Morphology, Performance and Fitness[1]

Stevan J. Arnold

*Department of Biology, University of Chicago,
Chicago, Illinois 60637*

Synopsis. Selection can be measured in natural populations by the changes it causes in the means, variances and covariances of phenotypic characters. Furthermore the force of selection can be measured in conventional statistical terms that also play a key role in theoretical equations for evolutionary change. The problem of measuring selection on morphological traits is simplified by breaking the task into two parts: measurement of the effects of morphological variation on performance and measurement of the effects of performance on fitness. The first part can be pursued in the laboratory but the second part is best accomplished in the field. The approach is illustrated with a hypothetical analysis of selection acting on the complex trophic morphology of snakes.

INTRODUCTION

My thesis in this paper is that it is possible to measure adaptive significance directly. In particular it is possible to characterize statistically the relationship between fitness and morphology in natural populations. One can argue that this statistical approach constitutes the highest grade of evidence for selection and adaptation. I will stress this direct approach to selection because of the unique insights it can offer and because it has often been neglected.

Despite its virtues, measurement of selection should not be considered a substitute for other modes of attack on adaptive significance. Direct analysis of selection will be most valuable when it is combined with analytical studies of function and with comparative studies that describe the scope of evolution. Likewise, inferences from functional and comparative studies will be strengthened by companion studies of selection in particular populations.

How can we measure the force of natural selection on specific aspects of morphology, physiology and behavior? The current practice, unfortunately, is to use *ad hoc* data analysis that is not related to any formal evolutionary theory. The strategy outlined here rests on recent advances in multivariate selection theory, which deals with the effects of selection acting simultaneously on multiple characters (Lande, 1979, 1980, 1982). These theoretical results, together with recent success in field measurement of fitness, indicate that selection can be measured in nature in the same terms that are used in equations for the evolutionary transformation of populations (Lande and Arnold, 1983). Multivariate selection theory is briefly reviewed here and a new result is introduced. This is the simple notion that when the selection acting on a trait is formally expressed as the statistical relationship between the trait and fitness, then selection can be factored into two parts: a performance gradient representing the effect of the trait on some aspect of performance (*e.g.*, the ability to swallow large prey) and a fitness gradient representing the effect of performance on fitness. The point of this distinction is that even when effects on fitness cannot be measured, it will often be possible to measure effects on performance. Thus, one contribution of the present paper is a statistical methodology for implementing the laboratory—and field—phased research program advocated by Bock (1977, 1980).

Although the approach advocated here is generally applicable to the study of adaptation, I will illustrate the strategy using only snake feeding as an example of performance. Such illustration, of course, makes the results relevant to the symposium and it may make the abstract concepts more tangible. Since no one has success-

---

[1] From the Symposium on *Adaptive Radiation Within a Highly Specialized System: The Diversity of Feeding Mechanisms of Snakes* presented at the Annual Meeting of the American Society of Zoologists, 27–30 December 1981, at Dallas, Texas.
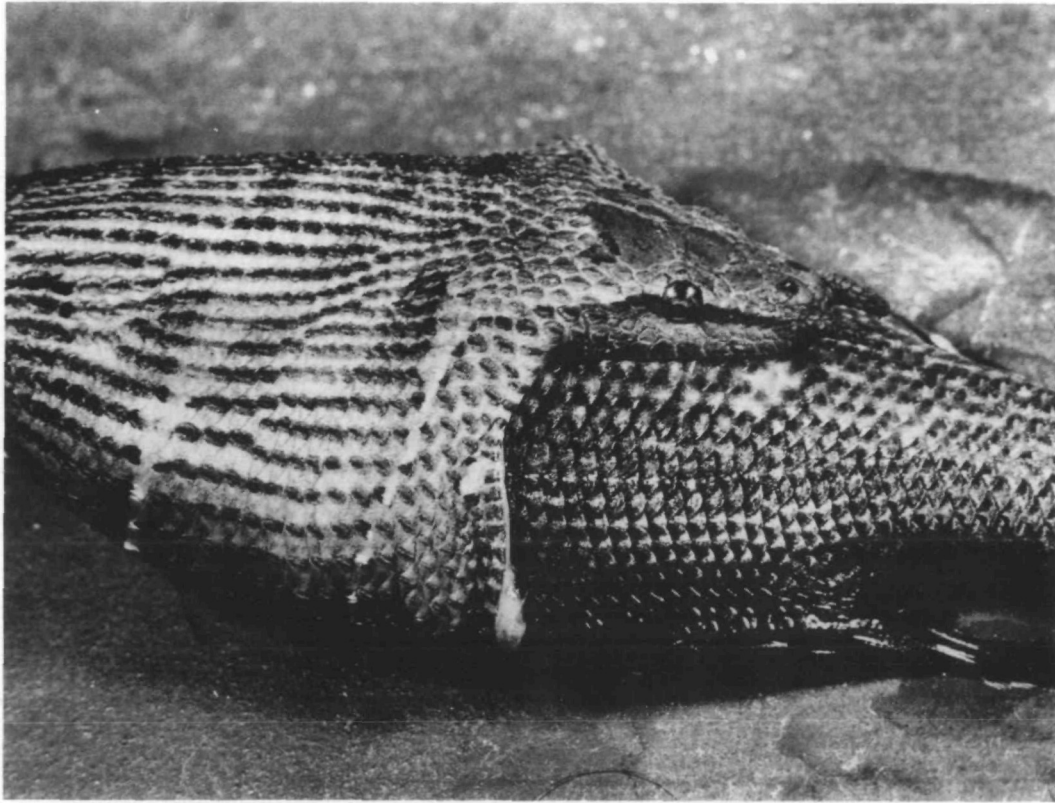
FIG. 1.   A tentacled snake (*Herpeton tentaculum* Lacépède) in the process of swallowing a fish. This large fish was close to the snake's breaking point and the ingestion process lasted 1½ hr.

fully implemented the entire methodology, some of my examples are necessarily hypothetical. My goal is to point out new, sometimes difficult directions for research rather than to display empirical results.

My discussion is restricted to the analysis of adaptation within populations of conspecifics. The main focus is on the analysis of selection that acts on the spectrum of rather minor variations that can be found within a local population. Such variation is usually ignored by morphologists. Although natural selection on phenotypic characters has been detected on numerous occasions during the past 100 years by documentation of subtle shifts in population means and variances (see review by Johnson, 1976), functional morphologists have tended to use other, less direct techniques to study adaptation. Nevertheless the analysis of population variation should become

a powerful weapon in the morphologist's arsenal: even adaptation in complex functional systems can be studied using recent theoretical advances. Of course, such studies are not a substitute for other approaches. Measurement of selection will be most informative when it is supplemented by observational, experimental and comparative work.

