An Introduction to Phylogenetically Based Statistical Methods, with a New Method for Confidence Intervals on Ancestral Values[1]

THEODORE GARLAND, JR.,[2] PETER E. MIDFORD, AND ANTHONY R. IVES

Department of Zoology, 430 Lincoln Drive, University of Wisconsin, Madison, Wisconsin 53706-1381

SYNOPSIS. Interspecific comparisons have played a prominent role in evolutionary biology at least since the time of Charles Darwin. Since 1985, the "comparative method" has been revitalized by new analytical techniques that use phylogenetic information and by increased availability of phylogenies (often from molecular data sets). Because species descend from common ancestors in a hierarchical fashion, related species tend to resemble each other (elephants look like elephants); therefore, cross-species data sets generally do not comprise independent and identically distributed data points. Phylogenetically based statistical methods attempt to account for this fact. Phylogenetic methods allow traditional topics in comparative and ecological physiology to be addressed with greater rigor, including the form of allometric relationships and whether physiological phenotypes vary predictably in relation to behavior, ecology or environmental characteristics, which provides evidence about adaptation. They can also address new topics, such as whether rates of physiological evolution have differed among lineages (clades), and where and when a phenotype first evolved. We present brief overviews of three phylogenetically based statistical methods: phylogenetically independent contrasts, Monte Carlo computer simulations to obtain null distributions of test statistics, and phylogenetic autocorrelation. In a new result, we show analytically how to use independent contrasts to estimate ancestral values and confidence intervals about them. These confidence intervals often exceed the range of variation observed among extant species, which points out the relatively great uncertainty inherent in such inferences. The use of phylogenies should become as common as the use of body size and scaling relationships in the analysis of physiological diversity.

INTRODUCTION

Many approaches can and should be used in evolutionary physiology (Feder *et al.*, 1987; Bennett, 1997; Gibbs, 1999; Koteja *et al.*, 1999), and combinations of approaches are often necessary (Garland and Carter, 1994; Leroi *et al.*, 1994; Hayes and Garland, 1995). Interspecific comparisons are the primary way to study long-term evolutionary changes in the phenotype or genotype, and they have undergone a renaissance in the last decade (Brooks and McLennan, 1991; Harvey and Pagel, 1991; Eggleton and Vane-Wright, 1994; Martins, 1996*a*). They are used both to generate and

to test hypotheses (Doughty, 1996; Larson and Losos, 1996), not only in evolutionary biology per se but also in such related fields as behavioral ecology (Krebs and Davies, 1997), functional morphology (Wainwright and Reilly, 1994), and ecological physiology (Garland and Adolph, 1994).

Comparative data sets are now routinely analyzed by phylogenetic methods (*e.g.,* Garland *et al.*, 1993; Miles and Dunham, 1993; Losos and Miles, 1994; Maddison, 1994; Maddison, 1995; Larson and Losos, 1996; Butler and Losos, 1997; Garland *et al.*, 1997; Schluter *et al.*, 1997; Barraclough *et al.*, 1998; Martins and Lamont, 1998; Pagel, 1998). Most traits studied by physiologists show continuous variation (*e.g.,* organ size, metabolic rate, blood hemoglobin level), and several phylogenetically based statistical methods are available for continuous-valued characters (Martins and Hansen, 1996, 1997). We provide an introduction to three of these: phylogenetically in-

[2] To whom correspondence should be addressed. Phone: 608-262-4437; Facsimile: 608-265-6320; E-mail: tgarland@facstaff.wisc.edu

dependent contrasts (Felsenstein, 1985; Garland et al., 1992; Purvis and Garland, 1993), computer simulation to obtain phylogenetically correct null distributions of test statistics (Martins and Garland, 1991; Garland et al., 1993), and phylogenetic autocorrelation (Cheverud et al., 1985; Gittleman and Kot, 1990). We also show how independent contrasts can be used to estimate phenotypes of hypothetical ancestors (and confidence intervals about those estimates). Elsewhere, we will show how to place a confidence or prediction interval (in the original data space) on regression equations derived from independent contrasts (Garland and Ives, in preparation). All of these computations are performed by the Phenotypic Diversity Analysis Program (PDAP), which is available on request from T.G.

PHYLOGENETICALLY INDEPENDENT CONTRASTS

Felsenstein (1985) proposed the first fully phylogenetic statistical method for analysis of comparative data. By fully phylogenetic, we mean that it can be applied to any topology and set of branch lengths. Although the original presentation of independent contrasts was couched in terms of a Brownian motion model of character evolution (Felsenstein, 1985), it can also be justified on first-principles statistical grounds (Grafen, 1989; Pagel, 1993). Felsenstein (1985) emphasized applications of independent contrasts to simple correlation and linear regression, but they can also be applied to almost any problem that requires such related statistical techniques as principal components analysis, multiple regression, path analysis, analysis of variance, and analysis of covariance (e.g., Garland, 1992, 1994; Garland et al., 1993; Martins, 1993; Díaz et al., 1996; Martin and Clobert, 1996; Bauwens and Díaz-Uriarte, 1997; Clobert et al., 1998; Wolf et al., 1998). As well, they can be used to compare single species with a set of others (Garland and Adolph, 1994, pp. 809–812; Martinez et al., 1995; McPeek, 1995; Eppley, 1996). Moreover, as with many other phylogenetic methods (Brooks and McLennan, 1991; Harvey and Pagel, 1991; Block et al., 1993;

Eggleton and Vane-Wright, 1994; Ryan and Rand, 1995; Gittleman et al., 1996; Martins, 1996a; Butler and Losos, 1997; Garland et al., 1997; Martins and Lamont, 1998; Pagel, 1998), independent contrasts can be used to address questions that are not accessible without phylogenetic information. For example, they can be used to compare rates of evolution across clades (Garland, 1992; Barbosa, 1993; Clobert et al., 1998).

When given accurate information on phylogenetic topology and branch lengths, the statistical power of independent contrasts for detecting a correlation between two traits is identical to that of a conventional correlation applied to nonphylogenetic data (Garland and Adolph, 1994). Computer simulations show that independent contrasts are reasonably robust with respect to violation of assumptions (e.g., deviations from Brownian motion character evolution, errors in branch lengths: Martins and Garland, 1991; Purvis et al., 1994; Díaz-Uriarte and Garland, 1996, 1998; Martins, 1996b). In some cases, and with caution, independent contrasts can also be applied to the analysis of discretely valued characters (see Garland et al., 1993; Martins, 1993; Grafen and Ridley, 1996; Ridley and Grafen, 1996). They can be applied with partial phylogenetic information (polytomies: Purvis and Garland, 1993; Losos, 1994; Stamps et al., 1997; Abouheif, 1998; Garland and Díaz-Uriarte, 1999) and with arbitrary branch lengths, appropriately checked for statistical adequacy (Grafen, 1989; Garland et al., 1992; Pagel, 1992; Clobert et al., 1998). Independent contrasts analyses can use covariates that are not phylogenetically inherited (see Wolf et al., 1998) and regressions weighted by within-species sample size (Bonine and Garland, 1999). Finally, information on variation among individuals within species can also be incorporated (Garland and Ives, in preparation; see also Martins and Lamont, 1998).

Computation of independent contrasts involves several steps (see Felsenstein, 1985; also Garland et al., 1992; Purvis and Garland, 1993). The goal is to produce, from the original N tip species, a set of N-1 stan-

dardized independent contrasts that are, in principle, independent and identically distributed. Thus, the procedure attempts to transform nonindependent data points into values that are suitable for use with conventional statistical procedures. A brief description of the actual computations follows, and a worked example can be found in Garland and Adolph (1994).

