Chapter 9

Phylogenetic comparison and artificial selection

Two approaches in evolutionary physiology

Theodore Garland, Jr.
Department of Biology, University of California, Riverside, CA, USA

Abstract: Interspecific comparison has a long and productive history in physiology. Conceptual and statistical advances over the last 15 years have demonstrated several ways in which comparisons can be enhanced by consideration of phylogenetic information, i.e., empirical estimates of the ways in which organisms are related (evolutionary trees). Choice of species to be compared should be informed by phylogenetic information. For example, a comparison of three species that inhabit high altitude with three that live at low altitude would be suspect if each of the two groups were composed of closely-related species (e.g., within single genera). To avoid such "phylogenetic pseudoreplication," one might instead study species from three different genera, each containing one high-altitude and one low-altitude inhabitant. Unfortunately, many studies have not been so carefully designed, sometimes because organisms were not accessible or because the studies incorporated data from the literature. Fortunately, several new statistical methods correct for problems caused by phylogenetic relatedness and descent with modification, the most common being phylogenetically independent contrasts. Another tool that can be used in comparative physiology is selective breeding, which has been practiced for millennia and applied in scientific contexts for over a century. In the last 20 years, ecological and evolutionary physiologists have begun using selection experiments to study processes of genetic adaptation in physiological and behavioral traits. For example, house mice have been maintained in the cold for multiple generations to see what adaptations may occur naturally in response to reduced ambient temperature ("laboratory natural selection"). Our own laboratory has used selective breeding to create four replicate lines of mice that exhibit high levels of voluntary wheel-running behavior, as well as various morphological and physiological characteristics that cause or allow the elevated locomotor activity. Similar experiments could be used to study adaptation to hypoxia.

Hypoxia: From Genes to the Bedside, edited by R.C. Roach *et al.*
Kluwer Academic/Plenum Publishers, New York, 2001.

Key words: adaptation; exercise; genotype-environment interaction; quantitative genetics; voluntary activity

INTRODUCTION

Evolutionary physiology is a subdiscipline that has developed since the late 1970s (12, 21, 29, 30, 38). It grew partly from the realization that many studies in ecological, environmental, and comparative physiology were somewhat naive with respect to the state-of-the-art in modern evolutionary biology. For example, although physiologists have often been interested in adaptation in the genetic/evolutionary sense (i.e., cross-generational genetic changes that occur as a result of natural selection [9]), many of them have made inferences about adaptation based on studies and data that would be considered insufficient for the purpose by evolutionary biologists (e.g., see 17, 25, 33, 34, 35, 37).

Another impetus for the development of evolutionary physiology was the idea that evolutionary studies could in many cases be strengthened by a more rigorous consideration of morphological, physiological, and biochemical mechanisms that account for variation at the organismal level (i.e., how organisms work). For many evolutionary biologists, the organism is left as something of a "black box." Peering inside the box has often been seen as unnecessary for understanding evolutionary (or ecological) phenomena. An expression sometimes heard is: "We shall assume that the organism works!" It is not that evolutionary biologists disdain mechanism; indeed, detailed and highly technically sophisticated studies of genetic mechanisms of evolutionary change are common. Moreover, studies in evolutionary morphology and biomechanics often examine mechanism in considerable detail (53, 59, 65, 70, 87, 90). Nevertheless, it remains true that evolutionary biologists relatively rarely study physiological mechanism (e.g., see 94).

Evolutionary physiologists also recognize that not all organisms are well suited for evolutionary studies, just as not all organisms are well suited for studies of physiological mechanisms (30, 38). For example, a long generation time places great constraints on the kinds of evolutionary studies that can be undertaken (e.g., nobody is going to do a selective breeding experiment with elephants), just as a small body size places constraints on the types of physiological studies that can be done (e.g., because of limits on the amount of blood that can be drawn or on the size of telemeter that can be carried during normal behavior). Hence, many evolutionary physiologists who once worked exclusively on wild animals have subsequently taken to studying laboratory "model organisms" (e.g., A. F. Bennett's studies of bacteria [10]; M. E. Feder's [28, 31] and R. B. Huey's [e.g., 46] studies of *Drosophila*; A. J. Zera's work on crickets [95]; my own laboratory's work on

domesticated house mice [e.g., 19, 26, 44, 47, 51, 56, 57, 75, 83, 84, 85, 86; and see below]) .

Although a shift to the use of model systems may not seem radical or like much of a sea change to either evolutionary biologists (witness the routine use of *Drosophila melanogaster*) or physiologists (routine use of rats, mice, and frogs), it can be almost anathema to some comparative and ecological physiologists. In the latter fields, one tradition has been to choose particular species for study because they are of inherent interest, often because they are "extreme" in some way (see 29, 37, 38). Examples would include organisms that live in environments with extreme physical conditions (deserts, polar regions, high altitude, great depth in the ocean), that are unusually small (shrew, hummingbird) or large (elephant, ostrich), that have peculiar behavior (e.g., feeding on blood, Galapagos marine iguanas that dive and feed on seaweed), or that have an unusual body shape (giraffe, snakes in general). general). A bias often exists against anything that is not seen as a "real organism." Our own laboratory has sometimes encountered this when trying to publish papers on domesticated house mice. Reviewers have, for example, wondered why we did not study a wild rodent, such as *Peromyscus*. Even when the topics are explicitly physiological or evolutionary, rather than ecological, we have encountered such comments as "so what to people who are interested in 'real' organisms?"

In any case, many evolutionary physiologists use species that, while unusual from the perspective of traditional ecological physiology, are routine for other fields. These model organisms can offer many advantages, such as a wealth of background information, the availability of molecular tools that have not yet been developed for wild organisms, and the ability to rear them in the laboratory as well as conduct cross-generational breeding studies (e.g., 54, 71, 82).

In this chapter, I present an overview of two general approaches that are having major impacts on the field of evolutionary physiology: phylogenetic comparison and selective breeding. Although the origins of these approaches are ancient, and they have been applied previously in physiology (38), applications in modern evolutionary physiology can probably be traced, respectively, to Huey and Bennett's (52) studies of the evolution of thermal physiology in Australian scincid lizards, and to Lynch's selective breeding on nest-building behavior in house mice (58, 61). Several other approaches are common in today's evolutionary physiology (30, 38), including formal and informal optimality models (e.g., 3, 22, 91, 92), but they will not be considered here.

PHYLOGENETIC COMPARISONS

Comparisons of species have been a mainstay of comparative physiology since its inception. They have allowed us to catalog the diversity of physiological processes exhibited by living organisms and also to discover general principles of organismal "design." For example, comparisons of animals of different body size have shown that most physiological rate processes vary in a predictable manner with body size, e.g., larger-bodied species generally have lower heart rates, lower rates of respiration, lower metabolic rates on a mass-specific basis, and longer life spans (13, 18, 81).

But the allometric generalities that have emerged from comparing species of different sizes do not apply in a completely general manner. That is, a single allometric equation does not fit all animals. Rather, we often see different relationships when comparing different evolutionary lineages (clades) of animals, such as squamates (lizards and snakes, which represent an evolutionary derivation from lizards) versus mammals versus birds. Sometimes the scaling exponents (i.e., the slope of a line fitted to double-log transformed data) of these relationships differ significantly, but more often we see shifts in the elevations of the lines. For example, squamates have lower standard metabolic rates and also lower maximal rates of oxygen consumption during forced exercise, even when measured at body temperatures (35-40 Celsius) that approximate those of mammals (88, 93). Differences in field metabolic rates, measured by doubly labeled water, are even greater, as they are also affected by variations in body temperatures and activity levels on both a daily and seasonal basis (69).

Lineage-specific allometric relationships that show similar slopes but different elevations (Y-intercepts) are referred to as "grade shifts," in evolutionary biology. Apparent grade shifts in physiological functions have long been known to comparative physiologists. Some relatively recent and widely cited examples include the putatively higher standard metabolic rates of passerine birds as compared with other birds (but see 41, 74), and the lower metabolic rates of marsupial as compared with placental (eutherian) mammals (18, 81, 93).