In the following sections I outline a methodology for studying adaptation. I begin with the problem of identifying appropriate phenotypic characters, review multivariate selection theory, outline the prospects for measuring selection on morphology and performance and, finally, discuss the limitations of the approach. Throughout I will use "morphology" as a shorthand for any measurable or countable aspect of structure, physiology or behavior.
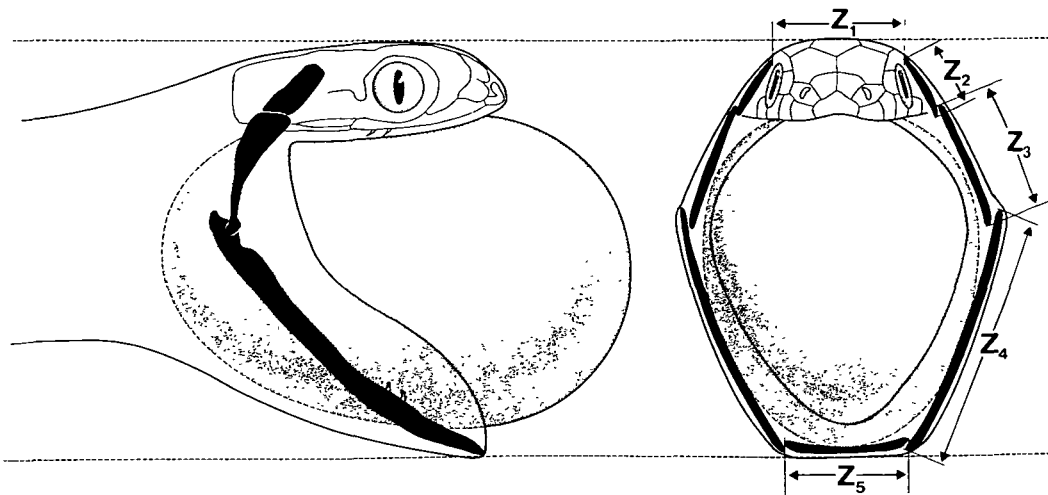
Fig. 2. A diagrammatic portrayal of an African egg-eating snake (*Dasypeltis* sp.) almost maximally distended during the ingestion of its prey (reconstructed from Rabb [1972] and Gans [1974]). Candidate structural elements contributing to swallowing ability are: width of the braincase (z_1) and lengths of the supratemporal (z_2), the quadrate (z_3), the mandible (z_4) and the mandibular symphysis (z_5).

IDENTIFICATION OF INTERESTING PHENOTYPIC CHARACTERS

Comparative, observational and experimental studies can often provide valuable clues about adaptive significance. Such work can generate hypotheses about adaptation that can then be tested with actual measurements of selection acting on variation within populations. For example, Mell (1929) compared genera and families of snakes and noted an association between number of scale rows girdling the body and tendency to eat large prey. Mell's observation makes functional sense because elastic skin between the scales stretches during prey ingestion (*e.g.*, Gans, 1974, and Fig. 1) and so the number of elastic elements will be proportional to the number of scale rows encircling a snake's body. An obvious next step is to see whether swallowing ability is associated with the differences in scale row number that can be found within snake populations, since this would permit measurement of part of the selective force impinging on scale row number. Ideally Mell's hypothesis should be tested in a variety of snake populations.

The mode of analysis is slightly more complex if comparative and observational

studies suggest that a whole suite of characters interact to confer a particular adaptive capacity. Consider, for example, the structural elements that are implicated as contributing to swallowing ability by observation, dissection and electromyography (Albright and Nelson, 1959; Gans, 1961; Kardong, 1979; Cundall, 1983; Pough and Groves, 1983). If we scrutinize a snake that is maximally distended during prey ingestion (Fig. 1) and visualize the location of bony elements by X-ray photography or dissection (Fig. 2) we can easily guess which bony elements contribute to swallowing ability. Thus from Gans' (1952, 1974) and Rabb's (1972) observations of egg-eating snakes (*Dasypeltis*), it appears that the braincase, supratemporal, quadrate, mandible and mandibular symphysis form a chain of elements that encircle the prey during ingestion (Fig. 2). Furthermore, at least some of these elements are proportionally large in taxa such as vipers that can swallow very large prey (Marx and Rabb, 1972). The five structural elements shown in Figure 2 will be used to illustrate the problem of measuring selection acting on an ensemble of traits, even though a variety of other characters might contrib-

ute to swallowing ability and should be included in any definitive study (*e.g.*, number of teeth on various elements, size and placement of cephalic muscles, length of palato-pterygoid arch).

Although comparative studies and functional analysis can suggest which traits participate in a particular adaptive complex, they do not provide direct evidence of selection or adaptation. For example, while functional and comparative studies implicate the structural elements shown in Figure 2 as contributors to swallowing performance, they do not tell us whether the length of the quadrate actually affects fitness through its effects on swallowing ability. Nor do functional and comparative studies tell us whether selection impinges more strongly on the quadrate or on the mandible. These kinds of questions *can* be approached with correlational studies that relate morphological variation to performance and fitness.

Such correlational work is not a poor cousin to experimental analysis. We can approach selection experimentally by ablating a character and measuring the resulting decrement in fitness. Unfortunately such experiments measure the effect on fitness along an artificial character scale, not along the actual spectrum of character variation that occurs in nature. Only the latter can be related to evolutionary theory for responses to selection. Thus experimental work (*e.g.*, ablation) can be informative when it is practical but it does not estimate the critical parameters most useful in evolutionary theory.

Before discussing how the forces of selection impinging on the five elements shown in Figure 2 might be measured, it will be useful to review recent progress in multivariate selection theory, in order to motivate the approach.

