EVOLUTIONARY PHYSIOLOGY

T. Garland, Jr. and P. A. Carter
Department of Zoology, 430 Lincoln Drive, University of Wisconsin, Madison, Wisconsin 53706

KEY WORDS: adaptation, allozyme, comparative methods, fitness, locomotion, natural selection, quantitative genetics

INTRODUCTION

"The objectives of comparative physiology are: (1) to describe the diverse ways in which different kinds of animals meet their functional requirements; (2) to elucidate evolutionary relationships of animals by comparing physiological and biochemical characteristics; (3) to provide the physiological basis of ecology . . . ; (4) to call attention to animal preparations particularly suitable for demonstrating specific functions; and (5) to lead to broad biological generalizations arising from the use of kind of animal as one experimental variable." (337, p. v)

"Physiological ecology is concerned with the way that physiological traits fit organisms for the ecological circumstances in which they live, so there is always, by definition, an implicit evolutionary component to it." (67, back cover)

"The field of physiological ecology...is...fundamentally evolutionary to the extent that it considers how organisms came to be the way they are and how they might change in the future." (39, p. 251)

The 1950 volume edited by Prosser outlined a broad agenda for comparative physiology (337). The purpose of the present paper is to alert physiologists to the development of a new subdiscipline, evolutionary physiology, which incorporates much of what is contained in Prosser's five objectives (see above) and a substantial fraction of what is generally termed physiological ecology (39, 61, 62, 67, 68, 135, 136, 155, 391, 421, 422). Following this introductory section, we highlight several crosscutting themes in evolutionary physiology, then describe and illustrate five major approaches for addressing such questions (ranging from quantitative genetics to interspecific comparison), and close with some ideas for integrating these approaches.
Our paper is selective in its coverage, and we devote little space to Prosser’s third objective, i.e. ecological implications of physiological and biophysical variation and evolution (see 1, 11, 24, 25, 65, 78, 81, 113, 114, 119, 137, 153, 189, 205, 209, 212, 220, 223, 229, 232, 236, 240, 241, 274, 283, 284, 310, 313, 320, 327, 329, 330, 346, 352, 377, 394, 405, 407, 422, 423, 428, 432, 431, 444).

Modern evolutionary physiology seems to have its origins in the late 1970’s, which witnessed debates concerning the metabolic and thermoregulatory status of dinosaurs and mammal-like reptiles (35, 42, 43, 101, 102, 303, 368, 383, 417). The next major impetus came from attempts to integrate quantitative genetic perspectives into behavioral and physiological ecology (9-11). These efforts were reflected in explicit attempts to document the magnitude and causes of physiological variation among individuals within populations (14, 15, 28, 30, 31, 33, 56, 58, 114, 143, 148, 149, 160, 161, 166, 167, 171, 189, 190, 216, 217, 226, 228, 256, 257, 311, 329, 376, 377, 401, 405, 426, 431, 462, 463), and whether this individual variation was correlated with behavior, life history traits, or ecology (56, 81, 143, 161, 190, 227, 249, 306, 311, 329, 332, 377, 401, 405, 435, 439, 447, 462, 463). Other studies tested whether individual variation in physiological traits had any genetic basis (55, 57, 58, 150, 154, 157, 158, 226, 260, 278-281, 402, 410, 424, 427), or could be molded by laboratory selective-breeding studies (37, 40, 41, 87, 179, 215, 218, 219, 278, 279, 360-362). Most recently, phylogenetically-based comparative studies have come to the fore (27, 50, 88, 164, 165, 213, 218, 265-267, 274, 336, 406, 437). Interestingly, the use of physiological information for reconstructing evolutionary relationships is not presently receiving much attention (but see 21, 289, and compare 106, 364).

Thinking about evolution is not new to physiologists (95, 135, 155, 242-245, 290, 316, 337, 371, 383, 456, 457, 460, 464; references therein). Nonetheless, at the risk of failing to appreciate sufficiently the accomplishments of past evolutionary-thinking physiologists, we see contemporary evolutionary physiology as fundamentally different from most of what came before. Many current practitioners began their studies as evolutionary biologists, or were formally trained in both evolutionary biology and physiology. Others represent physiologists who have moved forcefully in an evolutionary direction, often taking up formal collaborations with evolutionary biologists. Whatever their genesis, today’s evolutionary physiologists try to do state-of-the-art physiology and state-of-the-art evolutionary biology; the evolutionary interpretation is no longer an afterthought. As in behavioral ecology (345), part of this increase in evolutionary rigor came in response to Gould & Lewontin’s (175) criticisms of “the adaptationist programme” (36, 135, 262). Evolutionary physiologists now use a range of tools to test a priori
hypotheses. Previously, evolutionary conclusions usually were inductive and followed the accumulation of considerable data (e.g. the basal metabolic rates of many species); such an “encyclopedia” approach (135) has been criticized as “stamp collecting.” The switch from an inductive to a hypothetico-deductive model of analysis reflects some maturity in the field (61, 62, 135, 364); it is difficult to phrase non-trivial a priori hypotheses about processes until patterns have been thoroughly documented.

The 1987 publication of the results of a National Science Foundation-sponsored workshop on “New Directions in Ecological Physiology” (135), with an emphasis more evolutionary than ecological (e.g. 134), heralded the field (see also 34, 35, 39, 61, 62, 155, 214). The “Evolutionary Physiology” symposium held at the 1993 meetings of the Society for the Study of Evolution helped to advertise the field to nonphysiologists, and the formation in 1992 of a National Science Foundation panel in “Functional and Physiological Ecology,” now renamed “Ecological and Evolutionary Physiology,” will help to maintain this marriage (cf 135, 136, 237). Although current practitioners have various origins, many of the next generation will begin their graduate careers aspiring towards becoming evolutionary physiologists.

EVOLUTION FROM A PHYSIOLOGICAL PERSPECTIVE

AND VICE VERSA

Physiologists are interested in how organisms work (231, 380). A subset of physiologists also wants to know why organisms are designed to work in particular ways. Unless one assumes special creation of all organisms, an understanding of such why questions requires an evolutionary perspective (134, 299, 308, 460). In this section we briefly review some of the recurring evolutionary questions and related principles that have been considered in the physiological literature. The following five sections cover complementary approaches to studying physiological evolution.

How Do Different Kinds of Organisms Work?

Physiologists have always sought to discover general principles of organismal function, such as homeostasis, or the scaling of metabolic rate with body mass (25, 32, 100, 134, 290, 337, 339, 456, 457). Faced with the tremendous diversity of living organisms, both in terms of numbers of species and their behavioral and ecological variation, physiologists have asked what general principles apply to all or most organisms, how common are exceptions to the rules, and whether there exist multiple solutions to a given adaptive problem, such as life in hot, arid environments (24, 25). Because all organisms on this planet are descended from common ancestors (and perhaps from a single common ancestor), general biological principles
(e.g. use of DNA as a genetic material; structure of eukaryotic cell membranes; responses to changes in ambient temperature by mammals) are likely to occur in a strongly hierarchical—that is, phylogenetic—pattern.

Extremes of Adaptation, Model Species, and the August Krogh Principle

Identification of similarities among species allows the possibility that certain species may be able to serve as model systems for studying basic physiological processes (33, 62, 100, 456, 457). Krogh noted that for any physiological principle there exists an organism especially well suited for its study [259: e.g. giant axons of squid, ‘gas windows’ of crab legs (288); scallop muscles (292); rattlesnake tail muscles (375)]. Similarly, physiologists (including plant physiologists, 310) have long been aware that organisms living in extreme environments are especially likely to exhibit clear examples of evolutionary adaptation because of the presumably intense past selective pressures (24, 25, 155, 380). Organisms adapted to extreme environments can serve to illustrate the range of evolutionary possibilities (62, 456, 457), but we must be mindful that the organisms alive today—and hence available for physiological study—are but a small fraction of what has existed. Thus there is no guarantee that we can observe the range of possibilities even among the most extreme of living species. For example, the largest terrestrial mammal that ever lived (*Baluchitherium*) was much larger than living elephants, or mammoths, or mastodons (references in 165). Although the proposition is seductively appealing, little evidence exists that today’s species are any better adapted than those of a million or 100 million years ago (174, 224, 225). We must also remember that behavioral adaptation can go a long way towards ameliorating the need for physiological evolution (24, 25, 52, 209). In any case, species displaying extreme development of a particular physiological property can also prove useful as model systems (e.g. locomotor abilities: 153, 231).

