Evolution, 57(4), 2003, pp. 717–745

TESTING FOR PHYLOGENETIC SIGNAL IN COMPARATIVE DATA: BEHAVIORAL TRAITS ARE MORE LABILE

SIMON P. BLOMBERG,[1] THEODORE GARLAND, JR.,[1,2] AND ANTHONY R. IVES[3,4]
[1]*Department of Biology, University of California, Riverside, California 92521*
[2]*E-mail: tgarland@citrus.ucr.edu*
[3]*Department of Zoology, University of Wisconsin, Madison, Wisconsin 53706*
[4]*E-mail: arives@facstaff.wisc.edu*

Abstract.—The primary rationale for the use of phylogenetically based statistical methods is that phylogenetic signal, the tendency for related species to resemble each other, is ubiquitous. Whether this assertion is true for a given trait in a given lineage is an empirical question, but general tools for detecting and quantifying phylogenetic signal are inadequately developed. We present new methods for continuous-valued characters that can be implemented with either phylogenetically independent contrasts or generalized least-squares models. First, a simple randomization procedure allows one to test the null hypothesis of no pattern of similarity among relatives. The test demonstrates correct Type I error rate at a nominal $\alpha = 0.05$ and good power (0.8) for simulated datasets with 20 or more species. Second, we derive a descriptive statistic, *K*, which allows valid comparisons of the amount of phylogenetic signal across traits and trees. Third, we provide two biologically motivated branch-length transformations, one based on the Ornstein-Uhlenbeck (OU) model of stabilizing selection, the other based on a new model in which character evolution can accelerate or decelerate (ACDC) in rate (e.g., as may occur during or after an adaptive radiation). Maximum likelihood estimation of the OU (*d*) and ACDC (*g*) parameters can serve as tests for phylogenetic signal because an estimate of *d* or *g* near zero implies that a phylogeny with little hierarchical structure (a star) offers a good fit to the data. Transformations that improve the fit of a tree to comparative data will increase power to detect phylogenetic signal and may also be preferable for further comparative analyses, such as of correlated character evolution. Application of the methods to data from the literature revealed that, for trees with 20 or more species, 92% of traits exhibited significant phylogenetic signal (randomization test), including behavioral and ecological ones that are thought to be relatively evolutionarily malleable (e.g., highly adaptive) and/or subject to relatively strong environmental (nongenetic) effects or high levels of measurement error. Irrespective of sample size, most traits (but not body size, on average) showed less signal than expected given the topology, branch lengths, and a Brownian motion model of evolution (i.e., *K* was less than one), which may be attributed to adaptation and/or measurement error in the broad sense (including errors in estimates of phenotypes, branch lengths, and topology). Analysis of variance of log *K* for all 121 traits (from 35 trees) indicated that behavioral traits exhibit lower signal than body size, morphological, life-history, or physiological traits. In addition, physiological traits (corrected for body size) showed less signal than did body size itself. For trees with 20 or more species, the estimated OU (25% of traits) and/or ACDC (40%) transformation parameter differed significantly from both zero and unity, indicating that a hierarchical tree with less (or occasionally more) structure than the original better fit the data and so could be preferred for comparative analyses.

Key words.—Adaptation, behavior, body size, branch lengths, comparative method, constraint, physiology.

Received March 29, 2002. Accepted November 27, 2002.

A great triumph of biology has been the demonstration that organisms are descended from common ancestors and hence are related in a hierarchical fashion (Mayr 1982). An observation of almost equal importance is that phylogenetically related organisms tend to resemble each other for most aspects of the phenotype (e.g., hummingbirds look like hummingbirds, elephants look like elephants). We term this resemblance phylogenetic ''signal'' to avoid such terms as phylogenetic ''inertia'' or ''constraint,'' which are inconsistently defined, both conceptually and operationally (e.g., see Antonovics and van Tienderen 1991; McKitrick 1993; Wagner and Schwenk 2000; Orzack and Sober 2001; Reeve and Sherman 2001) and whose original use was rather different from current use (Blomberg and Garland 2002). As argued elsewhere (Blomberg and Garland 2002), we believe that the evolutionary processes implied by phylogenetic inertia and constraint are difficult, if not impossible, to estimate from comparative data alone (but see Hansen 1997; Orzack and Sober 2001; Schwenk and Wagner 2001). Therefore, we use phylogenetic signal simply to describe a tendency (pattern) for evolutionarily related organisms to resemble each other, with no implication as to the mechanism that might cause such resemblance (process). This use of phylogenetic signal is consistent with recent usage in systematic biology (e.g., Hillis and Huelsenbeck 1992) and is also similar to the ''phylogenetic effect'' of Derrickson and Ricklefs (1988) and the ''phylogenetic conservatism'' of Ashton (2001). We do not use ''phylogenetic conservatism,'' however, because it seems to connote less change than might be expected from the phylogenetic structure of the taxa in question; as argued below, the existence of phylogenetic signal does not require the existence of such processes as that may imply. We did not use the ''phylogenetic correlation'' of Gittleman et al. (1996a) because they used statistical approaches different from those presented here and because our methods do not involve correlation per se.

