
Commentary

Phylogenetic approaches in comparative physiology

Theodore Garland, Jr¹, Albert F. Bennett^{2,*} and Enrico L. Rezende¹

¹*Department of Biology, University of California, Riverside, CA 92521, USA* and ²*Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA*

*Author for correspondence (e-mail: abennett@uci.edu)

Accepted 13 June 2005

Summary

Over the past two decades, comparative biological analyses have undergone profound changes with the incorporation of rigorous evolutionary perspectives and phylogenetic information. This change followed in large part from the realization that traditional methods of statistical analysis tacitly assumed independence of all observations, when in fact biological groups such as species are differentially related to each other according to their evolutionary history. New phylogenetically based analytical methods were then rapidly developed, incorporated into ‘the comparative method’, and applied to many physiological, biochemical, morphological and behavioral investigations. We now review the rationale for including phylogenetic information in comparative studies and briefly discuss three methods for doing this (independent contrasts, generalized least-squares models,

and Monte Carlo computer simulations). We discuss when and how to use phylogenetic information in comparative studies and provide several examples in which it has been helpful, or even crucial, to a comparative analysis. We also consider some difficulties with phylogenetically based statistical methods, and of comparative approaches in general, both practical and theoretical. It is our personal opinion that the incorporation of phylogeny information into comparative studies has been highly beneficial, not only because it can improve the reliability of statistical inferences, but also because it continually emphasizes the potential importance of past evolutionary history in determining current form and function.

Key words: allometry, comparative method, evolutionary physiology, model of evolution, phylogeny, statistical analysis.

Introduction

Studies of organismal form and function rely on multiple types of scientific investigation, including theory, description, experimentation and comparison. Comparing species is an ancient human enterprise, done for a variety of reasons (Sanford et al., 2002). Since Charles Darwin, the ‘comparative method’ – comparing populations, species or higher taxa – has been the most common and productive means of elucidating past evolutionary processes (Harvey and Pagel, 1991; Brooks and McClennan, 2002). Comparative methods have been used extensively to infer evolutionary adaptation, that is, changes in response to natural selection (for alternate physiological meanings of ‘adaptation’, see Garland and Adolph, 1991; Bennett, 1997). They are most often promoted and criticized (e.g. Leroi et al., 1994) within this context. However, comparative methods are not used to infer adaptation alone (Garland and Adolph, 1994; Sanford et al., 2002), but are also employed to analyze the effects of sexual selection (e.g. Hosken et al., 2001; Nunn, 2002; Smith and Cheverud, 2002; Aparicio et al., 2003; Cox et al., 2003), which may be nonadaptive or even maladaptive with respect to natural selection. These methods can also be used to compare rates of

evolution across clades or the amount of morphospace occupied by clades or by ecologically defined groups (Garland, 1992; Clobert et al., 1998; Ricklefs and Nealen, 1998; Garland and Ives, 2000; Hutcheon and Garland, 2004; McKechnie and Wolf, 2004). Of particular interest for the present review, they are also widely used to explore trade-offs (e.g. Clobert et al., 1998; Vanhooydonck and Van Damme, 2001) and to examine functional (mechanistic) relationships among traits (e.g. Lauder, 1990; Iwaniuk et al., 1999; Mottishaw et al., 1999; Autumn et al., 2002; Hale et al., 2002; Gibbs et al., 2003; Johnston et al., 2003; Herrel et al., 2005), including allometric scaling with body size (e.g. Garland, 1994; Reynolds and Lee, 1996; Williams, 1996; Clobert et al., 1998; Garland and Ives, 2000; Nunn and Barton, 2000; Herrel et al., 2002; Perry and Garland, 2002; Rezende et al., 2002, 2004; Schleucher and Withers, 2002; McGuire, 2003; Al-kahtani et al., 2004; McKechnie and Wolf, 2004; Muñoz-García and Williams, in press).

Comparative methods have been radically restructured over the past two decades, and now routinely incorporate both phylogenetic information and explicit models of character

evolution. Indeed, Sanford et al. (2002) suggest that this new emphasis be termed the 'comparative *phylogenetic* method'. As outlined in Blomberg and Garland (2002), this revolution in comparative phylogenetic methodology followed from several conceptual advances: (1) adaptation should not be casually inferred from comparative data; (2) the incorporation of phylogenetic information increases both the quality and even the type of inference from comparative data alone; (3) because all organisms are differentially related to each other, taxa cannot be assumed to be independent of each other for statistical purposes; (4) statistical analyses of comparative data must assume some model of character evolution for effective inference; (5) taxa used in comparative analyses should be chosen in regard to their phylogenetic affinities as well as the area of functional investigation; and (6) even phylogenetically based comparisons are purely correlational and inferences of causation drawn from them can be enhanced by other approaches, including experimental manipulations.

To expand on some of these points, 'quality' in point 2 includes the simple fact that adding an independent estimate of phylogenetic relationships to a comparative analysis increases – often greatly – the amount of basic data that is brought to bear on a given question, whereas 'type' refers to analyses that are simply impossible without a phylogenetic perspective, such as reconstructing ancestral values or comparing rates of evolution among lineages. Although phylogenetic information and a suitable analytical method may allow any comparative data set to be 'rescued' from phylogenetic nonindependence (e.g. avoid inflated Type I error rates; point 3), phylogenetically informed choice of species (point 5) can accomplish more, such as actually increasing statistical power to detect relationships among traits (Garland et al., 1993; Garland, 2001). Finally, we note that point 6 was recognized long ago, but has been re-emphasized as phylogenetically explicit methods of statistical inference have been developed (e.g. see Lauder, 1990; Garland and Adolph, 1994; Leroi et al., 1994; Autumn et al., 2002).

The intent of this commentary is to provide a review of some advances that have occurred in the comparative method, with an emphasis on their place in comparative physiology. We examine the underlying reasons for the incorporation of phylogenetic information into comparative studies. In an Appendix, we give a brief overview of the three most commonly used and best understood phylogenetically based statistical methods: independent contrasts (IC; worked example in Fig. 5), generalized least-squares (GLS) models, and Monte Carlo computer simulations. These methods apply mainly to analysis of continuously varying (or at least quantitative) traits, which is the nature of most physiological traits (e.g. blood pressure, metabolic rate, enzyme activity). However, they can also easily incorporate independent variables that are treated as discrete categories, such as diet (e.g. insectivore, frugivore, sanguivore) or habitat (e.g. fresh or salt water). Discussions of methods for categorical traits and computer programs to implement them are available from Mark Pagel (e.g. see Pagel, 1999), in MacClade (Maddison and

Maddison, 2000), and in Mesquite (<http://mesquiteproject.org/mesquite/mesquite.html>; see also Paradis and Claude, 2002). For a general listing of phylogeny-related programs, see the website maintained by Joe Felsenstein (<http://evolution.genetics.washington.edu/phylip/software.html>).

We discuss when phylogenetically based statistical methods should be used and give some practical examples of where a phylogenetic perspective has improved our understanding of comparative data and evolutionary processes. We also discuss some of the practical and theoretical limitations of such methods. Throughout, we try to emphasize that the incorporation of phylogeny can greatly enhance comparative studies, deliver new insights, and open new areas for research. This is of necessity only a brief summary and readers are directed to more extensive discussions of the topics and issues raised here (e.g. Ridley, 1983; Lauder, 1981, 1982, 1990; Harvey and Pagel, 1991; Garland et al., 1992, 1999; Garland and Adolph, 1994; Harvey, 1996; Ricklefs and Nealen, 1998; Ackerly, 1999, 2000, 2004; Pagel, 1999; Purvis and Webster, 1999; Diniz-Filho, 2000; Feder et al., 2000; Garland and Ives, 2000; Maddison and Maddison, 2000; Garland, 2001; Rohlf, 2001; Autumn et al., 2002; Blomberg and Garland, 2002; Brooks and McLennan, 2002; Blomberg et al., 2002, 2003; Rezende and Garland, 2003; Housworth et al., 2004). We have intentionally not cited some 'forum' and 'perspective' type papers because we felt that their rhetoric was misleading, and in some cases they contain outright errors.

The empirical examples cited here are idiosyncratic, reflecting mainly our own research interests. Thus, we emphasize examples that involve physiological phenotypes, but include others when they are lacking. Our enthusiasm for phylogenetic approaches in comparative physiology should not be taken to imply, however, that we think they are more important than other approaches, such as measurement of selection acting in natural populations, experimental evolution (e.g. see Garland and Carter, 1994; Bennett and Lenski, 1999; Ackerly et al., 2000; Feder et al., 2000; Garland, 2001, 2003; Bennett, 2003; Swallow and Garland, 2005), or more purely mechanistic investigations (e.g. Mangum and Hochachka, 1998; Hochachka and Somero, 2002).

We are concerned that some of our discussion of assumptions and intricacies of phylogenetically based statistical methods may be off-putting to those who simply want to analyze their data (see also Felsenstein, 1985). However, it must be acknowledged that statistical analyses in general are not always simple and have underlying assumptions that cannot be ignored. Most of the tools that we use in everyday research (e.g. correlation, regression, analysis of variance, analysis of covariance) have been around for 50 years or even a century. Nonetheless, the field of statistics (both theoretical and applied) continues to refine these methods. Such questions as what type of line is best for describing functional relationships (e.g. Rayner, 1985; chapter 6 in Harvey and Pagel, 1991; Riska, 1991; McGuire, 2003; Garland et al., 2004), how to deal with non-linear relationships (Quader et al., 2004) or random effects in ANOVA models,

when to include or exclude interaction terms, how best to transform data, or when to employ nonparametric methods, still do not have simple, general answers. Moreover, new statistical methods continue to be developed, including computer-intensive approaches that were not possible 50 years ago (e.g. see Lapointe and Garland, 2001; Roff, in press). For many statistical parameters, including comparative methodologies, several different approaches (and attendant algorithms) may be used for estimation, none of which performs 'best' in all situations. We believe that it is important that a comparative biologist understand the assumptions and approaches underlying these methodologies, and does not just resort to their rote application, and that is the basis for our more detailed presentation.

Phylogeny and modern (statistical) comparative methods

The beginning of the transition into modern comparative phylogenetic methods is marked by the publication of Ridley's book on mating adaptations (Ridley, 1983) and by Felsenstein's article entitled 'Phylogenies and the comparative method' (Felsenstein, 1985). Both argued for the necessity of incorporating an explicitly phylogenetic perspective into analyses of comparative data. These authors were not the first to claim that comparative data generally violate the assumptions of conventional statistical methods (see Harvey and Pagel, 1991), but Felsenstein (1985) proposed the first *fully* phylogenetic method, i.e. one that could incorporate detailed information on topology and branch lengths, which he termed independent contrasts (IC). Although the full-blown IC method (see Appendix for description and worked example in Fig. 5) requires detailed information on phylogenetic topology, branch lengths (Fig. 1), and model of character evolution (Fig. 2) in order to be maximally reliable, Felsenstein (1985) also considered how one might make use of partial information, such as might be derived from a taxonomy that had some

resemblance to actual phylogenetic relationships, e.g. by comparing several pairs of species within a series of genera, an approach that is now commonly used (e.g. Monkkonen, 1995; for a review of plant examples, see Ackerly, 1999). Nonetheless, the requirements of the method seemed daunting to many, and its use in comparative physiology grew slowly. Indeed, one of us even helped to develop an alternative phylogenetic method, partly because of a lack of information on branch lengths (see Figs 1, 2) in a comparative study that seemed to preclude the use of IC (Huey and Bennett, 1987; and see extensions in Garland et al., 1991; Martins and Garland, 1991).

Concern about the possible influence of phylogeny in comparative and ecological physiology antedated Felsenstein's

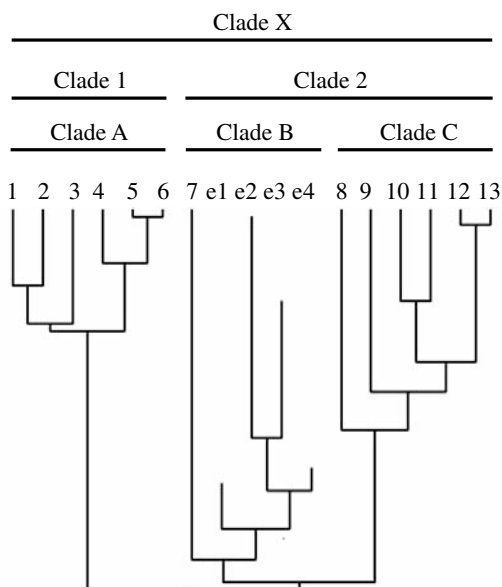


Fig. 1. Hypothetical evolutionary relationships among 17 species of organisms. Vertical axis represents time in relative units, with the top of the 'phylogenetic tree' representing the present. Hence, species 1-7 and 8-13 are alive now (extant), whereas species e1-e4 are extinct. Statements about phylogenetic relationships are based solely on recency of common ancestry. For example, species 1 and 2 are each other's closest relative because they share a more recent common history with each other than with any other species depicted in this figure. The horizontal axis is arbitrary, and note that nodes could be rotated for graphical convenience with no implication for evolutionary relationships. 'Clades' are hierarchically arranged, 'monophyletic' groups of species, including all species that have descended from a common ancestor as well as that basal ancestor. All species within a given clade are more closely related to each other (they share a more recent common ancestor) than to any species in another clade. In the strict sense, a clade includes all species that have ever existed within it. Thus, Clade B includes species 7 as well as e1-e4. However, as it is impossible to know of all extinct species within a given clade and as physiologists rarely include extinct taxa in their studies, the term 'clade' is often used in a relative way with respect to a particular collection of species that are included in a given study. Consider a comparative physiological study of species 1-13. Species 1-6 might be referred to as Clade 1, while species 7-13 might be referred to as Clade 2. However, note that species 7 is relatively distantly related to the other extant species in Clade 2 (i.e. species 8-13 shared a last common ancestor much more recently than the last common ancestor of them with species 7). Hence, a researcher studying species 1-13 might prefer to write in terms of Clades A, B and C in order to highlight the fact that, *a priori*, she would expect species 7 to be somewhat different from species 8-13. (Importantly, *a priori* hypotheses about particular single species can be tested with well-established phylogenetically based statistical methods, although they may not be convincing to some regardless of the level of statistical significance; see Garland et al., 1993; Garland and Adolph, 1994; Garland and Ives, 2000.) Branch lengths in this figure are proportional to divergence times. All phylogenetically based statistical methods use branch lengths in their calculations, although some assume (arbitrarily) that each branch segment is equal in length. Alternatively, under the commonly assumed Brownian motion model of character evolution (see Fig. 2), branch lengths are assumed to be in units proportional to (relative) divergence times and hence to the variance of character evolution along each branch segment (i.e. longer branches imply greater variance of character change; see Felsenstein, 1985).