MULTIVARIATE SELECTION THEORY

Recent progress in selection theory enables us to predict how much each of a whole series of characters will evolve each generation. However we must know both the patterns of inheritance and selection for the whole series. Using the symbol z with a subscript to denote each character

whose evolution is of interest, then the measurements that describe an individual can be represented by a column vector, **z**. For example, using the characters shown in Figure 2, **z** corresponds to the width of the braincase, length of supratemporal, etc., each measured in an animal of a particular age (*e.g.*, in a newborn snake) and arranged in a column. Likewise the population means of the characters can be represented by a column vector, $\bar{\mathbf{z}}$. Thus the first element in this vector is the average newborn braincase width in the population, $\bar{z}_1$. The evolutionary change in the multivariate mean from one generation to the next (the directional response to selection), can also be represented as a column vector, $\Delta\bar{\mathbf{z}}$. The first element in this vector is the shift in average newborn braincase width from one generation to the next. If the characters, **z**, are normally distributed (transformation of scale may be required to accomplish this) then, as Lande (1979) showed, the deterministic equation predicting one generation of evolution for the whole array of characters is surprisingly simple,

$$\Delta\bar{\mathbf{z}} = \mathbf{G}\ \mathbf{P}^{-1}\ \mathbf{S}. \qquad (1)$$

In this equation, **G** is the additive genetic variance-covariance matrix that describes the polygenic inheritance of characters, $\mathbf{P}^{-1}$ is the inverse of the matrix of phenotypic variances and covariances among characters and **S** is a column vector of selection differentials or shifts in character means due to selection within a generation. In other words, **G** describes inheritance, **P** describes character variation, **S** describes selection and $\Delta\bar{\mathbf{z}}$ gives the response to selection across generations.

In order to visualize these vectors and matrices, it may help if we focus on one character in the illustrated series (Fig. 2), namely braincase width, z_1, measured at a particular age, say in newborn animals. Thus the top lefthand element in the genetic matrix, **G**, is the additive genetic variance for braincase width, describing the correspondence between newborn braincase width in parents and their offspring. The other elements in the first row of this matrix are the genetic covariances between

braincase width and the other characters. These additive genetic covariances describe the genetic couplings between braincase width and the other characters, arising from pleiotropy and linkage disequilibrium. The corresponding elements in the phenotypic matrix, **P**, are respectively the phenotypic variance in newborn braincase width and the phenotypic covariances (unstandardized correlations) with the other characters. The top element in the vector **S** is the selection differential for newborn braincase width. This differential is the difference in mean braincase width of unselected (all of the newborn animals) and selected samples. The latter mean is calculated by weighting the newborn braincase width of each animal by its lifetime relative fitness and taking the average. Thus in order to predict the evolutionary response to selection on newborn braincase width, $\bar{z}_1$, we need to know the genetic and phenotypic variances of newborn braincase width, the genetic and phenotypic correlations with other characters and the shift in mean newborn braincase width induced by selection within a generation.

One important message from expression (1) is that inheritance can be treated as a separate issue from selection. Although evolution can be treated as a consequence of both inheritance and selection, they play distinct, separable roles in our equation for evolutionary change. This means that we can combine information from separate studies of inheritance and selection to make evolutionary predictions. Expression (1) implies that the selection differential, **S**, is the appropriate measure of selective force. It turns out, however, that another measure, the *selection gradient,* is even more useful (Lande, 1979; Lande and Arnold, 1983).

The selection differential and the selection gradient

The selection gradient, $\beta = \mathbf{P}^{-1}\,\mathbf{S}$, measures the *direct force of selection* on the characters, while the selection differential, **S**, measures these direct forces as well as the *indirect forces of selection* exerted through phenotypic correlations between characters. This distinction can be readily visu-

alized. The selection differential is the shift in character mean caused by selection within a generation. It is literally the difference between the character mean evaluated after selection (*e.g.,* the mean size of sparrows surviving an ice storm) and before selection (*e.g.,* the mean size of sparrows just before the ice storm). Now the mean of a particular character (*e.g.,* size) might shift because selection acts directly on that character (*e.g.,* larger individuals survive better because size itself confers a direct advantage) or because selection has acted on a correlated character (*e.g.,* it is actually physiological attributes correlated with size that confer the advantage) (Pearson, 1903).

The significance of the distinction between the selection differential and the selection gradient is that it is the selection gradient that plays the key role in Lande's equation for the evolution of mean phenotype,

$$\Delta\bar{z} = \mathbf{G}\,\beta. \qquad (2)$$

In other words, it is the selection gradient that encapsulates all the necessary information about selection needed to predict the directional response to selection. Thus the selection gradient emerges as a fundamental index of selection. How can it be measured?

Because the selection gradient is the partial regression of relative fitness on a character, it can be measured with conventional statistical techniques in any population in which fitness can be evaluated. This result follows from the observation that the selection differential, or shift in character mean, is equivalent to the covariance between relative fitness and the character (Robertson, 1966; Price, 1970, 1972). Since the selection differential is a covariance, the last two terms in expression (1) can be conveniently identified as a vector of partial regressions using standard statistical theory (Kendall and Stuart, 1979, p. 346). Thus the selection gradient is the partial regression of relative fitness on the character in question holding all other characters constant. It measures the change in relative fitness expected if that character were changed by a unit amount, but none of the other characters were varied. The
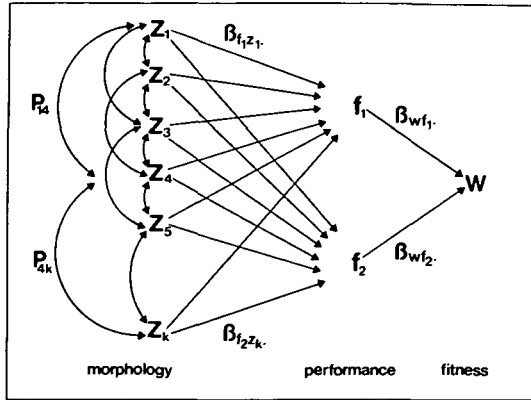
FIG. 3. The statistical relationships between morphology, performance and fitness can be represented with a path diagram. The morphological characters, z_1, z_2, ... z_5, are illustrated in Figure 2; z_k is some other character such as overall body size. The phenotypic covariances between these characters are represented by double headed arrows. For example P_{14} is the covariance between z_1 and z_4. The two performance variables, f_1 and f_2, represent ability to swallow two different kinds of prey. Relative fitness is represented by the symbol w. For simplicity here and in Figure 4, arrows indicating residual influences on performance and fitness are not shown.

identification of the selection gradient as a partial regression means that conventional statistical techniques can be used to estimate it. The selection gradient can also be viewed as the partial derivative of mean relative fitness with respect to the population mean of the character, holding all other character means constant (Lande, 1979).

