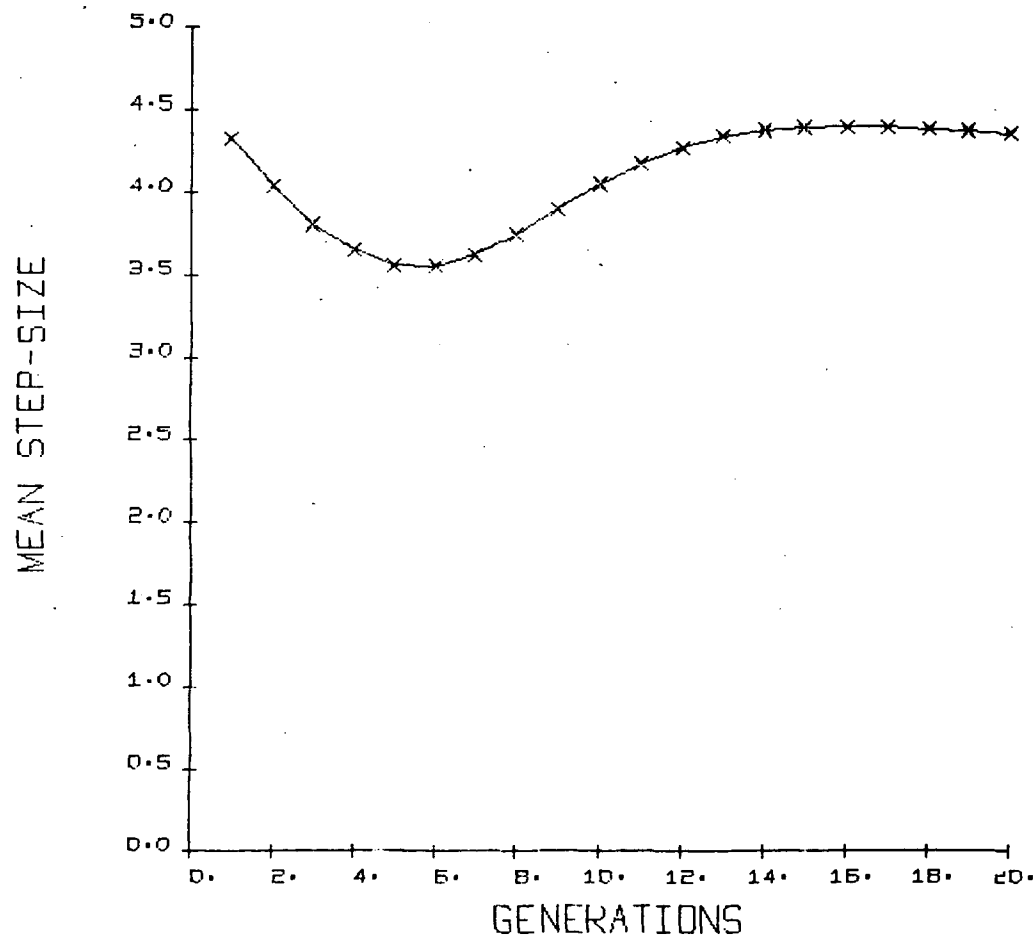
Figure 12.   Simulation model output: average "step-sizes"
        over time.   Command: meanstep.

Figure 13. The simulation model program listings.

PEAKGAME

```
v7=yes
echo 'enter PEAK function of TIME, tolrad, {pform, onepop} in Pascal:' >&2
read mline
echo 'off to work...' >&2
cat <<+ >sgame.p

(* end of Shell, start of Pascal *)
program stepgame (input,output,data,store,r1,r2);
label 13;
const
    fuzz= 1.0e-10;
    minp= 0;
    maxp= 100;
    players= 6;        (* constrained by plot symbols *)
    linelen= 128;
    blank= ' ';
    maxr= 0.5;
    kk= 1000;
type
    string = array[1..linelen] of char;
    fname = array[1..14] of char;
var
    flag1,flag2: boolean;
    choice,i,j,k,maxtime,new,old,p,s,ss,maxss,playin,tinc,time: integer;
    pform,linear,parabolic,onepop,lowp,highp,prevlow,prevhigh: integer;
    kids,pav,peak,prop,r,lambda,tolrad,totnum: real;
    letter: char;
    step: array[1..players] of integer;
    lastnum,num,sum: array[1..players] of real;
    pop: array[1..2,1..players,minp..maxp] of real;
    lab: array[1..players] of char;
    extinct: array[1..players] of boolean;
    blurb,f1: string;
    data,store,r1,r2: text;
    fn: fname;


#include "strings.i"


(* ========== *)
procedure setup(f:fname);
var        doit,i: integer;
    special: array[1..6] of char;
begin
special[1]:='o';special[2]:='u';special[3]:='t';
special[4]:='p';special[5]:='u';special[6]:='t';
doit:= 0;
```

```
for i:=1 to 6 do if special[i]<>f[i] then doit:= doit+ 1;
if doit>0 then rewrite(output,f);
writeln(output);
writeln(output,'        players    step-size');
for i:=1 to playin do writeln(output,i,step[i]);
writeln(output,'generations: ',time:1,' to ',maxtime:1,' = ',maxtime-time:1);
write(output,'description: ');
putstr(output,blurb,'1');
writeln(output,'Magic Line:');
writeln(output,' $mline ');
writeln(output);
rewrite(output,'/dev/tty');
end;

(* ========== *)
procedure saveprop(sfn:fname);
var        s,p:integer;
begin
rewrite(store,sfn);
writeln(store,time,playin);
for s:=1 to playin do write(store,step[s]); writeln(store);
for s:=1 to playin do
   begin
   for p:=minp to maxp do writeln(store,pop[old,s,p]:1:5);
   writeln(store,lastnum[s]:20);
   end;
end; (* saveprop *)


(* ========== *)
procedure readprop(fn:fname;switch:integer);
   (* switch: 0  read only the player distribution          *)
   (*         1  read player distribution and time and # players *)
var        temp:real;
begin
reset(data,fn);
if not eof(data)
   then if switch <>0
           then    begin
                   readln(data,time,playin);
                   for s:=1 to playin do read(data,step[s]);
                   readln(data);
                   end
           else    begin
                   readln(data);
                   readln(data);
                   end
   else writeln('Dummy The file is empty.');
temp:= 0;
for s:=1 to playin do
   begin
```

```
        for p:=minp to maxp do
                if not eof(data)
                        .then    begin
                                 readln(data,pop[old,s,p]);
                                 temp:= temp+ pop[old,s,p];
                                 end
                        else writeln('Error. File ran out at s=',s,' p=',p);
    readln(data,lastnum[s]);
    if abs(temp-lastnum[s])>fuzz
                then     begin
                         writeln('Error. Player total read in does not match that stored.');
                         writeln('temp= ',temp);
                         writeln('lastnum[s]= ',lastnum[s]);
                         end;
    temp:= 0;
    end;
end; (* readprop *)