First, pairs of species at the tips of a phylogeny are contrasted: the phenotype of one species is subtracted from the other, with the direction of subtraction arbitrary. Second, each such pair of tip species is pruned from the tree, and the phenotype of their ancestral node is estimated as the weighted mean of the descendants' phenotypes (weighting involves the inverse of branch lengths). Third, moving down the tree, further contrasts are computed, involving the values estimated for internal nodes. Fourth, each contrast is divided by its "standard deviation," which is the square root of the sum of its branch lengths, where branch lengths must be in units proportional to expected variance of evolution for the character being analyzed. (For all internal branches, a lengthening occurs to account for the fact that phenotypes of internal nodes are not data but estimates from the observed data. Thus, contrasts involving internal nodes need to be given less weight, and this devaluation is easily accomplished by increasing their standard deviations.) Fifth, sets of independent contrasts for different characters can be used in conventional statistical procedures, such as correlation and regression, with the constraint that all such relationships are computed through the origin (Garland et al., 1992).

Many empirical studies that use independent contrasts have now been published. Typically, significance levels are reduced (P values are higher) when analyzed by independent contrasts as compared with conventional analyses. In one such example, three of six correlations were significant at $P < 0.05$ by conventional methods but only one of six was significant by independent contrasts (Garland et al., 1991). This demonstrates the overly liberal nature of conventional statistical methods applied to comparative data. Nonetheless, convention-
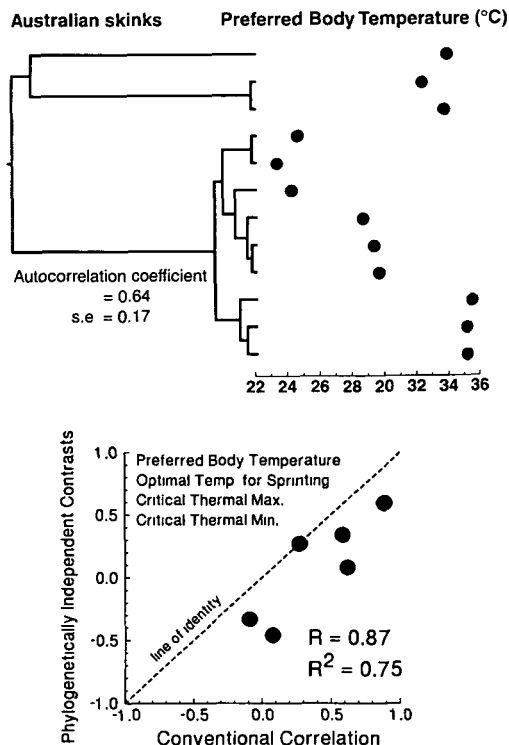

FIG. 1. Top shows hypothesized phylogenetic relationships for 12 species of Australian skinks (from Fig. 1 of Garland et al., 1991). Note how preferred body temperature appears to "follow phylogeny," as also indicated by a statistically significant phylogenetic autocorrelation coefficient of 0.64 (see Discussion: computed by the MRHO3 program of Cheverud and Dow [1985], modified by Miles and Dunham [1992]). Bottom shows relationship between pairwise Pearson product-moment correlation coefficients for four thermal traits (see Huey and Bennett, 1987), as estimated in the conventional manner and by phylogenetically independent contrasts (TIPS and FL1G, respectively, in Table 2 of Garland et al., 1991). Note how the conventional analyses tended to overestimate the strengths of relationships.

al and independent contrasts analyses are essentially trying to estimate the same thing (Martins and Garland, 1991; Pagel, 1993, 1998), and the correlation between the estimates is high (bottom of Figure 1: for other examples, see Martins, 1993; Westneat, 1995; Carrascal et al., 1999). Surveys of a broad range of studies show a similar pattern (Ricklefs and Starck, 1996; Price, 1997): conventional and independent contrasts correlations are correlated as about $r = 0.85$.

A related point is that independent contrasts can be applied to a star phylogeny (a giant polytomy with equal-length branches). When this is done, the resulting correlations, regressions, ANOVAs, root node estimates and confidence intervals (next section), etc., will be *exactly* the same as produced by a conventional statistical analysis (Purvis and Garland, 1993; see also Abouheif, 1998). This point also applies to computer-simulation approaches when simple Brownian motion evolution and a star phylogeny are used (see Garland *et al.*, 1993; Discussion). We are led to the somewhat perverse view that there is no such thing as a "nonphylogenetic" analysis. Rather, all analyses are phylogenetic, some just assume the special case of a star phylogeny!

CONFIDENCE INTERVALS ON THE MEAN VALUE OR ROOT NODE FOR A SET OF SPECIES

The most basic use of independent contrasts is to compute a phylogenetically weighted estimate of the mean value for a set of species. This is simply the estimated value at the root (basal) node of the phylogeny (Garland *et al.*, 1993; Schluter *et al.*, 1997). For most purposes, it is an extra piece of information; hence, it was not mentioned by Felsenstein (1985; for example, his Fig. 9 and Table 1 do not consider its computation).

The value at the root node can also be interpreted as an estimate of the phenotype of the hypothetical ancestor of all species in the data set. And, this value turns out to be exactly the same as the value reconstructed by squared-change parsimony (Maddison, 1991). For either interpretation of the root node, the confidence interval of its estimator may be desirable. For example, placing a confidence interval about the phylogenetically correct mean would allow tests of *a priori* hypotheses concerning the phenotype of the hypothetical ancestor (see also Garland *et al.*, 1997; Schluter *et al.*, 1997).

To compute a standard error and 95% confidence interval for the root node of a phylogeny, via the formalism of independent contrasts, perform the following steps (see Appendix for the proof and the PDAP documentation for a worked example):

1. compute the N-1 standardized independent contrasts, where N = the number of tips in the phylogeny (terminal taxa, which can be populations, species or higher taxa: see Discussion).
2. square the standardized independent contrasts.
3. sum the squared standardized independent contrasts.
4. divide quantity (3) by N-1. This is an estimate of *r*, the rate of character evolution.
5. multiply quantity (4) by $(v_1' \, v_2')/(v_1' + v_2')$ where v_1' and v_2' are the corrected lengths of the two branches that descend from the root (basal) node.
6. take the square root of quantity (5). This is a phylogenetically weighted estimate of the standard error of the root node or, equivalently, of the phylogenetically weighted mean of the tip data.
7. multiply quantity (6) by the critical value from the *t* distribution for $\alpha = 0.025$ and N-1 degrees of freedom, where N = the number of tip species. This yields the ± 95% confidence interval.

The foregoing procedure yields exactly the same estimates for the root node as described in Schluter *et al.* (1997), but without the use of maximum likelihood. But unlike the maximum likelihood approach, it does not apply to other internal nodes. The values computed by independent contrasts for other internal nodes are not optimal by any criterion: they are a type of "local parsimony" reconstruction, whereas the root node is a "global parsimony" reconstruction, which is the same as computed by squared-change parsimony (Maddison, 1991; Garland *et al.*, 1997). However, by rerooting a phylogenetic tree at a given internal node (which produces a trichotomy), the procedure described above can be used to obtain the appropriate estimates for any node. Our PDTREE program can be used for this and the values obtained will be identical to those produced by the ANCML program of Schluter *et al.* (1997) and to generalized least-squares estimators (Garland and Ives, in preparation).

The procedure of rerooting a phylogenetic tree also allows estimation of ancestral values *anywhere along a branch, i.e.,* not just at nodes. In other words, if we redraw a phylogenetic tree so that the root node is at an arbitrary point along any branch segment, then the mean and confidence intervals produced (*e.g.,* by PDTREE) estimate the value for the ancestor at that point in past evolutionary history (see PDAP documentation for worked examples).

A related problem is *predicting* the value of a hypothetical new species (extant or extinct). Again, rerooting can be used, but the confidence interval is replaced by a prediction interval, which includes a term (V_h) for the length of the unique branch leading to the new species (see end of Appendix).