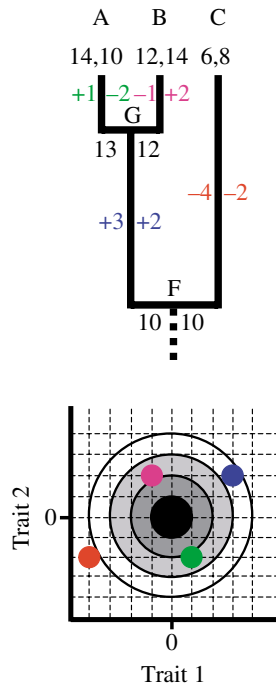


Fig. 2. Illustration of a Brownian motion (random walk in continuous time) model of character evolution, as might be implemented in a computer program (e.g. PDSIMUL of Garland et al., 1993). The goal is to simulate the evolution of two traits, beginning at the bottom of the phylogenetic tree and ending at the three tips, species A, B and C. A computer program begins at the bottom of the tree (internal node 'F') with user-specified starting values, in this example 10 and 10 for Traits 1 and 2, respectively. It then draws a random datum from a bivariate normal distribution of hypothetical evolutionary changes for the two traits. This distribution is illustrated by concentric rings proportional to density of data points in the z axis (projecting out of the page), with darker indicating a higher density of points; the tails of the distribution diminish to infinity). In this example, we assume that the means of this distribution are 0 for both traits, such that no general tendency for either to increase or decrease will be modeled. We also specify 0 correlation between them, such that they will 'evolve' independently, on average. For the amount of evolutionary change from node F to tip species C, we happen to draw values of -4 and -2 (red). Thus, species C has values of 6 and 8. For the change from node F to G, we draw $+3$ and $+2$ (blue). Above this, we draw two separate sets of changes: $+1$ and -2 leading to tip species A (green); -1 and $+2$ leading to tip species B (purple). Note that the amount of change tends to be greater for longer branches, reflecting a greater opportunity for evolutionary change. In practice, a computer program might achieve this by expanding or contracting the widths of the bivariate normal distribution for relatively longer or shorter branch lengths, respectively. Thus, under Brownian motion, for a given character, the variance of this distribution is set to be proportional to divergence time (along the length of each branch segment sequentially; Felsenstein, 1985, 1988). Note also that the distribution from which changes are drawn does not need to be (bivariate) normal (see Felsenstein, 1985, 1988). Moreover, the means of the distribution can be set to positive or negative values to impose directional trends in character evolution (Garland et al., 1993). These sorts of changes create models that are no longer simple Brownian motion.

(1985) publication. For example, explicit comparisons of marsupial with placental mammals (MacMillen and Nelson, 1969; Dawson and Hulbert, 1970) and of passerine with non-passerine birds (Lasiewski and Dawson, 1967) were motivated by cognizance of phylogeny, and some workers tried to partition the effects of phylogeny on physiological relationships (e.g. Andrews and Pough, 1985). Moreover, some workers voiced concerns about specific adaptive interpretations of characters shared more widely in their clades (e.g. Dawson and Schmidt-Nielsen, 1964; Dawson et al., 1977). What those earlier studies lacked was not necessarily a general perspective on the importance of phylogeny, but rather a formal logical and statistical methodology for incorporating detailed phylogenetic information. Analytical techniques have been greatly expanded and modified since 1985 (see below and Appendix), but Felsenstein's IC method is still the most widely used and his insights were pivotal to modernization of the comparative method. Moreover, the realization that IC is a special case of generalized least-squares (GLS) methods (see Appendix) means that the former can always serve as a useful entry point for the latter, and one that retains the major heuristic of 'tree thinking' (*sensu* Maddison and Maddison, 2000).

Traditional interspecific comparative analyses applied conventional statistical methods to test for associations between traits (e.g. metabolic rate and body size), or between a trait and an environmental variable (e.g. blood oxygen carrying capacity and altitude). This approach treats all data points (e.g. mean values for a series of species) as statistically independent of each other. Unfortunately, mean phenotypes of biological taxa usually will not be statistically independent because they are all related through their hierarchical

phylogenetic history. Empirically, more closely related species do indeed tend to resemble one another; put simply, hummingbirds look like hummingbirds, and turtles look like turtles, and the same is true for physiological traits (Blomberg et al., 2003; see below). This general tendency exists for several good biological reasons (Harvey and Pagel, 1991), including time lags for change to occur after speciation, occupation of similar niches by close relatives, and conservative phenotype-dependent responses to selection. Thus, the extent of these phylogenetic relationships – and hence the expected degree of resemblance – must also be figured into comparative analyses. Analytical techniques that do not incorporate phylogenetic information make the tacit statistical assumption that all the species studied are equally distantly related to each other, that is, that they descended along a 'star phylogeny' (Fig. 3A), when in fact their ancestral associations are hierarchical (Fig. 3C).

The foregoing statement requires substantial amplification. First, there is an alternative way to view the tacit statistical assumption. A star phylogeny, as shown in Fig. 3A, is usually drawn to imply that a set of species all originated from a

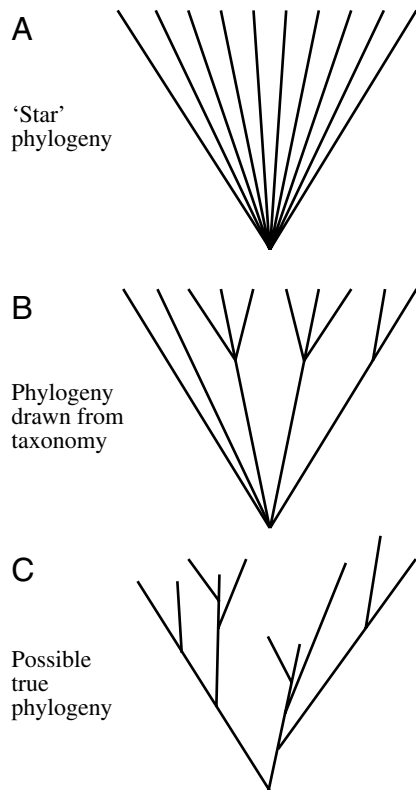


Fig. 3. (A) Illustration of what conventional statistical analyses assume when applied to comparative data ('star' phylogeny with equal-length branches). This model implies that values at tips of the tree (phenotypic means for some trait for 10 species) are statistically independent and identically distributed. (B) Phylogenetic tree that might be inferred from taxonomic information, e.g. if five genera within a single family were represented that contained, from left to right, one, one, three, three and two species in the data set. This assumes that the genera are an unrelated series of 'mini-stars' with no hierarchical structure within any of them. It also assumes that the taxa actually represent separate evolutionary lineages (monophyletic groups or clades), but such is not always the case for taxonomies. (C) Estimates of real phylogenies usually indicate hierarchical relationships and branches that do not necessarily line up along the tips of the tree. Non-contemporaneous tips can indicate that extinct taxa are included in the data set or that the rate of evolution has varied among branches. Real phylogenies like this cause various statistical problems, stemming from the non-independence of species' phenotypes, so phylogenetically based statistical methods are required for proper analyses. Modified from Garland (2001) and Rezende and Garland (2003).

common ancestor at virtually the same point in the past, i.e. that a 'big bang' of speciation occurred very rapidly for that particular set of species (and perhaps others not presently being considered). Alternatively, with respect to the assumption of statistical independence, it could imply that recent evolution of some character(s) has been so rapid that any evidence of successive speciation events is lost. In other words, some of the species may in fact be more closely related than others, phylogenetically speaking, but we would never know it just by

looking at the characters we are trying to study: no 'phylogenetic signal' remains. Extremely high measurement error could have a similar effect, but would also make us seriously doubt that the data were good enough for any sort of analysis. As discussed below, the issues of high trait lability and/or high measurement error can be addressed empirically, and recent studies have found that most traits do indeed exhibit phylogenetic signal, indicating that a star phylogeny does not provide a good fit to the data (Freckleton et al., 2002; Blomberg et al., 2003; Tieleman et al., 2003; Ackerly, 2004; Al-kahtani et al., 2004; Ashton, 2004a,b; Hutcheon and Garland, 2004; Laurin, 2004; Rezende et al., 2004; Rheindt et al., 2004; Ross et al., 2004; Muñoz-García and Williams, in press).

Second, it is important to consider what is meant by the 'branch lengths' of a phylogenetic tree that is used for analysis. In general, proponents of phylogenetically based comparative methods assume that analyses of physiological and other traits will involve use of a phylogenetic tree that was inferred from other data, such as variation in DNA sequences, which is presumed to be independent of the data being analyzed. Otherwise, it seems intuitively obvious that analyses may involve some circularity. However, this is actually a complicated subject and beyond the scope of the present paper (Felsenstein, 1985; de Queiroz, 2000). Leaving aside the general issue of having available a phylogeny that is independent of the characters under study, the branch lengths of the working phylogenetic tree are confounded with the model and rates of character evolution that will be assumed for statistical analyses of most real data sets (see Figs 1, 2). In other words, we usually do not have independent information on, for instance, divergence times *and* selective regimes that may have prevailed along various branches of the tree. In any case, all of the main phylogenetically based statistical methods require branch lengths in units proportional to expected variance of evolution for the character(s) under study (see Felsenstein, 1985, 1988; Garland et al., 1992, 1993, 1999; Garland and Ives, 2000; Rohlf, 2001; Blomberg et al., 2003; Housworth et al., 2004). Branch lengths essentially indicate our *a priori* expectations for how likely a given trait was to change (increase or decrease in value) from one node to another along a phylogenetic tree, and thus become an integral component of our statistical null model. Under a simple Brownian motion model, those branch lengths would necessarily be proportional to divergence times. Under any other model, such as the Ornstein–Uhlenbeck (OU) process, which is like Brownian motion while tethered to an elastic band and is used to model stabilizing selection or constraints on trait space (Felsenstein, 1988; Garland et al., 1993; Diaz-Uriarte and Garland, 1996; Martins and Hansen, 1997; Blomberg et al., 2003; Freckleton et al., 2003; Butler and King, 2004; Housworth et al., 2004), they would be more-or-less different from divergence times.

A simple hypothetical example can illustrate this distinction. Many traits evolve within limits set by physical or biological properties. Some of these are trivial. For example, body mass cannot evolve to be as small as 0 g. Others are more interesting.

Apparently, for example, activity body temperatures (T_b) of squamate reptiles (lizards and snakes) cannot evolve to be more than about 42°C. We do not know the ancestral activity T_b of squamates, but it was probably substantially lower than 42°C. Thus, during their initial radiation and diversification, T_b would have been free to evolve, perhaps in a fairly Brownian motion-like fashion, with an increase or decrease about equally likely to occur along any branch of the phylogeny. However, lineages that ‘explored’ the climate space towards higher T_b would eventually be constrained by the reduction in Darwinian fitness that can be caused by exceedingly high temperatures (e.g. *via* failure of spermatogenesis or outright death). Thus, if we were to depict a phylogenetic tree of squamates with branch lengths proportional to expected variance of T_b evolution, then we would need to know the T_b at the start of each branch segment and also have the branches be, in effect, different if the lineage was near a thermal limit, either upper or lower. That is, a lineage near an upper limit would have a low probability of evolving a higher T_b , but a ‘typical’ probability of evolving a lower T_b , and *vice versa*. It should be obvious that our ability to specify such detailed branch-length information for any trait in any group of wild organisms is severely limited. Thus, for simplicity and/or analytical tractability, phylogenetically based statistical methods usually begin with an assumption of Brownian motion evolution along whatever branch lengths are specified in a working phylogeny (e.g. Fig. 1). And in many cases (e.g. see reviews of published studies in Blomberg et al., 2003; Ashton, 2004a), these will be arbitrary values, such as setting all segments equal to unity in length or by some other simple rule (e.g. Fig. 4B). In such cases, it is often prudent to perform computations with more than one set of branches as a sensitivity analysis for the conclusions (e.g. see Ashton, 2004b; Hutcheon and Garland, 2004; Laurin, 2004). Similarly, some studies use multiple phylogenies (topologies) (e.g. Bauwens et al., 1995; Symonds and Elgar, 2002; Hodges, 2004).

As introduced above, for some models of evolution, including ones in which phenotypes respond essentially instantaneously (in evolutionary time) to changes in the selective regime, the appropriate branch lengths would be very long for those leading to tips of the tree and very short internally and near the base. In the limit, this becomes a star with no hierarchical structure (Fig. 3A). (A similar situation can arise if the tip data contain very large amounts of measurement error.) So, a conventional statistical analysis can be justified on first principles under *some* models of evolution, and computer simulations have confirmed this (Diaz-Uriarte and Garland, 1996; Price, 1997; Harvey and Rambaut, 2000; Martins et al., 2002). Furthermore, even if Brownian motion were an adequate descriptor of character evolution, we never have exact information on divergence times (and different characters likely evolve at different rates), so our branch lengths will always contain some amount of error. If that error were large enough, as in certain cases where evolution has been very much unlike Brownian motion, then we might be better off just assuming a star phylogeny, which can be accomplished by using conventional statistical methods. On the other hand,

even if traits evolve very rapidly in response to altered environmental conditions (selective regimes), environments can have a phylogenetic history (ecological or niche conservatism), which would confer phylogenetic structure on trait evolution (Harvey and Pagel, 1991; Desdevises et al., 2003). Most organisms do not have infinite mobility, and hence descendant generations are likely to live fairly near the haunts of their ancestors, and habitat selection can accentuate this ‘inheritance’ (see p. 30 in Garland et al., 1992). Indeed, several studies have shown that such traits as the latitude from which species (populations) were sampled can show significant phylogenetic signal (Freckleton et al., 2002; Hodges, 2004; Rezende et al., 2004; see also Desdevises et al., 2003).

These points have suggested to some that phylogenetically based analyses are so fraught with pitfalls that we should stick with non-phylogenetic ones. But a conventional statistical analysis actually has as many assumptions as a phylogenetic one. For example, it assumes that the species under analysis have not been interacting, e.g. as by character displacement (Hansen et al., 2000). It assumes that each species should be equally weighted, which is equivalent to saying that the heights of each branch from the root of the tree (assumed to be a star) are equal. And so forth.

In any case, it has become increasingly clear that, because we never know the true branch lengths and/or model of character evolution, we should pay careful attention to the branch lengths used, employing methods that can consider options ranging between a star and our working hierarchical phylogeny, and possibly something even more hierarchical. Thus, recent methods emphasize estimation of optimal branch length transformations as an essential part of phylogenetic analyses of comparative data (e.g. see Grafen, 1989; Diaz-Uriarte and Garland, 1996, 1998; Pagel, 1999; Harvey and Rambaut, 2000; Freckleton et al., 2002; Martins et al., 2002; Blomberg et al., 2003; Housworth et al., 2004). Although some researchers may be uneasy with such transformations of branch lengths, they are analogous to use of a Box–Cox procedure to find the optimal transformation of data (e.g. best approximation of normality) in conventional statistical procedures (for instance, use of a Box–Cox procedure to transform branch lengths; Reynolds and Lee, 1996). Moreover, aside from its benefits with computer-simulated data, such careful attention to branch lengths can sometimes improve statistical power to an important extent with real data (see below).