(* ======================================================================= *)


begin
(*(initializations *)

time:= 0;
pform:= 0;
linear:= 0; parabolic:= 1; (* dummies. would properly make a TYPE *)
onepop:= 0;  (* 0 means they are independent subpopulations *)
(* mline assigns peak function, form, and tolerance radius, and det. whether players share carrying capacity kk *)
$mline
old:= 1;
new:= 2;
totnum:= 0;
maxss:= 0;
prevlow:= minp;
prevhigh:= maxp;
flag1:= false;
flag2:= false;
for i:=1 to 14 do fn[i]:= blank;

writeln; writeln;
writeln('This run');
writeln('                                          is a new run -- 1');
writeln('                      is a continuation of the last (data in file pd) -- 2');
writeln('    is a continuation of some other run (data in file other than pd) -- 3');
writeln('                               uses a stored player distribution -- 4');
writeln('Enter 1,2,3 or 4: ');
read(choice);

case choice of
     1: flag2:= true;
```

```
        2: begin
      f1[1]:= 'p'; f1[2]:= 'd';
      flag1:= true;
      end;
        3: begin
      writeln('Enter name of file where data is stored: ');
      getstr(input,f1,'w');
      flag1:= true;
      end;
        4: begin
      writeln('Which file has the stored player distribution?');
      getstr(input,f1,'w');
      flag1:= true;
      flag2:= true;
      end;
   end;

   (* flag1 => is not a new run, stored player distribution *)
   (* flag2 => need prompting for players and step-size      *)

   if flag2
      then    begin
              writeln('Enter number of players (<=6) ');
              readln; read(playin);
              writeln('Enter the step-size of');
              for i:=1 to playin do
                      begin
                      writeln('player #',i:1,':');
                      readln; read(step[i]);
                      end;
              end;

   for j:=1 to playin do
      begin
      lastnum[j]:= 0;
      num[j]:= 0;
      sum[j]:= 0;
      for i:=1 to 2 do
              for k:=minp to maxp do pop[i,j,k]:= 0;
      end;

   i:=1;
   while (f1[i]<>blank)and(f1[i]<>'^')and(i<=14) do
      begin
      fn[i]:= f1[i];
      i:=i+1;
      end;

   if flag1
      then    begin
              if flag2 then readprop(fn,0)
```

```
                else    begin
                        readprop(fn,1);
          o             rewrite(r1,'res1');
                        rewrite(r2,'res2')
                        end;
            for s:=1 to playin do totnum:= totnum+ lastnum[s];
            end
    else    begin
            writeln;
            writeln('The initial tolerance range is between phenotypes ',peak-tolrad:1,' and ',peak+tolrad:1);
            writeln;
            writeln('Enter: player #, phenotype, numbers ');
            writeln('(O for player # when done)');
            readln; read(s,p,prop);
            while s<>0 do
                    begin
                    pop[old,s,p]:= prop;    (* if any s,p is used more than once, totals will be wrong *)
                    totnum:= totnum+ prop;
                    lastnum[s]:= lastnum[s]+ prop;
                    sum[s]:= sum[s]+ prop*p;
                    readln; read(s,p,prop);
                    end;
            end;

if flag2 then       begin
            rewrite(r1,'results1');
            rewrite(r2,'results2');
            end;


writeln;
writeln('Magic Line:');
writeln(' $mline ');
writeln('Write description of this run:');
readln;
getstr(input,blurb,'1');
writeln('Number of generations to be run?');
readln; read(tinc);  (* might not need the readln here *)
maxtime:= time+tinc;
setup('/dev/tty');
writeln('Does everything look O.K.? (y/n)');
readln; read(letter);
if letter='n' then goto 13;