Empirical example: estimating ancestral preferred body temperature of Australian scincid lizards

Huey and Bennett (1987) studied the thermal biology of 12 species of Australian scincid lizards. Here, we consider their data for preferred body temperatures (Fig. 1, top: see Garland *et al.* [1991] for a listing of the complete tip data set, topology, and branch lengths used here).

In their original analysis, Huey and Bennett (1987) arbitrarily set all branch lengths equal to one for squared-change parsimony computations. In a reanalysis, Garland *et al.* (1991) employed those branch lengths as well as estimates of divergence times (Fig. 1, top). Here, we also employ other sets of branch lengths. We used the arbitrary branch lengths suggested by Grafen (1989) and by Pagel (1992). The diagnostic test proposed by Garland *et al.*, 1992 (see also Díaz-Uriarte and Garland, 1996, 1998) suggested transformation of those branch lengths, and so we also employed them transformed by raising each branch segment length to an exponent. We also used the original time branch lengths raised to an exponent. Finally, as suggested by Garland *et al.*, 1992, p. 22), we constructed a pairwise distance matrix (with the new PDDIST program) from the preferred body temperature data and used the KITSCH87 program in J. Felsenstein's PHYLIP package (version 3.3) to fit branch lengths to our user-defined

tree (with negative branch lengths disallowed). Five of the estimated branch segment lengths were zero; two of these were for the branches leading to adjacent tip species (*Ctenotus taeniolatus* and *C. uber*), and hence could not be used for independent contrast calculations. Therefore, to each branch segment on the entire phylogenetic tree, we added the length of the shortest (non-zero) estimated branch segment, which was 0.12215. We used these branch lengths both directly and transformed by raising them to an exponent.

Figure 2 shows a conventional 95% confidence interval and several confidence intervals derived via independent contrasts computations, with the use of different branch lengths. As expected, the conventional confidence interval is generally narrower than those derived from phylogenetically independent contrasts, but exceptions occur. Also, some of the independent contrasts confidence intervals actually exceed the range of observed tip data (see Schluter *et al.* [1997] for other such examples). Note that the heuristic "sensitivity analysis" approach used by Huey and Bennett (1987) with the squared-change parsimony algorithm (see their Fig. 4) bears no correspondence to the formal confidence interval described above (see also Garland *et al.*, 1997). Point estimates of the root node value also differ, with the conventional mean falling within the range of estimates from independent contrasts (Fig. 2).

Empirical example: estimating ancestral body mass of Carnivora and ungulates

Garland *et al.* (1993) presented an example data set with body mass (and home range area) for 49 species of Carnivora and ungulates; phylogenetic branch lengths were estimated as divergence times (their Fig. 1). Figure 3 shows the body mass data on a logarithmic scale, along with conventional and independent contrasts means and 95% confidence intervals. Also shown is a probable approximate body mass for the last common ancestor of Carnivora and ungulates, as indicated by the fossil record (about 0.5 kg: see p. 271 and Appendix of Garland *et al.*, 1993). A directional evolutionary trend for increasing body size has

Root Node Estimate and 95% Confidence Interval
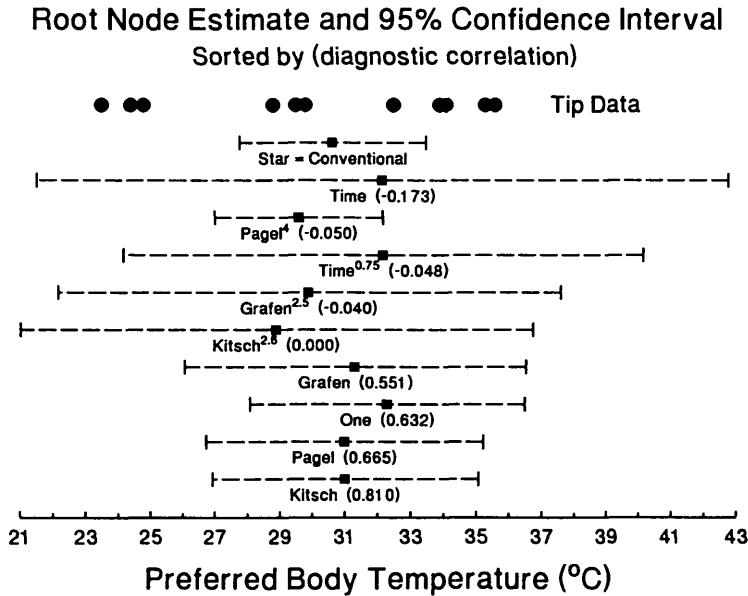### Sorted by (diagnostic correlation)

FIG. 2. Estimates of mean preferred body temperatures (tip data) for 12 species of Australian scincid lizards, as in Figure 1 (data from Huey and Bennett, 1987; Garland et al., 1991). A conventional mean and 95% confidence interval is indicated (these values are identical to an independent contrasts analysis performed on a star phylogeny). Below this are corresponding values computed by phylogenetically independent contrasts, using several different sets of branch lengths. The values in parentheses indicate the correlation between the absolute values of the standardized contrasts and their standard deviations, a branch-length diagnostic proposed by Garland et al. (1992; see also Díaz-Uriarte and Garland, 1996, 1998): the closer this number is to zero, the better.

occurred, which violates the assumption of simple Brownian motion character evolution and hence invalidates computations for making inferences about ancestral values.

The 49 species included in the example of Figure 3 are but a small sample of extant Carnivora and ungulates, and it would be of interest to redo such an analysis with all available data for extant species. Even more reliable results could be gained by adding information on body sizes of extinct species (cf. Hansen, 1997). They could be added directly to the phylogeny, with branches whose lengths terminate before the present, and independent contrasts computations done as usual. The effect of adding fossil species of small body size would be to "pull" the root-node estimate and confidence intervals towards smaller values, which could well include 0.5 kg. Of course, the option of adding fossil species to an analysis will rarely be available for physiological traits.

DISCUSSION

Interspecific comparisons always have been, and will always remain, an essential tool in comparative and ecological physiology. Phylogenetically based statistical methods are necessary for the analysis of comparative data and have helped to bring rigor to evolutionary physiology (Garland and Carter, 1994; Bennett, 1997). Existing methods can address a range of questions, including traditional ones about correlated character evolution (e.g., allometry) and about both proximate (mechanistic) and ultimate (e.g., adaptation) causes of physiological diversity. They can also address new questions that are inaccessible without phylogenetic information, such as character origins and rates of evolution. As compared with conventional analyses, phylogenetic analyses bring to bear a much greater amount of information, and so it makes sense that they should allow stronger and/or additional inferences.

Root Node Estimate and 95% Confidence Interval
49 Species of Carnivora and ungulates

FIG. 3. Illustration of adverse effect of a directional evolutionary trend on realism of an ancestral reconstruction. Data on body mass for 49 species of mammals and phylogenetic information (branch lengths in estimated divergence times) are from Garland *et al.* (1993). As discussed therein, fossil information suggests that the last common ancestor of the Carnivora and ungulates may have been about 0.5 kg. This value is below the range of the extant taxa in this data set (although well within the range of extant Carnivora). Thus, the independent contrasts root node estimate and 95% confidence interval is unrealistic. Nevertheless, these values do provide valid estimates with respect to the mean body mass of species included in the data set (see text).

Many examples now exist in which conclusions change when phylogenetic methods are applied. One surprising result has been the extent to which conventional and independent contrasts allometric equations, computed for avian species spanning a very wide body mass range (hummingbirds to ostriches), can differ significantly and hence lead to quite different predictive values (Reynolds and Lee, 1996; Williams, 1996).