An example of how phylogeny can affect statistical analyses

By overestimating the true number of independent observations, conventional statistical methods applied to comparative data typically lead to inflated Type I error rates, i.e. statistical significance is claimed too often (e.g. Grafen, 1989; Martins and Garland, 1991; Purvis et al., 1994; Diaz-Uriarte and Garland, 1996). A real example of the influence of phylogeny on interpretation of comparative data comes from a study that tested the hypothesis that the preferred body temperature and the

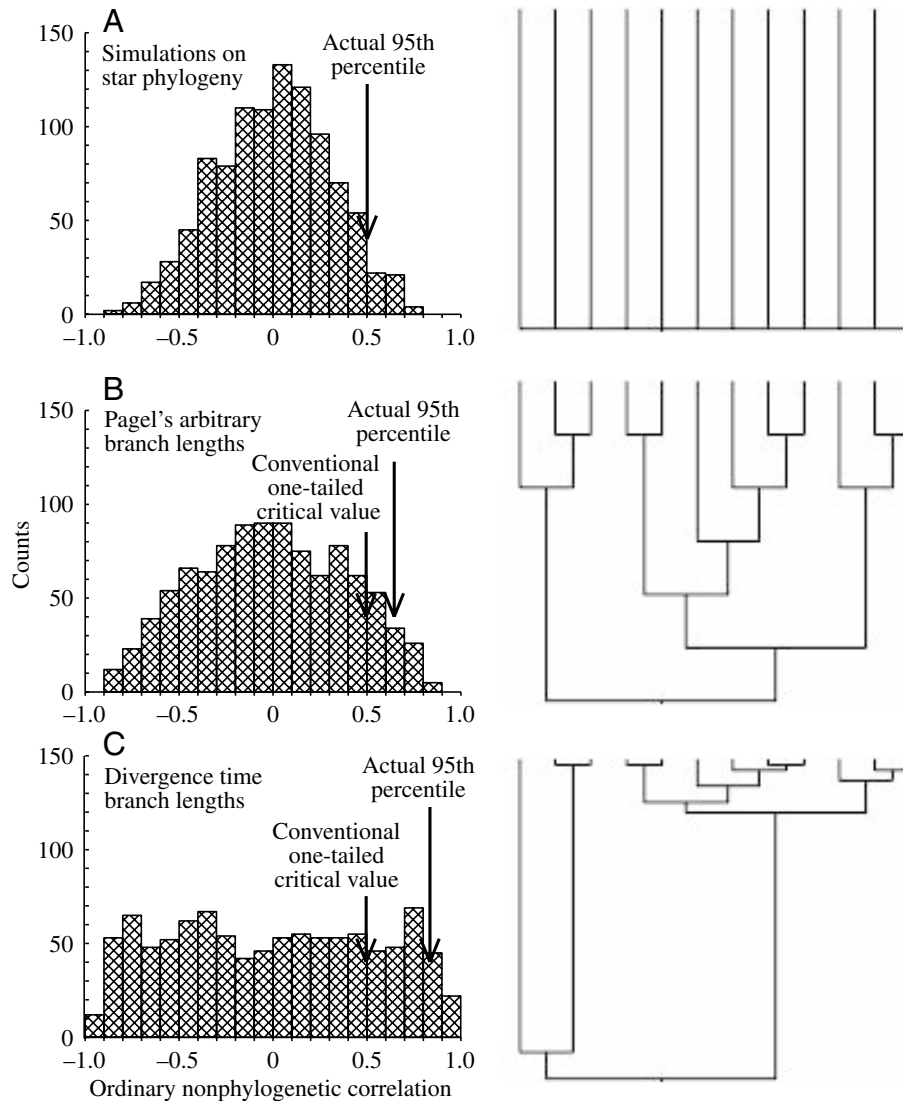


Fig. 4. Use of computer simulations to illustrate how Type I error rates for testing an association between two traits can be inflated by ignoring phylogenetic relationships. Shown are three distributions of ordinary, non-phylogenetic, Pearson product-moment correlations of tip data. In each of the three figures, data were simulated along the phylogeny shown, under a simple Brownian motion model of character evolution (see Fig. 2), with the correlation between the two traits set to zero (Martins and Garland, 1991; Garland et al., 1993). (A) Data simulated along a 'star' phylogeny, here depicted as a 'comb.' The upper 95th percentile is +0.504, which is statistically indistinguishable from the conventional critical value of +0.497. Compared with this distribution, the correlation for the real data on lizards (+0.585; see text) would be considered statistically significant at $P < 0.05$. (B) Data simulated along the hierarchical topology, but with less extreme branch lengths (arbitrary values as suggested by Pagel, 1992) than for the real branch lengths shown in (C). For these simulated data, the 95th percentile is +0.641, which is larger than the conventional critical value. Judged against this empirical null distribution, the correlation for the real data would be considered statistically non-significant ($P > 0.05$). (C) Simulations along the actual phylogeny used by Garland et al. (1991). The simulated data include an even greater number of sets for which the correlation is strongly positive, as compared with (B). The 95th percentile is +0.828, which is much larger than the nominal one-tailed critical value for testing a correlation coefficient (+0.497). If the phylogeny shown in C is close to reality, and if evolution has been similar to Brownian motion, then the results of C are more trustworthy than those of A.

optimal temperature for sprint running speed would be positively correlated among 12 species of lizards (Huey and Bennett, 1987; Garland et al., 1991). The ordinary Pearson correlation coefficient between these two temperatures, uncorrected for phylogenetic associations, is +0.585. Is this statistically significant, or might it have been obtained by chance sampling if the true correlation among all Australian skinks were zero?

The answer depends on what is assumed about the phylogenetic relationships of the 12 species. If we assume that species are unrelated, then we can refer to conventional tables of critical values for correlation coefficients. For a one-tailed test with 12 data points (and hence 10 degrees of freedom for testing a correlation), the critical value is +0.497, so a value of +0.585 would be considered significant at $P < 0.025$. If, instead, we want

to assume that the species are related in a hierarchical fashion, then we cannot use the conventional tables. Fortunately, however, we can incorporate phylogenetic information as follows (Martins and Garland, 1991; Garland et al., 1993). We can construct different working phylogenies, model the uncorrelated evolution of these traits by a Monte Carlo computer simulation that assumes random, Brownian-motion like trait change, and calculate a correlation coefficient for each simulated data set. We can then determine the critical 5% level for the correlation coefficient for each distribution. If we assume that all species are completely independent or related by a star phylogeny (Fig. 4A), then the one-tailed probability for obtaining a correlation as large as +0.585 is 0.023 (based on this particular set of 1000 simulated data sets), so the relationship is statistically significant at $P < 0.05$. In fact, if we do a very large number of simulations, then we will obtain *exactly* the same results as when referring to conventional tables.

If, however, we simulate data along our best estimate of the phylogeny of these lizards (Fig. 4C), then a correlation of +0.585 would be observed much more frequently than 5% of the time and would not be considered very unusual, hence not statistically significant ($P > 0.15$). If a hypothetical phylogeny with different branch lengths, involving fewer deep roots, were assumed (Fig. 4B), then a value of +0.585 would have a lower probability of being observed, but in this case would still be non-significant. Thus, the assumed pattern of the relationships among the species crucially affects the statistical significance of the observations: the more the phylogeny departs from a star, the lower is the number of effectively independent observations and the more likely we are to observe an extremely large (or small) correlation just by chance. If the working topology, branch lengths, and simulation model are somewhat realistic, then we will claim significance too often if we ignore phylogeny.

Another important point is that if the simulated data of Fig. 4B or C are analyzed with IC, using the corresponding phylogenies, then the resulting distribution of correlation coefficients will be the same as in Fig. 4A (results not shown). Thus, as discussed in the Appendix, the IC method uses the specified phylogenetic information to transform the data to make them independent and identically distributed. This then allows one to refer to conventional tables of critical values for hypothesis testing.

All of the simulations shown in Fig. 4 were done under a simple Brownian motion model (Fig. 2), but the model used can have a large effect on the resulting distributions of statistics (e.g. see Garland et al., 1993; Diaz-Uriarte and Garland, 1996; Price, 1997; Harvey and Rambaut, 2000; Martins et al., 2002; Freckleton et al., 2003). Brownian motion is a very simple model of character evolution, and its analytical tractability was exploited by Felsenstein (1985) to develop IC. It is a good model for traits that evolve solely by random genetic drift, and may also be adequate for some types of ‘fluctuating’ selection (i.e. when the direction of selection changes from generation to generation). As a basis for statistical methods to estimate and test character correlations, it may also be an adequate

model for traits that are subject to certain types of selection (Felsenstein, 1985, 1988; Grafen, 1989). But most traits probably evolve in ways that are too complicated and idiosyncratic to be modeled realistically by Brownian motion (Felsenstein, 1988; Hansen et al., 2000). Fortunately, simulations can use arbitrarily complex models of character evolution, limited only by one’s ability to write computer programs and imagination (e.g. Garland et al., 1993; Diaz-Uriarte and Garland, 1996, 1998; Price, 1997; Harvey and Rambaut, 2000; Freckleton et al., 2003). Of course, whether more complicated models lead to more accurate analyses depends on whether they are actually a better descriptor of past evolution by the characters under study, and that is something very difficult to know. Moreover, finding a model that fits a set of data reasonably well does not necessarily mean that it is the correct model, and other models can probably be found that would provide equally good fit (see also Blomberg et al., 2003). (As always, it is risky to attempt to infer process from pattern.) Given the near impossibility of knowing how traits actually evolved in the distant past, simulation studies are also used to gauge how robust analytical methods such as IC are to violation of their assumptions (e.g. Brownian motion, accurate knowledge of branch lengths), how diagnostic tests can alert one to such violations, and how well remedial measures (e.g. transformations of tip data or branch lengths) can rescue statistical performance when assumptions are violated (e.g. Martins and Garland, 1991; Purvis et al., 1994; Diaz-Uriarte and Garland, 1996, 1998; Harvey and Rambaut, 2000; Diniz-Filho and Torres, 2002; Freckleton et al., 2003). Still, physiologists may sometimes be able to improve the accuracy of assumed models by their knowledge of how organisms work (or could have worked), as in the case of limits to body temperature evolution discussed above. Similarly, paleontological information can be used to improve the realism of simulations (Garland et al., 1993).

We close this section by emphasizing that the use of computer simulations to obtain ‘phylogenetically correct’ (PC) null distributions for testing hypotheses about comparative data is a very general tool that can be used for virtually any analysis (Martins and Garland, 1991; Garland et al., 1993), including bivariate (e.g. Ricklefs and Nealen, 1998) or multivariate analyses of evolutionary diversification. For example, the PHYLOGR (available at <http://cran.r-project.org/>) program allows one to test hypotheses about canonical correlation or principal components analysis (PCA) in relation to computer-simulated data (R. Diaz-Uriarte and T. Garland, manuscript in preparation).

When and why to use phylogenetic information in comparative studies

What kinds of characters are amenable to comparative phylogenetically based analyses? Basically, any measurable trait can be studied with these methods. It can be a discrete character, such as presence or absence of a structure, or a continuously distributed trait, such as the length of a bone or

the value for a rate process. The methods are not confined to analysis of structure and function of individuals, but may also include aspects of ecology, such as home range size or environmental indices (e.g. mean annual temperature). However, for the analysis of evolutionary differences it is desirable to minimize the influence of the immediate environment on phenotypic characters prior to their measurement. That is, to the extent possible, all species should be exposed to common conditions (acclimated to the same environment) for some period of time prior to measurement. Ideally, they would be bred and raised under common conditions for one or two generations (Garland and Adolph, 1991). That is unfortunately not possible for many species. Moreover, given that the species under study probably vary in the conditions they inhabit, which set or sets of environmental conditions should be used? And what if the ordering of species phenotypes varies among those conditions because of genotype-by-environment interactions? These questions have no easy answers, but should be borne in mind during analysis and interpretation.

What kinds of characters demand a phylogenetic analysis? Although we may generally expect that most characters will tend to 'follow phylogeny', this is an empirical question. The simplest general test for whether related organisms actually do tend to resemble each other more than they resemble those that might be chosen randomly with respect to phylogenetic position uses randomization procedures (see also Abouheif, 1994; Ackerly, 2004; Laurin, 2004; Rheindt et al., 2004). Specifically, once phylogenetically IC have been computed, it is possible to calculate the variance of those contrasts. The lower the variance of the contrasts, the better the fit of the phylogeny (topology and branch lengths) to the character in question. To determine whether a given variance indicates the presence of statistically significant 'phylogenetic signal' (i.e. more closely related species tend to resemble each other more than they resemble randomly chosen species), one can compare it with the distribution of variances for a large number of data sets that have been randomized (shuffled) across the tips of the phylogeny (Blomberg and Garland, 2002; Blomberg et al., 2003). For studies with 20 or more species (for which statistical power should be reasonably high), more than 90% of the traits examined to date (including behavioral, physiological, morphological, life history and ecological/environmental traits) do exhibit a significant phylogenetic signal ($P < 0.05$: Blomberg et al., 2003; Al-kahtani et al., 2004; Ashton, 2004a,b; Rezende et al., 2004; Ross et al., 2004; Muñoz-Garcia and Williams, in press; see also Freckleton et al., 2002).

The empirical finding of pervasive phylogenetic signal implies that hierarchical phylogenies – as presented and used in numerous publications – provide a better fit to the data under analysis than does a star phylogeny (Figs 3A, 4A). This sends a strong message that we should routinely consider phylogenetic information in statistical analyses of comparative data. However, this does not necessarily mean that, for any given set of data, we should simply obtain a phylogenetic tree, perform an IC, GLS or Monte Carlo simulation analysis, and

automatically presume that the results will be more reliable than the comparable conventional statistical analysis. As we and others have emphasized for more than a decade, analyses using a given topology and branch lengths can perform relatively poorly if their assumptions are severely violated (e.g. see Grafen, 1989; Martins and Garland, 1991; Diaz-Uriarte and Garland, 1996, 1998; Price, 1997; Garland and Diaz-Uriarte, 1999; Harvey and Rambaut, 2000; Diniz-Filho and Torres, 2002; Martins et al., 2002; Freckleton et al., 2003; Housworth et al., 2004). Thus, we urge practitioners to apply robust tests for phylogenetic signal, diagnostic checks, and branch length transformations as warranted (for recent discussions and methods, see Freckleton et al., 2002; Blomberg et al., 2003), and sensitivity analyses by varying branch lengths and/or model of evolution (e.g. see Garland et al., 1993; Ashton, 2004b; Hutcheon and Garland, 2004; Laurin, 2004; Muñoz-Garcia and Williams, in press).

What sorts of branch lengths should be used? Given the uncertainties regarding branch lengths (see above), many workers have reported results with multiple branch lengths to explore consistency or the lack thereof (e.g. Ross et al., 2004). Although it is often the case that conclusions are relatively robust (insensitive) to the branch lengths used, this is not always true, and the importance of attempting to use 'optimal' branch lengths transformations can be illustrated with two empirical examples. Garland et al. (1993) analyzed home range areas in relation to body size for 49 species of carnivores and ungulates. Conventional analysis of covariance (ANCOVA) indicated a highly significant ($P < 0.001$) difference in size-adjusted home range areas of the two groups. Analysis *via* IC (or Monte Carlo simulations), however, revealed no statistically significant difference (two-tailed $P = 0.126$ for IC). The branch lengths used for the phylogenetic analyses were estimates of divergence times, derived from various sources. They passed the diagnostic 'lack of fit' tests as described in Garland et al. (1992). However, power to detect a difference is apparently improved by applying the transformations of branch lengths as proposed by Blomberg et al. (2003) to mimic particular models of character evolution. Using the branches transformed under the OU model for log body mass, the P value is reduced to 0.099, and using their Accelerating–Decelerating (ACDC) model the P value is reduced to 0.044, thus crossing the typical threshold of < 0.05 to be considered statistically significant (degrees of freedom were reduced by one in both cases to reflect the additional parameter estimated in these models; see also Diaz-Uriarte and Garland, 1996, 1998; Garland and Diaz-Uriarte, 1999). Similarly, in a recent comparison of the generic average body sizes of 'megabats' and 'microbats,' Hutcheon and Garland (2004) found statistical significance in the IC analysis only when using branch lengths transformed under the OU or ACDC models. We suspect that such increases in power may be more likely to occur in comparisons between groups that are fairly highly phylogenetically confounded (i.e. the independent variable of interest, such as diet, is highly clumped with respect to phylogeny, as in comparisons of clades; e.g. see

Garland et al., 1993; Vanhooydonck and Van Damme, 1999; Perry and Garland, 2002; Rezende et al., 2004) than in tests of correlations between two traits.