Knowing of the existence of clear differences among some evolutionary lineages, most modern comparative biologists would consider phylogeny as a possibly important factor when comparing species. Although they often do not think of such distinctions as passerine versus non-passerine birds, or marsupial versus placental mammals, as "phylogeny," these taxonomic categories nonetheless convey something about evolutionary relationships. Passerines are a particular lineage derived from within the avian family tree; marsupials and placentals are generally thought to represent sister lineages (at least if we ignore fossil groups).

1.1 An Example to Illustrate the Importance of Considering Phylogeny

At this point, it is useful to consider a real data set, one which illustrates the perils of ignoring phylogenetic relationships during data analysis. Figure 1A shows the log-log relationship between red blood cell count and

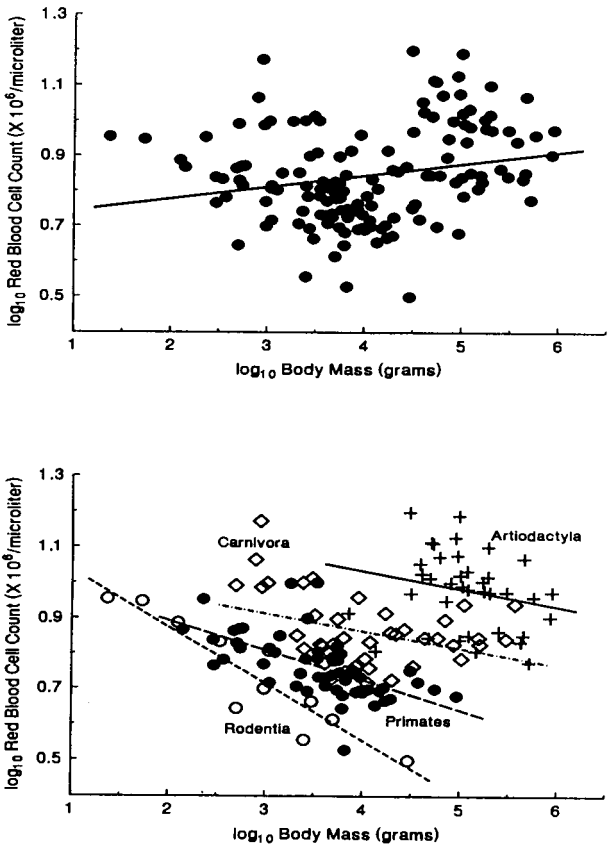

Figure 1. Example of how incorporating phylogenetic information can clarify patterns. Across species of mammals, the number of red blood cells (RBC) per unit volume of blood seems to increase with increasing body size (A). When phylogeny is considered, however (B), negative relationships are apparent within each of four clades, and clades tend to differ in average values. Data from Promislow (72 and pers. comm.).

mass for 146 species of mammals. These data are a subset of those used by Promislow (72). He used a data base compiled from animals living in the London Zoo, so most of the animals had been reared for some period of time under similar environmental conditions. As noted below, this cannot be strictly true for the range of species considered (e.g., mice, monkeys, wolves) because they eat such different diets and have such different housing requirements. Nevertheless, environmental conditions were presumably more similar across species than if each had been captured fresh from the field.

A conventional least-squares linear regression analysis, as is typically used in allometric studies, indicates a statistically significant (2-tailed $P = 0.0057$) positive relationship with a slope of 0.033 (S.E. = 0.012, 95% confidence interval = 0.010 - 0.057). Thus, larger-bodied species tend to have more red blood cells per unit volume of blood. This analysis makes absolutely no reference to the phylogenetic relationships of the 146 species under consideration.

However, as shown in Figure 1B, if we separate the 146 species into their taxonomic orders and perform a conventional analysis of covariance (ANCOVA), we obtain a negative pooled within-groups slope (2-tailed $P < 0.00005$) of -0.077 (S.E. = 0.012, 95% confidence interval = -0.100 - -0.054) and highly significant differences among the orders ($P < 0.00005$). Hence, our conclusions regarding the relationship between body size and the number of red blood cells per unit volume of blood changes when we consider at least a crude representation of phylogenetic relationships, i.e., separating species into their taxonomic orders.

A cautionary note must be added. Many taxonomies do not actually reflect phylogenetic relationships. In some cases, this is because names were assigned before any real knowledge of branching relationships was available. In other cases, names are retained because they have a long historical tradition and appear to convey useful common knowledge. The class Reptilia is a good example. As used traditionally, Reptilia includes living crocodilians, turtles, tuatara, and lizards (plus snakes and amphisbaenians, both of which appear to be evolutionary derivations from within the lizard tree), but excludes birds and mammals, both of which may have derived from within the basal group that gave rise to our "classic" reptiles (plus dinosaurs, pterosaurs, and various other extinct groups). The classes Aves and Mammalia are not sister lineages with the traditional class Reptilia. In this case, taxonomy would cause one to draw an incorrect phylogenetic tree.

Another example comes from the class Aves. Physiologists routinely compare the order Passeriformes with all other non-passerine birds (18, 81), as if the latter were the sister taxon representing one other order of birds. In reality, however, passerines are an evolutionary derivation from within other

birds, and those other birds comprise at least 22 recognized taxonomic orders (see 41, 74). Again, taxonomy misleads by implying that 23 or so orders of birds diverged simultaneously at some time in the distant past, rather than evolving by a sequence of bifurcations such that the orders themselves are hierarchically related.

Because comparative biologists have long appreciated that phylogenetic lineages may show general differences in various physiological and other traits, most of them would probably have done the sort of analysis illustrated in Figure 1B, or perhaps even further separated species into families. Nevertheless, recent conceptual and statistical developments in evolutionary biology have led to dramatic changes in the state-of-the-art with respect to choosing species for comparison, analyzing data from multi-species comparisons (including allometric relationships), and interpreting the results of statistical analyses.

The Statistical Perspective On Why Phylogeny Matters

At this point, it is necessary to take a more explicitly statistical perspective on the consideration of phylogenetic information when analyzing comparative data. First, we can ask what, if anything, a conventional regression analysis assumes about the phylogenetic relationships of the species in the data set? One might imagine that is assumes absolutely nothing, given that it makes no mention of phylogeny. In fact, however, a conventional regression analysis explicitly assumes that the species share no phylogenetic history, that they have descended from one "big-bang" speciation event at some point in the distant past. The diagrammatic representation of this assumption is referred to as a "star" phylogeny, as shown in Figure 2A. The star phylogeny is the phylogenetic translation of what conventional regression analysis assumes, as discussed in any statistics text: the data points represent a random sample from a homogeneous population, and the residuals (vertical deviations from the regression line) are independent and identically distributed.

Is it reasonable to make these assumptions when analyzing comparative data? In most cases, no. For example, when we separate the blood cell count data into their respective orders (which correspond, in this case, to separate evolutionary lineages or clades), we see a clear tendency for species to resemble other species within their own order. Moreover, only four mammalian orders are represented (e.g., no rabbits, bats or whales are included). Thus, the 146 species cannot be considered to represent a random sample of all mammals.