Are Species Differences in Physiology Adaptive?

“Four legs may be optimal, but we have them by conservative inheritance, not selected design.” (173, p. 44)

The neo-Darwinian synthesis (77b, 144, 350), including its emphasis on natural selection as the major driving force in evolution, led inevitably to the view that virtually all features of organisms are adaptive. Comparative physiologists have routinely viewed any differences among species as adaptations to their different life styles (60, 118, 126, 133, 172, 202, 290, 337-339, 382, 425, 449, 455, 467), and have provided many examples that clearly represent strong evidence for adaptation (32, 95, 371; see also 6).
Nonetheless, not all features of organisms represent adaptations to current environmental conditions (26, 36, 59, 175, 187, 272, 275, 454, 455, 460). Some, for example, represent simple inheritance from ancestors. A current thrust in evolutionary biology aims to develop rigorous methods for studying adaptation (see below), including ways to formally test hypotheses about the adaptive nature of organismal features (1, 6, 26, 50, 51, 59, 175, 182, 187, 239, 263, 265-267, 274, 275, 278, 307, 345, 349, 350, 394, 459).

The operational definition of evolutionary adaptation is quite controversial (36, 182, 345), but revolves mainly around the distinction between origin and maintenance of a trait. Some traits are maintained by current selection, but did not originate in response to the same selective pressures. For example, as noted by Darwin, the sutures between bones in the mammalian skull may now be adaptive for allowing the birth of relatively large-brained offspring in some species (humans), but they arose long ago in evolutionary history (144, p. 257). Other traits arose because of natural selection, but are now present for other reasons, such as "phylogenetic inertia," which may be attributable to developmental or genetic constraints in addition to selection (e.g. see 187). Thus the origin (evolutionary history) and/or the current maintenance (phenotype existence sensu 345) of a trait may be adaptive, and both are worthy topics of study. Phylogenetic comparative studies of interspecific variation can address historical origins, and measurement of selection within populations can address current maintenance (see below).

Imprecise usage of the term adaptation (36, 345, 454, 455) has led to many confusions in the physiological literature:

"Physiologists have used the term 'adaptation' in two entirely different ways. First, adaptation is used to describe compensatory, short-term changes to environmental or organismic disturbance. Such control systems are phenotypic and reveal the plasticity of physiological systems generally. Second, adaptation is used in the genetic and evolutionary sense of describing a trait or feature that has been cemented into the genotype through the pressure of natural selection. Comparing the acclimatory changes in erythrocyte phosphates of fish or humans with traits like the high-affinity Hb systems of the high-altitude-dwelling llamas, amphibians, and so on is clearly a category mistake." (454, p. 244)

Regardless of methodological or semantic debates, studies of adaptation will continue their prominence in both evolutionary biology and comparative physiology. Physiology can contribute something special to the study of adaptation: an understanding of mechanism (50, 70, 95, 239, 367, 371). An understanding of biochemical and biophysical mechanisms can help to define what is theoretically possible for organisms to achieve (the phenotype-set sensu 345), if selection were to favor it, and thus aid in the
identification of constraints on evolution [30, 71, 72, 100, 342 (but see
302); 367, 388, 396, 459, 465].

Are Organisms Optimally Designed?

"We do not think a functional explanation complete until we can show that a
structure or movement is optimal (by some plausible criterion) for the proposed
function." (5, p. 237)

In addition to presuming that most if not all features of organisms represent
adaptations, a common perspective in comparative physiology is to view
organisms as more-or-less optimally designed (3-5, 363). Reasons why
organisms may not be optimally designed and why optimality perspectives
in general may not be the best way to study evolution have been discussed
at length elsewhere (24, 25, 36, 77a, 79, 92, 124, 175, 182, 321, 361,
459; but see 241, 354, 390). With regard to this debate, we emphasize two
points. First, a fundamental reason for a lack of optimal solutions to adaptive
problems is that natural selection—the only known mechanism for adaptive
phenotypic evolution—can only work on existing phenotypic variation that
is at least partly heritable; thus some possible solutions to selective problems
will most likely never be accessible within a given lineage (36, 175, 222).
Second, a major problem with testing optimality predictions is the lack of
a suitable null model (see 182); using the optimal solution, e.g. perfect
matching, as a null model (see 109 on symmorphosis), is counter to the
way most biological and statistical inference is performed, e.g. no matching
as a null model.

The concept of symmorphosis is a recent example of an optimal design
perspective on animal morphology and physiology (451, 452). Favorable
discussions and criticisms of symmorphosis have appeared elsewhere (35,
p. 13; 65, 108-110, 117, 163, 231, 273, 381, 392), and casual references
appear frequently (35, 93, 134, 209). Qualitatively, the likelihood of
match between biological structures and their functional requirements is
intuitively obvious; exactly how good this match should be is less obvious.
Moreover, adequacy or sufficiency, rather than optimality, is the most likely
evolutionary outcome (11, 24, 25, 145, 147, 460, p. 17) because natural
selection tends to maximize relative, not absolute, fitness.

Trade-offs and Constraints: Why Do Traits Evolve in a
Correlated Fashion?

Constraints or trade-offs can be identified and studied at several different
levels of biological organization (12, 30, 67, 173, 218, 302, 374, 396,
459). The environment and its associated selective regime (26, 275) impose
constraints on what types of physiological variants and on what kinds of
organisms can survive and reproduce, both within populations at the level of individual variation, and over long term evolutionary time (258). Although selection sets ultimate limits on the dimensions of an allowable multidimensional phenotypic space, organisms will not necessarily ever reach those limits. Instead, fundamental properties of biological systems can preclude some variants, such as titanium in tortoise shells. Constraints set by inherent organismal properties can be elucidated through biochemistry and physiology [30, 71, 72, 342 (but see 302), 374, 388, 396], or through developmental biology. Alternatively, they can be evidenced as a lack of genetic variation in certain phenotypic dimensions. For example, non-zero genetic correlations are often interpreted as indicating evolutionary constraints (12), but genetic correlations (see below) can sometimes facilitate response to natural selection (154, 279) and can themselves be molded by selection (55-58); moreover, their interpretation can be very complicated (77a, 354). Wherever a constraint is identified, a fundamental question is whether it is absolute or could possibly be overcome (e.g. by a new mutation); physiology should address such questions.

**What is the Origin of Allometric Relationships?**

Allometric patterns and equations describing them have long fascinated physiologists (65, 66, 149, 187, 196, 274, 313, 320, 346, 351, 381, 414, 419, 450). Within a clade, and on a log-log scale, variation in body mass can explain more than 90% of the variation in a variety of physiological traits, such as basal metabolic or heart rate. Even with such a high coefficient of determination ($r^2$), variation of individual points about a linear regression can be substantial, both among individuals within species (38, 148, 160, 226) and among species (151, 304, 342, 399, 400). However, not all physiological traits vary strongly with body mass; examples include hematocrit, blood oxygen carrying capacity, and normal body temperatures of birds and mammals (160, 336, 381).

Ignoring variation about a regression line, an obvious question is whether the slope of the line itself (the allometric scaling exponent), or its intercept (196), represent general physical and biochemical constraints placed on organisms as opposed to adaptive evolutionary responses. For example, Weibel et al (452) categorize variation among species as "allometric...reflect(ing) intrinsic properties of the organism, particularly the size dependence of rate constants, such as stride frequency, heart rate, etc" or as "adaptive...relat(ing) to behavioral traits and to the ecological conditions to which the species are adapted by evolutionary selection..."