A diagrammatic view of the selection gradient may be helpful. In Figure 3, the arrows pointing from the characters to relative fitness are the selection gradients. The magnitude of these arrows represents the direct force of selection on the characters, β. The curved arrows connecting the characters represent the correlations or covariances among characters: the information captured in the matrix P.

Factoring the selection gradient: The performance and fitness gradients

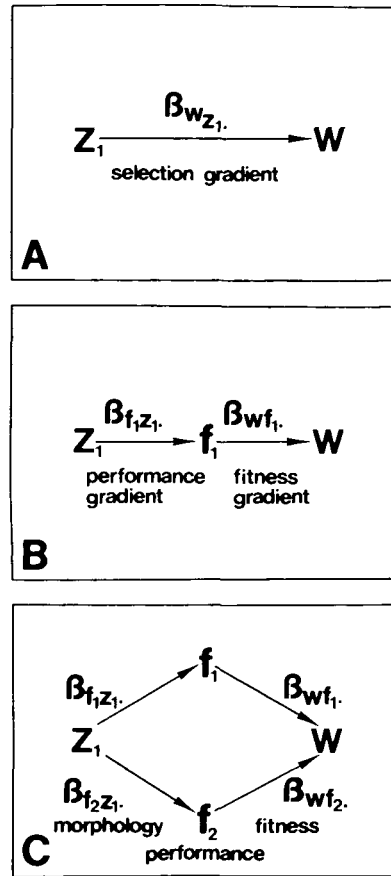The problem of measuring the selection gradient becomes manageable if we break

FIG. 4. A diagrammatic partitioning of the selection gradient. The selection gradient for a character (Fig. 4A) can be partitioned into two parts if the character affects a single performance variable, f_1: the performance gradient, $\beta_{f_1z_1}$, and the fitness gradient β_{wf_1} (Fig. 4B). If the character affects two performance variables, f_1 and f_2, the selection gradient can be partitioned into the paths $\beta_{f_1z_1}\beta_{wf_1}$ and $\beta_{f_2z_1}\beta_{wf_2}$ (Fig. 4C).

it into parts. If we focus on the selection gradient pertaining to a particular character (the arrow in Fig. 4A), we can imagine an intermediate variable that lies in between fitness and the character (Fig. 4B). Performance might be such a variable: the score in some ecologically relevant activity, such as running speed, thermoregulatory capacity or jumping ability (*e.g.*, Emerson, 1978; Bennett, 1980). Thus the problem can be broken down into two parts: measurement of the *performance gradient* (the effect of the character on performance) and measurement of the *fitness gradient* (the

effect of performance on fitness). There is a practical reason for recognizing these two components of the total selection gradient. To measure either the selection or the fitness gradient one would have to measure the relative fitness (relative reproductive success) of individuals in nature; this is often a difficult undertaking. In contrast the performance gradient could be measured in the laboratory and might yield valuable, if incomplete, information on adaptive significance.

Sewall Wright's method of path analysis (Wright, 1921, 1934, 1968) provides a convenient means for factoring the selection gradient into parts. The path coefficients in Wright's scheme are standardized partial regression coefficients. These coefficients have many convenient properties. One can make use of these properties and Wright's useful theorems and then translate the results back into results pertaining to selection gradients, which are ordinary partial regressions. One moves from one system of coefficients to the other using appropriate ratios of standard deviations. Thus *the selection gradient for character z_i* equals the path coefficient from z_i to relative fitness times the ratio of standard deviation of relative fitness to the standard deviation of character z_i,

$$\beta_{wz_i} = p_{wz_i} \frac{\sigma_w}{\sigma_{z_i}}. \tag{3}$$

In this notation the symbol z_i represents the value of the ith character holding all others constant. Using Wright's theorem for chains of causes, we can factor the total selection gradient into parts corresponding to paths toward and away from a performance variable that intercedes between the character z_i and fitness (Fig. 4B). From Wright's theorem, the total path coefficient can be expressed as a simple product of the connecting paths, provided there are no branches. Using expression (3), this same property holds for the selection gradient,

$$\beta_{wz_i} = \beta_{f_1 z_i} \beta_{wf_1}. \tag{4}$$

Thus the direct path from a character to relative fitness corresponding to the selec-

tion gradient can be factored into two parts corresponding to the *performance gradient*, $\beta_{f_1 z_i}$, and the *fitness gradient*, β_{wf_1}. The performance gradient is the partial regression of performance, f_1, on the character z_i, holding all other characters constant and the fitness gradient is the partial regression of relative fitness on performance, f_1, holding all other performance variables constant.

A character may influence more than one kind of performance or more than one component of fitness with branching paths as in Figure 4C. In such cases we can use Wright's theorem for correlated causes to partition the selection gradient into parts corresponding to influences exerted through different kinds of performance or fitness components. From Wright's second theorem, the total path connecting character z_i and relative fitness in Figure 4C is the sum of two paths, one through performance variable f_1 and one through performance variable f_2. From expression (3), the corresponding relationship in partial regression coefficients is

$$\beta_{wz_i} = \beta_{f_1 z_i} \beta_{wf_1} + \beta_{f_2 z_i} \beta_{wf_2}. \tag{5}$$

Thus the total selection gradient can be partitioned into parts, corresponding to branching paths of influence on fitness, as well as factored along paths. These elementary results can be applied to the analysis of selection in more complicated situations, as in Figure 3.

MEASURING THE PERFORMANCE GRADIENT

The sample

The sample that is tested for performance is of critical importance because the goal is to measure the *magnitude* of effects on performance, not merely whether the characters affect performance. Furthermore, the goal is to measure performance gradients that are representative of a particular natural population and to use the partial regression coefficient as the statistic of choice so that the results can be related to multivariate selection theory. This goal immediately implies several criteria for choosing a sample: (1) *Homogeneity.* We may get erroneous results if we pool specimens

from populations that differ in average performance or character values. For this reason it is critical to choose a particular natural reference population and sample from it (Klauber, 1941, 1945 gives herpetological examples). (2) *Sample before selection.* Differential survivorship in the field may change the means, variances and covariances of characters as well as their relationship to performance. On theoretical grounds, the measurements should be made on a sample that has not been shifted, eroded, or rotated by selection. One convenient way to accomplish this is to use newborn animals or a cohort raised in the laboratory under optimal conditions. (3) *Sample at random.* There is a temptation in laboratory work to use only healthy, "typical" animals. But the present goal is to replicate the spectrum of variation actually exposed to natural selection, so no arbitrary esthetic criteria should be used to choose a subset for tests. (4) *Sample size.* Multivariate work requires large samples. The standard errors of partial regressions have **n-k** degrees of freedom with **n** specimens and **k** characters. Thus the sample size should exceed the number of characters by a considerable amount to increase statistical power.