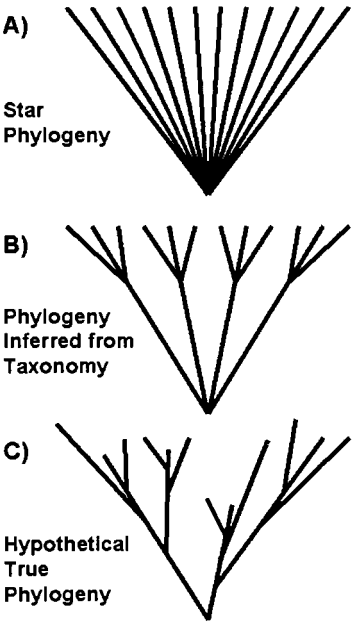

Figure 2. (A) Illustration of what conventional statistical analyses assume ("star" phylogeny) when applied to comparative data. (B) Moving beyond this, conventional statistical comparisons of taxa assume, in effect, that the taxa being compared are an unrelated series of "mini-stars" with no hierarchical structure within any taxon. assumes that the taxa actually represent separate evolutionary lineages [monophyletic groups or clades, in the language of evolutionary biology].) (C) Real phylogenies indicate hierarchical relationships and branches that do not necessarily line up along tips of the tree. Non-contemporaneous tips indicate that the rate of evolution has among branches. Real phylogenies like this cause various statistical problems, phylogenetically based statistical methods are required to analyze comparative data. Note that the horizontal axis in such diagrams is arbitrary and does not convey any information about, for example, degree of genetic or phenotypic differentiation among the species.

What are the consequences of ignoring phylogenetic relationships when analyzing comparative data? They are several and problematic, as has been demonstrated by a large number of both theoretical and empirical studies (e.g., 1, 2, 23, 32, 40, 41, 67, 73, 74, 89). First, Type I error rates will be inflated--significance will be claimed too frequently. Second, estimates of parameters, such as the slopes of scaling relationships, will be inaccurate. Third, statistical power will be affected.

The take-home message is that not attempting to incorporate phylogenetic information into analyses of comparative data is simply unacceptable. Although this message has not yet made it into some of the more mechanistically oriented fields, such as parts of comparative physiology, it is true nonetheless. I predict that in the future the use of phylogenetically based statistical methods will be as routine as the use of

statistics in general. Unlike the situation a century ago, few if any physiological studies of any kind can today be published without some use of statistics. Eventually, few if any multi-species comparative studies will be published without use of phylogenetic information. By analogy, few if any comparative studies are published today without due consideration of the possible confounding effects of body size. Today's comparative physiologists would never compare a shrew and an elephant without somehow controlling for effects of body size; in 50 years, they will never make such comparisons without attempting to control for phylogenetic effects.

The Importance of Common Rearing Conditions

Physiologists study phenotypes, which are the result of genetic effects, environmental effects, and their interactions during development and ontogeny. When species are compared, the usual presumption is that phenotypic differences among them reflect genetic differences, not just differences in the environmental conditions that they (or their mothers) have experienced. To ensure that species differences in physiology actually do reflect genetic differences, all species to be compared must be raised under common conditions (14, 15, 36, 37, 38). Unfortunately, it is typically difficult if not impossible to do so. For example, although it may be possible to capture adults (or juveniles) and to keep them under common conditions for some period of time (weeks?) prior to physiological measurement, it may not be possible to breed the animals and measure their offspring. If not, then maternal effects cannot be ruled out. Moreover, if a broad diversity of species is to be compared (e.g., including shrews and elephants), then it will simply be impossible to impose identical conditions for acclimation, let alone rearing (e.g., shrews eat animals whereas elephants eat plants).

The general rationale for including phylogenetic relationships in analyses of cross-species data sets is that genotypes and phenotypes are inherited from common ancestors. In addition, species may inherit environmental conditions (habitats and geographic localities) from their ancestors. Therefore, related species will tend to resemble each other, e.g., birds look like birds and bats look like bats. If common rearing conditions are not applied, then observed phenotypic differences among species may represent environmental effects rather than genetic differences. This would seem to weaken the specific justifications for overlaying the phenotypic data on a phylogenetic tree for analysis, but actually it would weaken any attempt to infer genetic (evolutionary) adaptation from the comparative data set. Lack of common-garden rearing conditions does not negate the importance of considering phylogenetic information, especially given that environmental

conditions (e.g., via behavioral habitat selection) may also tend to be inherited phylogenetically; rather, it weakens very generally the adaptive inferences than can be drawn from a data set. Hence, the results of any comparative study (whether of species populations) should be viewed with caution, unless it has implemented thorough common-garden controls (see also 14, 15, 36, 37, 38).

How Do We Account For Phylogenetic Relationships?

Given that ignoring phylogenetic relationships violates various assumptions of conventional statistical methods, what can we do about it? Performing an ANCOVA by order, as described above for the data shown in Figure 1, is a step in the right direction. It at least acknowledges the possibilities that orders may differ and that the within-order relationship of red blood cell count to body mass may be different (Figure 1B) from the one that appears when we ignore order as a factor (Figure 1A).

But the ANCOVA by orders is only partially phylogenetic. The analysis still assumes that each order contains no hierarchical relationships, i.e., that each can be represented by a star phylogeny, and these four stars are themselves connected as a star to the base (root) of the tree (Figure 2B). To take the analysis further and make it fully phylogenetic, we would need to specify the detailed hierarchical relationships of all 146 species.

Once we have specified the phylogeny (topology and branch lengths) of the species under study, we can employ statistical methods that use the phylogenetic information. Three main methods are available for incorporating full phylogenetic information into comparative analyses: phylogenetically independent contrasts, generalized least-squares, and Monte Carlo computer simulations. Recent work has shown that they are functionally equivalent, although quite distinct in implementation (see 41). A full description of these methods is beyond the scope of this chapter, but various reviews, with worked examples, are available (e.g., 1, 37, 40, 43, 73). In addition, free computer software is available from several sources (see Joe Felsenstein's website for a rather complete listing: http://evolution.genetics.washington.edu/phylip/software.html). For example, my colleagues and I have developed the Phenotypic Diversity Analysis Programs (available directly from me: see 40, 41, 43) as well as a new package named PHYLOGR (available at http://cran.r-project.org/).

Phylogenetic trees are estimates of (hypotheses about) the true, but unknown, evolutionary relationships of organisms. Hence, the results from any phylogenetic analysis of comparative data may be subject to modification if future information alters the arrangement of species (and possibly the branch lengths [e.g., estimated divergence times], although a

number of studies have shown that errors in branch lengths do not necessarily have fatal consequences [2, 23, 24]).

Choosing Species for Comparison

A long-standing tradition in comparative physiology (and in many other fields) has been to compare only two species, e.g., one from high altitude and the other from low, the latter offering a basis for comparison (a type of "control"). As discussed at length elsewhere, comparisons of only two species are generally inadequate for making inferences about genetic adaptation or even about mechanism (14, 15, 37, 38). However, as noted previously (37), one way to enhance the value of two-species comparisons would be to examine the results of several, perhaps completed by different workers. Indeed, as noted by Hopkins and Powell (50), comparisons of several pairs of "normal" with hypoxia-tolerant species (Canada with Bar-headed geese, Sherpa with lowland humans, mole rats with Norway rats, Green Sea with Loggerhead turtles, Mudskippers with Lungfish) all show higher P_{50} in the former of the pair. As these species pairs are only distantly related, they constitute essentially independent replicate comparisons. Thus, some sort of formal meta-analysis could be conducted to test the hypothesis that hypoxia tolerant vertebrates in general have high P_{50}. The literature probably contains data that would allow many other such paired-comparison tests (see also 32, 37).

A more common situation is that the species available for comparison are somewhat arbitrarily scattered across the phylogenetic tree. This is the typical outcome when species are chosen based on, for example, convenience, perhaps with the addition of data available in the literature. The point of this section is that the choice of particular species to be compared, whether intentional or unintentional, can have profound effects on both the utility and validity of a study. More specifically, I will argue that a phylogenetic perspective is a powerful and necessary part of choosing species for a comparative study.

Obviously, proper choice of species depends on the question asked, but determining what constitutes the best choice is actually a complicated and multi-faceted problem (e.g., 2; see also 42 pp. 437-445). For example, if one were interested in cataloging the diversity of metabolic rates across all mammals, then one would want to sample from all (or at least most) major lineages. If body size affects the trait of interest, then selecting species as different in size as possible ("mouse to elephant") is obviously an important consideration as well. Of course, actually sampling many groups may be logistically difficult as not all groups will be available either locally or through animal suppliers. Moreover, if a broad diversity of species is

studied, then it will be exceedingly difficult if not impossible to implement common rearing conditions, or even common acclimation conditions, prior to making measurements.