How do we choose species for study? Traditionally, animals were chosen for comparative studies for any number of reasons, including convenience (e.g. local availability or an existing literature data base), possession of an interesting biological trait (e.g. the long neck of the giraffe), occupation of an extreme environment (e.g. a hot dry desert, the Arctic), or characteristics that make it well suited to study a particular physiological process (Garland and Adolph, 1994; Garland and Carter, 1994; Bennett, 2003). Frequently, a particular species or group living in a particular environment has been the key that originally sparked interest in the project. It is now clear that phylogenetic information should also be considered when choosing species for study (for a simulation study on the effects of taxon sampling when testing for correlated evolution, see also Purvis and Webster, 1999; Ackerly, 2000).

To increase analytical power, it is a good idea to include other species that experience a very broad range of the environmental ('independent') variable. You might then randomly sample species from a broad taxon (e.g. mammals) or focus exclusively on a particular lineage, such as bats or rodents. From a design perspective, the latter strategy is preferable because the broader comparison will involve distant relatives that vary in many traits, potentially complicating the analysis of particular traits of interest. From a phylogenetic point of view, comparisons of distant relatives are like an experiment with multiple uncontrolled variables (Garland and Adolph, 1994; Garland, 2001). To quote Felsenstein (1985, p. 465), '*Comparative biologists tend to suspect comparisons of distantly related species; they hope to base their comparisons on recent evolutionary events that have not been overlaid by much subsequent change*'. In principle, it might be possible to control for confounding traits that differ in distant relatives by including additional independent variables in the analysis, but it is often difficult to know *a priori* what those traits might be, let alone actually obtain quantitative data for them. In any case, an example in which casting too broad a net seems to reduce statistical power is provided by a study of body mass evolution in birds (Garland and Ives, 2000, p. 354). A comparison of passerines with their sister clade indicates that the former have significantly smaller log body masses, on average, whereas a comparison of passerines with all birds (including their sister clade) does not. (It should be noted that the identity of the sister clade of passerines is controversial, and the foregoing example may well change as improved phylogenetic information becomes available.) A related topic is whether one might *a priori* exclude certain subclades from a comparative analysis because they are 'unusual' as compared with the larger clade in general. For example, many studies of lizards (e.g. Perry and Garland, 2002) exclude snakes. A recent study by Bininda-Emonds and Gittleman (2000) suggests that this sort of *a priori* data exclusion may be less warranted than is often presumed.

A particularly powerful comparative design is one that has several different pairs of closely related species that differ in

the variable of interest (e.g. high and low temperature) and has these species pairs relatively distantly related to each other (i.e. in different branches of the phylogeny). As noted by Garland (2001), a particularly favorable distribution of this sort has the power to detect significant associations even when conventional statistical methods fail to do so. However, some workers choose species in this way, but then analyze only the pairs of tip species rather than performing a full analysis of the entire phylogeny (e.g. Lavergne et al., 2004). If that is done, Type I error rates should be correct, and the analysis should be robust with respect to errors in branch lengths and/or model of evolution, but statistical power will likely be lost (see Ackerly, 2000). A more extreme analytical variant is to perform a sign test on the tip pairs (Felsenstein, 1985), thus not using any information on branch lengths, but this comes at the extreme loss of statistical power (Ackerly, 2000).

The worst, that is, the least powerful, comparative design is one in which all species on one side of the root of the tree share, say, high values for an independent variable of interest (e.g. high temperature) and those on the other side of the root share low values (e.g. low temperature) (e.g. Garland et al., 1993; Garland, 2001). Although some methods can enhance inferential power in such situations (e.g. Schondube et al., 2001), it is not an attractive comparative scenario.

How many species or other taxa need to be included in a comparative study? In general, the statistical power of phylogenetically based analyses, when applied with an accurate phylogeny and model of character evolution, is the same as for conventional statistical methods, so standard power calculations can be employed (e.g. see fig. 5 in Garland and Adolph, 1994). However, it is also true that phylogenetic analyses sometimes uncover relationships that were not apparent in conventional analyses (see below).

Examples of the utility of incorporating a phylogenetic perspective

One obvious utility of phylogenetically based analytical methods is their ability to estimate ancestral values (Schluter et al., 1997; Martins and Lamont, 1998; Cunningham et al., 1998; Garland et al., 1999). Such estimations would be impossible without phylogenetic information. Using phylogenetic methods, one can ask where a particular trait arose within the evolutionary diversification of a group and whether it arose once or multiple times (e.g. Schondube et al., 2001; Reznick et al., 2002; Johnston et al., 2003; Espinoza et al., 2004; Hodges, 2004; Lane et al., 2004; Rezende et al., 2004; Berenbrink et al., 2005). Inferences about nodal values permit the estimation of evolutionary change along each branch segment of an evolutionary tree and hence analysis of features correlated with that change, including inferences regarding the order of evolution of the components of a complex trait (e.g., see Lauder, 1980, 1981, 1990; Pagel, 1999; Autumn et al., 2002; Pérez-Barbería et al., 2002; Oakley, 2003; Berenbrink et al., 2005).

We will now review just a few examples from the literature

in which phylogenetic information has added to our understanding and interpretation of comparative data. The first two of these deal with the evolution of lower metabolic rate in endotherms as part of their adaptation to desert environments.

It has long been recognized that low metabolic rates, low body temperatures, and an ability to become torpid would be beneficial to endotherms in hot, arid environments, in order to minimize heat load and energy (and water) demands in environments of low productivity (e.g. Dawson and Bartholomew, 1968; Dawson and Hudson, 1970; Williams, 1996; Tieleman et al., 2003; Rezende et al., 2004). When these traits were first discovered in desert caprimulgid birds (e.g. poorwills and nighthawks), they were initially interpreted as adaptations to desert conditions (e.g. Bartholomew et al., 1962). That is, they were seen as part of the evolutionary changes that permitted the occupation of hot, arid environments. Subsequently, however, it was recognized that other species of caprimulgids from more mesic and even tropical areas also had low metabolic rates and could also become torpid (e.g. Dawson and Schmidt-Nielsen, 1964; Lasiewski et al., 1970). A recent phylogenetic analysis (Lane et al., 2004) confirms that some of these traits exist in even the most basal members of the clade (but not the sister group, owls). Therefore, this constellation of thermoregulatory traits might be more general for this group and not an evolutionary adaptation to desert environments *per se*. Thus, while the possession of low metabolic rates and torpor ability may have facilitated the occupation of arid environments by caprimulgids, and thus constituted a 'preadaptation', these traits do not appear to have evolved as adaptations to them.

The hypothesis of low resting metabolic rates during adaptation to desert environments was also tested in the group Procyonidae (raccoons and their relatives). Chevalier (1991) measured metabolic rates of individuals from desert and mesic populations of ringtails, and from single populations of four other procyonids. These data can be analyzed *via* a conventional least-squares linear regression of log metabolic rate on log body mass. The procedure is to exclude the desert ringtail population, fit the regression line, and then compute the one-tailed 95% prediction interval for a new observation (see fig. 4B in Garland and Ives, 2000). The desert ringtail population falls below the regression line, consistent with the hypothesis of adaptation, but not outside the prediction interval, and thus not 'significantly' so. If the same procedure is followed with phylogenetically independent contrasts, the ringtail datum falls *far* outside the prediction interval (fig. 4C in Garland and Ives, 2000; see also Garland and Adolph, 1994) and thus the low metabolic rate of the desert population can be associated with desert occupation by this group.

Why the large difference in results? With a conventional analysis, each of the five data points is weighted equally for both computing the regression line and the prediction interval, and the place of the datum to be predicted (the desert ringtail's metabolic rate) is not considered in the sense that a star phylogeny (e.g. Figs 3A, 4A) is assumed, mathematically speaking. In the phylogenetic approach, two differences occur.

First, the data points are weighted differentially when the regression line is computed, so it differs somewhat from the conventional line. Second, for computing the prediction interval, the algebra specifically recognizes that the desert ringtail population has a very close relative, the mesic ringtail population, which has a fairly high metabolic rate, and thus the prediction is 'pulled' to a higher value. This makes intuitive sense because we would generally expect a close relative to be a better predictor of an unmeasured organism's phenotype as compared with the prediction derived from one (or several) less closely related species. The two effects together, but particularly the second (see fig. 4 in Garland and Ives, 2000), weight the comparison to be primarily between the desert and mesic ringtail populations (i.e. between the tip of interest and its closest relative in the data set), whereas the conventional analysis just compares the focal tip to all other values in a general, unprincipled way, thus losing statistical power.

We realize that sometimes it seems that phylogenetic methods only reduce analytical power and may obscure real relationships. However, the procyonid example is an instance where incorporating phylogenetic information actually supports an adaptive hypothesis that would not be found with a conventional, non-phylogenetic analysis. For another example, see Al-kahtani et al. (2004) on the correlation between size-corrected kidney mass and habitat aridity in rodents.

Turning from an analysis of specific adaptive patterns to more general issues in comparative physiology, phylogenetically based methods can be equally useful there too. Perhaps one of the most familiar relationships in comparative data is that between basal metabolic rate (BMR) and body size, the famous 'mouse-to-elephant' curve. The allometric slope of this relationship and its interpretation have been debated endlessly in the literature. However, most calculations of that relationship are based on the same incorrect assumption of the independence of observations that historically characterized other comparative data. This can be particularly problematic for such data sets as that on BMR, where certain groups (e.g. rodents) tend to be over-represented in the observations and others (e.g. cetaceans) tend to be greatly under-represented. As discussed above, failure to account for the relationships among the taxa will overestimate effective sample size and underestimate error limits. This situation is not unique to the allometry of metabolism, and equally applies to all compilations of scaling relationships (e.g. Calder, 1984; Peters, 1984). Recent phylogenetically based recalculations of these relationships (e.g. Garland and Ives, 2000; Cruz-Neto et al., 2001; Hosken et al., 2001; Symonds and Elgar, 2002) often differ significantly from those of conventional analyses, including the value for slope of the allometric equation relating BMR to body size. For instance, conventional statistical methods produce a (log-log) slope of 0.670 for 254 species of birds, but four different calculations involving phylogenetically independent contrasts have slopes ranging from 0.709 to 0.759, and none of the 95% confidence intervals for these latter slopes include the value of 0.670 (but see reanalysis of a refined data set by McKechnie and Wolf,

2004). Clearly, debates about the interpretation of allometric slopes rest on infirm ground if the values in question have been incorrectly calculated (see also Nunn and Barton, 2000).

In further regard to the allometry of avian metabolism, there has been a longstanding debate (e.g. Lasiewski and Dawson, 1967) as to whether separate equations should be used for the allometric relationship of BMR in passerine and non-passerine birds (note that the later taxon is paraphyletic and liable to cause apoplexy among cladists). (This is just one of many examples in which possible 'grade shifts' are of interest, i.e. differences among clades; see also Garland et al., 1993; Ackerly, 1999; Purvis and Webster, 1999; Ackerly et al., 2000; Nunn and Barton, 2000; fig. 1 in Garland, 2001.) The incorporation of phylogeny into data analysis has resolved that debate: these two groups do not show a statistically significant difference in mass-corrected BMR (Reynolds and Lee, 1996; see also Rezende et al., 2002). In addition, further analysis of those data (Garland and Ives, 2000) revealed a very interesting evolutionary pattern in the passerines: the rates of evolution of both body size and size-corrected metabolic rate within this group are significantly less than those in other birds (see also McKechnie and Wolf, 2004). This finding may indicate that passerines are under more size and energetic constraints than other avian taxa in general. This result is an example of the utility of phylogenetic methods to our understanding of the evolution of physiological characters; uncovering this result would have been impossible without them.

Here it is worth noting as an aside that ANCOVAs and related techniques (reviewed in Harvey and Pagel, 1991) were traditionally applied to examine metabolic scaling and whether 'grade shifts' may be present. These analyses are 'phylogenetic' in the sense that taxonomic groupings, such as families or orders, are used as factors (main effects). If one presumes that these taxa are separate evolutionary lineages (clades), then phylogeny is being partly considered. However, orders, families, and even genera themselves contain hierarchical relationships of their constituent species, and so a taxonomy-derived ANCOVA cannot capture the entire richness of phylogenetic information that may be available. This is why we consider Felsenstein's IC (Felsenstein, 1985) to be the first *fully* phylogenetic comparative statistical method.

A final example involves the dietary, latitudinal and climatic correlates of BMR and maximal metabolic rate (under cold exposure) in rodents (Rezende et al., 2004). Although conventional multiple regression analyses indicated that diet, latitude and temperature could explain significant amounts of the interspecific variation in mass-corrected BMR, a phylogenetic analysis indicated that only latitude was a significant predictor. As most traits showed substantial phylogenetic signal, the latter analyses should be more reliable. Those authors point out that whereas several interspecific comparisons of mammalian BMR have reported a significant association with diet using conventional statistics, this association has not yet been supported with phylogenetically based methods. Diet, at least when scored in crude categories, tends to be strongly associated with phylogeny in mammals,

including rodents, so it is conceptually and statistically difficult to analyze dietary effects separate from phylogeny. As noted by Garland et al. (1993), quantitative information on diet should increase statistical power to detect associations with other traits. Indeed, a significant relationship between BMR and diet, scored quantitatively, was found in a recent phylogenetic analysis of the Carnivora (Muñoz-García and Williams, in press).

Some notes of caution

We hope that the foregoing sections have communicated our enthusiasm for incorporating a phylogenetic perspective into comparative analyses. Indeed, we believe it is a necessity. However, all techniques and methods have their shortcomings and difficulties. Here we point out some of the practical and theoretical limitations of phylogenetic comparative analyses.

First, one needs a phylogeny. If one does not already exist, then you need to derive it for your organisms of interest. This is obviously no trivial matter, especially if your principal interest is physiology and not systematics. One possible first step is to infer phylogeny from taxonomy, but this is especially risky for groups where the existing taxonomy was not derived from actual phylogenetic information (i.e. information about the branching order of past branching events). Even if existing taxonomic information is not positively misleading with respect to phylogeny, it will generally lead to working phylogenies that contain numerous soft polytomies – unresolved nodes depicting several taxa differentiating simultaneously rather than as a series of discrete bifurcations (Fig. 3B). Because they reflect uncertainty in our phylogenetic information, soft polytomies cause analytical problems for phylogenetic approaches. Analytical methods that adjust degrees of freedom (Purvis and Garland, 1993; Garland and Diaz-Uriarte, 1999; empirical example in Tieleman et al., 2003) or employ more sophisticated computer simulations (Housworth and Martins, 2001) are available, but result in lowered statistical power as compared with an analysis that used a fully resolved tree. Of course, it is also possible, perhaps in collaboration with a *bona fide* systematist, to construct a phylogeny using appropriate data and well-defined inferential procedures (e.g. Felsenstein, 2004). Indeed, the ready availability of DNA sequencing technology and computer programs (e.g. see <http://evolution.genetics.washington.edu/phylip/software.html>) has democratized the process, and many physiologists are now making their own trees (e.g. Block et al., 1993; Johnston et al., 2003; Tieleman et al., 2003).