In the following sections the snake example will assume that measurements have been made on a cohort or sample of like-aged animals. If the characters change with age, as in the snake example, age-specific measurements are required to separate the effects of ontogeny and selection (Lande, 1982).

Choosing the measures of performance

The most immediate question is also the hardest to answer: What is the best measure of performance? Unfortunately it probably can not be answered *a priori* unless there has already been enough exploratory work to establish clearly a preferred measure. There are, however, at least two good *a posteriori* criteria. (1) The measure should be ecologically relevant. In other words we might choose a performance measure on the grounds that it shows a higher fitness gradient or is subject to stronger stabilizing selection than other measures. (2) The

measure should be phylogenetically interesting. For example, if two measures show negative correlation then both might be included on the grounds that there might be some interesting structural basis for the tradeoff or because the corresponding genetic correlation might affect evolutionary trajectories. Likewise, comparative work might show that higher taxa differ more in some types of performance than in others, thus identifying the most variable measure for scrutiny at the population level. Both of these criteria suggest that the best course of action is to cast the net broadly by measuring performance in a variety of ways so that correlational analysis or measurements of effects on fitness can objectively narrow the field for future work.

A variety of measures of snake swallowing performance have been used or proposed, but at present we have no real basis for choosing among them. The proposed performance measures include: (1) the number of maxillary movements executed during prey ingestion (Pough and Groves, 1983), (2) the maximum caliber or circumference of prey that can be ingested, the so called breaking point (Arnold, 1982), (3) the prey weight specific rate of ingestion (Arnold, 1982), (4) the maximum mass of prey that can be ingested or digested (Pough and Groves, 1983), (5) the energetic cost of prey handling (Feder and Arnold, 1982). One difficult, but informative way to choose among these measures would be to compare their selection gradients. This would require measuring performance in several ways on each individual in a large sample and then assaying fitness in the field as outlined in the next section.

Swallowing performance is likely to vary dramatically among prey taxa, adding another dimension to the measurement of performance. Some prey have effective antipredator defenses that can thwart ingestion (*e.g.*, Arnold, 1982) but other prey are relatively defenseless and easy to swallow. Furthermore prey taxa of the same mass may differ markedly in shape, so that prey configuration may be a major dimension affecting swallowing performance

(Greene, 1983). One approach to these complications is to define prey specific measures of performance (e.g., a salamander swallowing performance, mouse swallowing performance, etc.). Correlational analysis can then be used to collapse the variable set down to a manageable number of categories or to identify important tradeoffs in prey-handling abilities.

Interpreting the performance gradient

Once performance measures have been determined, one can attempt to measure the performance gradients of the characters. For the sake of illustration, let us suppose that the chosen measure is the maximum circumference of a particular prey type that can be successfully ingested, defined as the snake's breaking point. More exactly a breaking point could be established for each snake in a sample by a sequential feeding scheme that determines the prey circumference at which ingestion and prey rejection are equiprobable (Shaffer, unpublished, uses logistic regression for a similar purpose). Additionally the morphological characters must be measured on each specimen. The performance gradients could then be calculated as the partial regression of performance on each character, holding all other characters constant (many computer statistical packages can perform the necessary calculations).

Calculation of the performance gradients enables us to express the swallowing performance of the average specimen as a sum of contributions from the list of morphological characters. For example, using the character list from Figure 2,

$\bar{f}_1$ = average swallowing performance
= $\beta_{f_1/1}$ $\bar{z}_1$ = average contribution of braincase width
+ $\beta_{f_1/2}$ $\bar{z}_2$ = average contribution of supratemporal length
+ $\beta_{f_1/3}$ $\bar{z}_3$ = average contribution of quadrate length
+ $\beta_{f_1/4}$ $\bar{z}_4$ = average contribution of mandibular length
+ . . . = contribution from other elements.

Each of the contributions is a product of a

performance gradient and an average character value. Thus the characters could be scaled according to their contributions to swallowing performance. On naive grounds we might expect the mandible to contribute more to swallowing performance than the supratemporal. After all, the mandible is longer and would seem to play a greater role in encircling the prey (Fig. 2). This argument, however, might be dead wrong. As every morphologist knows, dissection and measurements are no substitute for analysis of function. In the present case, we let actual swallowing performance determine the functional contributions. Although the mandible is larger than the supratemporal ($\bar{z}_4 > \bar{z}_2$), the supratemporal might actually make a greater contribution to swallowing performance if its performance gradient is sufficiently steep ($\beta_{f_1/2} \gg \beta_{f_1/4}$).

Note that if a set of characters affects only one performance variable then there is a fitness gradient that is the final common path for all the selection gradients pertaining to the characters. As a consequence, the performance gradients would be proportional to the total selection gradients, so the characters could be scaled with respect to the total force of directional selection acting on them, without actually measuring either the selection or the fitness gradients. Unfortunately, it is probably rare for a suite of characters to affect only one aspect of performance.

MEASURING THE FITNESS GRADIENT

The fitness gradient for a performance variable can be measured as the partial regression of relative fitness on the performance variable in question holding other performance variables constant. The fitness gradients for a set of performance variables could be measured without simultaneously measuring the performance or selection gradients for a set of morphological characters, but the preceding notes regarding samples would still apply.

In order to measure the fitness and selection gradients, one must actually score the fitness of individuals. Ideally one would like to score the lifetime fitness of individuals, which is usually taken as the number of

offspring surviving to an age corresponding to that of the parent (Crow and Kimura, 1970). Measurement of lifetime fitness is difficult to accomplish, but field workers are increasingly successful at attaining this goal (*e.g.*, Tinkle, 1967; McGregor *et al.*, 1981). But even when it is impractical to score lifetime fitness, it may still be possible to score some major component of fitness such as juvenile survivorship or mating success. In such cases one can calculate the selection gradients and differentials corresponding to parts of total fitness. These partial gradients are often of interest in their own right even if they do not describe the total effect of selection.