To study adaptation in the genetic, evolutionary sense, one first identifies an independent variable of interest, such as temperature or oxygen concentration. In general, to enhance statistical power, the species to be compared should represent a broad range of this independent variable. For example, we might study adaptation to hypoxia with a set of mammalian species (or populations), each of which occupies a fairly restricted elevational range along the continuum from sea level to very high altitude. We might then sample randomly from all terrestrial mammals. Alternatively, we might choose a particular lineage, such as rodents or primates, and sample only within that. For a given affordable sample size (number of species), the latter strategy may avoid complications caused by comparing distant relatives, which may be analogous to apples and oranges (or "chalk and cheese" in the United Kingdom)(37). In other words, distant relatives are likely to differ in many traits, not just those related to the independent variable of interest (e.g., altitude), such that a comparison of distant relatives is like an experiment with multiple uncontrolled variables.

Even within a given lineage, such as rodents, it is often possible to choose species such as to allow a sort of paired-comparisons analysis, as depicted in Figure 3C. When analyzed with phylogenetically based statistical methods, this sampling scheme actually affords higher statistical power than would be offered by the sampling scheme shown in Figure 3A (e.g., picking very distantly related species). The worst possible sampling scheme would be to choose species such that half occupied one end of the environmental continuum and the other half occupied the other end, but each of these sets of species were close relatives. When analyzed with phylogenetic statistics, this design yields low statistical power (see 40, 89).

In summary, given that one employs phylogenetically informed statistical analyses, which are necessary to guard against inflated Type I error rates and inaccurate estimates of statistical parameters (such as allometric slopes), statistical power can be greatly affected by the phylogenetic positions of the species in the analysis (89). On the positive side, choosing species so as to include multiple instances of high-low differences between close relatives (Figure 3C) can yield statistical power that is higher than that obtained by random sampling of species. Similarly, if one wishes to compare a particular species of interest (e.g., the bar-headed goose, the giraffe) with an allometric standard, statistical power can be enhanced by the use of phylogenetic methods that allow one to specify the position of the species on the phylogenetic tree (41). Thus, phylogeny can be your friend. On the negative side, however, choosing species that are phylogenetically clumped with

respect to the independent variable of interest (Fig. 3B) can lead to greatly reduced statistical power.

Recent computer-simulation studies provide additional guidance on choosing species (2). For example, the comparative physiology tradition of choosing animals that are known to be either extreme in one or more phenotypic characteristics, or that live in extreme habitats, can enhance statistical power, so long as they are not so unusual as to become a statistical outlier (perhaps because of other adaptations that we did not consider), and so long as we know where they fit on the phylogenetic tree (37).

In practice, relatively few multi-species studies in comparative physiology or physiological ecology have been designed from the ground up. Instead, organisms are included partly because they should be on statistical and/or phylogenetic grounds, partly because they are of particular (and perhaps irrational!) interest on the part of the investigator, and partly just because they were available in that part of the world (sometimes from the local pet shop). As well, many studies include data that has already been published in the literature, often by other researchers. The resulting mix of species, if analyzed in toto and by conventional statistical methods, may tend to mislead the investigator. Fortunately, if we can specify with some degree of accuracy the phylogenetic relationships of even the motleyest crew of species, then modern phylogenetically based analytical methods have great potential to "rescue" an analysis from phylogenetic obfuscation.

Finally, I would like to dispel what seems a common misunderstanding. Many biologists seem to think that if they employ phylogenetically based statistical methods most of their hard-won statistically significant results will go away. This also seems to be one reason that some workers have been resistant to their use. (Other reasons include just not wanting to be bothered with phylogeny, which is not legitimate, and lacking phylogenetic information, which is legitimate but can sometimes be overcome [e.g., see 11, 23, 24, 39].)

Biologists fear what seems to be a reduction in statistical power. However, that apparent reduction (when phylogenetically based methods are

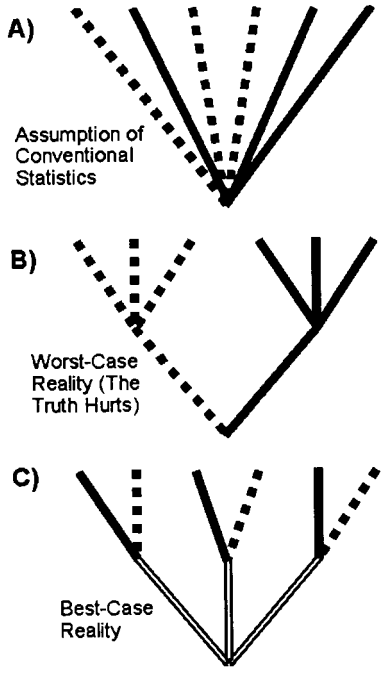

Figure 3. Three of the many possible evolutionary relationships of six species (or of six populations within a given species, e.g., Homo sapiens), three of which occur in one habitat type and three of which inhabit another. For example, the dashed branches might represent species inhabiting low altitudes (LA), whereas the solid branches represent species inhabiting high altitudes (HA). A goal of a comparative study might be to compare one or more phenotypes (e.g., body size, blood hemoglobin concentration) between the LA and HA species. (A) If the six data points were to be analyzed with conventional statistics, then the implicit assumption would be that none of the species share any evolutionary history beyond what they all share (i.e., since they diverged from a common ancestor). This can be termed a "star" phylogeny. (B) If the investigator were to determine the phylogenetic relationships of the six species, then they might turn out to represent just two distinct lineages, one containing all three of the LA species and the other containing the HA species. This situation might be termed "phylogenetic pseudoreplication." Instead of six independent data points, something closer to two are available. If the data were analyzed with phylogenetically correct ("PC") statistical methods, then Type I error rates could be protected to retain the a priori level for accepting statistical significance (e.g., $P < 0.05$), but the statistical power to detect a difference between the LA and HA species (if one existed) would be greatly reduced. (C) Alternatively, the evolutionary tree for the six species might be such that the closest included relative of each HA population was a different LA population. If the data were analyzed with phylogenetically based statistical methods, then not only would Type I error rates be protected, but in addition the statistical power to detect a difference between the LA and HA species would be higher than if analyzed with phylogenetically uninformed ("PU") statistics. If possible, comparative studies should be designed so that the independent variable of interest (e.g., altitude) maps onto the phylogenetic tree something like what is shown in (C), rather than what is shown in (B).

used) is illusory. Power of alternate statistical methods can only be compared when the Type I error rates of the methods are the same. Conventional statistical tests applied to phylogenetically structured data may yield seemingly higher power by allowing inflated Type I error rates. A scientist may be willing to accept statistical significance at an alpha level greater than the traditional 0.05, but he or she would want to do so intentionally and a priori, not because the species being studied were related in a hierarchical fashion. When Type I error rates are made comparable, power of phylogenetic methods can actually be higher (e.g., 67). In addition, the power of phylogenetically based statistical methods applied to phylogenetically structured data (and assuming that the topology, branch lengths, and model of character evolution are known) is actually identical to the power of conventional statistical tests applied to non-structured data (37).

Relevance to the Study of Human Altitude Adaptation

As discussed by Brutsaert (14, 15), considering the evolutionary relationships of human populations is important for understanding whether any of those native to high altitudes show evidence of genetic adaptation to high altitude. Fortunately, several human populations are native to high altitude, including some in Tibetan, South America, Ethiopia, Kenya, and Papua New Guinea. Moreover, these appear to represent at least three independent invasions of high altitude. Hence, it is possible to design a comparative study that resembles the situation depicted in Figure 3C, in which each high-altitude native population is contrasted with their close relatives that do not live at high altitude. As noted above, this sort of comparison should afford adequate statistical power for detecting altitude adaptations that actually exist, at least if the same trend is exhibited in each of the independent comparisons. Of course, the usual caveats about the need for common rearing conditions, etc., would apply, as has been discussed above and elsewhere (14, 15, 36, 37, 38).