Felsenstein's (1985) method of independent contrasts is the best justified, best understood, and most versatile of existing phylogenetically based statistical methods. Given what we know so far, it must be viewed as the method of choice for comparative analyses of continuous-valued characters. Appropriately, it is being widely applied in ecological and evolutionary morphology and physiology (*e.g.*, Sessions and Larson, 1987; Losos, 1990; Garland *et al.*, 1991; Promislow, 1991; Moreno and Carrascal, 1993; Walton, 1993; Garland, 1994; Bauwens *et al.*, 1995; Westneat, 1995; Díaz *et al.*, 1996; Dutenhoffer and Swanson, 1996; Eppley, 1996; Irschick *et al.*, 1996;

Reynolds and Lee, 1996; Ricklefs *et al.*, 1996; Williams, 1996; Zani, 1996; Autumn *et al.*, 1997; Harris and Steudel, 1997; Carrascal *et al.*, 1999). Other methods exist (Martins and Hansen, 1996) and will continue to be developed (*e.g.*, Hansen, 1997; Martins and Hansen, 1997; Pagel, 1998), but we predict that independent contrasts will persist as a routine, and often sufficient, analytical tool.

Estimation of means versus ancestral values

We have shown how independent contrasts can be used to estimate phylogenetically correct means, standard errors, and confidence intervals for a given characteristic of a set of species. Such inferences depend on the quality of the phenotypic (tip) data, on the accuracy of the topology and branch lengths used in analyses, and on the assumption that evolutionary change has been akin to Brownian motion (see also Martins and Hansen, 1996). Past directional trends (at least simple ones) do not necessarily harm estimates of phylogenetically correct means and standard errors because, as noted by Grafen (1989, p. 148), "The logic of [independent contrasts] has no truck with history." Rather, independent contrasts can be viewed simply as a way to transform comparative data into values that are, in principle, independent and identically distributed.

The same estimates (means, standard errors, confidence intervals) can also be interpreted as characteristics of the hypothetical last common ancestor of all species in the data set. For this interpretation, however, past directional trends can be fatal. For example, if a phenotypic trait had undergone an evolutionary trend towards either increasing or decreasing values, but all other assumptions were valid, then the point estimate as an estimate of the ancestral value could be very far off (*e.g.*, Fig. 3).

Because the point estimate for the root node is based on global parsimony (see also Garland *et al.*, 1997; Schluter *et al.*, 1997), it will always falls within the range of values for species included in the analysis (*e.g.*, Fig. 3). As shown in Figure 2, however, when phylogenetic information is in-

corporated, confidence intervals about the root-node estimate can exceed the range of variation observed among extant species (Schluter et al., [1997] provide another example).

Although past directional trends harm inferences about ancestral states, they do not necessarily compromise estimation of relationships between characters, such as allometric slopes (see also Felsenstein, 1985; Grafen, 1989; Díaz-Uriarte and Garland, 1996, p. 30). And, the independent contrasts mean value can also be used to position regression equations computed by independent contrasts, and hence lacking a Y-intercept, back onto the original data space (Garland et al., 1993; example in Williams, 1996). Use of such equations for description and prediction also is not compromised by past directional trends (Garland and Ives, in preparation).

Standard errors (and hence confidence intervals) generally will be wider when phylogenetic relationships are incorporated in statistical computations. Contrary to common intuition, however, this difference is *not* because the phylogenetic computations use fewer degrees of freedom. This may at first seem puzzling, because the complication that hierarchical (*i.e.*, non-star) phylogenetic relationships introduce to comparative data is often referred to as a "degrees of freedom problem," which follows from the idea that a set of species' mean values does not constitute values that are independent in the statistical sense. As Pagel (1993, see also 1998) demonstrates, however, the real problem is that nonphylogenetic methods do not properly partition variance among the species.

As is apparent from Figure 2, confidence intervals on the root node can be quite sensitive to the branch lengths used for computations. Although no one has yet studied this sensitivity formally, it seems to be greater than for the estimation of correlated character evolution (*e.g.*, see Grafen, 1989; Garland et al., 1992; for empirical examples, see Garland et al., 1991; Martins, 1993; Westneat, 1995). Irrespective, the correctness of the branch lengths used for computations is clearly an important component of the overall data set. Various di-

agnostics for branch-length adequacy have been suggested, but only the one proposed by Garland et al. (1992) has been formally studied (see Díaz-Uriarte and Garland, 1996, 1998), and then only from the perspective of Type I error rates for estimating bivariate character correlations. Figure 2 indicates four different sets of branch lengths that would be judged adequate by the diagnostic of Garland et al. (1992), yet they vary widely in the 95% confidence interval (Pagel[4], Time$^{0.75}$, Grafen$^{2.5}$, Kitsch$^{2.6}$). This variation points to the need for further studies of estimation and testing of branch lengths for phylogenetically based statistical methods (see also Martins and Hansen, 1997).

Schluter et al. (1997) recently developed a method for deriving estimates and confidence intervals for the values of a trait at *all* internal nodes of a phylogenetic tree. This method uses maximum likelihood techniques to develop implicit formulae from which estimates and confidence intervals can be obtained numerically (their ANCML program). Although we have discussed our independent contrasts method primarily in terms of estimating the value of a trait at the base of a phylogenetic tree, the method can also be used to obtain estimates and confidence intervals for *any* internal node. This is done by reconfiguring the tree so that the node in question is at the root (see PDAP documentation). The resulting estimates and confidence intervals are identical to those obtained by the method of Schluter et al. (1997). Therefore, although couched in the terminology of independent contrasts and derived by use of a simpler methodology, our results are functionally equivalent to theirs. Note that the confidence intervals shown in Figure 8 of Schluter et al. (1997) are not *joint* confidence intervals. That is, they are not the appropriate confidence intervals for making simultaneous inferences about the entire set of internal nodes.

As developed by Martins and Hansen (1997), estimates of trait values at the basal node can also be obtained using generalized least-squares models (Judge et al., 1985). As will be shown in detail elsewhere (Garland and Ives, in preparation), generalized

least-squares estimators produce identical results to the explicit equations derived by use of the independent contrasts formalism. Although the generalized least-squares approach gives estimates only for the basal node, estimates of trait values at other internal nodes can be obtained by reconfiguring the phylogenetic tree to place the node in question at the base, as with the independent contrasts approach presented here.

Populations, species, and higher taxa as data points

A typical comparative data set consists of estimates of the average values for one or more phenotypic traits for each of several species. Sometimes data are available for multiple subspecies or populations within a species. If the phylogenetic relationships of these are known (or assumed) and gene flow is low, then the populations can add to the overall sample size just as if they were separate "species" (Garland *et al.*, 1992; Garland and Adolph, 1994; Foster and Cameron, 1996; Bauwens and Díaz-Uriarte, 1997; Pierce and Crawford, 1997). If gene flow occurs in a complicated fashion across the multiple populations, then difficulties arise because most analytical methods presume that phylogenetic relationships are divergent rather than reticulate (but see below on the phylogenetic autocorrelation approach). In the simplest case of two populations from a given species, gene flow between them only (potentially) shortens the branch lengths that should be used for analyses. Thus, we have suggested that a good design for a comparative study might be to include pairs of populations from each of a series of species (Garland *et al.*, 1992). The population differences would then provide information on microevolutionary (within-species) phenomena, whereas differences among species and higher nodes would inform about macroevolutionary phenomena.

When data sets are derived from the literature, it may only be possible to obtain a composite estimate of the average value for a species. That is, information from several different references, often involving different populations, may be averaged or otherwise combined to yield a single value that is then used to represent the species (*e.g.*, Clobert *et al.*, 1998). Obviously, this sort of procedure must be undertaken cautiously, because population differences are common both at the level of genotype and phenotype (Garland and Adolph, 1991). At higher taxonomic levels, and under the assumption that the taxa are monophyletic, data for different species are sometimes combined to yield an estimate of the average value for, say, a genus (*e.g.*, Moreno and Carrascal, 1993). In any case, a typical comparative data set consists of "average" values for a set of populations, subspecies, species or even higher taxa. For simplicity, we have referred to all such data points as "species."