In addition to the statistical and analytical concerns with regard to branch lengths and models of character evolution mentioned above, the basic accuracy of topological information will affect results. Phylogenies are only estimates of (hypotheses about) true but unknown (and probably unknowable) branching relationships. Any conclusions drawn from a study are ever susceptible to future modification or falsification by a revision of the phylogenetic hypothesis. This is not simply a theoretical concern. For example, the study on

lizard running speed and body temperature discussed above (Garland et al., 1991) revised the conclusions of an earlier (and partially phylogenetic) study (Huey and Bennett, 1987), partly because of a subsequent phylogenetic revision. It is possible that the conclusions will be revised again if the phylogeny is further revised. Along these lines, recent methods are making it possible to incorporate phylogenetic uncertainty directly into comparative analyses that include simultaneous estimation of the phylogeny from DNA sequence data (Huelsenbeck and Rannala, 2003; see also p. 693 in Butler and King, 2004), as was originally suggested by Felsenstein (1985).

Next is the issue of the number of taxa appropriate for a comparative analysis. One of us co-authored a paper entitled 'Why not to do two-species comparative studies ...' (Garland and Adolph, 1994), which pointed out that the absolute minimum number of taxa required is three, in order to provide at least one degree of freedom for hypothesis testing (and to allow deduction of the direction of character evolution). In practice, many more taxa will generally be required to achieve adequate statistical power and the desired level of coverage of both phylogenetic and putatively adaptive states (for discussions, see Garland and Adolph, 1994; Garland et al., 1997; Garland, 2001). Some of these taxa may be difficult or impossible to obtain. For example, one of Bennett's students (Eppley, 1996) wanted to include crab plovers for a study of the ontogeny of endothermy in charadriiform birds. Unfortunately (for multiple reasons), crab plovers nest by the Persian Gulf in Iraq and Iran, which were then at war. These and such other issues, such as obtaining collecting permits for many species in different geographical locations, can make comparative studies difficult in practical terms. In addition, a trade-off must exist between the number of species that can be studied and the depth of investigation for each species, although the latter can be overcome with time if investigators publish their methods and raw data in sufficient detail to allow subsequent cumulative comparative studies that combine new data with data mined from the literature (Mangum and Hochachka, 1998). When a comparative study incorporates literature data for multiple traits, it is often the case that missing data severely limit analyses (e.g. see Bininda-Emonds and Gittleman, 2000). Some recent comparative analyses have employed fairly sophisticated methods for dealing with missing data (e.g. Fisher et al., 2003), but phylogenetically based methods for maximizing effective sample size with missing data need to be developed (S. P. Blomberg, personal communication; see also related methods in Garland et al., 1999; Garland and Ives, 2000).

Although confidence intervals can be computed for estimates of ancestral states under simple evolutionary models, when the number of species is small these can be so wide that they include or even exceed the range of observed states at the tips of the phylogeny (see fig. 8 in Schluter et al., 1997; fig. 2 in Garland et al., 1999). Narrower limits can be calculated for larger phylogenies (e.g. see Laurin, 2004; K. E. Bonine, T. T. Gleeson and T. Garland, Jr, manuscript submitted for publication), but it must be kept in mind that most analytical

procedures assume a simple evolutionary model, such as Brownian motion. If this assumption is invalid, as when evolutionary trends have occurred, then estimates may be quite misleading (for some empirical examples, see Garland et al., 1999; Oakley and Cunningham, 2000; Webster and Purvis, 2002). Comparative historical analysis can thus never really know ancestral or intermediate states, but only conjecture about them. Only experimental evolutionary analyses, which establish ancestral state and observe intermediate states, can have that certainty (e.g. Bennett and Lenski, 1999; Oakley and Cunningham, 2000; Garland, 2001). However, it is also possible to include fossil taxa directly in a phylogenetic analysis (e.g. see Polly, 2001; Laurin, 2004; Ross et al., 2004; Hone et al., 2005), although rarely if ever for physiological traits. Although reaching decisions about inclusion/exclusion of taxa can be problematic (Garland et al., 1997), as in any comparative study, the inclusion of fossil taxa has great potential to increase both the accuracy and precision of estimates of ancestral states (e.g. see Laurin, 2004). For example, 'fossil' taxa can be added anywhere on a phylogeny, with branches of any length, including length of zero. Thus, one can, if desired, set the value of a trait at any node on a phylogenetic tree by adding to it what amounts to a 'ghost node' with a specified tip value. All of this can be done in our PDTREE program, and we encourage further theoretical and empirical work in this area (see also Laurin, 2004).

Conventional statistical analyses of comparative data typically treat mean values for species (or populations) as if they were estimated without error. This is often unavoidable if the data set includes values from the literature, as often only mean values have been reported. Nonetheless, it can cause problems. For example, as noted above, allometric slopes are often of interest in comparative physiology (e.g. see Calder, 1984; Peters, 1984; Reynolds and Lee, 1996; Clobert et al., 1998; Ricklefs and Nealen, 1998; Garland and Ives, 2000; Hosken et al., 2001; Symonds and Elgar, 2002; McGuire, 2003; McKechnie and Wolf, 2004; Muñoz-García and Williams, in press), and it is well known that least-squares linear regressions will tend to underestimate the slope when the independent variable (e.g. log body mass) includes error variance (e.g. see Rayner, 1985; Riska, 1991; Nunn and Barton, 2000). If information on the within-species variation is available (e.g. estimates of standard errors associated with each tip value), then 'measurement error models' can be employed (e.g. Fuller, 1987). An important area of current research is developing such methods for phylogenetic analyses (chapter 6 in Harvey and Pagel, 1991; Christman et al., 1997; Martins and Hansen, 1997; Martins and Lamont, 1998; Felsenstein, 2004; Garland et al., 2004; Housworth et al., 2004). In the context of phylogenetically independent contrasts, it is possible, in effect, to use estimates of tip standard errors to first lengthen terminal branches of the working phylogeny, then perform analyses (Garland et al., 2004). As has been noted by several workers (e.g. Purvis and Rambaut, 1985; Ricklefs and Starck, 1996; Purvis and Webster, 1999; Nunn and Barton, 2000), contrasts between

two tips (as opposed to those involving deeper nodes, whose branches are lengthened as part of the contrasts algorithm) that are connected by short branches fairly commonly appear as 'outliers' in analyses. Rather than reflecting a truly high rate of evolution since the tips in question diverged, such a pattern may instead reflect disproportionate effects of measurement error in the tip values (or errors in estimates of the branch lengths). Incorporation of information on the error associated with estimates of tip values thus has the potential to reduce this problem and hence allow greater confidence in terminal contrasts. This is important because many comparative studies intentionally include one or more species (or populations) of particular interest plus their closest available relatives, which are necessarily connected by relatively short branches. Interpretation may hinge critically on whether a given contrast between two tips is large in magnitude [e.g. see the ringtail example of Chevalier (1991) as discussed in Garland and Adolph (1994) and Garland and Ives (2000)]. Another point is that data quality is particularly important when close relatives are compared (Purvis and Webster, 1999).

Finally, we must remember that all comparative studies (phylogenetic or not) are inherently correlational and, taken alone, cannot demonstrate causality of relationships (but see Autumn et al., 2002). Only experiments can demonstrate causality. This also raises the issue of inference when the 'independent' variable of interest is highly confounded with phylogeny (clumped in particular parts of the tree). This situation often arises in studies of diet, which is usually categorized fairly crudely, e.g. as carnivore, omnivore, herbivore (e.g. see Garland et al., 1993; Perry and Garland, 2002; Rezende et al., 2004; but see also Muñoz-Garcia and Williams, in press). This sort of diet categorization often shows a high degree of phylogenetic clumping. Diet is also often significantly associated with some 'dependent' variable, such as body size-corrected metabolic rate, in a conventional statistical analysis, while a phylogenetically based statistical analysis shows much less support for the relationship (i.e. a higher P value). Two things must be noted. First, strong phylogenetic clumping of an independent variable (e.g. when diet tends to be uniform within clades but differ among clades) leads to low statistical power to detect its effect on a dependent variable (Garland et al., 1993; Vanhooydonck and Van Damme, 1999). Second, even if an effect is detected (e.g. $P < 0.05$), it cannot be logically attributed to diet without good reason to dismiss possible effects of other shared derived features (synapomorphies) of one or more of the clades. In the limit, a comparison of two clades suffers from many of the same inferential problems as does a comparison of two single species (Garland and Adolph, 1994).

Summary and perspectives

The comparative method has been progressively refined from simple analogy to a highly quantitative and statistically sophisticated scientific methodology. A century ago, for

instance, comparative studies rarely involved sufficient replication, let alone statistical evaluation of their results. Today, a comparative study that did not have these features would not be publishable. We are now in the midst of another progressive refinement of the comparative method, this time one that includes the historical (evolutionary) relationships of the organisms involved. If it is admitted that evolution has occurred and that different groups of organisms are differentially related to each other, then theoretical considerations and statistical models have shown that this information must be taken into account in analyses that involve multiple species. Otherwise, phylogenetic relationships enter the analysis implicitly as an uncontrolled variable that may lead to incorrect conclusions. We believe that one day application of phylogenetic methods in comparative physiology will be as routine as the use of statistical analysis in general, or the use of allometric equations and analysis of covariance to control for effects of body size. Some recent physiology textbooks support this prediction (as do evolution texts, e.g. Freeman and Herron, 2004). For example, Spicer and Gaston (1999) mention the use of IC and related methods, and note some physiological studies in which conclusions are altered by their application. Bradshaw (2003) includes two pages of text discussion about 'The comparative method' and an appendix with a partially worked example of phylogenetically independent contrasts. Willmer et al. (2005) also discuss phylogenetic perspectives, although they do not provide an example of calculations for independent contrasts.

We further believe that it is important to concentrate on the positive aspects of including phylogeny in the comparative method. It is the best way to remind ourselves continually that all functional characters are the products of evolution. This is the essence of evolutionary physiology: characters are not stable through time but are continually susceptible to modification. Phylogenetic comparative methods are a principal tool of evolutionary physiology to examine patterns and to infer processes of evolutionary change. They are uniquely positioned to permit us to speculate about ancestral conditions, as well as rates and patterns of evolution in historical time. Incorporation of phylogeny into the comparative method can and has served to expand the kinds of questions that biologists are capable of studying.

A final point is that phylogenetically based statistical analyses often suggest that evolutionary adaptation is just not as common as we once thought in comparative physiology (or at least it is hard to find strong empirical support for it). If these results are confirmed once adequate meta-analyses are performed (e.g. P. Carvalho, J. A. F. Diniz-Filho and L. M. Bini, personal communication), including due consideration of effects of errors in phylogenies, it could lead to an important reorientation of perspectives, given that earlier generations of comparative physiologists routinely *assumed* the presence of evolutionary adaptation in most if not all traits they studied, rather than seeing adaptation as one possible explanation for possession of a character (Feder et al., 1987, 2000; Garland

and Adolph, 1994; Garland and Carter, 1994; Bennett, 1997; Autumn et al., 2000).

Appendix

Phylogenetically based statistical methods in a nutshell

Although others exist (e.g. Thorpe et al., 1996; Rochet et al., 2000; Diniz-Filho and Torres, 2002; Paradis and Claude, 2002; Desvignes et al., 2003; Butler and King, 2004; Housworth et al., 2004), the three main phylogenetically based statistical methods are independent contrasts (IC), generalized least-squares models (GLS; Grafen, 1989; Martins and Hansen, 1997; Garland and Ives, 2000; Rohlf, 2001), and Monte Carlo computer simulations (as introduced above in the section entitled 'An example of how phylogeny can affect statistical analyses'). All of them can be applied to a wide range of analyses, including correlation, regression, analysis of variance and covariance, and principal components analysis. (However, various multivariate methods, such as canonical correlation and discriminant analysis, are poorly developed and represent important areas for future work.) They all share the same basic assumptions about the correctness of the topology and branch lengths (but see Huelsenbeck and Rannala, 2003). [They also share the assumption that 'measurement error' is an unimportant part of the among-species variation, although all can be extended to use information on such sources of variation (e.g. see chapter 6 in Harvey and Pagel, 1991; Christman et al., 1997; Martins and Hansen, 1997; Martins and Lamont, 1998; Felsenstein, 2004; Garland et al., 2004; Housworth et al., 2004).] Both IC and GLS analyses share the assumption that character evolution can be modeled as Brownian motion, or some analytically tractable variation thereof, whereas the Monte Carlo simulation approach is more flexible in this regard (e.g. Garland et al., 1993; Diaz-Uriarte and Garland, 1996, 1998; Price, 1997; Harvey and Rambaut, 2000). The methods can also be combined. For example, one could compute a correlation by IC and test its significance relative to simulated data. Indeed, for some of the more sophisticated phylogenetic comparative methods, computer simulations are the most reasonable or only way to perform hypothesis testing (e.g. see Pagel, 1999; Blomberg et al., 2003; Housworth et al., 2004; see also the PHYLOGR package available at <http://cran.r-project.org/>). Further details, advantages and limitations of these methods are beyond the scope of this paper and the reader is referred to the original papers. We now briefly discuss IC and GLS approaches.

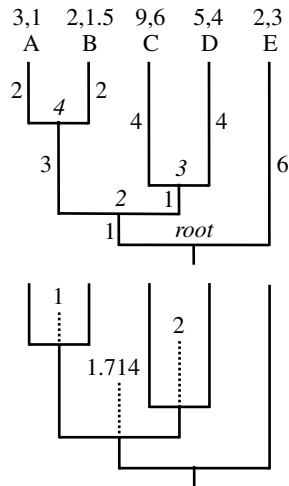
The IC method is an algorithm that transforms data to account for the differential relatedness of the taxa within a study (Felsenstein, 1985). It also turns out that analyses with contrasts yield numbers that are mathematically identical to those obtained through equivalent GLS analyses (Garland and Ives, 2000; Rohlf, 2001; see below). Thus, IC can be viewed simply as a clever algorithm to avoid the need to invert large matrices (Freckleton et al., 2003), and the algorithm was originally developed by Felsenstein (1973) in the context of attempting to estimate phylogenetic trees from continuous-valued characters.

Independent contrasts converts the original N measurements (which were non-independent of each other if they represent mean values for hierarchically related taxa) into $N-1$ contrasts of the measurements between pairs of related taxa or (estimated) ancestral nodes in the phylogeny. Computations are done for one trait at a time, as shown in Fig. 5. If multiple traits are involved in the analysis, then the contrasts calculated separately for each trait are used to compute a correlation, regression, multiple regression, etc. Worked examples for bivariate correlations can be found in Garland (1994, fig. 11.2) and Garland and Adolph (1994, fig. 2; also reproduced in box 9.2, pp. 348-349, of Freeman and Herron, 2004). Readers are cautioned that some published examples of the calculations (including in text books) are incorrect because they are oversimplified and do not properly use branch lengths as described in Felsenstein (1985). In addition, some publications have failed to calculate correlations and regressions through the origin (see below). Thus, readers are encouraged to validate any new computer program by running through one of the published examples listed above.