SELECTION EXPERIMENTS

No matter how well designed, nor how interesting the study organisms, comparative studies have limitations. For example, environmental effects can be very difficult to control and may confound adaptive interpretations, especially if genotype-by-environment interactions occur. In addition, comparative studies provide data that are purely correlational. As well, they do not allow one to study evolution in action (in "real time"), nor easily to

determine the sequence of evolutionary changes that may occur during adaptation (e.g., whether behavior adapts more rapidly than underlying morphological and physiological traits [see 38, 45, 52, 55, 68]). Hence, a number of evolutionary physiologists have turned to selective breeding experiments (e.g., 10, 46, 95).

Selection experiments are one subset of approaches in the overall area of quantitative genetics (8, 14, 15, 27, 38, 45, 49, 64, 77, 78). In fact, they are the oldest "approach" in quantitative genetics in the sense that they have been occurring since humans first began developing agricultural practices, including the gradual process of domesticating various animals. Although currently less in vogue than some more recently developed types of genetic manipulations (e.g., knockouts, transgenesis), selection experiments offer a major advantage over these in being more representative of genetic changes that occur in nature in response to natural or sexual selection. In the wild, selection acts directly on complex phenotypes (e.g., life history traits, behavior), most of which are highly polygenic (affected by many genes, most of which probably have relatively small effects). Hence, allele frequencies at many loci may change in response to selection. In contrast, a typical transgenic experiment alters one or at most a few genetic loci. Accordingly, selection experiments still find an important place in the biomedical sciences (e.g., 20, 66).

At least two general kinds of selection experiments can be distinguished (38, 45, 79): artificial selection and laboratory natural selection. Traditional artificial selection involves barnyard or lab populations in which each individual in each generation is scored for some phenotypic trait or combination of traits. Some bottom or top proportion of individuals is then chosen to become the parents off the next generation. This is called truncation selection. One variation on this theme is taking at least one male and female from within each family, then allowing them to mate with other individuals in their line but outside of their own family. This is termed within-family selection, and it increases the effective population size, reduce the rate of inbreeding, and helps to eliminate maternal effects.

C. B. Lynch used within-family artificial selection to alter nesting behavior of laboratory house mice. She maintained a total of six lines, two bred randomly as controls, two selected for large nests, and two selected for small nests. Replication of experimental lines, and consistency of response, is crucial in order that ensuing differences can be attributed to the effects of selection rather than founder effects and/or random genetic drift, perhaps in combination with the occurrence of unique mutations. Lynch's overall goal has been to understand the evolution of thermoregulatory phenotypes (behavioral, morphological, and physiological), viewed as an integrated suite of interacting traits (an "adaptive syndrome"). Her selected lines have been

used in many subsequent studies, and informative parallels have been drawn with clinal variation in wild (introduced) populations of house mice in North America (16, 61, 62, 63).

In laboratory natural selection, freely breeding populations are exposed to altered husbandry conditions, which could favor altered demographic schedules, or to altered environmental conditions, such as different temperatures. An example of this approach is Barnett and Dickson's experiments in which wild house mice were captured and used to establish two breeding colonies, one housed at approximately room temperature and the other in the cold. They performed two such experiments, for 9-14 generations, once in Scotland (7: average room temperatures of 21 and -3 C) and then again in Australia (23 and +3 C) (4, 5). In both experiments, various changes were observed, at least some of which seemed to represent the evolution of genetic adaptations to the cold (6). The results were rather complicated, however, and in both cases only a single line was kept in either the "control" (room temperature) or "experimental" (cold) condition. This lack of replication makes it difficult if not impossible to determine with confidence whether apparently adaptive changes are really so (i.e., the result of an altered selective regime) or the result of random genetic drift.

Artificial Selection for High Wheel Running in Mice

Our laboratory has conducted an artificial selection experiment to increase levels of voluntary wheel-running behavior in mice, and we are monitoring correlated changes in other behavioral, morphological, physiological, and biochemical traits. The overall goal is to understand how increased activity evolves, at levels ranging from motivation to exercise physiology (56, 57, 51, 75, 83, 84, 85, 86)

The original progenitors were outbred, genetically variable laboratory house mice (*Mus domesticus*) of the Hsd:ICR strain, purchased from Harlan Sprague Dawley in 1993 (83). Genetic variation in the base population is similar to variation among individuals in wild populations of *Mus domesticus* (19, 76; and references therein). After two generations of random mating, mice were randomly paired and assigned to 8 closed lines (10 pairs in each). In each subsequent generation, when the offspring of these pairs were 6-8 weeks old, they were housed individually with access to a running wheel for 6 days and a computer recorded wheel revolutions in 1-min intervals (1.12-m circumference, attached to standard clear plastic housing cages via a stainless steel tube inserted into a hole in the wall of the cage). In 4 "selected" lines, the highest-running (quantified as total number of revolutions run on days 5 and 6 of the six day test) male and female from each family were chosen as breeders to propagate the lines to the next

generation. In the 4 "control" lines, a male and a female were randomly chosen from each family. Within all lines, the chosen breeders were randomly paired except that matings between siblings were disallowed.

The purpose of maintaining replicate selected and control lines is to account for random genetic changes, such as founder effects and drift, which can cause lines to diverge even in the absence of selection. Any particular genetic or phenotypic difference between a given selected line and a given random-bred control line may or may not be related to the phenotype that was actually under selection. Inferences about the causal factors underlying phenotypic changes in a selected line are greatly strengthened if replicate lines are maintained (48).

After 16 generations, revolutions/day had increased 2.7-fold (mainly by increased running speed) and reached an apparent selection limit (Figure 4 shows data through 24 generations). This limit appears to correspond to the maximal aerobic speed estimated in the base population (56), and neither maximal oxygen consumption, when measured a week prior to wheel testing, nor basal metabolic rate has responded to selection by generation 22 (unpublished results). (Maximal oxygen consumption may show differences between the selected and control lines at other ages and/or under different housing conditions (84).)

Both sexes, but especially, females, have primarily increased their average running speed rather than the amount of time spent running (Figure 5: see also 56, 75, 83). Various morph-physiological differences between selected and control lines exist, some of which may represent genetic adaptations for sustained exercise (44). For example, mice from selected lines have higher insulin-stimulated glucose uptake in some hindlimb muscles (26) and more symmetrical hindlimb bone lengths (T. Garland and P. A. Freeman, unpublished results). Selected-line mice are smaller in body mass (85) and have less body fat than controls, at least under some conditions (86). When housed with vs. without access to running wheels for 8 weeks, suborganismal training responses (e.g., increases in hematocrit, citrate synthase activity of hindlimb muscle) are often greater in mice from selected lines (genotype-by-environment interaction), presumably because they run more (51). Motivation is now under study, and pharmacological experiments suggest altered dopaminergic function in the brains of selected-line mice.

Aside from the above-mentioned consistent differences between the selected and control lines, the four replicate selected lines show statistically significant differences in a number of traits, including wheel running itself. Of particular interest, two of the four selected lines now contain a high frequency (approximately 50%) of individuals with small muscles, in which the gastrocnemius exhibits an almost 50% reduction in mass, along with an

approximate doubling of mass-specific oxidative capacity (44). Comparisons of parents and offspring suggest that this phenotype is inherited as an autosomal recessive allele (unpublished observations). Moreover, population-genetic model fitting (in collaboration with Martin Morgan and Patrick A. Carter) provides evidence that the allele must have been under positive selection in the two selected lines. (Presumably, the other two selected lines lost the allele, which was rare in the base population, by chance either at founding or shortly thereafter by genetic drift.) Our working hypothesis is that these "mighty mini-muscles" are adaptive for sustained, relatively high-speed running, perhaps because of shorter diffusion distances. In collaboration with Helga Guderley and Philippe Houle-Leroy, we are now testing this possibility.