The dichotomy between allometric and adaptive variation is imprecise (and analogous to the quantitative genetic problem of trying to separate genetic from environmental effects when genotype-by-environment interac-
Inter-specific allometric patterns can themselves be adaptive in the sense that they are maintained by natural selection. In some cases, the adaptiveness of allometric relationships is subject to experimental evolutionary tests. For example, Weber (448) observed a fairly tight allometric relationship between wing proportions and body size among populations and species of *Drosophila*. Did this represent an ineluctable constraint on the shapes of fruit flies that could be produced? Did the existing phenotypic diversity indicate what was possible? To test this, Weber artificially selected on wing proportions and succeeded in producing lines of flies that exceeded the limits seen in nature (GS Wilkinson, personal communication, has done similar experiments with another species of fly). Thus the allometric pattern seen in nature did not reflect an unconditional genetic or developmental constraint on the shape of flies. Instead, it must have been maintained by natural selection.

Artificial selection experiments (see below) will not be practical for studying most examples of allometry in physiological traits. Another way to alter allometric relationships, one that relies on proximate rather than ultimate mechanisms, is "allometric engineering," as applied by Sinervo and colleagues to lizard eggs to alter characteristics of hatchling lizards (393–396). Alternatively, an understanding of physiological mechanisms can help in determining whether a particular pattern of phenotypic variation or covariation (e.g. an allometric relationship) represents what could possibly exist or just what selection has allowed.

**QUANTITATIVE GENETIC ANALYSES**

"...species represent the product of evolution, whereas the process can only be studied within species...For the physiological ecologist, heritability is the most useful piece of genetic information since it is both descriptive and predictive." (278, p. 497)

Most traits studied by physiologists show continuous variation (body temperature, metabolic rate, blood pressure, blood hemoglobin levels, enzyme activities). Quantitative genetics was developed early in this century for studying the genetic basis of traits that showed more-or-less continuous, as opposed to discrete, variation (references in 9, 10, 13, 51, 58, 127, 132, 180, 183, 261, 262, 297, 323, 453). The general assumption is that such traits are affected by alleles segregating at many loci and that each locus has a relatively small effect on the phenotype. Quantitative genetics has a long history of application in plant and animal breeding, but a recent revival for application to problems in evolutionary biology began in about 1980 (14, 260–262, 357, 402).
Quantitative genetics uses observed phenotypic variation among individuals of known relationship (e.g., parents and their offspring, full- and half-sibs) to estimate the relative magnitudes of genetic and environmental effects on the phenotypic variation observed within a population (i.e., variation among individuals). Thus it traditionally does not attempt to identify the effects of variation at single gene loci. Instead, it allows estimation of summary statistics, such as heritability, that are both descriptive and predictive. When more than one trait is studied, estimates of shared genetic and environmental effects are available. Quantitative genetics is actually a broad collection of tools that can address topics including whether the phenotypic variation in a single trait (e.g., differences in resting blood pressure among individuals within a population) is to any extent genetically based; whether selection on a trait could lead to improvement; whether different traits are genetically coupled such that selection on one would necessarily produce a correlated response in another (including the nature of allometric relationships, e.g., 448); how many genes are responsible for the phenotypic differences between populations of the same or of closely related species; whether one or more “major genes” are segregating within a population; whether gene action is entirely additive or includes dominance or epistasis; and the direction and magnitude of past natural selection. Recent marriages of quantitative genetics with molecular marker techniques are allowing the actual identification of specific loci with relatively large effects, termed quantitative trait loci (230, 324).

Only if a trait under selection is to some extent genetically based will natural selection result in evolution. The equation, \( R = h^2 s \), is used to describe the evolutionary response of the mean value of a single phenotypic trait to natural (or artificial) selection, where \( R \) = the genetic response to selection, \( s \) = the selection differential (difference in means of selected and unselected individuals), and \( h^2 \) = the narrow-sense heritability (ratio of additive genetic variance to total phenotypic variance, ranging between zero and one). Narrow-sense heritability can be estimated by measuring the phenotypes of related individuals, such as offspring and their parents (13, 51, 58, 132, 180, 297, 323, 410, 453), or through artificial selection experiments. A multivariate version of the foregoing equation substitutes a matrix of additive genetic variances and covariances for \( h^2 \) (13, 51, 56–58, 261, 262).

Physiological traits are highly susceptible to a variety of environmental effects, both acute and chronic (e.g., acclimation and acclimatization): “Much of the intraspecific physiological variation encountered by physiologists is a result of short-term physiological acclimation...” (62, p. 203). Therefore, one might expect that physiological traits would often exhibit low or even zero narrow-sense heritabilities. Until recently, comparative physiologists...
simply did not address such questions empirically (e.g. discussion following Reference 25). However, available empirical studies indicate that physiological traits often do show substantial heritabilities (37, 40, 41, 58, 127, 132, 158, 178–180, 203–205, 215, 219, 260, 278–281; 312, 328, 355, 360–362, 402, 410, 420, 461). Some measures of locomotor performance in lizards and snakes (55, 57, 150, 226, 424, 427) and even maximal oxygen consumption in garter snakes (157) seem to show very high heritabilities (0.5–0.9) based on comparisons of families of full-siblings. Unfortunately, these estimates can be inflated by non-additive genetic effects or maternal effects, and more sophisticated breeding designs are required to estimate narrow-sense heritabilities. Other physiological traits, such as basal metabolic rate in house mice (278, 279; T Garland, Jr et al, unpublished), show very low (< 0.05) narrow-sense heritabilities.

Genetic correlations are the two-trait analogue of heritabilities; they indicate the extent to which the phenotypic covariance of two traits is genetically based. Although genetic correlations can be caused by linkage disequilibrium (for example, from physical linkage of genes), pleiotropy is a more typical cause. Pleiotropy simply refers to one gene affecting more than one trait. Shared biochemical, physiological, or developmental pathways are likely to be reflected as pleiotropic gene action. Thus genetic correlations may be particularly interesting to physiologists because of what they can suggest about physiological mechanisms (10, 279, 360). Conversely, knowledge of physiological mechanisms can be used to predict genetic correlations (154, 318). For example, Garland (150) predicted a necessary trade-off between speed and stamina, based on the fast-twitch, slow-twitch dichotomy (an oversimplification) of muscle fiber types. In contrast, the genetic correlation estimated from measurements of full-sibling families of garter snakes was actually positive (for possible explanations, see 154, 436). Rather than constrain, this positive correlation could facilitate genetic response if natural selection favored increased overall locomotor abilities (150, 154, 166; cf 279, 360). Dohm & Garland (115) looked to differences in developmental timing to predict genetic correlations between numbers of scales in different regions of a snake's body; empirically, this prediction was only partially supported. In D. melanogaster, artificial selection on desiccation resistance increased longevity and depressed early fecundity both in stocks that were originally selected for delayed reproduction and in control lines (360, 362). Other physiological performance characters also responded to these selection regimes, including flight duration and ethanol tolerance (increased by desiccation but not by starvation selection). Graves et al (179) studied the underlying mechanisms and found that desiccation tolerance was reduced substantially in flies depleted of glycogen reserves by flight in both selected and control lines; thus variation in the
amount of glycogen reserves was at least partly responsible for the correlation between seemingly unrelated physiological traits. This is an example of how antagonistic pleiotropy between genes that have differential effects on early fitness can create physiological correlations observed during selection.

Both heritabilities and genetic correlations can be used to predict responses to selection, although the number of generations over which such predictions will be accurate depends on the constancy of the genetic parameters (12, 51, 132, 261). Traits with low heritabilities are often inferred to have been subject to intense past selection, although a number of tenuous assumptions are involved (51, 158). The presence of dominance or epistasis can also be used to draw inferences about how past selection has acted on a trait (54, 132, 158, 193, 197, 279, 297). To investigate the nature of genetic variation, such as whether it is entirely additive or includes dominance or epistasis, crosses of selected lines, inbred lines, subspecies, or even closely related species can be used (278, 279, 297, 347). Such crosses also form the basis for estimating the minimum number of independently segregating genetic factors required to account for a difference in phenotypic mean (469).

ARTIFICIAL SELECTION EXPERIMENTS

Selection experiments are one form of genetic manipulation (51, 127, 132, 180, 183, 200, 201, 297, 323, 353, 453) that has intuitive appeal for evolutionary physiologists. They are in some sense more natural than are more modern alternatives such as the production of transgenic organisms (136, 263; J Breslow, this volume; M Paul et al, this volume). Of course, selection in the laboratory cannot entirely mimic selection in nature because the former generally involves much more specific targets of selection (single characters), higher selection intensities (and often truncation selection), smaller populations, and much shorter time scales (360, 361, 397), although these limitations do not necessarily apply to study organisms such as microorganisms.