setup('setup');
lab[1]:='x'; lab[2]:='+'; lab[3]:='r';
lab[4]:='1'; lab[5]:='u'; lab[6]:='d';
writeln(r1,'@label=on');
writeln(r1,'@xlab="generations" ');
writeln(r1,'@ylab="numbers" ');
writeln(r2,'@label=on');
```

```
writeln(r2,'@xlab="phenotype" ');
writeln(r2,'@ylab="generations" ');
writeln(r2,'@ysort');
for i:=1 to playin do    (* useful for data echo only -- doesn't appear on plots *)
   begin
   if lastnum[i]>0 then extinct[i]:= false else extinct[i]:= true;
   if step[i]>maxss then maxss:= step[i];
   writeln(r1,'@ptitle=',lab[i],':step-size_of_',step[i]:1);
   writeln(r2,'@ptitle=',lab[i],':step-size_of_',step[i]:1);
   if not flag1 then
                begin
                writeln(r1,lab[i],time,'     ',lastnum[i]:1:3);
                writeln(r2,lab[i],(sum[i]/lastnum[i]):1:3,'    ',time);
                sum[i]:= 0;
                end;
   end;


(* the per-generation stuff ================================================== *)
while time<maxtime do
begin
$mline     (* peak as a function of time *)
lowp:= round(peak-tolrad+0.5);
highp:= round(peak+tolrad-0.5);
if lowp<minp then lowp:= minp;    (*  allows peak to start with part of its   *)
if highp>maxp then highp:= maxp; (* entire tolerance range outside minp-maxp *)
if prevlow<lowp then lowp:= prevlow;
if prevhigh>highp then highp:= prevhigh;
for s:=1 to playin do if not extinct[s] then
   for p:= lowp to highp do
            begin
            if pform=parabolic then
                    lambda:= 1-sqr((p-peak)/tolrad) (* parabolic fn of dist to peak *)
            else     lambda:= 1-abs(p-peak)/tolrad;   (* function of distance from peak *)
            if lambda<0 then lambda:= 0;
            if onepop=0      (* effects of *)
                    then kids:= pop[old,s,p]*lambda*(1+maxr*(1- lastnum[s]/kk))/3  (* independent crowding *)
                    else kids:= pop[old,s,p]*lambda*(1+maxr*(1- totnum/kk))/3;   (* joint crowding *)
            ss:= step[s];
            pop[new,s,p]:= pop[new,s,p]+ kids;
            num[s]:= num[s]+ kids;
            sum[s]:= sum[s]+ kids*p;
            if (p+ss)<=maxp then
                    begin
                    pop[new,s,p+ss]:= pop[new,s,p+ss]+ kids;
                    num[s]:= num[s]+ kids;
                    sum[s]:= sum[s]+ kids*(p+ss);
                    end;
            if (p-ss)>=minp then
                    begin
                    pop[new,s,p-ss]:= pop[new,s,p-ss]+ kids;
                    num[s]:= num[s]+ kids;
```

```
                    sum[s]:= sum[s]+ kids*(p-ss);
                    end;
            pop[old,s,p]:= 0;
            prevlow:= round(peak-tolrad+0.5)- maxss;
            if prevlow<minp then prevlow:=minp;
            prevhigh:= round(peak+tolrad-0.5)+ maxss;
            if prevhigh>maxp then prevhigh:=maxp;
            end;
time:= time+ 1;
(* output and plot generation stuff ,clear sum and num *)
writeln(r2,'n',peak:1:3,time);
    (* time is number of generations gone *)
    (* readout is state at the end of timeth generation *)
writeln(output);
writeln(output,'*** ',time);
totnum:= 0;
for i:=1 to playin do
    begin
    write(output,num[i]:1:3,' ');
    if (num[i]=0) and (not extinct[i]) then
            begin
            writeln('Player ',i,' has gone extinct. Time: ',time);
            extinct[i]:= true;
            end;
    writeln(r1,lab[i],time,'   ',num[i]:1:3);
    if num[i]=0 then pav:=0 else pav:= sum[i]/num[i];
    writeln(r2,lab[i],pav:1:3,'   ',time); (* extinct players will have phenotype zero *)
    totnum:= totnum+ num[i];
    lastnum[i]:= num[i];
    num[i]:=0;
    sum[i]:=0;
    end;
old:= new;
new:= 3-old;
end;
(* end while *)

saveprop('pd');
13: end.