The best estimates of fitness are obtained from longitudinal studies in which individuals are followed through time. Longitudinal samples are especially critical if the characters of interest change during ontogeny (Lande and Arnold, 1983). Such is the case with mensural characters, *e.g.*, those in Figure 2. In such cases, and particularly if generations overlap, it is useful to follow a large cohort through time. The complete research program would consist of: (1) measuring the characters and estimating **P**, (2) scoring performance and calculating the performance gradients, (3) individually marking the specimens and releasing them, (4) recapturing the specimens and estimating fitness or at least some of its major components, (5) measuring anew any characters that change with ontogeny during the recapture program so that ontogenetic trajectories can be used as characters and finally (6) calculating the selection gradients for the characters and for the performance variables.

The feasibility of combining laboratory studies of performance with field estimates of fitness depends on the study organism. Tinkle's (1967) and Fox's (1975) success with longitudinal estimates of fitness and selection with populations of the lizard *Uta stansburiana* suggests that the more complex multivariate program described here may very well be feasible with short-lived, abundant lizards. Likewise recent field studies of garter snakes (*Thamnophis*) suggest that multivariate studies of selection

are feasible (Kephart, 1981). The ideal organism is abundant enough to permit large samples, can be individually marked without imposing selection (colored paints are not appropriate), has populations that are isolated or discrete so that recapture is feasible and results are not confounded by immigration or emigration, and finally it must be feasible to score fitness either by equating differential recapture with differential survivorship and/or by scoring reproductive success. The scoring of fitness is much simplified if lifetimes are short and generations do not overlap (Lande, 1982).

DISCUSSION

The goal of this paper is to reconcile laboratory measurement of performance with two recent developments in evolutionary biology. One of these is growing sophistication in field measurement of fitness (*e.g.*, Howard, 1979) and the other is multivariate generalization of the selection theory used by quantitative geneticists (Smith, 1936; Hazel, 1943; Falconer, 1960; Lande, 1979). These two developments indicate that it is possible to measure the force of natural selection on whole suites of morphological characters (*e.g.*, Boag and Grant, 1981) and that selection in nature can be characterized in the same terms that are used in theoretical equations (*e.g.*, Lande and Arnold, 1983). A resulting optimism is that empiricists and theoreticians will increasingly speak a common language, contribute to each other's enterprise and together move out of the electrophoretic doldrum described by Lewontin (1974). The connection between this endeavor and functional morphology lies in the measurement of performance, an important stepping stone between structural variation and fitness. As Bock (1980) has pointed out, we need to measure both the effects of morphology on performance (*e.g.*, Emerson and Diehl, 1980) and the effect of performance on fitness in order to gain a full picture of selection and adaptation. I know of no studies that accomplish this.

This paper has stressed the directional

aspects of selection, but stabilizing aspects are equally interesting and important. The theoretical setting and prospects for measuring stabilizing selection are thoroughly discussed by Lande and Arnold (1983), but the following sketch gives some of the main results. The short term consequences of both stabilizing and directional selection can be appreciated by visualizing a selection surface. The results can be seen with a three dimensional surface in which two horizontal axes represent two continuous phenotypic characters and the vertical axis is fitness. The directional force of selection (the selection gradient) is proportional to the slope of the surface, while the stabilizing force of selection is proportional to the curvature of the surface. Both slope and curvature are evaluated at the point on the surface corresponding to the population mean. The measurement of curvature and gradient can be accomplished with curvilinear, multiple regression. Such analysis can indicate whether the population is maladaptive (population mean located away from a fitness peak), whether there is strong directional selection on particular characters (steep slope in particular directions) or whether there is strong stabilizing (convex curvature) or disruptive selection (concave curvature). Thus, even when there is no directional selection on a character, analysis of stabilizing selection can reveal whether the population is perched on a peak or sitting on a flat selection surface.

The statistical view of selection used here may help define research problems as well as perform the strictly utilitarian function of showing how best to measure selection. Consider the familiar concepts of functional conflict and incompatible adaptation, the notion that there may be a tradeoff between functional capacities. A first step in testing such a proposition is to identify measures of performance that characterize the two capacities thought to be in conflict. The covariance between the performance measures then provides a test for conflict: if there is a tradeoff the covariance should be negative and its magnitude indicates the strength of conflict. A second, critical step, however, is to determine whether both the selection gradients for the two performance measures are positive (or have the same sign), for only in this case is there a conflict in selection.

Just as the alchemist yearned for a formula that would turn lead into gold, so the modern biologist longs for a conceptual framework that will make any data set coruscate with revelations. The framework described here falls considerably short of such expectations, but then, modern chemistry has not fulfilled the alchemists' most ardent dream. In the spirit of revising expectations to fit the tools at hand, I will review some of the more pressing objections to the methodology, several of which were politely aired at the symposium:

The first few objections deal with the perception of impossible tasks. (1) *The methodology cannot be applied to every taxon.* This is certainly true. Many interesting taxa are sufficiently difficult to obtain that the present approach, requiring large samples, cannot be implemented. Nevertheless work on more common taxa may reveal generalizations about selection that can be extrapolated to rare taxa that can be directly studied only with comparative and functional techniques. (2) *Fitness cannot be measured.* Fitness is not an easy thing to measure, but it *can* be estimated in a surprising array of taxa. Birds, for example, present considerable difficulty to field measurement of selection because of their vagility, yet ornithologists have been increasingly successful in measuring fitness and even the inheritance of fitness components (*e.g.*, Perrins and Jones, 1974; Noordwijk *et al.*, 1980; McGregor *et al.*, 1981; Smith, 1981). (3) *The approach is worthless if effects on lifetime fitness cannot be measured.* The truth in this statement is that the *total* force of selection on a trait can only be measured by assessing effects on lifetime fitness. But many of the questions that concern evolutionary biologists simply deal with effects on parts of lifetime fitness: Is the mandible adapted for prey capture? Does larger size promote survivorship? Such questions can be rigorously approached by measuring effects on performance or on fitness components. Effects on

total fitness are important but they are not the only issue.