The goal of IC is thus to transform the original data into independent and equally distributed contrasts (assuming its assumptions are met) that are then amenable to standard statistical comparisons and analyses. The only constraint on statistical analyses of contrasts is that all models are forced through the origin (Garland et al., 1992). But this is actually just a requirement of the IC algorithm (Rohlf, 2001), and so it is also possible to calculate, for example, phylogenetically correct y -intercepts for regression equations (Garland et al., 1993; Garland and Ives, 2000). Although first presented in the context of correlation and regression, and most commonly used for such analyses, the algebra of IC also allows such univariate analyses as computing phylogenetically correct mean values (and standard errors) for clades (also interpretable as hypothetical ancestors: Garland et al., 1999), comparing average values of clades or ecologically defined groups (Garland et al., 1993; Rezende et al., 2004), and comparing average rates of evolution between clades (Garland, 1992; Garland and Ives, 2000; Hutcheon and Garland, 2004; McKechnie and Wolf, 2004). IC analyses can also be used for multivariate purposes, such as principal components analysis (PCA; e.g. Clobert et al., 1998; see also Ricklefs and Nealen, 1998).

Although not intuitively obvious, if one collapses a phylogeny to be a star by shortening all internal branches to zero length, while lengthening all terminal branches so that all tips remain or become contemporaneous (as in Figs 3A and 4A), then all of the results of IC calculations will be identical to those of conventional 'non-phylogenetic' analyses (Purvis and Garland, 1993). In fact, this is a good exercise to perform to verify that a particular computer program for computing IC is actually working correctly. In any case, a conventional analysis, which mathematically assumes a star phylogeny with contemporaneous tips, can be viewed as a special case of a phylogenetic analysis (Garland et al., 1999).

Grafen (1989) first introduced GLS models (his 'standard



Phylogenetically independent contrasts calculations for trait 1

Node name	Contrast name	Contrast values	Raw contrast	Standard deviation	Standardized contrast	(Corrected) length of each branch	
						Left	Right
<i>4</i>	A–B	3–2	1.0	2.0	0.5	2.0	2.0
<i>3</i>	C–D	9–5	4.0	2.828	1.414	4.0	4.0
<i>2</i>	4–3	2.5–7	–4.5	2.646	–1.701	[4.0]	[3.0]
<i>root</i>	2–E	5.071–2	3.071	2.952	1.040	[2.714]	6.0

Fig. 5. Worked example of Felsenstein's phylogenetically independent contrasts (IC; Felsenstein, 1985). For the upper tree: letters at tips (terminal nodes) are names of five species (A–E); numbers at tips are hypothetical data for two phenotypic traits (e.g. body mass, bill length); numbers adjacent to branches indicate lengths in units of expected variance of character evolution (which would be proportional to divergence times under a simple Brownian motion model of character evolution; see Fig. 2 and text); names for internal nodes are in italics. For the bottom tree: broken lines and numbers indicate amounts by which internal branches are lengthened during calculations. The goal of the IC algorithm is to use phylogenetic information to transform the N original, non-independent species values into $N-1$ independent and identically distributed contrasts. These values can then be used in conventional statistical procedures, with the constraint that all calculations (e.g. correlation, regression) are computed through the origin. The algorithm begins at the tips of the phylogenetic tree with sister taxa (typically species). In this example, these would be species A and B as well as C and D. For these two contrasts (which can also be identified by their basal node (4 and 3, respectively), the difference in phenotypes between the pair members is first computed. These 'raw contrasts' are then divided by their 'standard deviations,' which are the square roots of the sums of the branch lengths. This yields 'Standardized Contrasts,' which are the values actually used for subsequent analyses.

Contrasts can also be calculated for deeper nodes in the tree if we estimate the phenotypes of hypothetical ancestors. For node 4, the appropriate value (assuming Brownian motion evolution with no directional trend) would be the average of its daughter species A and B, i.e. a value of 2.5 for trait 1. Similarly, for node 3, the value is 7. The raw contrast (difference) between nodes 4 and 3 is thus a value of $2.5-7=-4.5$. This value would then be divided by its standard deviation. However, this contrast is different from those

calculated at the tips of the tree. The contrast of node 4 vs 3 uses two estimated phenotypes, not actual measured data. Therefore, we should devalue it relative to the tip contrasts. This is accomplished by lengthening the branches leading to internal nodes. As explained in Felsenstein (1985), Under Brownian motion, the amount of lengthening is computed as: (daughter branch length 1 \times daughter branch length 2) / (daughter branch length 1 + daughter branch length 2). Thus, the branch leading to node 4 is lengthened by 1, giving a corrected length of 4. The branch to node 3 is lengthened by 2, for a corrected length of 3. These corrected values are then used to compute the standard deviation of contrast 4 vs 3.

To compute the contrast between internal node 2 and terminal node E, the value at node 2 must be estimated. It is taken as a weighted average, with weights inversely proportional to the lengths of the branches leading to its daughter nodes 3 and 4. The value computed for node 2 will, therefore, be more similar to the value at node 3 because the branch leading to it (corrected length=3) is shorter than the one leading to node 4 (corrected length=4). The branch leading to node 2 is also lengthened, with computations that use the already-lengthened branches above node 2.

The IC algorithm involves iterative calculations best accomplished by computer programs, such as the PDTREE module of the Phenotypic Diversity Analysis Programs (available from T.G.) or the corresponding module in Mesquite (<http://mesquiteproject.org/mesquite/mesquite.html>). Readers are cautioned that some computer programs do not properly lengthen internal branches or have other errors. One way to check whether a program is performing the correct calculations is to collapse the internal branches of the phylogeny to be length zero, and then make all of the terminal branches (those leading to tips) equal in length. This yields a 'star' phylogeny (e.g. Figs 3A and 4A). This star phylogeny can then be used to compute statistics with independent contrasts, such as the Pearson correlation (through the origin)

between two traits. This value should be identical to the value obtained with a conventional statistical program (not through the origin). If the values differ, the IC program is doing something incorrectly.

Most programs for independent contrasts will output the contrasts and some simple statistics, such as the correlation between two sets of contrasts. For more complicated analyses, such as multiple regression or principal components analysis, one will generally need to write out the contrasts and then read them back into a commercial statistics package, most of which will allow the user to specify that calculations of correlations, regressions, etc., be done through the origin.

If standardized contrasts are calculated for the second trait shown in the figure, then the Pearson correlation (through the origin) with trait 1 is 0.85289. This would be considered not statistically significant as compared with a critical value of 0.878 for a two-tailed test with three degrees of freedom. For comparison, the conventional Pearson correlation of the tip data (mathematically equivalent to performing the contrasts calculations on a star phylogeny) is 0.88024, which would be considered marginally significant. Obviously, one should not get too exorcised about such small differences in parameter estimates or *P* values, but this example serves to make the point that results will always be somewhat different when the phylogeny is specified as being hierarchical rather than a star. For the least-squares linear regression slope, values are 0.55003 for IC and 0.60057 for conventional; corresponding *y*-intercepts are 1.00724 (calculated as described in Garland et al., 1993; Garland and Ives, 2000) and 0.57759.

Note that the value at the very bottom of the tree, the basal or *root* node, can also be estimated, but for this data set there is nothing with which to contrast it. Therefore, Felsenstein (1985) did not show calculation of the *root* node value in his example. However, this value is of interest because it represents (1) the phylogenetically correct estimate of the mean value for all five species and also (2) an estimate of the value for the hypothetical ancestor. As discussed elsewhere (e.g. Garland et al., 1999), both interpretations hinge on the Brownian motion assumption, and the latter also requires that no directional trend in character evolution has occurred. One can also compute standard errors and confidence intervals for the *root* node (Garland et al., 1999). For trait 1, the value computed by our PDTREE program is 4.115 with a standard error (S.E.) of ± 1.7055 and a 95% confidence interval of -0.6204 to 8.8499 . Conventional values (which are identical to those obtained after collapsing the phylogeny to be a star) are 4.200 with S.E. ± 1.3191 and 95% CI of 0.5376 to 7.8624 . Thus, the phylogenetic point estimate is slightly different from the conventional one, but the S.E. and CI are substantially wider. As also noted in the text, this general pattern of wide S.E. values and CIs for estimates at basal nodes has been pointed out before (Garland et al., 1999), and emphasizes that it is often hard to infer ancestral values with much confidence (Schluter et al., 1997; Cunningham et al., 1998). Finally, note that the values given above, which are derived literally by the algebra of phylogenetically independent contrasts in the PDTREE program (Garland et al., 1999), are the same as those obtained by GLS (e.g. see box 3 in Cunningham et al., 1998), and can also be obtained from our REGRESSION.M Matlab program (Blomberg et al., 2003).

Importantly, for nodes not at the base of the phylogeny, the values estimated during the IC algorithm have no special meaning. Rather, they should be viewed merely as intermediate steps in the calculations. Only the value for the root node is equivalent to the maximum likelihood estimate, which can be obtained by GLS

or by the 'squared-change parsimony' algorithm (Schluter et al., 1997; Cunningham et al., 1998; Garland et al., 1999). However, an interesting property of the IC algorithm is that if you reroot the tree at an internal node (as can be done in our PDTREE program) then the value obtained for that node will be the same as the GLS value (Garland et al., 1999; empirical examples there and in Laurin, 2004). Although such rerooting when performing the IC algorithm affects the values estimated at nodes, it does not affect the estimated correlation or slope between two traits.

For this example data set, the diagnostic Pearson correlation (not through the origin) between the absolute values of the standardized contrasts and their standard deviations (Garland et al., 1992) is 0.663 for trait 1 and 0.144 for trait 2, neither of which is statistically significant (two-tailed critical value for two d.f.=0.950). This would suggest that the branch lengths are adequate for the Brownian motion assumption of independent contrasts, but we should be cautious with so few data points. (Note also that the hypothetical tip values shown in the figure were just made up, not actually produced by a Brownian motion process.)

Blomberg et al. (2003) derived a statistic, termed *K*, which indicates the amount of 'phylogenetic signal' in the tip data relative to the expectation for a trait that evolved by Brownian motion along the specified topology and branch lengths. A value of 1.00 indicates exactly the amount expected. Values < 1 indicate less resemblance of relatives than expected, and values > 1 indicate stronger-than-expected phylogenetic 'clumping' of trait values. For trait 1, $K=1.084$, indicating that the tendency for related species to resemble each other is, averaged over the entire tree, just about what one would expect if the traits had actually evolved by a process similar to Brownian motion. The PHYSIG.M Matlab program can be obtained from T.G. on request.

As discussed in the text, Blomberg et al. (2003) also present a randomization test for the presence of phylogenetic signal (also implemented in PHYSIG.M). For trait 1, 247 of 1000 randomized data sets (tip values shuffled randomly) yielded a mean squared error that was less than or equal to the value of 8.0264 for the data in their correct position. Thus, the presence of signal would not be considered statistically significant at the usual $P < 0.05$. However, computer simulations demonstrate that approximately 20 species are required to achieve a statistical power of 80% for detecting signal when it is present (Blomberg et al., 2003), so this result should not be taken to suggest that phylogenetic signal is actually lacking for these example data. (Note that the *K* statistic does not depend on sample size and is a valid descriptive statistic for the amount of signal even for small data sets such as this.)

Finally, the matrix derived from this tree that would be used in GLS calculations (see text) is as follows (as produced by our PDDIST program as a DSC file):

	A	B	C	D	E
A	6	4	1	1	0
B	4	6	1	1	0
C	1	1	6	2	0
D	1	1	2	6	0
E	0	0	0	0	6

Other examples of this sort of matrix can be found in Cunningham et al. (1998, box 3) and Freckleton et al. (2002, fig. 1).

regression') as a way to incorporate phylogenetic information, and noted that they were a generalization of IC. But his paper attempted to be more general by assuming that (1) the 'working' phylogeny included soft polytomies, representing uncertainty about true branching order, and (2) that no starter branch lengths were available, so they had to be assigned by some arbitrary rule, one of which he offered. He also proposed a type of branch length transformation to optimize fit of the tree to the tip data. Grafen offered computer code written in GLIM, a commercial package that is not widely used by comparative biologists, to implement what he termed the 'phylogenetic regression'. As a result, his method was neither adequately appreciated nor widely used. Subsequently, phylogenetic GLS methods were promoted by other workers (Martins and Hansen, 1997; Butler et al., 2000; Garland and Ives, 2000; Rohlf, 2001) and are now becoming a routine comparative tool (e.g. see Diniz-Filho and Torres, 2002; Freckleton et al., 2002; Blomberg et al., 2003; Housworth et al., 2004).

It is important to reiterate that given the same tip data, phylogenetic information (topology and branch lengths), and assumed model of evolution, IC and GLS methods yield identical results. IC is thus a special case of GLS models. Worked examples of a univariate GLS analysis can be found in Cunningham et al. (1998, box 3) and Freckleton et al. (2002). Because phylogenetic GLS analyses are not as familiar as IC, we will briefly explain how they work. First, the phylogenetic tree is converted to a symmetrical matrix that is intended to represent the expected variances and covariances of the tip data or, if in a regression model, those of the residuals. The diagonals of this matrix indicate the expected variances, and are taken simply as the total branch length distance from the root to each tip. If the tree has contemporaneous tips (e.g. as in Fig. 4), then all these values will be the same. These values thus represent the putative total opportunity for evolutionary change that each species has experienced (since the basal split of the tree). The off-diagonals are taken as the amount of branch length that is *shared* by any two tips, i.e. from the root of the tree to last common ancestor.

Once the phylogenetic tree has been converted into a variance-covariance matrix, its incorporation into standard statistical analyses is actually rather intuitive. For example, in a standard linear regression model it is assumed that the expected variance-covariance matrix of the residuals is the identity matrix, which has values of unity for all diagonals and values of zero for all off-diagonal elements. Standard 'weighted regression' is simply a variant of this in which the diagonal elements need not be the same value, which has the effect of giving the data points different 'pull' in computing the regression equation. Many common statistical packages allow one to perform weighted regression. The phylogenetic GLS regression is essentially the same, except that the specified matrix can now have off-diagonal elements that are not all zeros. Note that if one specifies the phylogeny to be a star, then the matrix is the identity matrix, so GLS methods – as with IC – can yield standard statistical results.

Both IC and GLS analyses can be modified to incorporate some other models of character evolution. In essence, this is done by transforming the branch lengths of the working phylogenetic tree. As with data transformations in conventional statistics, transformations of branch lengths can be done from a purely statistical perspective (e.g. Grafen, 1989; Garland et al., 1992; Diaz-Uriarte and Garland, 1996, 1998; Freckleton et al., 2002) or in a fashion intended to mimic some model of character evolution, such as the Ornstein-Uhlenbeck (OU) process, a more complex model that has been used to mimic characters under stabilizing selection (see Felsenstein, 1988; Garland et al., 1993; Diaz-Uriarte and Garland, 1996; Martins and Hansen, 1997; Blomberg et al., 2003; Housworth et al., 1994). No studies have yet examined whether one approach or the other yields better statistical performance, but transforms designed to mimic explicit models facilitate inferences with respect to parameters of those models.

Although they are functionally equivalent for most purposes, IC and GLS approaches differ in how intuitive they are for certain analyses, and also in terms of what software is available for their implementation. For example, with IC, 'tree thinking' is retained, which facilitates graphical analyses, identification of places (bifurcations) in a phylogeny where rapid evolutionary events occurred, and also suggests more intuitively such procedures as rerooting along branches to reconstruct hypothetical ancestors as direct descendants or to predict values of unmeasured species (see Garland et al., 1999; Garland and Ives, 2000; Reynolds, 2002; Ross et al., 2004). With IC, it is easier to employ different sets of branch lengths for different traits (e.g. Garland et al., 1992; Lovegrove, 2003; Rezende et al., 2004), which may be particularly useful when one trait does not actually show phylogenetic signal (e.g. Tieleman et al., 2003; Rheindt et al., 2004) and/or for traits that are only nuisance variables, such as details of measurement or calculation methods that differ among studies (e.g. Wolf et al., 1998; Perry and Garland, 2002; Rezende et al., 2004). With GLS, on the other hand, once the phylogenetic variance-covariance matrix has been constructed, a variety of commercial and free statistical and matrix algebra packages can be used (e.g. SAS, as in Butler et al., 2000; the Matlab PHYSIG package of Blomberg et al., 2003; PHYLOGR and APE in the R language, available at <http://cran.r-project.org/>).