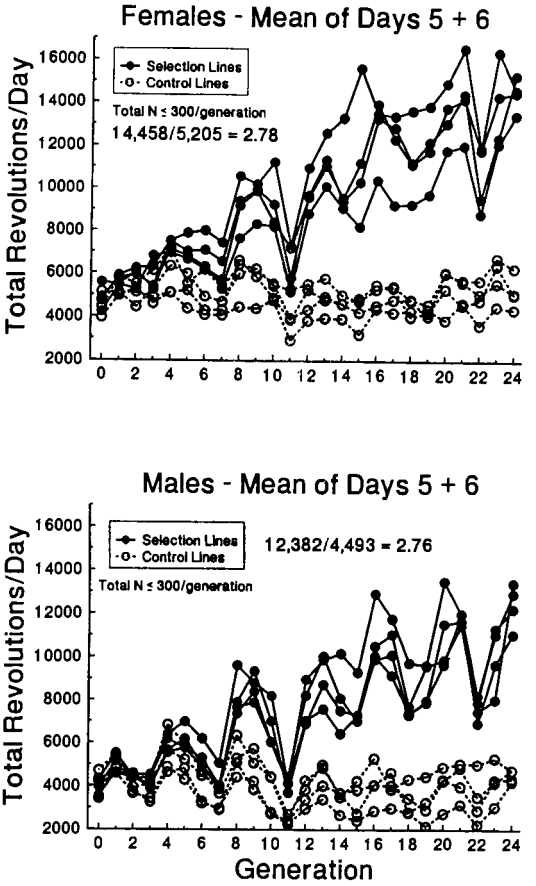

Figure 4. Wheel running (line means) of 8 lines of house mice either selected for high wheel running or bred randomly as controls. Dips in wheel running that seem to occur approximately every four generations (especially notable in males) correspond to summer generations, during which elevated humidity (and sometimes temperature) may cause reduced activity. Note that females always run more than males, but that the response to selection, relative to control lines, is similar in the two sexes.

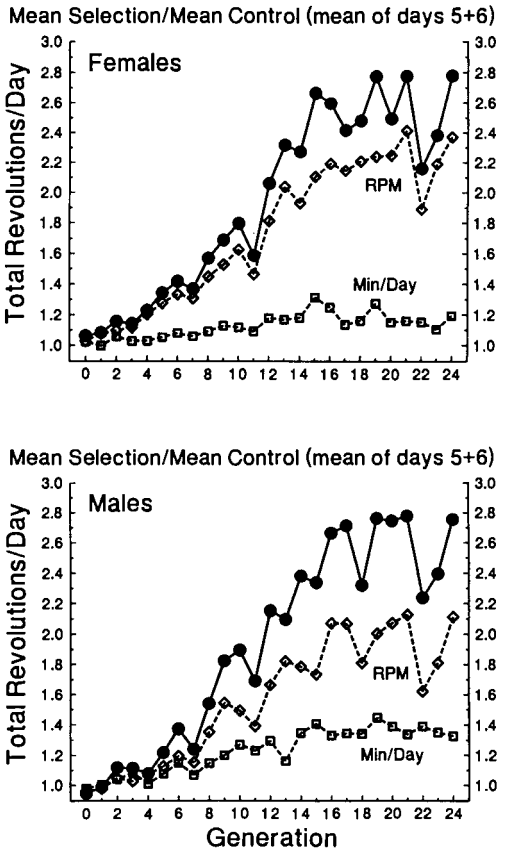

Figure 5. Ratio of mean wheel running for the selected compared to the control lines. Mice in the selected lines, especially for females, have accomplished more total revolutions (closed circles and solid line) per day mainly by increasing their average running speed (RPM, calculated as total revolutions divided by the number of 1-min intervals during which any revolutions were recorded), rather than the number of minutes spent running (Min/Day).

Prospects for the Use of Selection Experiments to Elucidate Adaptation to Hypoxia

Applications of selection experiments to the study of hypoxia adaptation are easy to envisage. For example, in the mode of laboratory natural selection, one might establish replicate lines of mice or rats at high altitude or in hypobaric chambers, while also maintaining control lines at sea level. The mice could be allowed to breed freely (within lines) for perhaps 10 generations (as in Barnett's studies of cold adaptation in mice: 4, 5, 6, 7). As generations passed, various other candidate phenotypes (traits thought to enhance physiological function under conditions of low partial pressure of

oxygen) could be monitored, such as hemoglobin levels, shape of the oxy-hemoglobin dissociation curve, lung capacity, pulmonary diffusing capacity, metabolic rate or activity levels.

Alternatively, one could implement selective breeding. Lines could be established in any convenient animal quarters. An expeditious test of whole-animal hypoxia tolerance could be devised, such as time to lose the righting response when the partial pressure of oxygen is lowered acutely. The most tolerant individuals would then be chosen as parents to produce the next generation. Again, candidate phenotypes would be monitored, i.e., subordinate traits hypothesized to facilitate hypoxia tolerance.

In either type of selection experiment, it would be predicted that one or more key phenotypes would change consistently in the selected lines (e.g., hemoglobin levels). Most likely, this would represent a trait that was somehow limiting to hypoxia tolerance, or at least to optimal function under hypoxia. If one or more traits did change consistently, then they would represent putative adaptations for hypoxia.

Of course, it is also possible that replicate lines will show a similar overall response, e.g., at the whole-animal level, but that the details of underlying adaptive mechanisms will differ, as appears to be the case in our replicate lines of house mice that have been selected for high wheel running. Interestingly, Tibetan and Andean human populations show what appear to be different adaptations to high altitude -- if, in fact, they are adaptations (see 14, and references therein)!

In any case, mechanisms can be complicated and slippery things. Therefore, one might subsequently propose to do the reciprocal experiment (see 38). That is, one could establish new lines, from the same original base population, and select directly on the putative adaptation. Thus, one might select directly for higher hemoglobin levels (mice have been successfully selected for hematocrit [80]). If hypoxia tolerance then increased across generations, then this would strengthen the original interpretation that elevated hemoglobin levels were an adaptation for hypoxia. Finally, as noted previously (38), selection experiments could be used to test predictions of symmorphosis (25, 35, 91, 92), i.e., that multiple components of a physiological pathway should be matched in capacity, and hence that they should all change approximately in parallel when selection is imposed at the level of organismal performance.

ACKNOWLEDGMENTS

Supported in part by U.S. N.S.F. grants IBN-9728434 (to T.G.) and DEB-9981967 (to T.G. and A. R. Ives). I thank R. B. Huey for many helpful

discussions over the years, and for allowing me to use material from one of our unpublished manuscripts. D. E. L. Promislow kindly provided me with the data used in his 1991 paper. I. A. Girard and S. P. Blomberg commented on the manuscript. I also thank S. R. Hopkins, F. L. Powell, and T. D. Brutsaert for sharing with me drafts of their manuscripts and/or presentations.