Two general kinds of selection experiments can be distinguished (360). Traditional artificial selection involves laboratory or barnyard populations in which each individual in each generation is scored for some phenotypic trait (or combination of traits) of interest. Some top or bottom percentage of individuals from the distribution of phenotypic scores is then selected as the parents for the next generation; this is termed truncation selection (for physiological examples, see 74, 90, 91, 203–205, 215, 378). Variations on this theme are used routinely in plant and animal breeding, such as taking at least one male and female from each family to reduce inbreeding (132, 200, 201, 297, 353: e.g. 278, 279). In laboratory natural selection, freely breeding populations are exposed to intentionally altered environmental
conditions, such as different temperatures, or to a laboratory or other
environment that is novel as compared with nature (22, 23, 37, 40, 41,
219, 269, 362, 420; references therein), or to husbandry conditions changed
so as to favor altered demographic schedules, such as delayed reproduction
(179, 360–362; references therein).

Artificial selection can be a sharper experimental instrument because it
is more precise and allows one to select a particular physiological trait (cf
36, 218, 219, 360). It is also useful for estimating realized heritabilities
and genetic correlations (132, 200, 201, 261, 278, 279, 297, 323, 353).
Laboratory natural selection, on the other hand, may allow clearer insight
as to what might occur in nature; only the environment is specified, and
the adaptive solution is left up to the organism (36). In theory, either
protocol can yield multiple solutions; in practice, the number of ways in
which a selective problem will be solved in a particular organism is an
empirical issue that has been little studied.

Selection experiments will reveal traits that evolve as correlated responses,
thus indicating the interdependence of aspects of the phenotype (51, 278,
279). For example, artificial selection on maximal sprint running speed
might divulge that leg length evolved as a correlated response. Such a result
would suggest that leg length was causally related to speed (6). Our
laboratory is currently conducting artificial selection for voluntary wheel-
running behavior in mice and will monitor correlated changes in physiolog-
ical and hormonal traits.

Mechanistic inferences derived from correlated responses to selection can
be greatly strengthened by subsequently doing the converse experiment; for
example, selecting directly on leg length to see if sprint speed evolves as
a correlated response. Experiments of this nature have been done on aging
and its correlates in D. melanogaster (179, 358, 360–362, 384), and this
genetic model species has served as subject for a number of other long-term
selection experiments involving physiological characteristics (128, for 250–
300 generations: some bacterial selection experiments have exceeded 2,000
generations, 269). In these selection experiments, replicate control and
selected lines are required (i.e. at least two of each) in order to make
inferences about correlated responses (194, 279, 362). In general, the design
of selection experiments is complicated (see references cited herein).

Some physiological traits are too difficult to measure on hundreds of
individual organisms each generation. Some measurements may not be
sufficiently reproducible to allow effective artificial selection. Others require
sacrifice of the organism (e.g. heart size, although even this might be
accomplished nondestructively with ultrasonic imaging techniques). For traits
that require destructive sampling (e.g. brain mass: references in 261),
artificial selection is still possible, for example, through the use of sibling
selection (see 132, 297). But many physiological measurements can be automated (215), or are relatively simple; mice, for example, have been successfully selected for hematocrit (378) and for thyroid function (74).

An interesting question for physiologists is whether selection yields repeatable results at the level of physiological mechanism. If selection to increase some organismal trait is imposed on several different replicate lines drawn from the same homogeneous base population, does the trait increase via the same physiological or morphological mechanism? For example, would all lines of mice selected for higher sprint speed respond with an increase in leg length, or an increase in the percentage of fast twitch muscle fibers, or an increase in muscle mass? Alternatively, would different solutions appear in each line? The optimality perspective on physiological evolution might suggest a single solution (but see 408), as appears to have occurred in early comparisons between stocks of *D. melanogaster* selected for delayed reproduction (276, 357). Most geneticists and evolutionary biologists, on the other hand, would not be surprised to see multiple solutions and unpredictable responses (82, 90, 91, 179, 203, 204, 269, 299, 319, 362); in the jargon of quantitative genetics, the answer will depend on whether genetic correlations between the trait being selected and other traits remain the same in all replicate selection lines. If several lower-level traits change in response to organismal selection, then another intriguing question is whether they all change in parallel (as symmorphosis would suggest: cf 134), or one at a time (suggesting a sequential series of limiting factors). A third question of interest is whether evolution follows the principle of "last hired, first fired." That is, if one selects for improved performance, then relaxes or reverses selection, do the mechanistic components decrease in the same order as they increased?

Evidence from *D. melanogaster* can address the foregoing questions. Service et al (385) utilized reverse selection to examine the nature of genetic and phenotypic correlations in stocks produced by selection for delayed reproduction. In reverse-selected lines (selected for early reproduction from delayed regime), longevity fell while early reproduction increased; starvation resistance also fell, while ethanol and desiccation tolerance remained unaltered during the first 20 generations of reverse selection. Graves et al (179) reexamined these same stocks after 100 generations of reversed selection, and found that starvation, desiccation, and ethanol resistance had dropped further than at generation 20. Leroi et al (270) have now found a shifting of the nature of the original genetic correlations uncovered in the Rose postponed-aging stocks [e.g. (357); early fecundity B > O]. After ten years of laboratory evolution, the pattern of early fecundity observed in the standard assay environment now favors the O lines. The B lines still preserve their early fecundity advantage in the B culture regime, which is slightly
different from that used to maintain the O stocks. In addition, in the standard environment, the development time of the delayed reproduction vs early reproduced control group now favors the early reproduced line, such that the early fitness trade-off now resides in the development time component as opposed to early fecundity. These results seem to indicate that although life history trade-offs may be inevitable, the nature of the genetic correlations that control them are plastic, such that selection may have more leeway than we imagine to create solutions to adaptive problems.

In most laboratory experiments, selection operates on preexisting genetic variation. However, for those experiments (e.g. involving microorganisms) that extend hundreds or thousands of generations and/or involve very large population sizes, new mutations can also be important. Changes in the frequencies of preexisting genes can lead to changes in genetic correlations (and heritabilities), but where new mutations are possible, changes in genetic correlations are particularly likely.

BIOCHEMICAL AND PHYSIOLOGICAL STUDIES OF ALLOZYME VARIATION

"Metabolic control theory, including both experimental and theoretical extensions, provides a 'glue' to hold physiology and genetics together." (86, p. 193)

Following the advent of protein electrophoresis in the mid-1960's, numerous studies demonstrated correlations between genotype or allele frequencies (actually, protein phenotypes representing alternative alleles) and environmental or ecological factors, such as habitat temperature, seasonality, latitude, or altitude (271). Many workers interpreted such correlations indicating the action of natural selection. These interpretations were criticized for being based on correlational data and for not assigning a more important role to random mutation and genetic drift (169, 175, 271). Several research groups, therefore, developed a biochemical and physiological approach to studying the evolutionary significance of genetic variation at specific loci (89, 137, 247–250, 331, 334, 439, 441, 443, 444). Examples include work by Powers and colleagues on killifish lactate dehydrogenase (Ldh) (98, 111, 112, 322, 332–334); Watt and colleagues on sulfur butterfly phosphoglucone isomerase (Pgi) (439, 440, 445–447); Koehn and colleagues on mussel leucine aminopeptidase (Lap) (29, 199, 248, 251–254); van Delden and many others on Drosophila alcohol dehydrogenase (Adh) (2, 8, 48, 73, 80, 268, 315, 429); Hartl, Dykhuizen, and Dean on several loci in E. coli (103, 104, 121, 123, 124, 185, 186); Burton and colleagues on copepod glutamate-pyruvate transaminase (Gpt) (63, 64); Hoffman on sea anemone Pgi (206–208); and Snyder and colleagues on Peromyscus hemoglobins (75, 76, ...
EVOLUTIONARY PHYSIOLOGY 593

Many other studies have correlated multi-locus heterozygosity with measures of physiological performance in an attempt to explain levels of genetic variation, without considering functional differences among allozymes; these have been reviewed elsewhere (169). Here we attempt to clarify several issues concerning the rationale, assumptions, and empirical measures involved in functional studies of variation at single loci.