Other methods: Monte Carlo computer simulations

All analytical procedures have limitations. One limitation that independent contrasts, as typically applied, shares with conventional statistical procedures is a reliance on conventional null distributions of test statistics (see Crowley, 1992). One way around such limitations is to use Monte Carlo computer simulations to create empirically scaled null distributions of whatever test statistic is of interest. If these simulations are done along a specified phylogeny, then the resulting null distributions will incorporate phylogenetic effects (Martins and Garland, 1991; Garland *et al.*, 1993; Reynolds and Lee, 1996). A computer program (*e.g.*, our PDSIMUL) starts at the root of a user-specified phylogenetic tree and simulates character evolution up the tree to create a set of tip data. These simulated data are then analyzed in the same way as the real data. The test statistic can be almost anything of interest, such as an estimate of a correlation between two traits. Many simulated data sets can be analyzed, typically 1,000, thus allowing creation of a phylogenetically correct null distribution of the test statistic. Because every detail of the evolutionary model must be specified, including the starting values at the root of the phylogenetic tree and, if desired, limits to how far the characters can evolve, the simulation approach makes assumptions completely explicit.

Computer simulation is a very general method for hypothesis testing, and can be combined with other phylogenetic methods, such as independent contrasts or squared-change parsimony (Garland et al., 1991; Martins and Garland, 1991; Moreno and Carrascal, 1993; Westneat, 1995; Butler and Losos, 1997). It can also incorporate uncertainties in phylogenetic topology (Losos, 1994; Stamps et al., 1997; Abouheif, 1998). Its main limitation is that it can be unwieldy for complicated analytical designs. Nevertheless, the method has been applied many times and seems especially useful for phylogenetic analysis of variance and covariance (e.g., Garland et al., 1993; Ferguson et al., 1996; Martin and Clobert, 1996; Reynolds and Lee, 1996; Harris and Steudel, 1997; Perez-Barberia and Gordon, 1999).

Other methods: phylogenetic autocorrelation

Cheverud and colleagues (Cheverud and Dow, 1985; Cheverud et al., 1985) developed a method, based on general network autocorrelation procedures, that uses phylogenetic information to partition the observed phenotypic mean value for each species into two parts, a phylogenetic component (the part that has been inherited from ancestors) and a specific component (representing independent evolution). The regression model is:

$$y = \rho W y + e$$

where y is the vector of observed phenotypic species' mean values (standardized), ρ is the autocorrelation coefficient to be estimated, which ranges from -1 to $+1$ and measures the correlation between the phenotypic trait vector y and the purely phylogenetic value Wy, W is the $N \times N$ connectivity matrix, indicating hypothesized or expected "phylogenetic similarity" (row normalized), and e is the residual vector indicating the specific component of y (independent evolution). Thus, the value for each species is predicted by a linear combination of the scores of all related species, which means either *all* species or a subset, depending on the elements of W. The elements of W are roughly comparable to the branch lengths required by independent contrasts (which must be in units of expected variance of character evolution). Diagonal elements of W are set to zero (contra the example in Harvey and Pagel, 1991, p. 135). However, different workers have derived the off-diagonal elements in various ways, so care must be taken when comparing results across both empirical (cf. Cheverud et al., 1985; Gittleman and Kot, 1990; Miles and Dunham, 1992; Martins, 1993; Westneat, 1995) and simulation studies (Purvis et al., 1994; Martins, 1996b). Most simply, the elements of W can be taken as the inverses of the pairwise phylogenetic (pathlength or "patristic") distances from each species to each other (e.g., Westneat, 1995; Pierce and Crawford, 1997); alternatively, the inverses of the number of nodes separating each pair of species can be used (Miles and Dunham, 1992; Westneat, 1995). Or, the amount of phylogenetic branch length (evolutionary history) shared by two species can be divided by the average total height of the branches from root to each of the two tips (Martins, 1993). Maximum likelihood is used to estimate the regression model because the dependent variable, y, appears on both sides of the equation. Gittleman and Kot (1990) extended the method to allow estimation of a parameter to transform the pairwise distances in W (their α parameter), which is similar to transformation of branch lengths that can be used with independent contrasts (Grafen, 1989; Garland et al., 1992; Reynolds and Lee, 1996).

The purpose of the phylogenetic autocorrelation method is two-fold. First, in principle, the estimated residual values, e, should be free of phylogenetic resemblance and, hence, independent (and identically distributed) in the statistical sense. If so, then they can be used in ordinary statistical procedures. In this sense, the putatively phylogeny-free residuals are very much analogous to Felsenstein's (1985) standardized independent contrasts. The crucial difference is that, whereas contrasts retain 100% of the variance in the original species' data, the autocorrelation residuals account for only a fraction of the total variance. Other than statistical convenience, no-

body has ever given a good reason why some of the among-species variance should be ignored for testing hypotheses about function or adaptation (Harvey and Pagel, 1991, pp. 137, 170: see also Martins, 1996*b*; Martins and Hansen, 1996). A further complication is that effects of phylogenetic inheritance and adaptation may not be fully separable (Hansen, 1997). Simulation studies indicate that autocorrelation residuals often perform less well than independent contrasts for estimating and testing correlations between characters (Purvis *et al.*, 1994), and that the former can be especially unreliable with sample sizes fewer than about 40 (Martins, 1996*b*). Still, results from the two methods often are remarkably similar (Garland *et al.*, 1993; Martins, 1993; Westneat, 1995; Pierce and Crawford, 1997; Carrascal *et al.*, 1999).

Second, the autocorrelation coefficient reflects the extent to which related species tend to resemble each other, and this statistic can be tested for statistical significance. No comparable statistic is simply available from independent contrasts (but see Price *et al.*, 1997). Positive values of ρ indicate that relatives are similar (*e.g.*, Fig. 1, top), and this is what is usually found in empirical studies (*e.g.*, Westneat, 1995; Pierce and Crawford, 1997). Negative values occur if relatives tend to be dissimilar (Cheverud and Dow, 1985), *e.g.*, because of character displacement, and are rarely found (but see Miles and Dunham, 1992). Another statistic from the autocorrelation method is (Cheverud *et al.*, 1985, p. 1343):

$$R^2 = 1-(\text{variance of } e/\text{variance of } y)$$

which is the proportion of variance accounted for by phylogeny. Either ρ or R^2 can be useful for comparing the "phylogenetic lability" of different characters (*e.g.*, Miles and Dunham, 1992; Gittleman *et al.*, 1996). If Gittleman and Kot's (1990) transformation parameter α is implemented, then the comparison of traits would need to consider both it and ρ or R^2. Another positive feature of the autocorrelation approach is that it can be applied with phylogenies that include hybridization events (*e.g.*, Cullum 1997).

Advice for comparative studies

We close with several pieces of advice for planning and conducting comparative studies. First, study more than two species (Garland and Adolph, 1994). When choosing species for study, begin with close relatives (perhaps two populations of the same species) and move out phylogenetically as far as necessary to encompass the appropriate range of phenotypic and ecological diversity. (The risk here is expanding the study so far, in terms of phylogenetic distance, that chalk and cheese starts being compared.) If resources allow, include enough outgroups to explore the phylogenetic generality of results and to aid in reconstruction of ancestral states.

In publications, present the raw tip data and adequate methods describing how the tip data were obtained (*e.g.*, where organisms were collected, detailed measurement conditions). Nothing is more frustrating than searching the literature, finding a data point for some difficult-to-study, rare or even recently extinct species, and not being able to include it in a review because the physiological methods are inadequately described! We need to ensure that comparative studies can be a cumulative enterprise, as are studies of phylogeny reconstruction by DNA sequencing. When gathering phylogenetic information, collaborate with systematists as necessary. Always present the phylogeny used (topology and branch lengths) so that others can reanalyze the data if and when improved phylogenetic information becomes available.