+
: end of Pascal, restart of Shell
pi sgame.p
mv obj game
game
```

```
********** steplot par [par2 par3 ...]
v7=yes

:Explanation of parameters:
: 1          plots population sizes versus generations of the original run
: 2          plots average phenotypes over generations for original run
: 3          does 1 for continuation run
: 4          does 2 for continuation run
: 5          does 1 for original and continuation runs together
: 6          does 2 for original and continuation runs together
:Prior to a second continuation run, and each subsequent continuation run,
:the continuation data must be appended to the original data by issuing
:the command "tack", because only one set of continuation data is maintained
:at any time.

cat results1 > r1.plot
cat results2 > r2.plot

for i in $*
do
    case $i in
    1) (cat setup; qplot -join r1.plot) | wrap 42 | pr -3 | lpr;;
    2) (cat setup; qplot -join r2.plot) | wrap 42 | pr -3 | lpr;;
    3) (cat setup; qplot -join res1) | wrap 42 | pr -3 | lpr;;
    4) (cat setup; qplot -join res2) | wrap 42 | pr -3 | lpr;;
    5) grep -v @ res1 | cat >> r1.plot;
       (cat setup; qplot -join r1.plot) | wrap 42 | pr -3 | lpr;;
    6) grep -v @ res2 | cat >> r2.plot;
       (cat setup; qplot -join r2.plot) | wrap 42 | pr -3 | lpr;;
    *) echo $i 'is a wierd parameter';;
esac

cat results1 > r1.plot
cat results2 > r2.plot


done


********** fitnesses
v7=yes

: uses peakgame output file results1 to calculate and
: plot relative fitnesses over time

cat <<+ >fit.plot
@xlab="generations"
@ylab="relative mean fitness"
@label=on
```

```
@join
+

awk -f fitawk.a results1 >> fit.plot

qplot fit.plot | wrap 42 | pr -3 | lpr


********** meanstep
v7=yes

: uses peakgame output file results1 to calculate and plot
: the average step size for each generation.

cat <<+ >mstep.plot
@ylab="mean step-size"
@xlab="generations"
@join
+

awk -f avstep.a results1 >>mstep.plot
qplot mstep.plot | wrap 42 | pr -3 | lpr


********** fitawk.a
BEGIN {last=1. ; sum=0. ; flag=0. ; maxf= 0.}

NF==3. && $2 = last {for(i in num) {
                        prop[i]= num[i]/sum
                        if (flag=0. && prev[i]=0.) fabs[i]= prop[i]/prev[i]
                               else fabs[i]= 0.
                        if (fabs[i]>maxf) maxf= fabs[i]
                        prev[i]=prop[i]
                        }
                if (flag=0.) for (i in fabs) print sym[i], last-1., fabs[i]/maxf
                maxf= 0.
                last= $2
                sum= 0.
                flag= 1.
                }

NF==3. && $2==last {sym[$1]=$1; num[$1]=$3; sum += $3}

END {for(i in num) {
        prop[i]= num[i]/sum
        if (flag=0.     prev[i]=0.) fabs[i]= prop[i]/prev[i]
               else fabs[i]= 0.
        if (fabs[i]>maxf) maxf= fabs[i]
        prev[i]=prop[i]
        }
    if (flag=0.) for (i in fabs) print sym[i], last-1., fabs[i]/maxf
```

```
  }


********** avstep.a
BEGIN {flag=0.}

NF==1. { if (index($1,":") = 0.){
            s= index($1,"=")
            step[substr($1,s+1.,1.)]= substr($1,length($1),1.)
            }
      }

NF==3. && flag==0. { flag=1.; last=$2 }

NF==3. && $2 = last { for (i in prod) sum+= prod[i]
                      print last,sum/cnt
                      cnt=0.
                      sum=0.
                      last=$2
                      }

NF==3. && $2==last {sym[$1]=$1; prod[$1]=$3 * step[$1]; cnt+= $3}

END {for (i in prod) sum+= prod[i]; print last, sum/cnt}


********** tack
v7=yes

:appends continuation data to original run data, for additional
:continuation runs.

grep -v @ res1 | cat >> results1
grep -v @ res2 | cat >> results2
```