For genotypic variation to affect fitness, two requirements must be met. First, the allozymes produced by the different genotypes must exhibit functional differences; second, the functional differences among allozymes must cause biochemical or physiological differences that are detectable at the organismal level (Figure 1). Thus a convincing research program must

![Diagram](image)

**Figure 1** The centrality of organismal performance paradigm, much of which was developed in relation to studies of locomotor performance (see 9, 39, 154, 166, 210, 220, 431). Both genetic and environmental effects act through development and ontogeny to determine an organism’s primary phenotypic characteristics, which often are categorized as biochemical, physiological or morphological. Acting in concert, these traits determine whole-organism performance abilities, such as maximal sprint running speed or stamina (153, 166, 231) or perhaps ability to raise offspring (110). In practice, behavior, e.g. motivation, can also affect measurements of performance, which are typically obtained in the laboratory by forcing the organism to perform at its maximum capacity for the trait of interest (6, 34, 137, 231). Performance defines the extent or limits of an organism’s capabilities, whereas behavior indicates how an organism actually uses (or fails to use) these capacities. Selection acts most directly on behavior, but behavior is limited by performance. Thus, genotypic or biochemical variation (e.g. allozyme variants identified by protein electrophoresis) should only be subject to selection if they have effects at the level of organismal performance and hence behavior (89, 248). It is absolutely critical that the appropriate ecological context of the organism be considered when determining a performance to measure in the laboratory; the performance must relate to behavior in the field to have any relevance to natural selection (9, 14, 34, 35, 113, 114, 137, 153, 154, 158, 166, 220, 227, 439, 442). Natural selection is defined operationally as a correlation between fitness and phenotype. Dashed arrows indicate the possibility of direct environmental effects on performance (e.g. the effects of substrate on sprinting ability) or behavior (e.g. temperature-dependent switches in antipredator behavior) (references in 58, 154, 166). The inseparability of physiology, behavior, and environment has long been a central tenant of physiological ecology (1, 24, 25, 52, 78, 81, 205, 209, 210, 212, 223, 241, 258, 284, 291, 313, 327, 329, 330, 380, 382, 421, 422, 432, 435).
first measure the appropriate characteristics of enzyme function and then relate the observed variation to whole-organism performance and/or behavior.

**Functional Differences Among Allozymes**

Three measures of enzyme function are common, each of which can be influenced by structural differences among allozymes. The Michaelis constant, $K_m$, is the substrate concentration that yields a reaction velocity equal to one half of the maximum reaction velocity, $V_{\text{max}}$. $K_m$ is generally considered to be a measure of substrate binding affinity (202, but see 140). The catalytic rate constant, $K_{\text{cat}}$, is a measure of the amount of product produced per active site on an enzyme per unit time. Thus it is a measure of the speed with which an enzyme functions, standardized to the number of active sites contained by the enzyme. The third measure, $V_{\text{max}}$, is the product of $K_{\text{cat}}$ and enzyme concentration, $[E]$, and is the maximum reaction velocity at saturating levels of substrate. Values of these enzymatic parameters are specific to the reaction conditions used; thus in vitro reaction conditions must be chosen that match or approximate the ecological context indicated by field studies. For example, it was vital to Powers’ work on Ldh in killifish (see references above) that the biochemical studies be conducted at the temperatures suggested by his field studies: as it turned out, no differences in Ldh function existed at temperatures convenient for biochemical work (25°C), but significant differences existed at temperatures relevant to the natural environment of killifish (5 and 30°C)!

For selection to act on allozymes, they must differ in at least one of the foregoing functional parameters. For a given set of reaction conditions, differences among allozymes in $K_m$ or in $K_{\text{cat}}$ can only be caused by differences in structure; therefore, if selection acts on either one of these functional properties, it also acts on the locus that produces the allozymes. However, two different factors can influence $V_{\text{max}}$.

First, $V_{\text{max}}$ can be affected by differences in allozyme structure. As noted above, allozyme differences in $K_{\text{cat}}$ are caused by differences in allozyme structure; any such changes in $K_{\text{cat}}$ would result in a change in $V_{\text{max}}$ (332). $V_{\text{max}}$ can also differ among allozymes because of allozyme-specific differences in $[E]$ caused by differential stability of the allozymes [an enzyme’s stability is a function of its structure (202)]; any such changes in $[E]$ would also result in a change in $V_{\text{max}}$ (440). In both instances, selection acting on $V_{\text{max}}$ would be acting directly on the locus of interest.

Second, $V_{\text{max}}$ can be affected by changes in $[E]$ caused by genetic factors unrelated to allozyme structure. Allozyme-specific differences in $[E]$ can be the result of a control locus that differentially affects the allozymes. Laurie et al (268) demonstrated that some activity differences among Adh allozymes in *Drosophila* are caused not by the Adh alleles themselves, but by linked
variants of controlling regions of the chromosome. In this example, selection on $V_{\text{max}}$ might actually be acting on the controlling gene(s) rather than on the Adh locus.

Despite such potential complexities in interpreting the origins of variation in $V_{\text{max}}$, it is a most useful measure of enzymatic activity. First, $V_{\text{max}}$, as compared to $K_{\text{cat}}$, is the better measure of enzymatic activity with regard to the whole-organism phenotype simply because it measures maximal activity, not just substrate turnover/active site. (The converse of this is also true; $K_{\text{cat}}$ is a better measure of activity at the level of the enzyme molecule itself because it does not include [E].) Second, the ratio of $V_{\text{max}}$ to $K_m$ is a good approximation of enzyme velocity at low substrate concentrations, which are usually the physiologically-relevant conditions (440). In fact, $V_{\text{max}}/K_m$ ratios have typically been used as the measure of enzymatic effectiveness (332, 440). Finally, questions about the origins of variation in $V_{\text{max}}$ can be addressed by measuring $K_{\text{cat}}$ (440); in this way, differences in $V_{\text{max}}$ among allozymes can be ascribed to differences in either $K_{\text{cat}}$ or [E]. Determination of $K_{\text{cat}}$ requires a completely purified enzyme, which is a non-trivial task. In any case, $V_{\text{max}}$ is an enzymatic measure of potentially great functional and hence evolutionary significance. $V_{\text{max}}$ has routinely been measured by physiological ecologists (and by exercise physiologists) as a simple indicator of biochemical functional capacities (83, 84, 87, 88, 148, 158, 160, 171, 184, 202, 228, 257, 291, 293, 329, 330, 339, 467).

Organismal Effects of Differences in Allozyme Function

For selection to “see” differences in functional characteristics of allozymes, they must cause variation at the level of the whole organism (Figure 1). In other words, allozymes that differ in function must also cause differences in the rate of flux, the efficiency of flux, and/or amounts of a given substrate (86, 250). Such differences in metabolic pathway characteristics can directly or indirectly influence fitness through their impact on energy supply, availability, and/or use (86, 442, 443).

The traditional view of metabolic pathways suggests that all control of flux through a pathway resides only with rate limiting regulatory enzymes (18, 19). However, quantitative theories of metabolic control, in which flux control potentially resides at all steps in a pathway, have been developed by Kacser (metabolic control analysis: 233–235; see also 125, 192), Savageau (biochemical systems theory: 372–374), Crabtree (flux oriented theory: 96, 97), and their colleagues. Each of these theories has its own array of assumptions and characteristics, and excellent reviews are available elsewhere (94, 138).

In metabolic control theory, control of flux through a pathway can be shared by all enzymes in the pathway; control coefficients can be calculated
for each enzymatic step in a pathway for any given set of reaction conditions. The control coefficient of an enzyme actually measures the sensitivity of the flux through the pathway to any changes in the functional capabilities of that enzyme, and is inversely proportional to the $V_{\text{max}}/K_m$ ratio. Furthermore, control coefficients of all enzymes in a pathway are interrelated, so that changing the control coefficient of one enzyme will change the control coefficients of one, some, or all of the other enzymes in the pathway. But, a change in the functional characteristics of an enzyme may or may not result in changes in its own control coefficient; this is a question that must be answered empirically for each enzyme in a pathway and for every set of reaction conditions of interest.