Finally, analyze data with conventional statistics and with phylogenetically independent contrasts and other appropriate methods. Differences among the methods may prove to be informative, and this is an area that requires more study (Martins and Garland, 1991; Martins and Hansen, 1996; Ricklefs and Starck, 1996; Butler and Losos, 1997; Price, 1997; Abouheif, 1998).

LITERATURE CITED

Abouheif, E. 1998. Random trees and the comparative method: a cautionary tale. Evolution 52:1197–1204.

Autumn, K., C. T. Farley, M. Emshwiller, and R. J. Full. 1997. Low cost of locomotion in the banded gecko: a test of the nocturnality hypothesis. Physiol. Zool. 70:660–669.

Barbosa, A. 1993. Morphometric variation of the hindlimb of waders and its evolutionary implications. Ardeola 40:65–75.

Barraclough, T. G., S. Nee, and P. H. Harvey. 1998. Sister-group analysis in identifying correlates of diversification. Evol. Ecol. 12:751–754.

Bauwens, D. and R. Díaz-Uriarte. 1997. Covariation of life-history traits in lacertid lizards: a comparative study. Am. Nat. 149:91–111.

Bauwens, D., T. Garland, Jr., A. M. Castilla, and R. Van Damme. 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 W. H. Dantzler (ed.), Handbook of physiology. Section 13: comparative physiology. Vol. I, pp. 3–16. Oxford University Press, New York.

Block, B. A., J. R. Finnerty, A. F. R. Stewart, and J. Kidd. 1993. Evolution of endothermy in fish: mapping physiological traits on a molecular phylogeny. Science 260:210–214.

Bonine, K. E., and T. Garland, Jr. 1999. Sprint performance of phrynosomatid lizards, measured on a high-speed treadmill, correlates with hindlimb length. Journal of Zoology, London 247:in press.

Brooks, D. R. and D. A. McLennan. 1991. Phylogeny, ecology, and behavior: a research program in comparative biology. University of Chicago Press, Chicago. 434 pp.

Butler, M. and J. B. Losos. 1997. Testing for unequal amounts of evolution in a continuous character on different branches of a phylogenetic tree using linear and squared-change parsimony: an example using Lesser Antillean Anolis lizards. Evolution 51:1623–1635.

Carrascal, L. M., E. Moreno, and I. Mozetich. 1999. Locomotion mode as a link between leg morphology and habitat preferences. A phylogenetic and ecomorphological study with Passeroidea (Aves: Passeriformes). Evol. Ecol. (In press).

Cheverud, J. M. and M. M. Dow. 1985. An autocorrelation analysis of genetic variation due to lineal fission in social groups of rhesus macaques. Am. J. Phys. Anthro. 67:113–121.

Cheverud, J. M., M. M. Dow, and W. Leutenegger. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. Evolution 39:1335–1351.

Clobert, J., T. Garland, Jr., and R. Barbault. 1998. The evolution of demographic tactics in lizards: a test

of some hypotheses concerning life history evolution. J. Evol. Biol. 11:329–364.

Crowley, P. H. 1992. Resampling methods for computation-intensive data analysis in ecology and evolution. Annu. Rev. Ecol. Syst. 23:405–447.

Cullum, A. 1997. Comparisons of physiological performance in sexual and asexual whiptail lizards (genus Cnemidophorus): Implications for the role of heterozygosity. Am. Nat. 150:24–47.

Díaz, J. A., D. Bauwens, and B. Asensio. 1996. A comparative study of the relation between heating rates and ambient temperatures in lacertid lizards. Physiol. Zool. 69:1359–1383.

Díaz-Uriarte, R. and T. Garland, Jr. 1996. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. Syst. Biol. 45:27–47.

Díaz-Uriarte, R. and T. Garland, Jr. 1998. Effects of branch lengths errors on the performance of phylogenetically independent contrasts. Syst. Biol. 47:654–672.

Doughty, P. 1996. Statistical analysis of natural experiments in evolutionary biology: Comments on recent criticisms of the use of comparative methods to study adaptation. Am. Nat. 148:943–956.

Dutenhoffer, M. S. and D. L. Swanson. 1996. Relationship of basal to summit metabolic rate in passerine birds and the aerobic capacity model for the evolution of endothermy. Physiol. Zool. 69:1232–1254.

Eggleton, P. and R. I. Vane-Wright (eds.) 1994. Phylogenetics and ecology. Linnean Society Symposium Series Number 17. Academic Press, London.

Eppley, Z. A. 1996. Charadriiform birds in Antarctica: behavioral, morphological, and physiological adjustments conserving reproductive success. Physiol. Zool. 69:1502–1554.

Feder, M. E., A. F. Bennett, W. W. Burggren, and R. B. Huey (eds.) 1987. New directions in ecological physiology. Cambridge University Press, New York. 364 pp.

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Ferguson, S. H., J. A. Virgl, and S. Lariviere. 1996. Evolution of delayed implantation and associated grade shifts in life history traits of North American carnivores. Ecoscience 3:7–17.

Foster, S. A. and S. A. Cameron. 1996. Geographic variation in behavior: a phylogenetic framework for comparative studies. In E. P. Martins (ed.), Phylogenies and the comparative method in animal behavior, pp. 138–165. Oxford University Press, Oxford.

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 L. J. Vitt and E. R. Pianka (eds.), Lizard ecology: Historical and experimental perspectives, pp. 237–259 (+ references). Princeton University Press, Princeton.

Garland, T., Jr. and S. C. Adolph. 1991. Physiological

differentiation of vertebrate populations. Annu. Rev. Ecol. Syst. 22:193–228.

Garland, T., Jr. and S C. Adolph. 1994. Why not to do two-species comparative studies: limitations on inferring adaptation. Physiol. Zool. 67:797–828.

Garland, T., Jr. and P. A. Carter. 1994. Evolutionary physiology. Annu. Rev. Physiol. 56:579–621.

Garland, T., Jr. and R. Díaz-Uriarte. 1999. Polytomies and phylogenetically independent contrasts: an examination of the bounded degrees of freedom approach. Syst. Biol. (In press).

Garland, T., Jr., P. H. Harvey, and A. R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. Syst. Biol. 41:18–32.

Garland, T., Jr., R. B. Huey, and A. F. Bennett. 1991. Phylogeny and thermal physiology in lizards: a reanalysis. Evolution 45:1969–1975.

Garland, T., Jr., K. L. M. Martin, and R. Díaz-Uriarte. 1997. Reconstructing ancestral trait values using squared-change parsimony: plasma osmolarity at the amphibian-amniote transition. *In* S. S. Sumida and K. L. M. Martin (eds.), *Amniote origins: Completing the transition to land,* pp. 425–501. Academic Press, San Diego.

Garland, T., Jr., A. W. Dickerman, C. M. Janis, and J. A. Jones. 1993. Phylogenetic analysis of covariance by computer simulation. Syst. Biol. 42:265–292.

Gibbs, A. G. 1999. Laboratory selection for the comparative physiologist. J. Exp. Biol. In press.

Gittleman, J. L. and M. Kot. 1990. Adaptation: Statistics and a null model for estimating phylogenetic effects. Syst. Zool. 39:227–241.

Gittleman, J. L., C. G. Anderson, M. Kot, and H.-K. Luh. 1996. Phylogenetic lability and rates of evolution: a comparison of behavioral, morphological and life history traits. *In* E. P. Martins (ed.), *Phylogenies and the comparative method in animal behavior,* pp. 166–205. Oxford University Press, Oxford.

Grafen, A. 1989. The phylogenetic regression. Phil. Trans. Royal. Soc. Lond. B 326:119–157.

Grafen, A. and M. Ridley. 1996. Statistical tests for discrete cross-species data. J. Theor. Biol 183: 255–267.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51 1341–1351.