REFERENCES

1. Ackerly DD. Comparative plant ecology and the role of phylogenetic information. In: *Physiological plant ecology*, edited by Press MC, Scholes JD, and Braker MG. Oxford, UK: Blackwell Science, 1999, p. 391-413.
2. Ackerly DD. Taxon sampling, correlated evolution, and independent contrasts. *Evolution* 54: 1480-1492, 2000.
3. Alexander R McN. *Optima for animals*. (Revised ed.) Princeton: Princeton Univ. Press, 1996.
4. Barnett SA, and Dickson, RG. Changes among wild house mice (*Mus musculus*) bred for ten generations in a cold environment, and their evolutionary implications. *J Zool, Lond* 203: 163-180, 1984a.
5. Barnett SA, and Dickson, RG. Milk production and consumption and growth of young of wild mice after ten generations in a cold environment. *J Physiol* 346: 409-417, 1984b.
6. Barnett SA., and Dickson, RG. Wild mice in the cold: some findings on adaptation. *Biol Rev* 64: 317-340, 1989.
7. Barnett SA., Munro KMH, Smart JL, and Stoddart RC. House mice bred for many generations in two environments. *J Zool, Lond* 177: 153-169, 1975.
8. Bell, G. *Selection: the mechanism of evolution*. New York: Chapman & Hall, 1997.
9. Bennett AF. Adaptation and the evolution of physiological characters. In: *Handbook of physiology. Section 13: comparative physiology. Vol. I.* , edited by Dantzler WH. New York: Oxford Univ. Press, 1997, p. 3-16.
10. Bennett AF., and Lenski RE. Experimental evolution and its role in evolutionary physiology. *Am Zool* 39: 346-362, 1999.
11. Blomberg S. Fels-Rand: an Xlisp-Stat program for the comparative analysis of data under phylogenetic uncertainty. *Bioinformatics* 16: 1010-1013, 2000.
12. Bradley TJ, and Zamer W. Introduction to the Symposium: What is evolutionary physiology? *Am Zool* 39: 321-322, 1999.
13. Brown JH, and West GB, eds. *Scaling in biology*. New York: Oxford Univ. Press, 2000.
14. Brutsaert TD. Limits on inferring genetic adaptation to high altitude in Himalayan and Andean populations. *High Alt Med Biol* 2: in press, 2001.
15. Brutsaert TD. Genetic and environmental adaptation in high altitude natives: conceptual, methodological, and statistical concerns, 2001. This volume.
16. Bult A, and Lynch CB. Breaking through artificial selection limits of an adaptive behavior in mice and the consequences for correlated responses. *Behav Gen* 30: 193-206, 2000.
17. Burggren WW, and Bemis WE. Studying physiological evolution: paradigms and pitfalls. In: *Evolutionary innovations*, edited by Nitecki MH. Chicago: Univ. Chicago Press, 1990, p. 191-238.
18. Calder WA. *Size, function and life history*. Cambridge: Harvard Univ. Press, 1984.

19. Carter PA, Garland T Jr., Dohm MR, and Hayes JP. Genetic variation and correlations between genotype and locomotor physiology in outbred laboratory house mice (*Mus domesticus*). *Comp Biochem Physiol A* 123: 155-162, 1999.
20. Crabbe JC, Young ER, Deutsch CM, Tam BR, and Kosobud A. Mice genetically selected for differences in open-field activity after ethanol. *Pharmacol Biochem Behav* 27: 577-581, 1987.
21. Daniels CB, and Orgeig S. Proceedings of the satellite symposium for the Australian Physiological and Pharmacological Society: the evolution of physiological processes. *Clinical Exp Pharmacol Physiol* 25: 715, 1998. (and following papers)
22. Diamond JM. Evolutionary physiology. In: *The logic of life: the challenge of integrative physiology*, Boyd CAR, and Noble D. Oxford: Oxford Univ. Press, 1993, p. 89-111.
23. Díaz-Uriarte R, and Garland T Jr. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. *Syst Biol* 45: 27-47, 1996.
24. Díaz-Uriarte R, and Garland T Jr. Effects of branch length errors on the performance of phylogenetically independent contrasts. *Syst Biol* 47: 654-672, 1998.
25. Dudley R, and Gans C. A critique of symmorphosis and optimality models in physiology. *Physiol Zool* 64: 627-637, 1991.
26. Dumke CL, Swallow JG, Rhodes JS, Garland T, Maslowski E, Gazdag AC, and Cartee GD. Effects of genetic selection and voluntary wheel running on glucose transport in mice. *Med Sci Sports Exercise* 31(5 Supplement): S127, 1999. (Abstract)
27. Falconer DS, and Mackay TFC. *Introduction to quantitative genetics*. (4th. ed.) Essex, England: Longman, 1996.
28. Feder ME. Engineering candidate genes in studies of adaptation: the heat-shock protein Hsp70 in *Drosophila melanogaster*. *Am Nat* 154(Supplement): S55-S66, 1999.
29. Feder ME, Bennett AF, Burggren WW, and Huey RB, eds. *New directions in ecological physiology*. New York: Cambridge Univ. Press, 1987.
30. Feder ME, Bennett AF, and Huey RB. Evolutionary physiology. *Annu Rev Ecol Syst* 31: 315-341, 2000.
31. Feder ME, and Hofmann GE. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu Rev Physiol* 61: 243-282, 1999.
32. Felsenstein J. Phylogenies and the comparative method. *Am Nat* 125: 1-15, 1985.
33. Freeman S, and Herron JC. *Evolutionary analysis*. Upper Saddle River, New Jersey: Prentice Hall, 1998.
34. Futuyma DJ. *Evolutionary biology*. (3rd ed.) Sunderland, Mass.: Sinauer Associates, 1998.
35. Garland T Jr. Testing the predictions of symmorphosis: conceptual and methodological issues. In: *Principles of animal design: the optimization and symmorphosis debate*, edited by Weibel ER, Bolis L, and Taylor CR. New York: Cambridge Univ. Press, 1998, 40-47.
36. Garland T Jr., and Adolph SC. Physiological differentiation of vertebrate populations. *Annu Rev Ecol Syst* 22: 193-228, 1991.
37. Garland T Jr., and Adolph SC. Why not to do two-species comparative studies: limitations on inferring adaptation. *Physiol Zool* 67: 797-828, 1994.
38. Garland T Jr., and Carter PA. Evolutionary physiology. *Annu Rev Physiol* 56: 579-621, 1994.
39. Garland T Jr., and Díaz-Uriarte R. Polytomies and phylogenetically independent contrasts: an examination of the bounded degrees of freedom approach. *Syst Biol* 48: 547-558, 1999.

40. Garland T Jr., Dickerman AW, Janis CM, and Jones JA. Phylogenetic analysis of covariance by computer simulation. *Syst Biol* 42: 265-292, 1993.
41. Garland T Jr., and Ives AR. Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. *Am Nat* 155: 346-364, 2000.
42. Garland T Jr., Martin KLM, and Diaz-Uriarte R. Reconstructing ancestral trait values using squared-change parsimony: plasma osmolarity at the origin of amniotes. In: *Amniote origins: completing the transition to land*, edited by Sumida SS, and Martin KLM. San Diego: Academic Press, 1997, p. 425-501.
43. Garland, T Jr, Midford PE, and Ives AR. An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral values. *Am Zool* 39: 374-388, 1999.
44. Garland T Jr., Swallow JG, Girard I, Rhodes JS, Houle-Leroy P, Guderley H, Freeman PW, Dumke CL, Cartee GD, Koteja P, McAleer MW, Hosack GR, Belter JG, and Carter PA. Exercise adaptations in lines of house mice genetically selected for high voluntary wheel-running behavior. *The Physiologist* 43: 328, 2000. (Abstract)
45. Gibbs AG. Laboratory selection for the comparative physiologist. *J Exp Biol* 202: 2709-2718, 1999.
46. Gilchrist GW, Huey RB, and Partridge L. Thermal sensitivity of *Drosophila melanogaster*: evolutionary responses of adults and eggs to laboratory natural selection at different temperatures. *Physiol Zool* 70: 403-414, 1997.
47. Hayes JP, Garland T Jr., and Dohm MR. Metabolic rates and reproduction of *Mus*: are energetics and life history linked? *Funct Ecol* 6: 5-14, 1992.
48. Henderson ND. Spurious associations in unreplicated selected lines. *Behav Genet* 27: 145-154, 1997.
49. Hill WG, and Caballero A. Artificial selection experiments. *Annu Rev Ecol Syst* 23: 287-310, 1992.
50. Hopkins SR, and Powell FL. Common themes of adaptation to hypoxia. 2001. This volume.
51. Houle-Leroy P, Garland T Jr., Swallow JG, and Guderley H. Effects of voluntary activity and genetic selection on muscle metabolic capacities in house mice *Mus domesticus*. *J Appl Physiol* 89: 1608-1616, 2000.
52. Huey RB, and Bennett AF. Phylogenetic studies of coadaptation: preferred temperatures versus optimal performance temperatures of lizards. *Evolution* 41: 1098-1115, 1987.
53. Hutchinson JR, and Gatesy S. Adductors, abductors, and the evolution of archosaur locomotion. *Paleobiol* 26: 734-751, 2000.
54. Kellogg EA., and Shaffer HB. Model organisms in evolutionary studies. *Syst Biol* 42: 409-414, 1993.
55. Kohane MJ, and Parsons PA. Domestication: evolutionary changes under stress. *Evol Biol* 23: 31-48, 1988.
56. Koteja P, Swallow JG, Carter PA, and Garland T Jr. Energy cost of wheel running in house mice: implications for coadaptation of locomotion and energy budgets. *Physiol Biochem Zool* 72: 238-249, 1999.
57. Koteja P, Swallow JG, Carter PA, and Garland T Jr. Maximum cold-induced food consumption in mice selected for high locomotor activity: implications for the evolution of endotherm energy budgets. *J Exp Biol* 204: 1177-1190, 2001.
58. Lacy RC, Lynch CB, and Lynch GR. Developmental and adult acclimation effects of ambient temperature on temperature regulation of mice selected for high and low levels of nest-building. *J Comp Physiol* 123: 185-192, 1978.