Much resistance to metabolic control theory has come from those unwilling to discard the traditional idea of metabolic control by one or a few key regulatory enzymes (18, 19). It is important to realize that metabolic control theory does not necessarily preclude the traditional view of rate limiting regulatory enzymes (198). Metabolic control theory does, however, provide a methodology by which control of flux can be empirically measured at the different steps in a metabolic pathway; the traditional view of metabolic control is but one possible evolutionary outcome. That metabolic control theory is tenable and useful is shown by the fact that some of its parameters have been empirically estimated for a variety of loci in several pathways in diverse taxa (44, 103, 168, 191, 246, 398, 466) and that it has been used successfully to develop quantitative genetic analyses of metabolic pathways (85, 104, 238, 413, 438) (see also 110 on a possible link to symmorphosis).

Several different, but not exclusive, scenarios describe how selection might affect flux-dependent measures of organismal performance. Selection might simply affect the rate of flux through a pathway, and so select for allozymes that either maximize flux or do not limit it. Empirical measurements in various systems have shown allozyme-dependent rates of flux (103, 122, 446). Selection might also affect the different impact of allozymes on the efficiency of flux (86, 250); in this untested scenario, the selective advantage of a high rate of flux in a pathway is tempered by the energetic cost of the maintenance of enzyme pools used in that pathway. Finally, selection can affect the differential impact that allozymes can have on pools of substrates in a pathway; deleterious effects of enzyme deficiency diseases are usually caused by substrate accumulation (86, 460).

MEASURING SELECTIVE IMPORTANCE IN THE FIELD

"Natural selection acts on phenotypes, regardless of their genetic basis, and produces immediate phenotypic effects within a generation that can be measured
without recourse to principles of heredity or evolution. In contrast, evolutionary response to selection, the genetic change that occurs from one generation to the next, does depend on genetic variation... Upon making this critical distinc-
tion...precise methods can be formulated for the measurement of phenotypic natural selection.” (262, p. 1210)

Natural selection has been defined in various and sometimes overly complex ways (129). The simplest and operationally most useful definition of natural selection is variation in Darwinian fitness that is correlated with variation in one or more phenotypic traits. This definition emphasizes that natural selection is a purely phenotypic phenomenon that occurs and can be measured within generations (58, 99, 166, 177, 262, 309, 344, 379, 394, 430). Moreover, it emphasizes that selection acts on phenotypic variation, without regard to its genetic basis, and thus can be futile in the sense of leading to no improvement in a population. The realization that repeatable, individual variation is the most fundamental requirement for natural selection to occur has stimulated many recent studies of the magnitude and correlates of individual variation in physiological, performance, and behavioral traits (see references in Introduction).

Quantifying selection in nature requires measurement of individual differences in fitness, e.g. lifetime reproductive success, and in some trait of interest, e.g. standard metabolic rate. A correlation between fitness and the phenotypic trait equals selection. Because true fitness is exceedingly difficult to measure, such components of fitness as survivorship or clutch size are usually measured as a substitute (129, 262, 345). Incomplete measures of fitness will limit inferences that can be drawn, but are an important first step. To date, only a handful of studies have specifically addressed whether natural selection acts on individual variation in physiological traits in natural populations. For example, Jayne & Bennett (227) demonstrated a correlation between survivorship and speed or stamina in garter snakes (see also 56, 229, 232, 352: reviews in 35, 39, 58, 166).

The foregoing approach to quantifying selection in nature is a “black box” in the following sense. Into the black box goes a known number of individuals with a known distribution of phenotypes (e.g. sprint speeds of hatchling lizards) and out comes a smaller number of individuals with a possibly altered distribution. The alteration of the distribution is attributable to the effects of natural selection, assuming no differential immigration or emigration with respect to the phenotypic trait being studied, no ontogenetic changes in the phenotype, etc (129, 262). The nature of the selective agent(s) is, however, unknowable from such information. For example, if faster lizards survived longer, it would not be known whether this was because (a) they were better able to escape from predators, (b) they were better able to catch insects and hence less likely to starve to death, or (c) maximal
sprint speed was phenotypically correlated with some other trait (stamina?) that was the actual target of selection. Thus correlational studies of selection in nature are an important first step, but they are incomplete with respect to understanding the causes of selection. Elucidation of the causes of selection requires additional information, such as direct observations of animals in nature (e.g. observations of predator-prey interactions) (39, 153, 166, 180, 227, 329: see also 24, 25). Once the mechanism of selection is understood, more concrete interpretations can be made about the original field data suggesting selection on genotype-related behaviors (Figure 1). Furthermore, predictions about other selective effects can be made and tested (in the case of allozyme variation, see 440, 444, 447).

Correlational studies of selection in the wild (129) can be enhanced by experimental manipulations (166, 309). Because selection in nature will often be weak, extremely large sample sizes can be required to detect its action. This is a problem of statistical power. A standard way to increase power in correlation or regression analyses is to increase the range of variation in the independent variable (e.g. maximal sprint speed). Sinervo and colleagues (393–396) have used this approach, via experimental manipulation of eggs and dams, to study selection on offspring size and clutch size in lizards. Artificial selection experiments could also be used to extend variation beyond the natural range (318), and crossbreeding or direct genetic manipulation might be used to alter expression of allozymes, followed by release of manipulated individuals into natural populations. Finally, field transplants (241) can be performed to determine the relative fitness of varying phenotypes under different environmental conditions.

INTERSPECIFIC COMPARATIVE METHODS

"...we must learn to treat comparative data with the same respect as we would treat experimental results..." (298, p. vii)

Interspecific comparisons are a long-standing tradition in physiology (6, 20, 126, 172, 338, 380, 382, 383, 464). For example, broad surveys of data compiled primarily from the literature, then plotted on log-log axes vs body mass, have provided a plethora of descriptive and predictive interspecific allometric equations (see references in What is the Origin of Allometric Relationships?).

At the opposite extreme in terms of sample size, two species, differing in behavior or ecology often have been chosen and compared to determine whether they show phenotypic differences that could be interpreted as adaptations to the presumably different selective regimes (26, 275) imposed
by the differences in behavior or ecology. Garland & Adolph (156) have argued that extreme caution must be used when attempting to infer adaptation from two-species comparative studies. Any two species are likely to show differences in almost any phenotypic trait that one might choose to measure. These differences are almost guaranteed by random mutation and genetic drift acting independently in the lineages leading to the two study species, unless counteracted by uniform stabilizing selection. Thus given a sufficient sample size of individuals from each of the two study species, a statistically significant difference will probably be found; that is, the null hypothesis of no difference in physiology will likely be rejected. About 50% of the time the difference will, by chance, be in the same direction as that predicted by the alternative hypothesis (adaptation to the environmental factor). Thus the chance of rejecting the null hypothesis of no adaptive differences in physiology can be as high as 50%; $\alpha$, the Type I error rate, may be closer to 0.50 than to the nominal 0.05!

Garland & Adolph (156) argue that, at a minimum, three species are required for a comparative study that aims to make inferences about adaptation, and the more species the better. But multi-species data sets bring with them numerous statistical complications. In brief, species (and sometimes populations within a species) related by a hierarchical phylogeny cannot be assumed to represent statistically independent data points (Figure 2). Species inherit both their genome and their environment (unless a dispersal event or rapid climatic change has occurred) from their immediate ancestor. Closely related species (i.e. species that diverged relatively recently) will therefore tend to be quite similar with respect to most aspects of their genotype, environment, and phenotype (139, 176, 187).