This work was supported by NSF grants: DEB-0196384 to T.G. and A. R. Ives; DEB-0416085 to D. N. Reznick, T.G. and M. S. Springer; IBN-9905980 and IBN-0091308 to A.F.B. We thank S. P. Blomberg, A. Diniz-Filho, and K. Phillips for comments on earlier versions of the manuscript, and R. E. Ricklefs for extensive comments on the original submitted version.

References

- Abouheif, E.** (1999). A method to test the assumption of phylogenetic independence in comparative data. *Evol. Ecol. Res.* **1**, 895-909.
- Ackerly, D. D.** (1999). Comparative plant ecology and the role of phylogenetic information. In *Physiological Plant Ecology. The 39th Symposium of the*

- British Ecological Society, University of York 7-9 September 1998 (ed. M. C. Press, J. D. Scholes and M. G. Braker), pp. 391-413. Oxford, UK: Blackwell Science.
- Ackerly, D. D.** (2000). Taxon sampling, correlated evolution, and independent contrasts. *Evolution* **54**, 1480-1492.
- Ackerly, D. D.** (2004). Adaptation, niche conservatism, and convergence: Comparative studies of leaf evolution in the California chaparral. *Am. Nat.* **163**, 654-671.
- Ackerly, D. A., Dudley, S. A., Sultan, S. E., Schmitt, J., Coleman, J. S., Linder, R., Sandquist, D. R., Geber, M. A., Evans, A. S., Dawson, T. E. et al.** (2000). The evolution of plant ecophysiological traits: recent advances and future directions. *BioScience* **50**, 979-995.
- Al-kahtani, M. A., Zuleta, C., Caviedes-Vidal, E. and Garland, T., Jr** (2004). Kidney mass and relative medullary thickness of rodents in relation to habitat, body size, and phylogeny. *Physiol. Biochem. Zool.* **77**, 356-365.
- Andrews, R. M. and Pough, F. H.** (1985). Metabolism of squamate reptiles: Allometric and ecological relationships. *Physiol. Zool.* **58**, 214-231.
- Aparicio, J. M., Bonal, R. and Cordero, P. J.** (2003). Evolution of the structure of tail feathers: implications for the theory of sexual selection. *Evolution* **57**, 397-405.
- Ashton, K. G.** (2004a). Comparing phylogenetic signal in intraspecific and interspecific body size datasets. *J. Evol. Biol.* **17**, 1157-1161.
- Ashton, K. G.** (2004b). Sensitivity of intraspecific latitudinal clines of body size for tetrapods to sampling, latitude and body size. *Integr. Comp. Biol.* **44**, 403-412.
- Autumn, K., Ryan, M. J. and Wake, D. B.** (2002). Integrating historical and mechanistic biology enhances the study of adaptation. *Quart. Rev. Biol.* **77**, 383-408.
- Bartholomew, G. A., Hudson, J. W. and Howell, T. R.** (1962). Body temperature, oxygen consumption, evaporative water loss, and heart rate in the poor-will. *Condor* **64**, 117-125.
- Bauwens, D., T., Garland, A. M., Jr, Castilla and Van Damme, R.** (1995). Evolution of sprint speed in lacertid lizards: morphological, physiological, and behavioral covariation. *Evolution* **49**, 848-863.
- Bennett, A. F.** (1997). Adaptation and the evolution of physiological characters. In *Handbook of Physiology. Section 13, Comparative Physiology*, Vol. I (ed. W. H. Dantzler), pp. 3-16. New York: Oxford University Press.
- Bennett, A. F.** (2003). Experimental evolution and the Krogh Principle: generating biological novelty for functional and genetic analyses. *Physiol. Biochem. Zool.* **76**, 1-11.
- Bennett, A. F. and Lenski, R. E.** (1999). Experimental evolution and its role in evolutionary physiology. *Am. Zool.* **39**, 346-362.
- Berenbrink, M., Koldkjær, P., Kepp, O. and Cossins, A. R.** (2005). Evolution of oxygen secretion in fishes and the emergence of a complex physiological system. *Science* **307**, 1752-1757.
- Bininda-Emonds, O. R. P. and Gittleman, J. L.** (2000). Are pinnipeds functionally different from fissiped carnivores? The importance of phylogenetic comparative analyses. *Evolution* **54**, 1011-1023.
- Block, B. A., Finnerty, J. R., Stewart, A. F. R. and Kidd, J.** (1993). Evolution of endothermy in fish: mapping physiological traits on a molecular phylogeny. *Science* **260**, 210-214.
- Blomberg, S. P. and Garland, T., Jr** (2002). Tempo and mode in evolution: Phylogenetic inertia, adaptation and comparative methods. *J. Evol. Biol.* **15**, 899-910.
- Blomberg, S. P., Garland, T., Jr and Ives, A. R.** (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* **57**, 717-745.
- Bradshaw, S. D.** (2003). *Vertebrate Ecophysiology: An Introduction to its Principles and Applications*. Cambridge: Cambridge University Press.
- Brooks, D. R. and McLennan, D. A.** (2002). *The Nature of Diversity: An Evolutionary Voyage of Discovery*. Chicago: University of Chicago Press.
- Butler, M. A. and King, A. A.** (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* **164**, 683-695.
- Butler, M. A., Schoener, T. W. and Losos, J. B.** (2000). The relationship between sexual size dimorphism and habitat use in Greater Antillean *Anolis* lizards. *Evolution* **54**, 259-272.
- Calder, W. A.** (1984). *Size, Function, and Life History*. Cambridge, MA: Harvard University Press.
- Chevalier, C. D.** (1991). Aspects of thermoregulation and energetics in the Procyonidae (Mammalia: Carnivora). PhD thesis, University of California, Irvine, USA. 202 pp.
- Christman, M. C., Jernigan, R. W. and Culver, D.** (1997). A comparison of two models for estimating phylogenetic effect on trait variation. *Evolution* **51**, 262-266.
- Clobert, J., Garland, T., Jr and Barbault, R.** (1998). The evolution of demographic tactics in lizards: a test of some hypotheses concerning life history evolution. *J. Evol. Biol.* **11**, 329-364.
- Cox, R. M., Skelly, S. L. and John-Alder, H. B.** (2003). A comparative test of adaptive hypotheses for sexual dimorphism in lizards. *Evolution* **57**, 1653-1669.
- Cruz-Neto, A. P., Garland, T., Jr and Abe, A. S.** (2001). Diet, phylogeny, and basal metabolic rate in phyllostomid bats. *Zoology* **104**, 49-58.
- Cunningham, C. W., Omland, K. E. and Oakley, T. H.** (1998). Reconstructing ancestral character states: a critical reappraisal. *Trends Ecol. Evol.* **13**, 361-366.
- Dawson, W. R. and Bartholomew, G. A.** (1968). Temperature regulation and water economy of desert birds. In *Desert Biology*, Vol. I (ed. G. W. Brown), pp. 357-394. New York: Academic Press.
- Dawson, W. R. and Hudson, J. W.** (1970). Birds. In *Comparative Physiology and Thermoregulation*, Vol. I (ed. G. C. Whittow), pp. 223-310. New York: Academic Press.
- Dawson, T. J. and Hulbert, A. J.** (1970). Standard metabolism, body temperature and surface areas of Australian marsupials. *Am. J. Physiol.* **218**, 1233-1238.
- Dawson, W. R. and Schmidt-Nielsen, K.** (1964). Terrestrial animals in dry heat: Desert birds. In *Handbook of Physiology*, **4**, *Adaptation to the Environment* (ed. D. B. Dill), pp. 481-492. Washington DC: American Physiological Society.
- Dawson, W. R., Bartholomew, G. A. and Bennett, A. F.** (1977). A reappraisal of the aquatic specializations of the Galapagos marine iguana (*Amblyrhynchus cristatus*). *Evolution* **31**, 891-897.
- Desdevises, Y., Legendre, P., Azouzi, L. and Morand, S.** (2003). Quantifying phylogenetically structured environmental variation. *Evolution* **57**, 2647-2652.
- de Queiroz, K.** (2000). Logical problems associated with including and excluding characters during tree reconstruction and their implications for the study of morphological character evolution. In *Phylogenetic Analysis of Morphological Data* (ed. J. J. Wiens), pp. 192-212. Washington, DC: Smithsonian Institution Press.
- Diaz-Uriarte, R. and Garland, T., Jr** (1996). Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. *Syst. Biol.* **45**, 27-47.
- Diaz-Uriarte, R. and Garland, T., Jr** (1998). Effects of branch length errors on the performance of phylogenetically independent contrasts. *Syst. Biol.* **47**, 654-672.
- Diniz-Filho, J. A. F.** (2000). *Metodos Filogeneticos Comparativos*. Ribeirao Preto, Brazil: Holos.
- Diniz-Filho, J. A. F. and Torres, N. M.** (2002). Phylogenetic comparative methods and the geographic range size-body size relationship in new world terrestrial carnivora. *Evol. Ecol.* **16**, 351-367.
- Eppley, Z. A.** (1996). Charadriiform birds in Antarctica: behavioral, morphological, and physiological adjustments conserving reproductive success. *Physiol. Zool.* **69**, 1502-1554.
- Espinoza, R. E., Wiens J. J. and Tracy, C. R.** (2004). Recurrent evolution of herbivory in small, cold-climate lizards: breaking the ecophysiological rules of reptilian herbivory. *Proc. Natl. Acad. Sci. USA* **101**, 16819-16824.
- Feder, M. E., Bennett, A. F., Burggren, W. W. and Huey, R. B. (eds)** (1987). *New Directions in Ecological Physiology*. New York: Cambridge University Press.
- Feder, M. E., Bennett, A. F. and Huey, R. B.** (2000). Evolutionary physiology. *Annu. Rev. Ecol. Systemat.* **31**, 315-341.
- Felsenstein, J.** (1973). Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Gen.* **25**, 471-492.
- Felsenstein, J.** (1985). Phylogenies and the comparative method. *Am. Nat.* **125**, 1-15.
- Felsenstein, J.** (1988). Phylogenies and quantitative characters. *Annu. Rev. Ecol. Syst.* **19**, 445-471.
- Felsenstein, J.** (2004). *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates.
- Fisher, D. O., Blomberg, S. P. and Owens, I. P. F.** (2003). Extrinsic versus intrinsic factors in the decline and extinction of Australian marsupials. *Proc. R. Soc. Lond. B* **270**, 1801-1808.
- Freckleton, R. P., Harvey, P. H. and Pagel, M.** (2002). Phylogenetic analysis and comparative data: A test and review of evidence. *Am. Nat.* **160**, 712-726.