59. Lauder GV. Function of the caudal fin during locomotion in fishes: kinematics, flow visualization, and evolutionary patterns. *Am Zool* 40: 101-122, 2000.
60. Lynch CB. Response to divergent selection for nesting behavior in *Mus musculus*. *Genetics* 96: 757-765, 1980.
61. Lynch CB. Genetic basis of cold adaptation in laboratory and wild mice, *Mus domesticus*. In: *Living in the cold: physiological and biochemical adaptations*, edited by Heller HC, Musacchia XJ, and Wang LCH. New York: Elsevier Science, 1986, p. 497-504.
62. Lynch CB. Clinal variation in cold adaptation in *Mus domesticus*: verification of predictions from laboratory populations. *Am Nat* 139: 1219-1236, 1992.
63. Lynch CB. Evolutionary inferences from genetic analyses of cold adaptation in laboratory and wild populations of the house mouse, *Mus domesticus*. In: *Quantitative genetic studies of behavioral evolution*, edited by Boake CRB. Chicago: Univ. Chicago Press, Chicago., 1994, p. 278-301.
64. Lynch M, and Walsh JB. 1998. *Genetics and analysis of quantitative traits*. Sunderland, Mass.: Sinauer Associates, 1998.
65. Marden JH., O'Donnell BC, Thomas MA, and Bye JY. Surface-skimming stoneflies and mayflies: The taxonomic and mechanical diversity of two-dimensional aerodynamic locomotion. *Physiol Biochem Zool* 73: 751-764, 2000.
66. Marley RJ, Arros DM, Henricks KK, Marley ME, and Miner LL. Sensitivity to cocaine and amphetamine among mice selectively bred for differential cocaine sensitivity. *Psychopharmacol* 140: 42:51, 1998.
67. Martins EP, and Garland T Jr. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. *Evolution* 45: 534-557, 1991.
68. Mayr E. 1963. *Animal species and evolution*. Cambridge, Mass.: The Belknap Press of Harvard Univ. Press, 1963.
69. Nagy KA, Girard IA, and Brown TK. Energetics of free-ranging mammals, reptiles, and birds. *Annu Rev Nutr* 19: 247-277, 1999.
70. Nishikawa KC. Neuromuscular control of prey capture in frogs. *Phil Trans R Soc B* 354: 941-954, 1999.
71. Powell JR. *Progress and perspectives in evolutionary biology: the Drosophila model*. New York: Oxford Univ. Press, 1997.
72. Promislow DEL. The evolution of mammalian blood parameters: patterns and their interpretation. *Physiol Zool* 64: 393-431, 1991.
73. Purvis A., and Webster AJ. Phylogenetically independent comparisons and primate phylogeny. In: *Comparative primate socioecology*, edited by Lee PC. Cambridge, U.K.: Cambridge Univ. Press, 1999, p. 44-70.
74. Reynolds PS, and Lee RM III. Phylogenetic analysis of avian energetics: passerines and nonpasserines do not differ. *Am Nat* 147: 735-759, 1996.
75. Rhodes JS, Koteja P, Swallow JG, Carter PA, and Garland T Jr. Body temperatures of house mice artificially selected for high voluntary wheel-running behavior: repeatability and effect of genetic selection. *J Therm Biol* 25: 391-400, 2000.
76. Rice MC, and O'Brien SJ. Genetic variance of laboratory outbred Swiss mice. *Nature* 283: 157-161, 1980.
77. Robertson A., ed. *Selection experiments in laboratory and domestic animals*. Farnam Royal, Slough, U.K.: Commonwealth Agricultural Bureau, 1980.
78. Roff DA. *Evolutionary quantitative genetics*. New York: Chapman & Hall, 1997.
79. Rose MR, Graves JL Jr., and Hutchinson EW. The use of selection to probe patterns of pleiotropy in fitness characters. In: *Insect life cycles: genetics, evolution and coordination*, edited by Gilbert F. London: Springer-Verlag, 1990, p. 29-42.

80. Schlager G, and Weibust RS. Selection for hematocrit percent in the house mouse. *J Heredity* 67: 295-99, 1976.
81. Schmidt-Nielsen K. *Scaling: why is animal size so important?* Cambridge: Cambridge Univ. Press, 1984.
82. Silver LM. *Mouse genetics. Concepts and applications.* New York.: Oxford Univ. Press, 1995.
83. Swallow JG, Carter PA, and Garland T Jr. Artificial selection for increased wheel-running behavior in house mice. *Behav Genet* 28: 227-237, 1998a.
84. Swallow JG, Garland T Jr., Carter PA, Zhan W-Z, and Sieck GC. Effects of voluntary activity and genetic selection on aerobic capacity in house mice (*Mus domesticus*). *J Appl Physiol* 84: 69-76, 1998b.
85. Swallow JG, Koteja P, Carter PA, and Garland T Jr. Artificial selection for increased wheel-running activity in house mice results in decreased body mass at maturity. *J Exp Biol* 202: 2513-2520, 1999.
86. Swallow JG, Koteja P, Carter PA, and Garland T Jr. Food consumption and body composition in mice selected for high wheel-running activity. *J Comp Physiol B* In press, 2001.
87. Thomason JJ, ed. *Functional morphology in vertebrate paleontology.* Cambridge: Cambridge Univ. Press, 1995.
88. Thompson GG, and Withers PC. Standard and maximal metabolic rates of goannas (Squamata: Varanidae). *Physiol Zool* 70: 307-323, 1997. Correction 71: 126, 1998.
89. Vanhooydonck B, and Van Damme R. Evolutionary relationships between body shape and habitat use in lacertid lizards. *Evol Ecol Res* 1: 785-805, 1999.
90. Wainwright PC, and Turingan RG. Evolution of pufferfish inflation behavior. *Evolution* 51: 506-518, 1997.
91. Weibel ER. *Symmorphosis: on form and function in shaping life.* Cambridge, Mass: Harvard Univ. Press, 2000.
92. Weibel ER, Taylor CR, and Bolis L, eds. *Principles of animal design: the optimization and symmorphosis debate.* Cambridge, U.K.: Cambridge Univ. Press, 1998.
93. Withers PC. *Comparative animal physiology.* Fort Worth, Saunders College, 1992.
94. Zera AJ, and Harshman LG. Physiology of life history trade-offs. *Annu Rev Ecol Syst* 32: in press, 2001.
95. Zera AJ, Sanger T, and Cisper GL. Direct and correlated responses to selection on JHE activity in adult and juvenile *Gryllus assimilis*: implications for stage-specific evolution of insect endocrine traits. *Heredity* 80: 300-309, 1997.