The most obvious problem with species non-independence is that it lowers the degrees of freedom available for hypothesis testing. For example, suppose one wished to correlate blood hemoglobin level with altitude for a series of three species. Assume that species mean values were available for both hemoglobin level and altitude (perhaps we would be dealing with species that presently exist as only single populations, each with a narrow altitudinal distribution). The sample size is thus three species’ means, and the null hypothesis would be no correlation between hemoglobin level and altitude. The 1-tailed alternative hypothesis would be a positive correlation between hemoglobin level and altitude. Because altitude is the independent variable (the presumed selective regime), we could compute a least-squares linear regression of hemoglobin level on altitude. Assuming we judge statistical significance at an a priori $\alpha = 0.05$, then with one degree of freedom, the critical value for the $t$ statistic is 6.314 (from Table 12 of 356) or 39.9 for the equivalent $F$ statistic (from Table 16 of 356). (In terms of hypothesis testing, we could equivalently look up the critical value for the Pearson
What Conventional Statistics Assumes

What Evolution Provides

Figure 2  Diagrammatic representation of the statistical problems caused by the hierarchical nature of evolutionary relationships and descent with modification. Typically, the field of statistics assumes that data points are independent, as would be the case if we studied 10 species that were related as shown on the left; here, instantaneous speciation resulted in 10 independent lineages that led to 10 living species that might be studied by a physiologist. Thus if we were to test for a correlation between species mean values for two phenotypes (e.g. size-corrected heart mass and maximal oxygen consumption) of these 10 species, or perhaps between one phenotype and an environmental factor (e.g. blood hemoglobin concentration and altitude), we could claim the nominal N-2 = 8 degrees of freedom for hypothesis testing. If, instead, the 10 species were actually related as shown on the right, we would have something fewer than 8 d.f. available for hypothesis testing. Although no simple correction factor for degrees of freedom is available, various methods exist that explicitly use the phylogenetic topology and branch lengths to allow valid hypothesis testing (see text).

product-moment correlation coefficient with 1 d.f., which is ±0.988 for the 1-tailed test) (from Table B.16 of 468).

If our three study species were the result of one ancestral lineage splitting simultaneously into three daughters, the foregoing procedure would be perfectly acceptable. If, on the other hand, two of our species were very close relatives, then we would have something fewer than three independent data points and hence something less than one degree of freedom. In the limit, if two of our species had diverged only yesterday, then we would have only two independent data points and hence no degrees of freedom for hypothesis testing! In effect, this brings us back to a two-species comparison and illustrates another perspective on why two-species comparisons are inadequate for inferring adaptation-inadequate d.f. (156).

Interspecific Comparisons in a Phylogenetic Context

"If we assume that the...cladogram...is correct, we can then hypothesize what the particular common ancestor must have been like." (20, p. 14)

Incorporation of a phylogenetic perspective into comparative studies is essential from a statistical perspective (see above, below, and Figure 2) and
moreover allows one to address questions that simply cannot be considered in the absence of phylogenetic information (59, 187, 211, 213, 265–267, 275, 302). For example, if one has data for the mean phenotypes (character states) of a series of species and some estimate of their evolutionary relationships, then one can use a parsimony algorithm (59, 142, 213, 274, 285–287, 296, 409, 411, 458) to infer the likely phenotype of ancestors, that is, nodes on the phylogenetic tree (78, 119, 213, 465). Thus parsimony reconstructions allow one to infer where in a clade a particular feature arose (35, 62, 69, 71, 72, 107, 130, 294, 365, 366, 370), if it has arisen more than once and, if so, the minimum number of times it has arisen (49, 105, 130, 264, 301, 349, 387), such as how many times air breathing evolved in fishes (62), toe fringes evolved in lizards (277), or the ability to produce benzaldehyde arose and/or was lost in tiger beetles (7). Once nodal values have been estimated, the inferred changes that have occurred along each branch segment of the phylogenetic tree can be computed, thus allowing inferences about the directions of past evolutionary change, tests for correlations in the changes of two or more characters (59, 78, 164, 165, 187, 213, 274, 277, 335), elucidation of the sequence of changes that occurred during the evolution of a complex trait (49, 134, 239, 265–267, 301), and tests for whether the presence of a particular state in one character or environmental feature predisposes some other trait to change in a particular direction (7, 187, 285, 287). If associations between characters and environmental factors are established (see 162, pp. 29, 30), then inferences about adaptation are possible (11, 26, 50, 59, 156, 175, 182, 187, 239, 265–267, 274, 275, 277, 302, 307, 345, 349, 350, 415, 416). If independent information on divergence times is available, then rates of evolution can be studied (152, 210, 465: of 408).

**Phylogenetically-Based Statistical Methods**

The foregoing uses of phylogenetic information to study variation among species are not statistical in any formal sense; that is, P-values or confidence intervals are not being assigned to the estimates of ancestral states, inferred changes, or correlations of inferred changes across traits. But phylogenetic analyses of character evolution can also be explicitly statistical with formal estimation and/or hypothesis testing. Since 1985 a number of phylogenetically-based statistical methods have been proposed. Of the available alternatives, Felsenstein’s (139) method of phylogenetically independent contrasts is the best understood and is applicable to a wide range of questions, including correlation, principal components analysis, regression, multiple regression, ANOVA, and ANCOVA (152, 159, 162, 164, 176, 295, 296, 341). This method was designed for use with traits exhibiting continuous variation, such as body size or metabolic rate (for applications with phys-
iological traits see 27, 88, 156, 164, 188, 213, 218, 274, 336, 386, 406, 437). The simplest use of phylogenetically independent contrasts is to study correlated evolution, such as the allometry of some trait in relation to body size (165). Both PC-based (159, 296) and Macintosh-based (162, 340) computer programs to conduct independent contrasts analyses and various other comparative methods are available.

In many cases, statistical analyses done by a method that allows for phylogenetic non-independence will indicate that relationships between variables (as judged by correlation or regression: 164, 187, 188), or differences among groups [as judged by analysis of variance (ANOVA) or covariance (ANCOVA): 159], indicate weaker and hence less significant relationships. Such is not always the case, however (27, 156, 274, 348, 437).

Can one predict whether a phylogenetic statistical analysis will yield an answer that is different from a conventional, non-phylogenetic method? In general, if the phenotypic data being analyzed (e.g. basal metabolic rate, sprint speed) follow the phylogeny—if species strongly resemble their close relatives for the traits being tested for a correlation—then a statistical method that allows for phylogenetic relationships will indicate a weaker relationship than one that assumes all species to be related as by a star (left in Figure 2). In other words, if it is phylogenetic resemblance of species values that is driving an apparently significant correlation between traits, then a phylogenetically based statistical method, such as independent contrasts, will indicate a weaker and less statistically significant relationship.

Other phylogenetically based statistical methods, including squared-change parsimony and some techniques for discrete traits, are discussed elsewhere (152, 170, 187, 213, 274, 275, 282, 285-287, 296, 307, 335). Estimates of phylogenetic relationships are becoming more widely available (131, 389, 412). Most of the methods can deal with unresolved nodes in phylogenies (see 159, 170, 176, 187, 341).

INTEGRATING MICRO- AND MACROEVOLUTIONARY APPROACHES

To understand the hows and whys of evolutionary change at the phenotypic level (e.g. physiological traits), multiple valid approaches exist that can converge on the same endpoint. Because microevolutionary (within-species) phenomena can be studied experimentally, as through artificial selection experiments, physiologists may find them particularly attractive. But motivation can come from either direction. The senior author, for example, undertook quantitative genetic analyses of basal metabolic rates of mice in hopes of better understanding the (in)famous mouse-elephant curve (cf 261). Similarly, Bennett’s bacterial selection experiments (37, 40, 41) and Huey’s
Drosophila selection experiments (215, 218, 219) were preceded by studies of interspecific variation in thermal physiology of lizards (31, 164, 195, 210, 213: cf 39, 213, 214, 218), which are not ideally suited to artificial selection. Finally, microevolutionary analyses of the correlated evolution of snake color patterns and antipredator behaviors, including locomotor abilities, were motivated by interspecific patterns (see 55–58).

Comparative approaches focus on the endpoints of evolutionary processes. Because of the non-independence of species values caused by hierarchical descent with modification, statistical methods that allow for phylogenetic relatedness are required to determine whether the (co)variation observed among species represents more of a pattern than could have arisen simply by chance processes, such as random mutation and genetic drift. If an appropriate statistical method confirms that the observed pattern is really unusual, then—and only then—do the data call for an explanation, such as adaptation by natural selection.

Given that a statistically significant pattern is observed for among-species variation (e.g. group differences) or covariation (e.g. correlations between two physiological traits), then at least three processes (mechanisms) might account for it: (a) selection acting within species; (b) genetic couplings of characters; and (c) higher-level phenomena such as species selection or lineage sorting (references in 77b, 120, 134, 144, 166, 305, 350, 418, 459). Mechanisms (a) and (b) are familiar to physiologists, but (c) covers phenomena that are less well understood. In some cases biogeographic, paleoclimatological, and/or fossil information can be used to help construct scenarios for the evolution of physiological (or other) traits (78, 107, 225, 284, 302, 314, 342, 367, 369).