The next two, related objections arise from conceptual confusion. (4) *There can be no selection without inheritance.* The problem here is a confusion between selection and response to selection. Selection itself is a purely phenotypic process that occurs within generations. The farmer that culls his inbred chicken flock exerts selection (changes in the average phenotype or its variance) even if there is no genetic variation in the barnyard. In contrast, response to selection, or the change across generations, certainly does depend on genetic variation. (5) *Selection must be measured by its effects on genetic variation.* Here again, evolution is confused with selection. Selection *can* be measured by its effect on phenotypic variation, as Haldane (1954) and Wright (1931, 1980) have advocated. The evolutionary response to such selection, however, will depend on genetic variation (and covariation) and cannot be predicted from selection alone. The fact that selection is not equivalent to evolution does not invalidate the phenotypic study of selection, it merely tempers the evolutionary conclusions that can be drawn. Of course, it is desirable to know something about inheritance as well as selection and in many cases it may be possible to combine studies of selection and inheritance, so that sound evolutionary inference is possible (*e.g.*, Boag and Grant, 1981). If families are used in studies of character variation and performance, then genetic parameters (*e.g.*, heritabilities and genetic correlations) can be obtained as a bonus to the measurement of selection. Albert Bennett and I recently measured morphological variation and locomotory performance in 15 litters (full-sib families) of newborn garter snakes (*Thamnophis radix*). We then individually marked all the animals and released them in an attempt to measure fitness and selection gradients by mark-and-recapture.

The final set of objections deals with technical difficulties, imagined and real. (6) *Morphology is not detectably variable within populations.* It is tempting to imagine that *all* variation in measurable characters, such

as skull width and mandible length, is size related, but studies of population variation do not bear out this impression. Klauber (1938), for example, described intrapopulation variation in the head dimensions of rattlesnakes that was independent of body size. Likewise, age-invariant meristic characters such as scale counts typically vary within lizard and snake populations and show coefficients of variation that range from 2–6% (*e.g.*, Klauber, 1941; Kerfoot, 1970; Kerfoot and Kluge, 1971).

(7) *Learning, ontogenetic and environmental variation will hopelessly complicate the measurement of selection.* Each of these kinds of variation may lessen the force of selection that acts in nature but they do not invalidate its measurement. Learning, for example, may compensate for differences in trophic morphology by behavioral adjustment in prey handling techniques. The result will be a reduced intensity of selection on morphology. However, since the gradient connecting performance with fitness is measured in the field, it will account for such learning effects and other environmental complications.

Ontogenetic change in morphology presents another kind of complication. If individuals have different growth trajectories then morphological differences assessed at one age will not reflect future differences. Geneticists have approached this problem by using the slopes and intercepts of individual growth trajectories (ontogenetic coefficients of allometry) as characters (*e.g.*, Cock, 1966; Kidwell *et. al.*, 1979; Atchley and Rutledge, 1980). In principle one could measure selection on the coefficients that describe individual growth or learning curves, but aside from Weldon (1901) and di Cesnola's (1906) studies this seems not to have been attempted.

Consideration of environmental variation introduces the ecological aspect of selection measurement. Random differences in the microenvironment encountered by individuals within a population may lessen the force of selection on morphology. For this reason the gradient to fitness must be measured in the field. Nonrandom association between phenotypes

and their habitats introduces a more serious complication. If phenotypes are correlated with microenvironment (e.g., larger individuals monopolize the safest territories) then selection on behavioral attributes may be erroneously attributed to morphology. In principle one could take such environment × phenotype correlations into account but their magnitudes must be known. This problem is closely related to the next one.

(8) Unmeasured characters may be the actual focus of selection. This is a major limitation of the methodology. The selection gradient is a partial regression and consequently its magnitude depends on which characters are measured. The resulting ambiguity in interpretation of selection gradients can be minimized but it cannot be completely eliminated. (This same ambiguity, incidentally, complicates the interpretation of any correlation or regression.) The problem can be minimized by studying correlations among a large array of characters that are candidates for natural selection. Any character that is uncorrelated with the character in question will not affect its selection gradient and will not introduce a complication if it is not measured. Thus the biometric study of character intercorrelations is an important adjunct to the measurement of selection.

(9) The agents of selection are never identified. "Selection" may call forth images of birds pouncing on pepper moths (Biston betularia). Agents of selection, however, are notoriously difficult to identify and it is the magnitude of selection, rather than the identity of agents, that figures in equations for evolutionary change. For this reason the preceding discussions have dwelt entirely on the statistical characterization of selection's phenotypic effects. In some circumstances it is possible to identify the agents of selection while measuring their impact. This can be accomplished, for example, by measuring selection during short time intervals that are dominated by identifiable selective agents. For example, the selection that Bumpus (1898) measured in house sparrows can be attributed

to an ice storm that occurred one evening in Providence, Rhode Island.

CONCLUSION

I have argued that selection and adaptation can be studied directly by statistically characterizing the relationships between morphology, performance and fitness. The approach is timely. It is not enough to complain that adaptation is often invoked without critical evidence (Williams, 1966; Lewontin, 1978, 1979; Gould and Lewontin, 1979). We also need an analytical approach that emphasizes what can be accomplished. The strategy outlined here is a step in the right direction.

REFERENCES

Albright, R. G. and E. M. Nelson. 1959. Cranial kinetics of the generalized snake Elaphe quadrivattata II. Functional morphology. J. Morphol. 105:241–292.

Arnold, S. J. 1982. A quantitative approach to antipredator performance: Salamander defense against snake attack. Copeia 1982:247–253.

Atchley, W. R. and J. J. Rutledge. 1980. Genetic components of size and shape I. Dynamic components of phenotypic variability and covariability during ontogeny in the laboratory rat Evolution 34:1161–1173.

Bennett, A. F. 1980. The thermal dependence of lizard behaviour. Anim. Behav. 28:752–762.

Boag, P. T. and P. R. Grant. 1981. Intense natural selection in a population of Darwin's finches (Geospizinae) in the Galapagos. Science 214:82–85.

Bock, W. J. 1977. Toward an ecological morphology. Vogelwarte 29 (Sonderheft):127–135.

Bock, W. J. 1980. The definition and recognition of biological adaptation. Amer. Zool. 20:217–227.

Bumpus, H. C. 1898. The elimination of the unfit as illustrated by the introduced sparrow, Passer domesticus. Woods Hole Marine Biol. Stat. 6:209–226

Cock, A. G. 1966. Genetical aspects of metrical growth and form in animals. Quart. Rev. Biol. 41:131–190.

Crow, J. F. and M. Kimura. 1970. *An introduction to population genetics theory.* Burgess Publ. Co., Minneapolis.

Cundall, D. 1983. Activity of head muscles during feeding by snakes: A comparative study. Amer. Zool. 23:383–396.

Di Cesnola, A. P. 1906. A first study of natural selection in "*Helix arbustorum*" (Helicogena). Biometrika 5:387–399.

Emerson, S. 1978. Allometry and jumping in frogs: Helping the twain to meet. Evolution 32:551–564.

Emerson, S. and D. Diehl. 1980. Toe pad morphology and mechanisms of sticking in frogs. Biol. J. Linn. Soc. 13:199–216.