Harris, M. A. and K. Steudel. 1997. Ecological correlates of hind-limb length in the Carnivora. J. Zool., London 241:381–408.

Harvey, P. H. and M. D. Pagel. 1991. *The comparative method in evolutionary biology.* Oxford University Press, Oxford. 239 pp.

Hayes, J. P. and T. Garland, Jr. 1995. The evolution of endothermy: testing the aerobic capacity model. Evolution 49:836–847.

Huey, R. B. and A. F. Bennett. 1987. Phylogenetic studies of coadaptation: preferred temperatures versus optimal performance temperatures of lizards. Evolution 41:1098–1115.

Irschick, D. J., C. C. Austin, K. Petren, R. N. Fisher, J. B. Losos, and O. Ellers. 1996. A comparative analysis of clinging ability among pad-bearing lizards. Biol. J. Linn. Soc. 59:21–35.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lutkepohl, and T.-C. Lee. 1985. *The theory and practice of econometrics.* John Wiley and Sons, New York.

Koteja, P., J. G. Swallow, P. A. Carter, and T. Garland, Jr. 1999. Energy cost of wheel running in house mice: Implications for coadaptation of locomotion and energy budgets. Physiol. Biochem. Zool. 72: 238–249.

Krebs, J. R. and N. B. Davies (eds.) 1997. *Behavioural ecology: An evolutionary approach.* 4th ed. Blackwell Science, Oxford. 456 pp.

Larson, A. and J. B. Losos. 1996. Phylogenetic systematics of adaptation. *In* M. R. Rose and G. V. Lauder (eds.), *Adaptation,* pp. 187–220. Academic Press, San Diego.

Leroi, A. M., M. R. Rose, and G. V. Lauder. 1994. What does the comparative method reveal about adaptation? Am. Nat. 143:381–402.

Losos, J. B. 1990. Ecomorphology, performance capability, and scaling of West Indian *Anolis* lizards: An evolutionary analysis. Ecol. Monogr. 60:369–388.

Losos, J. B. 1994. An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. Syst. Biol. 43:117–123.

Losos, J. B. and D. B. Miles. 1994. Adaptation, constraint, and the comparative method: Phylogenetic issues and methods. *In* P. C. Wainwright and S. M. Reilly (eds.), *Ecological morphology: Integrative organismal biology,* pp. 60–98. University of Chicago Press, Chicago.

Maddison, D. R. 1994. Phylogenetic methods for inferring the evolutionary history and process of change in discretely valued characters. Annu. Rev. Entomol. 39:267–292.

Maddison, W. P. 1991. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters. Syst. Zool. 40:304–314.

Maddison, W. P. 1995. Calculating the probability distributions of ancestral states reconstructed by parsimony on phylogenetic trees. Syst. Biol. 44:474–481.

Martin, T. E. and J. Clobert. 1996. Nest predation and avian life history evolution in Europe versus North America: a possible role of humans? Am. Nat. 147:1028–1046.

Martinez del Rio, C., K. E. Brugger, J. L. Rios, M. E. Vergara, and M. Witmer. 1995. An experimental and comparative study of dietary modulation of intestinal enzymes in European starlings (*Sturnus vulgaris*). Physiol. Zool. 68:490–511.

Martins, E. P. 1993. A comparative study of the evolution of *Sceloporus* push-up displays. Am. Nat. 142:994–1018.

Martins, E. P. (ed.) 1996a. *Phylogenies and the comparative method in animal behavior.* Oxford University Press, Oxford. 415 pp.

Martins, E. P. 1996b. Phylogenies, spatial autoregression and the comparative method: A computer simulation test. Evolution 50:1750–1765.

Martins, E. P. and T. Garland, Jr. 1991. Phylogenetic analyses of the correlated evolution of continuous

characters: A simulation study. Evolution 45:534–557.

Martins, E. P. and T. F. Hansen. 1996. The statistical analysis of interspecific data: a review and evaluation of comparative methods. *In* E. P. Martins (ed.), *Phylogenies and the comparative method in animal behavior*, pp. 22–75. Oxford University Press, Oxford.

Martins, E. P. and T. F. Hansen. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149:646–667.

Martins, E. P. and J. Lamont. 1998. Estimating ancestral states of a communicative display: A comparative study of *Cyclura* rock iguanas. Anim. Behav. 55:1685–1706.

McPeek, M. A. 1995. Testing hypotheses about evolutionary change on single branches of a phylogeny using evolutionary contrasts. Am. Nat. 145:686–703.

Miles, D. B. and A. E. Dunham. 1992. Comparative analyses of phylogenetic effects in the life history patterns of iguanid reptiles. Am. Nat. 139:848–869.

Miles, D. B. and A. E. Dunham. 1993. Historical perspectives in ecology and evolutionary biology: the use of phylogenetic comparative analyses. Annu. Rev. Ecol. Syst. 24:587–619.

Moreno, E. and L. M. Carrascal. 1993. Ecomorphological pattern of aerial feeding in oscines (Passeriformes: Passeri). Biol. J. Linn. Soc. 50:147–165.

Pagel, M. D. 1992. A method for the analysis of comparative data. J. Theor. Biol. 156:431–442.

Pagel, M. D. 1993. Seeking the evolutionary regression coefficient: An analysis of what comparative methods measure. J. Theor. Biol. 164:191–205.

Pagel, M. 1998. Inferring evolutionary processes from phylogenies. Zoologica Scripta 26:331–348.

Perez-Barberia, F. J., and I. J. Gordon. 1999. The functional relationship between feeding type and jaw and cranial morphology in ungulates. Oecologia 118:157–165.

Pierce, V. A. and D. L. Crawford. 1997. Phylogenetic analysis of glycolytic enzyme expression. Science 276:256–259.

Price, T. 1997. Correlated evolution and independent contrasts. Phil. Tran. Roy. Soc. Lond. B 352:519–529.

Price, T. D., A. J. Helbig, and A. D. Richman. 1997. Evolution of breeding distributions in the Old World leaf warblers (Genus *Phylloscopus*). Evolution 51:552–561.

Promislow, D. E. L. 1991. The evolution of mammalian blood parameters: Patterns and their interpretation. Physiol. Zool. 64:393–431.

Purvis, A. and T. Garland, Jr. 1993. Polytomies in comparative analyses of continuous characters. Syst. Biol. 42:569–575.

Purvis, A., J. L. Gittleman, and H.-K. Luh. 1994. Truth or consequences: Effects of phylogenetic accuracy on two comparative methods. J. Theor. Biol. 167:293–300.

Reynolds, P. S. and R. M. Lee, III. 1996. Phylogenetic analysis of avian energetics: Passerines and non-passerines do not differ. Am. Nat. 147:735–759.

Ricklefs, R. E. and J. M. Starck. 1996. Applications of phylogenetically independent contrasts: A mixed progress report. Oikos 77:167–172.

Ricklefs, R. E., M. Konarzewski, and S. Daan. 1996. The relationship between basal metabolic rate and daily energy expenditure in birds and mammals. Am. Nat. 147:1047–1071.

Ridley, M. and A. Grafen. 1996. How to study discrete comparative methods. *In* E. P. Martins (ed.), *Phylogenies and the comparative method in animal behavior*, pp. 76–103. Oxford University Press, Oxford.

Ryan, M. J. and A. S. Rand. 1995. Female responses to ancestral advertisement calls in Tungara frogs. Science 269:390–392.

Schluter, D., T. Price, A. O. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51:1699–1711.

Sessions, S. K. and A. Larson. 1987. Developmental correlates of genome size in plethodontid salamanders and their implications for genome evolution. Evolution 41:1239–1251.

Stamps, J. A., J. B. Losos, and R. M. Andrews. 1997. A comparative study of population density and sexual size dimorphism in lizards. Am. Nat. 149:64–90.