Some organisms are particularly suitable both for comparative phylogenetic analyses [including comparisons of natural populations, subspecies, or laboratory strains with known relationships (16, 141, 155, 162)] and for quantitative genetic analyses, including artificial selection experiments [Drosophila (82–85, 87, 88, 128, 179, 184, 203–205, 214, 215, 218, 219, 255, 317, 355, 358–362, 384, 420, 448); and Mus (16, 17, 22, 23, 45–47, 141, 143, 190, 230, 260, 278–281, 300, 324, 325, 410)]. Whatever the choice as to organism, analytical mode or physiological system, a plurality of approaches will be necessary to understand any large question in evolutionary physiology (51, 146, 166, 218, 223, 275, 308, 318, 329, 418, 432, 433, 434, 465). Comparative studies, for example, indicate what did happen during evolution, but not necessarily what had to happen; similarly, “selection experiments indicate what might happen in nature, but not necessarily what will happen” (219, p. 755). Understanding the ultimate causes and proximate mechanisms of the evolution of endothermy is a good example of a long-standing problem in evolutionary physiology that calls for multi-
disciplinary approaches (B Block, this volume; 35, 42, 43, 50, 53, 101, 102, 130, 148, 160, 189, 221, 225, 303, 326, 368, 383, 406, 417, 437; references therein). For example, our laboratory is currently analyzing the genetic correlation between minimal and maximal rates of oxygen consumption and beginning selection experiments on voluntary activity levels in Mus (see also 143, 190, 347).

CONCLUSION AND FUTURE PROSPECTS

Evolution and physiology have much to offer each other (36, 61, 62, 71, 72, 108, 135, 155, 214, 339, 342, 343, 388). Knowledge of physiological mechanisms can allow much deeper insight into possible reasons for evolutionary correlations and constraints than is possible for many of the traits typically studied by evolutionary biologists (e.g. morphology). A comparative perspective can even enlighten biomedical and clinical issues (460). For example, Rose and colleagues have provided clear evidence that an evolutionary perspective can (or at least should!) alter accepted views on aging (179, 357–362, 384, 460). Similarly, Kluger's (242–245) studies of fever and White's (456, 457) comparative perspective on acid-base balance during hypothermia have affected the way physicians view and treat human patients. "Those who see the body as a machine designed by a careless engineer are prone to therapeutic hubris. The antidote is a deep understanding of each organ's phylogeny and functions, as well as its ontogeny and structure." (460, p. 18)

We see evolutionary physiology moving forward on many fronts during the next decade. Which of the several promising areas, such as phylogenetically-based comparative studies, artificial selection studies in the laboratory, or physiological analyses of single-gene products will yield the greatest insights is difficult to predict. Perhaps the most illuminating studies will be those that apply several complementary approaches (35, 36, 39, 43, 51, 88, 155, 166, 214, 218, 302, 308, 345, 423, 432) to an ecologically and phylogenetically well-known group of species that is tractable for physiological studies. Such studies will not be easy, quick, or inexpensive, but they may yield understanding that is greater than any equivalent series of piecemeal studies done on several different species.

The tools now exist to permit comprehensive studies of physiological evolution. Such studies will be greatly facilitated by interactions of physiologists with biochemists, morphologists, ethologists, ecologists, geneticists, and systematists. We envision studies in which knowledge of biochemistry, physiology, biomechanics, and/or developmental biology is first used to predict trade-offs between various physiological functions. These hypothesized constraints are then tested in at least two ways, by quantification of
EVOLUTIONARY PHYSIOLOGY

33. Bennett AF. 1987. Inter-individual variability: an underutilized resource. See Ref. 135, pp 147–69
40. Bennett AF, Lenski RE. 1993. Evolutionary adaptation to temperature II. Thermal niches of experimental lines of *Escherichia coli*. *Evolution* 47:1–12
ness measurements and predictions under conditions with no alcohol stress. Genetics 120:743–53


74. Chai CK. 1970. Selective breeding for
1131 I thyroid uptake in mice. Genetics 64:29–40
deer mice (Peromyscus manicolor): physiology of beta-globin variants and
76. Chappell MA, Snyder LRG. 1984. Biochemical and physiological corre-
lates of deer mouse alpha-chain hemog-
77a. Charlesworth B. 1990. Optimization
models, quantitative genetics, and mu-
tation. Evolution 44:520–38
of variation in energy storage in
56:49–56
evolutionary analysis of metabolism in
nine Drosophila species. Evolution. In
press
89. Clarke B. 1975. The contribution of
ecological genetics to evolutionary the-
ory: Detecting the direct effects of
atural selection on particular polymor-
phic loci. Genetics 79:101–13
90. Cohan FM, Hoffmann AA. 1989. Uni-
form selection as a diversifying force in
evolution: evidence from Drosoph-
91. Cohan FM, Hoffmann AA, Gayley
TW. 1989. A test of the role of
epistasis in divergence under uniform
selection. Evolution 43:766–74
92. Congdon JD, Gibbons JW. 1987. Mor-
phological constraint on egg size: A
challange to optimal egg size theory?
93. Cooper DM, Weller-Ravell D, Whipp
BJ, Wasserman K. 1984. Growth-re-
lated changes in oxygen uptake and
heart rate during progressive exercise in
94. Cornish-Bowden A, Cardenas ML, ed.
New York: Plenum
95. Cowles RB. 1958. The evolutionary
significance of the scrotum. Evolution
12:417–18
96. Crabtree B, Newsholme EA. 1983. A
quantitative approach to metabolic con-
systematic approach to describing and
analyzing metabolic control systems.
Trends Biol. Sci. 12:4–12
98. Crawford DL, Powers DA. 1989. Mo-
lecular basis of evolutionary adaptation
at the lactate dehydrogenase-B locus
in the fish Fundulius heteroclitus. Proc.
Natl. Acad. Sci. USA 86:9365–69
99. Crespi BJ. 1990. Measuring the effects
of natural selection on phenotypic in-
teraction systems. Am. Nat. 135:32–47
100. Crews D. 1992. Behavioural endocri-
nology and reproduction: an evolution-
ary perspective. In Oxford Reviews of
Reproductive Biology, ed. SR Milligan,
101. Crompton AW, Taylor CR, Jagger JA.
1978. Evolution of homeothermy in
in long-necked dinosaurs; did the
sauroptod have bird lungs? Comp.
Biochem. Physiol. 101A:43–46
104. Dean AM, Dykhuzien DE, Hartl DL. 1988. Theories of metabolic control in quantitative genetics. See Ref. 453, pp. 536–48
106. de Queiroz A, Wimberger PH. 1993. The usefulness of behavior for phylogeny estimation: levels of homoplasy in behavioral and morphological characters. Evolution 47:46–60
134. Feder ME. 1987. The analysis of physiological diversity: the prospects for pattern documentation and general questions in ecological physiology. See Ref. 135, pp. 38–70


164. Garland T Jr, Huey RB, Bennett AF.
scription and experimental studies of clinal variation in birds. Am. Zool. 31:694–705


244. Knowles RG, Pogson CI, Salter M. 1990. Application of control analysis to the study of amino acid metabolism. See Ref. 94, pp. 377–84


250. Koehn RK, Zera AJ, Hall JG. 1983. Enzyme polymorphism and natural se-


EVOLUTIONARY PHYSIOLOGY 615


314. Norris KS. 1958. The evolution and


376. Schall JJ, Bennett AF, Putnam RW. 1982. Lizards infected with malaria:
spring sprint speed in the philopatric sex. *Am. Nat.* In press


430. Wake DB, Roth G, eds. 1989. *Complex Organismal Functions: Integration and
Evolution in Vertebrates. Chichester UK: Wiley & Sons


Watt WB, Carter PA, Donohue K. 1986. Females' choice of "good genotypes" as mates is promoted by an insect mating system. Science 233:1187–90


Wilson MA, Gatten RE Jr, Greenberg N. 1989. Glycolysis in Anolis caro-