Falconer, D. S. 1960. *Introduction to quantitative genetics.* Ronald Press, London.

Feder, M. E. and S. J. Arnold. 1982. Anaerobic metabolism and behavior during predatory encounters between snakes (*Thamnophis elegans*) and salamanders (*Plethodon jordani*). Oecologia. 53:93–97.

Fox, S. F. 1975. Natural selection on morphological phenotypes of the lizard *Uta stansburiana.* Evolution 29:95–107.

Gans, C. 1952. The functional morphology of the egg-eating adaptations in the snake genus *Dasypeltis.* Zoologica 37:209–244.

Gans, C. 1961. The feeding mechanism of snakes and its possible evolution. Amer. Zool. 1:217–227.

Gans, C. 1974. *Biomechanics, An approach to vertebrate biology.* Lippincott Co., Philadelphia.

Gould, S. J. and R. C. Lewontin. 1979. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. Proc. R. Soc. London B 205:581–598.

Greene, H. W. 1983. Dietary correlates of the origin and radiation of snakes. Amer. Zool. 23:431–441.

Haldane, J. B. S. 1954. The measurement of natural selection. Caryologia Vol. Suppl.:480–487.

Hazel, L. N. 1943. The genetic basis for constructing selection indexes. Genetics 28:476–490.

Howard, R. D. 1979. Estimating reproductive success in natural populations. Amer. Natur. 114: 221–231.

Johnson, C. 1976. *Introduction to natural selection.* Univ. Park Press, Baltimore.

Kardong, D. V. 1979. "Protovipers" and the evolution of snake fangs. Evolution 33:433–443.

Kendall, M. and A. Stuart. 1979. *The advanced theory of statistics.* Vol. 2, *Inference and relationship.* MacMillan Publ. Co., New York.

Kephart, D. G. 1981. Population ecology and population structure of *Thamnophis elegans* and *Thamnophis sirtalis.* Ph.D. Diss., Univ. Chicago.

Kerfoot, W. C. 1970. The effect of functional changes upon the variability of lizard and snake body scale numbers. Copeia 1970:252–260.

Kerfoot, W. C. and A. G. Kluge. 1971. Impact of the lognormal distribution on studies of phenotypic variation and evolutionary rates. Syst. Zool. 20:459–464

Kidwell, J. F., P. W. Gregory, and H. R. Guilbert.

1952. A genetic investigation of allometric growth in hereford cattle. Genetics 37:158–174.

Kidwell, J. F., J. G. Herbert, and H. B. Chase. 1979. The inheritance of growth and form in the mouse. V. Allometric growth. Growth 43:47–57.

Klauber, L. M. 1938. A statistical study of the rattlesnakes. V. Head dimensions. Occ. Pap. San Diego Soc. Nat. Hist. 4:2–53.

Klauber, L. M. 1941. Four papers on the application of statistical methods to herpetological problems. I. The frequency distributions of certain herpetological variables. Bull. Zool. Soc. San Diego 17: 5–31.

Klauber, L. M. 1945. Herpetological correlations. I. Correlations in homogeneous populations. Bull. Zool. Soc. San Diego 21:5–101.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: Body size allometry. Evolution 33:402–416.

Lande, R. 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. Evolution 34:292–305.

Lande, R. 1982. A quantitative genetic theory of life history evolution. Ecology 63(3):607–615.

Lande, R. and S. J. Arnold. 1983. Measuring selection on correlated characters. Evolution. (In press)

Lewontin, R. C. 1974. *The genetic basis of evolutionary change.* Columbia Univ. Press, New York, 346 pp.

Lewontin, R. C. 1978. Adaptation. Scientific American 239:156–169.

Lewontin, R. C. 1979. Sociobiology as an adaptationist program. Behav. Sci. 24:1–10.

Marx, H. and G. B. Rabb. 1972. Phyletic analysis of fifty characters of advanced snakes. Fieldiana 63: 1–321.

McGregor, P. K., J. R. Krebs, and C. M. Perring. 1981. Song repertoires and lifetime reproductive success in the great tit (*Parus major*). Amer. Natur. 118:149–159.

Mell, R. 1929. *Beiträge zur Fauna Sinica* IV. *Grundzüge einer Ökologie der chinesischen Reptilien und einer herpetologischen Tiergeographie Chinas.* Walter de Gruyter, Berlin.

Pearson, K. 1903. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection of the variability and correlation of organs. Philos. Trans. Roy. Soc. London Ser. A 200:1–66.

Perrins, C. M. and P. J. Jones. 1974. The inheritance of clutch size in the great tit (*Parus major* L.). Condor 76:225–228.

Pough, F. H. and J. D. Groves. 1983. Specializations of the body form and food habits of snakes. Amer. Zool. 23:443–454.

Price, G. R. 1970. Selection and covariance. Nature 227:520–521.

Price, G. R. 1972. Extension of covariance selection mathematics. Ann. Hum. Genet. London 35:485–490.

Rabb, G. 1972. Snakes and eggs. Brookfield Banderlog 39:7–10.

Robertson, A. 1966. A mathematical model of the culling process in dairy cattle. Anim. Prod. 8:93–108.

Smith, H. G. 1936. A discriminant function for plant selection. Ann. Eugen. London 7:240–250.

Smith, J. N. M. 1981. Does high fecundity reduce survival in song sparrows? Evolution 35:1142–1148.

Tinkle, D. W. 1967. The life and demography of the side-blotched lizard, Uta stansburiana. Misc. Publ. Mus. Zool. Univ. Michigan 132:182 pp.

Van Noordwijk, A. J., J. H. van Balen, and W. Scharloo. 1980. Heritability of ecologically important traits in the great tit, Parus major. Ardea 68:193–203.

Weldon, W. F. R. 1901. A first study of natural selection in Clausilia laminata (Montagu). Biometrika 1:109–124.

Williams, G. S. 1966. Adaptation and natural selection, a critique of some current evolutionary thought. Princeton Univ. Press, Princeton.

Wright, S. 1921. Correlation and causation. J. Agric. Research 20:557–585.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 10:97–159.

Wright, S. 1934. The method of path coefficients. Ann. Math. Stat. 5:161–215.

Wright, S. 1968. Evolution and the genetics of population. Vol. 1, Genetic and biometric foundations Univ. Chicago Press, Chicago.

Wright, S. 1980. Genic and organismic selection. Evolution 34:825–843.

.