Wainwright, P. C. and S. M. Reilly (eds.) 1994. *Ecological morphology: Integrative organismal biology*. University of Chicago Press, Chicago. 367 pp.

Walton, B. M. 1993. Physiology and phylogeny: the evolution of locomotor energetics in hylid frogs. Am. Nat. 141:26–50.

Westneat, M. W. 1995. Feeding, function, and phylogeny: Analysis of historical biomechanics in labrid fishes using comparative methods. Syst. Biol. 44:361–383.

Williams, J. B. 1996. A phylogenetic perspective of evaporative water loss in birds. Auk 113:457–472.

Wolf, C. M., T. Garland, Jr., and B. Griffith. 1998. Predictors of avian and mammalian translocation success: Reanalysis with phylogenetically independent contrasts. Biol. Conservation 86:243–255.

Zani, P. A. 1996. Patterns of caudal-autotomy evolution in lizards. J. Zool., London 240:201–220.

Corresponding Editor: Todd Gleeson

Appendix

The presentation is made in two parts. First, we analyze the simplest possible phylogenetic tree with only two tips. Second, we develop the case of an arbitrarily complex tree with multiple tips.

Consider a phylogenetic tree with two tips having characters described by the random variables X_1 and X_2, and let the value at the basal node be given by the random variable X_z. Under the assumption that evolution proceeds in a process like Brownian motion, X_1 and X_2 are normally distributed with variances v_1 r and v_2 r, respectively, where v_1 and v_2 are branch lengths

on the phylogenetic tree and r is the rate of character evolution. The estimator of the expectation of X_z, denoted $\hat{X}_z$, is (Felsenstein, 1985)

$$\hat{X}_z = \frac{v_2}{v_1 + v_2}X_1 + \frac{v_1}{v_1 + v_2}X_2. \quad (A1)$$

This expression is the average of X_1 and X_2 weighted by a factor that is proportional to the inverse of their variances.

Define the estimator of r as the squared difference between the two tip values weighted by the inverse of the sum of the branch lengths (Felsenstein, 1985):

$$\hat{r} = \frac{1}{v_1 + v_2}(X_1 - X_2)^2 = r\left(\frac{X_1 - X_2}{\sqrt{(v_1 + v_2)r}}\right)^2. \quad (A2)$$

Because X_1 and X_2 are independent, the contrast (X_1-X_2) is normally distributed with mean 0 and variance $(v_1 + v_2)r$. Therefore, by definition $(X_1 - X_2)/\sqrt{(v_1 + v_2)r}$ is a χ^2 distribution with one degree of freedom. This gives the 2α confidence intervals for r as

$$\left(\frac{\hat{r}}{\chi^2_{1-\alpha,1}}, \frac{\hat{r}}{\chi^2_{\alpha,1}}\right). \quad (A3)$$

The distribution of $\hat{X}_z$ is calculated by noting that

$$\frac{\hat{X}_z - x_z}{\sqrt{\frac{v_1 v_2}{v_1 + v_2}\hat{r}}} = \frac{\frac{\hat{X}_z - x_z}{\sqrt{v_1 v_2 r/(v_1 + v_2)}}}{\sqrt{\chi^2_1}} \quad (A4)$$

is a t distribution with one degree of freedom. Here, x_z denotes the true value of the character at the basal node. Eq. A4 follows from the fact that

$$\hat{X}_z - x_z = \left(\frac{v_2}{v_1 + v_2}\right)X_1 + \left(\frac{v_1}{v_1 + v_2}\right)X_2 - x_z \quad (A5)$$

is a normal distribution with mean 0 and variance $(v_1 v_2/(v_1 + v_2))$, and by definition a Student t distribution with one degree of freedom is a normal distribution (with mean 0 and variance 1) divided by the square root of a χ^2 distribution with one degree of freedom.

The case of an arbitrarily complex phylogenetic tree is similar. Let X_z be the value of the character at the basal node, and let X_1 and X_2 be the values at the nodes immediately above. The estimator of the expectation of X_z is now (Felsenstein, 1985)

$$\hat{X}_z = \frac{v_2'}{v_1' + v_2'}\hat{X}_1 + \frac{v_1'}{v_1' + v_2'}\hat{X}_2 \quad (A6)$$

where $\hat{X}_i$ is the normally distributed estimator of X at node i. Normality of $\hat{X}_i$ follows from the normality of successive estimators working down from the tips of the phylogenetic tree. As before, the values of $\hat{X}_i$ are weighted by the inverse of their variances $v_i'r$, where v_i' is the corrected branch length below node i. v_i' has two components (Felsenstein, 1985). The first, v_i, is the length of the branch between the basal node and node i which accounts for Brownian motion evolution.

The second is $v_j'v_k'/(v_j' + v_k')$, where j and k are the nodes above node i. This accounts for the variance in the estimator of X_i as derived in Eq. A5.

The estimator of r for a phylogenetic tree with N tips is defined as (Felsenstein, 1985)

$$\hat{r} = \frac{1}{N - 1} \sum_{contrasts} \frac{1}{v_j' + v_k'}(\hat{X}_j - \hat{X}_k)^2$$

$$= r\frac{1}{N - 1}\sum_{contrasts}\left(\frac{\hat{X}_j - \hat{X}_k}{\sqrt{(v_j' + v_k')r}}\right)^2 \quad (A7)$$

where the summation is taken over the $N-1$ contrasts between adjacent nodes. This estimator of r is the sum of squared differences between adjacent nodes weighted by the corrected branch lengths separating them. $\sum_{contrasts}(\hat{X}_j - \hat{X}_k)/\sqrt{(v_j' + v_k')r}$ is the sum of squares of $N - 1$ independent normal distributions with mean 0 and variance 1, so by definition it is a χ^2 distribution with $N - 1$ degrees of freedom. Independence of the normal distributions in Eq. A7 follows from the assumption that rates of evolution down sister branches of the phylogenetic tree are identical and independent.

As in the case of the two-tipped tree, the distribution of $\hat{X}_z - x_z$ is normal with mean 0 and variance $v_1' v_2' r/(v_1' + v_2')$. Therefore,

$$\frac{\hat{X}_z - x_z}{\sqrt{\frac{v_1' v_2'}{v_1' + v_2'}\hat{r}}} = \frac{\frac{\hat{X}_z - x_z}{\sqrt{v_1' v_2' r/(v_1' + v_2')}}}{\sqrt{\frac{\chi^2_{N-1}}{N - 1}}} \quad (A8)$$

is by definition a Student t distribution with N-1 degrees of freedom, and the 2α confidence interval for the estimate of x_z is

$$\hat{X}_z \pm t_{\alpha,N-1}\sqrt{\frac{v_1' v_2'}{v_1' + v_2'}\hat{r}}. \quad (A9)$$

To obtain a formula for the prediction interval for the value of x for a new species h, x_h, first consider the case in which the branch leading to species h is rooted at the base of the phylogenetic tree and has length v_h. The mean of the estimate of x_h, $\hat{X}_h$, equals the mean of $\hat{X}_z$, the estimate of x at the base of the tree. The variance of $\hat{X}_h$ equals the variance of $\hat{X}_z$ plus the variance due to evolution along the branch leading to species h: $v_1' v_2' r/(v_1' + v_2') + rv_h$. Therefore, the 2α prediction interval for $\hat{X}_h$ is

$$\hat{X}_z \pm t_{\alpha,N-1}\sqrt{\hat{r}\left(\frac{v_1' v_2'}{v_1' + v_2'} + v_h\right)}. \quad (A10)$$

This can be generalized for a species located anywhere on the phylogenetic tree by re-rooting the tree so that the hypothetical branch leading to the new species h stems directly from the base (root). Note that this re-rooting can even be in the middle of a branch (also see text).

PDTREE calculates the estimated basal node value (Eq. A6) and its confidence interval (Eq. A9). PDINSTR.DOC provides a worked example.