- Freckleton, R. P., Pagel, M. and Harvey, P. H.** (2003). Comparative methods for adaptive radiations. In *Macroecology: Reconciling Divergent Perspectives on Large Scale Ecological Processes* (ed. K. J. Gaston and T. M. Blackburn). Oxford, UK: Blackwell Science.
- Freeman, S. and Herron, J. C.** (2004). *Evolutionary Analysis*. 3rd edn. Upper Saddle River, NJ: Prentice Hall.
- Fuller, W. A.** (1987). *Measurement Error Models*. New York: John Wiley and Sons.
- Garland, T., Jr** (1992). Rate tests for phenotypic evolution using phylogenetically independent contrasts. *Am. Nat.* **140**, 509-519.
- Garland, T., Jr** (1994). Phylogenetic analyses of lizard endurance capacity in relation to body size and body temperature. In *Lizard Ecology: Historical and Experimental Perspectives* (ed. L. J. Vitt and E. R. Pianka), pp. 237-259. Princeton, NJ: Princeton University Press.
- Garland, T., Jr** (2001). Phylogenetic comparison and artificial selection: Two approaches in evolutionary physiology. In *Hypoxia: From Genes to the Bedside. Advances in Experimental Biology and Medicine*, Vol. 502 (ed. R. C. Roach, P. D. Wagner and P. H. Hackett), pp. 107-132. New York: Kluwer Academic/Plenum Publishers.
- Garland, T., Jr** (2003). Selection experiments: an under-utilized tool in biomechanics and organismal biology. In *Vertebrate Biomechanics and Evolution* (ed. V. L. Bels, J.-P. Gasc and A. Casinos), pp. 23-56. Oxford, UK: BIOS Scientific Publishers.
- Garland, T., Jr and Adolph, S. C.** (1991). Physiological differentiation of vertebrate populations. *Annu. Rev. Ecol. Syst.* **22**, 193-228.
- Garland, T., Jr and Adolph, S. C.** (1994). Why not to do two-species comparative studies: limitations on inferring adaptation. *Physiol. Zool.* **67**, 797-828.
- Garland, T., Jr and Carter, P. A.** (1994). Evolutionary physiology. *Annu. Rev. Physiol.* **56**, 579-621.
- Garland, T., Jr and Diaz-Uriarte, R.** (1999). Polytomies and phylogenetically independent contrasts: examination of the bounded degrees of freedom approach. *Syst. Biol.* **48**, 547-558.
- Garland, T., Jr and Ives, A. R.** (2000). Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. *Am. Nat.* **155**, 346-364.
- Garland, T., Jr, Huey, R. B. and Bennett, A. F.** (1991). Phylogeny and thermal physiology in lizards: A reanalysis. *Evolution* **45**, 1969-1975.
- Garland, T., Jr, Harvey, P. H. and Ives, A. R.** (1992). Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst. Biol.* **41**, 18-32.
- Garland, T., Jr, Dickerman, A. W., Janis, C. M. and Jones, J. A.** (1993). Phylogenetic analysis of covariance by computer simulation. *Syst. Biol.* **42**, 265-292.
- Garland, T., Jr, Midford, P. E. and Ives, A. R.** (1999). An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral values. *Am. Zool.* **39**, 374-388.
- Garland, T., Jr, Ives, A. R. and Midford, P. E.** (2004). Within-species variation and measurement error in phylogenetic comparative methods. *Integr. Comp. Biol.* **44**, 556. (abstract).
- Gibbs, A. G., Fukuzato, F. and Matzkin, L. M.** (2003). Evolution of water conservation mechanisms in *Drosophila*. *J. Exp. Biol.* **206**, 1183-1192.
- Grafen, A.** (1989). The phylogenetic regression. *Phil. Trans. R. Soc. Lond. B* **326**, 119-157.
- Hale, M. E., Long, J. H., Jr, McHenry, M. J. and Westneat, M. W.** (2002). Evolution of behavior and neural control of the fast-start escape response. *Evolution* **56**, 993-1007.
- Hansen, T. F., Armbruster, W. S. and Antonson, L.** (2000). Comparative analysis of character displacement and spatial adaptations as illustrated by the evolution of *Dalechampia* blossoms. *Am. Nat.* **156**, S17-S34.
- Harvey, P. H.** (1996). Phylogenies for ecologists. *J. Anim. Ecol.* **65**, 255-263.
- Harvey, P. H. and Pagel, M. D.** (1991). *The Comparative Method in Evolutionary Biology*. Oxford: Oxford University Press.
- Harvey, P. H. and Rambaut, A.** (2000). Comparative analyses for adaptive radiations. *Phil. Trans. R. Soc. Lond. B* **355**, 1599-1605.
- Herrel, A., O'Reilly, J. C. and Richmond, A. M.** (2002). Evolution of bite performance in turtles. *J. Evol. Biol.* **15**, 1083-1094.
- Herrel, A., Podos, J., Huber, S. K. and Hendry, A. P.** (2005). Evolution of bite force in Darwin's finches: a key role for head width. *J. Evol. Biol.* **18**, 669-675.
- Hochachka, P. W. and Somero, G. N.** (2002). *Biochemical Adaptation-Mechanism and Process in Physiological Evolution*. Oxford: Oxford University Press.
- Hodges, W. L.** (2004). Evolution of viviparity in horned lizards (*Phrynosoma*): testing the cold climate hypothesis. *J. Evol. Biol.* **17**, 1230-1237.
- Hone, D. W. E., Keeseey, T. M., Pisani, D. and Purvis, A.** (2005). Macroevolutionary trends in the Dinosauria: Cope's rule. *J. Evol. Biol.* **18**, 587-5195.
- Hosken, D. J., Jones, K. E., Chipperfield, K. and Dixon, A.** (2001). Is the bat os penis sexually selected? *Behav. Ecol. Sociobiol.* **50**, 450-460.
- Houseworth, E. A. and Martins, E. P.** (2001). Random sampling of constrained phylogenies: Conducting phylogenetic analyses when the phylogeny is partially known. *Systemat. Biol.* **50**, 628-639.
- Housworth, E. A., Martins, E. P. and Lynch, M.** (2004). The phylogenetic mixed model. *Am. Nat.* **163**, 84-96.
- Huelsenbeck, J. P. and Rannala, B.** (2003). Detecting correlation between characters in a comparative analysis with uncertain phylogeny. *Evolution* **57**, 1237-1247.
- Huey, R. B. and Bennett, A. F.** (1987). Phylogenetic studies of coadaptation: Preferred temperatures versus optimal performance temperatures of lizards. *Evolution* **41**, 1098-1115.
- Hutcheon, J. M. and Garland, T., Jr** (2004). Are megabats big? *J. Mamm. Evol.* **11**, 257-276.
- Iwaniuk, A. N., Pellis, S. M. and Whishaw, I. Q.** (1999). Is digital dexterity really related to corticospinal projections?: re-analysis of the Heffner and Masterson data set using modern comparative statistics. *Behav. Brain Res.* **101**, 173-187.
- Johnston, I. A., Fernandez, D. A., Calvo, J., Vieira, V. L. A., North, A. W., Abercomby, M. and Garland, T., Jr** (2003). Reduction in muscle fibre number during the adaptive radiation of notothenioid fishes: a phylogenetic perspective. *J. Exp. Biol.* **206**, 2595-2609.
- Lane, J. E., Brigham, R. M. and Swanson, D. L.** (2004). Torpor in Whip-poor-wills. *Physiol. Biochem. Zool.* **77**, 297-304.
- Lapointe, F.-J. and Garland, T., Jr** (2001). A generalized permutation model for the analysis of cross-species data. *J. Classif.* **18**, 109-127.
- Lasiewski, R. C. and Dawson, W. R.** (1967). A re-examination of the relation between standard metabolic rate and body weight in birds. *Condor* **69**, 13-23.
- Lasiewski, R. C., Dawson, W. R. and Bartholomew, G. A.** (1970). Temperature regulation in the Little Papuan Frogmouth. *Condor* **72**, 332-338.
- Lauder, G. V.** (1981). Form and function: structural analysis in evolutionary morphology. *Paleobiology* **7**, 430-442.
- Lauder, G. V.** (1982). Historical biology and the problem of design. *J. Theor. Biol.* **97**, 57-67.
- Lauder, G. V.** (1990). Functional morphology and systematics: studying functional patterns in an historical context. *Annu. Rev. Ecol. Syst.* **21**, 317-340.
- Laurin, M.** (2004). The evolution of body size, Cope's Rule and the origin of amniotes. *Syst. Biol.* **53**, 594-622.
- Lavergne, S., Thompson, J. D., Garnier, E. and Debussche, M.** (2004). The biology and ecology of narrow endemic and widespread plants: a comparative study of trait variation in 20 congeneric pairs. *Oikos* **107**, 505-518.
- Leroi, A. M., Rose, M. R. and Lauder, G. V.** (1994). What does the comparative method reveal about adaptation? *Am. Nat.* **143**, 381-402.
- Lovegrove, B. G.** (2003). The influence of climate on the basal metabolic rate of small mammals: a slow-fast metabolic continuum. *J. Comp. Physiol. B* **173**, 87-112.
- MacMillen, R. E. and Nelson, J. E.** (1969). Bioenergetics and body size in dasyurid marsupials. *Am. J. Physiol.* **217**, 1246-1251.
- Maddison, D. R. and Maddison, W. P.** (2000). *MacClade 4, Analysis of Phylogeny and Character Evolution*. Sunderland, MA: Sinauer Associates.
- Mangun, C. P. and Hochachka, P. W.** (1998). New directions in comparative physiology and biochemistry: mechanisms, adaptations, and evolution. *Physiol. Zool.* **71**, 471-484.
- Martins, E. P. and Garland, T., Jr** (1991). Phylogenetic analyses of correlated evolution of continuous characters: a simulation study. *Evolution* **45**, 534-557.
- Martins, E. P. and Hansen, T. F.** (1997). Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* **149**, 646-667. Erratum *Am. Nat.* **153**, 448.
- Martins, E. P. and Lamont, J.** (1998). Estimating ancestral states of a communicative display: a comparative study of *Cyclura* rock iguanas. *Anim. Behav.* **55**, 1685-1706.
- Martins, E. P., Diniz-Filho, J. A. F. and Housworth, E. A.** (2002). Adaptive

- constraints and the phylogenetic comparative method: a computer simulation test. *Evolution* **56**, 1-13.
- McGuire, J. A.** (2003). Allometric prediction of locomotor performance: an example from Southeast Asian flying lizards. *Am. Nat.* **161**, 337-349.
- McKechnie, A. E. and Wolf, B. O.** (2004). The allometry of avian basal metabolic rate: good predictions need good data. *Physiol. Biochem. Zool.* **77**, 502-521.
- Monkkonen, M.** (1995). Do migrant birds have more pointed wings? A comparative study. *Evol. Ecol.* **9**, 520-528.
- Mottishaw, P. D., Thornton, S. J. and Hochachka, P. W.** (1999). The diving response mechanism and its surprising evolutionary path in seals and sealions. *Am. Zool.* **39**, 434-450.
- Muñoz-García, A. and Williams, J. B.** (in press). Basal metabolic rate in carnivores is associated with diet after controlling for phylogeny. *Physiol. Biochem. Zool.* **78**.
- Nunn, C. L.** (2002). Spleen size, disease risk and sexual selection: a comparative study in primates. *Evol. Ecol. Res.* **4**, 91-107.
- Nunn, C. L. and Barton, R. A.** (2000). Allometric slopes and independent contrasts: a comparative test of Kleiber's Law in primate ranging patterns. *Am. Nat.* **156**, 519-533.
- Oakley, T. H.** (2003). On homology of arthropod compound eyes. *Integr. Comp. Biol.* **43**, 522-530.
- Oakley, T. H. and Cunningham, C. W.** (2000). Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny. *Evolution* **54**, 397-405.
- Pagel, M. D.** (1992). A method for the analysis of comparative data. *J. Theor. Biol.* **156**, 431-442.
- Pagel, M.** (1999). Inferring the historical patterns of biological evolution. *Nature* **401**, 877-884.
- Paradis, E. and Claude, J.** (2002). Analysis of comparative data using generalized estimating equations. *J. Theor. Biol.* **218**, 175-185.
- Pérez-Barbería, F. J., Gordon, I. J. and Pagel, M.** (2002). The origins of sexual dimorphism in body size in ungulates. *Evolution* **56**, 1276-1285.
- Perry, G. and Garland, T., Jr** (2002). Lizard home ranges revisited: effects of sex, body size, diet, habitat, and phylogeny. *Ecology* **83**, 1870-1885.
- Peters, R. H.** (1984). *The Ecological Implications of Body Size*. Cambridge: Cambridge University Press.
- Polly, P. D.** (2001). Paleontology and the comparative method: ancestral node reconstructions versus observed node values. *Am. Nat.* **157**, 596-609.
- Price, T.** (1997). Correlated evolution and independent contrasts. *Phil. Trans. R. Soc. Lond. B* **352**, 519-529.
- Purvis, A. and Garland, T., Jr** (1993). Polytomies in comparative analyses of continuous characters. *Syst. Biol.* **42**, 569-575.
- Purvis, A. and Rambaut, A.** (1995). Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analysing comparative data. *Comp. Appl. Biosci.* **11**, 247-251.
- Purvis, A. and Webster, A. J.** (1999). Phylogenetically independent comparisons and primate phylogeny. In *Comparative Primate Socioecology* (ed. P. C. Lee), pp. 44-70. Cambridge: Cambridge University Press.
- Purvis, A., Gittleman, J. L. and Luh, H.-K.** (1994). Truth or consequences: effects of phylogenetic accuracy on two comparative methods. *J. Theor. Biol.* **167**, 293-300.
- Quader, S., Isvaran, K., Hale, R. E., Miner B. G. and Seavy, N. E.** (2004). Nonlinear relationships and phylogenetically independent contrasts. *J. Evol. Biol.* **17**, 709-715.
- Rayner, J. M. V.** (1985). Linear relations in biomechanics: the statistics of scaling functions. *J. Zool.* **206**, 415-439.
- Reynolds, P. S.** (2002). How big is a giant? The importance of methodology in estimating body size of extinct mammals. *J. Mammal.* **83**, 321-332.
- Reynolds, P. S. and Lee, R. M.** (1996). Phylogenetic analysis of avian energetics: Passerines and nonpasserines do not differ. *Am. Nat.* **147**, 735-759.
- Rezende, E. L. and Garland, T., Jr** (2003). Comparaciones interespecificas y métodos estadísticos filogenéticos. In *Fisiología Ecológica and Evolutiva. Teoría y casos de estudio en animales* (ed. F. Bozinovic), pp. 79-98. Santiago: Ediciones Universidad Católica de Chile.
- Rezende, E. L., Swanson, D. L., Novoa, F. F. and Bozinovic, F.** (2002). Passerines versus nonpasserines: so far, no statistical differences in the scaling of avian energetics. *J. Exp. Biol.* **205**, 101-107.
- Rezende, E. L., Bozinovic, F. and Garland, T., Jr** (2004). Climatic adaptation and the evolution of basal and maximum rates of metabolism in rodents. *Evolution* **58**, 1361-1374.
- Reznick, D. N., Mateos, M. and Springer, M. S.** (2002). Independent origins and rapid evolution of the placenta in the fish genus *Poeciliopsis*. *Science* **298**, 1018-1020.
- Rheindt, F. E., Grafe, T. U. and Abouheif, E.** (2004). Rapidly evolving traits and the comparative method: how important is testing for phylogenetic signal? *Evol. Ecol. Res.* **6**, 377-396.
- Ricklefs, R. E. and Nealen, P.** (1998). Lineage-dependent diversification. *Funct. Ecol.* **12**, 871-885.
- Ricklefs, R. E. and Starck, J. M.** (1996). The application of phylogenetically independent contrasts: a mixed progress report. *Oikos* **77**, 167-172.
- Ridley, M.** (1983). *The Explanation of Organic Diversity: The Comparative Method and Adaptations for Mating*. Oxford, UK: Clarendon.
- Riska, B.** (1991). Regression models in evolutionary allometry. *Am. Nat.* **138**, 283-299.
- Rochet, M. J., Cornillon, P. A., Sabatier, R. and Pontier, D.** (2000). Comparative analysis of phylogenetic and fishing effects in life history patterns of teleost fishes. *Oikos* **91**, 255-270.
- Roff, D. A.** (in press). *Introduction to Computer-intensive Methods of Data Analysis in Biology*. Cambridge, UK: Cambridge University Press.
- Rohlf, F. J.** (2001). Comparative methods for the analysis of continuous variables: Geometric interpretations. *Evolution* **55**, 2143-2160.
- Ross, C. F., Henneberg, M., Ravosa, M. J. and Richard, S.** (2004). Curvilinear, geometric and phylogenetic modeling of basicranial flexion: is it adaptive, is it constrained? *J. Hum. Evol.* **46**, 185-213.
- Sanford, G. M., Lutterschmidt, W. I. and Hutchison, V. H.** (2002). The comparative method revisited. *BioScience* **52**, 830-836.
- Schleucher, E. and Withers, P. C.** (2002). Metabolic and thermal physiology of pigeons and doves. *Physiol. Biochem. Zool.* **75**, 439-450.
- Schluter, D., Price, T., Mooers, A. O. and Ludwig, D.** (1997). Likelihood of ancestor states in adaptive radiation. *Evolution* **51**, 1699-1711.
- Schondube, J. E., Herrera-M., L. G. and Martínez del Río, C.** (2001). Diet and the evolution of digestion and renal function in phyllostomid bats. *Zoology* **104**, 59-73.
- Smith, R. J. and Cheverud, J. M.** (2002). Scaling of sexual size dimorphism in body mass: a phylogenetic analysis of Rensch's rule in primates. *Int. J. Primatol.* **23**, 1095-1135.
- Spicer, J. I. and Gaston, K. J.** (1999). *Physiological Diversity and its Ecological Implications*. Oxford: Blackwell Science.
- Swallow, J. G. and Garland, T., Jr** (2005). Selection experiments as a tool in evolutionary and comparative physiology: insights into complex traits – An introduction to the symposium. *Integr. Comp. Biol.* **45**, 387-390.
- Symonds, M. R. E. and Elgar, M.** (2002). Phylogeny affects estimation of metabolic scaling in mammals. *Evolution* **56**, 2330-2333.
- Thorpe, R. S., Black, H. and Malhotra, A.** (1996). Matrix correspondence tests on the DNA phylogeny of the Tenerife lacertid elucidate both historical causes and morphological adaptation. *Syst. Biol.* **45**, 335-343.
- Tieleman, B. I., Williams, J. B. and Bloomer, P.** (2003). Adaptation of metabolism and evaporative water loss along an aridity gradient. *Proc. R. Soc. Lond. B* **270**, 207-214.
- Vanhooydonck, B. and Van Damme, R.** (1999). Evolutionary relationships between body shape and habitat use in lacertid lizards. *Evol. Ecol. Res.* **1**, 785-805.
- Vanhooydonck, B. and Van Damme, R.** (2001). Evolutionary trade-offs in locomotor capacities in lacertid lizards: are splendid sprinters clumsy climbers? *J. Evol. Biol.* **14**, 46-54.
- Webster, A. J. and Purvis, A.** (2002). Testing the accuracy of methods for reconstructing ancestral states of continuous characters. *Proc. R. Soc. Lond. B* **269**, 143-150.
- Williams, J. B.** (1996). A phylogenetic perspective of evaporative water loss in birds. *Auk* **113**, 457-472.
- Willmer, P., Stone, G. and Johnston, I.** (2005). *Environmental Physiology of Animals*. 2nd edn. Oxford, UK: Blackwell Science.
- Wolf, C. M., Garland, T., Jr and Griffith, B.** (1998). Predictors of avian and mammalian translocation success: reanalysis with phylogenetically independent contrasts. *Biol. Conserv.* **86**, 243-255.