A crucial point that has not been sufficiently appreciated in the literature is that phylogenetic signal will occur to some extent under most simple, stochastic models of character evolution along a tree with any amount of hierarchical structure. Under such models as Brownian motion, evolutionary changes are simply added to values present in the previous generation or at the previous node on a phylogenetic tree. Thus, members of lineages that have only recently diverged will necessarily (on average) tend to be similar, as compared with more distantly related lineages. This effect carries on to the

values for taxa at the tips of a phylogenetic tree (e.g., extant species). The presence of phylogenetic signal does not require a failure to evolve (Wake 1991) along certain branches, but only that the branch lengths of a phylogenetic tree are proportional to the expected variance of evolution for the character in question. Hence, significant phylogenetic signal does not require appeal to such processes as inertia or constraint (Blomberg and Garland 2002).

Although phylogenetic signal is a necessary property of stochastic evolution along a hierarchical tree, whether phylogenetic signal can be detected for a given trait will, of course, depend on the sample size of a dataset, the power of the statistical test, the accuracy of the phylogenetic tree (e.g., topological errors will tend to obscure the signal), and the accuracy of the trait data (high measurement error will also tend to obscure signal). Moreover, a test for phylogenetic signal can also be viewed as a test for hierarchical tree structure, if one assumes something like Brownian motion to have been in effect. Or, it can be interpreted as a test for Brownian motion if one assumes the phylogenetic topology and branch lengths are known without error. (Related to the issue of detecting phylogenetic signal is the issue of quantifying and comparing evolutionary rates [e.g., see Garland 1992; Martins 1994]. For example, traits that evolve slowly are often identified as having phylogenetic ''inertia'' or being subject to phylogenetic ''constraint'' [review in Gittleman et al. 1996a; see also Blomberg and Garland 2002].)

Although it is simply a necessary consequence of stochastic evolution along a hierarchical tree, the existence of phylogenetic signal has important practical consequences. First, the characteristics (either phenotypic or genetic) of a species that has not yet been studied can often be rather well predicted simply from knowledge of its phylogenetic position and the characteristics of some close relatives. For example, the ability to detect species that deviate significantly from general allometric equations can be greatly enhanced by the use of phylogenetically informed statistical procedures (Garland and Ives 2000). Second, average values for the phenotypes of a set of species generally do not represent independent and identically distributed datapoints for a statistical analysis. (They also typically do not represent random samples from the ''population'' of all organisms or even of a more restricted lineage.) Hence, conventional statistical methods are usually inappropriate for the analysis of comparative data (e.g., Ridley 1983; Felsenstein 1985; Grafen 1989; Harvey and Pagel 1991; Garland et al. 1993, 1999; Reynolds and Lee 1996; Martins and Hansen 1997; Ackerly 2000; Rohlf 2001; Swiderski 2001). This realization has led to a large body of work aimed at the development of phylogenetically based statistical methods, spurred in particular by the publication of two different methods in 1985 (Cheverud et al. 1985; Felsenstein 1985). Both analytical and simulation studies have shown that, under a fairly wide range of phylogenies and models of character evolution, related species do tend to resemble each other and that this tendency will inflate Type I error rates for testing hypotheses (e.g., whether two traits are correlated across species) with comparative data (e.g., Grafen 1989; Martins and Garland 1991; Purvis et al. 1994; Díaz-Uriarte and Garland 1996; Harvey and Rambaut 1998; Martins et al. 2002).

Theory is one thing, but the real world is another. Hierarchical phylogenetic relationships and a tendency for relatives to resemble each other are not without exception. For example, interspecific hybridization and horizontal gene transfer violate the general idea that phylogenetic trees are always divergent. Convergent evolution leads to distantly related organisms resembling each other more closely than expected (Wake 1991), whereas character displacement causes closely related organisms to resemble each other less than expected (Losos 2000). Other modes of evolution can also lead to amounts of phylogenetic signal that are lower than expected from a given topology and branch lengths, including limits to evolution (e.g., see Garland et al. 1993; Díaz-Uriarte and Garland 1996) and the Ornstein-Uhlenbeck process (which models stabilizing selection that can obliterate the resemblance of descendant species and their ancestors, e.g., Felsenstein 1988; Garland et al. 1993; Hansen and Martins 1996; Martins et al. 2002). Therefore, several workers have suggested that it might not always be necessary to incorporate phylogenetic information in statistical analyses of comparative data (e.g., Westoby et al. 1995; Bjorklund 1997; Price 1997; Mazer 1998; Losos 1999; Ashton 2001). For a given set of species and a given phenotypic trait, the amount of phylogenetic signal might not be very strong. Thus, perhaps one should first test for phylogenetic signal in a dataset and apply a phylogenetically based statistical method only if it is statistically significant (e.g., Gittleman and Kot 1990; Gittleman et al. 1996a; Abouheif 1999). As a relatively early example, Gittleman and Luh (1992, p. 401) concluded that ''it is mandatory initially to diagnose the comparative data at hand. . . . If phylogenetic correlation is not observed, then comparative method procedures should not be adopted.'' Following this suggestion, some empirical papers have not used phylogenetically based statistical methods, after one or more diagnostic tests indicated a lack of significant phylogenetic signal (e.g., Irschick et al. 1997).

This perspective suggests a dichotomy between phylogenetic and nonphylogenetic analyses, but this dichotomy is both illusory and unnecessary. As noted previously (e.g., Purvis and Garland 1993), a conventional (nonphylogenetic) statistical analysis is just one that assumes the special case of a star phylogeny, that is, a hard polytomy with equal-length branches (and implicitly assumed character evolution by something like Brownian motion with stochastically equal rates along each branch). The three best-justified phylogenetically based statistical methods can all be applied with a star phylogeny, in which case they yield results that are exactly the same as a conventional analysis (phylogenetically independent contrasts: Felsenstein 1985; generalized least squares: Grafen 1989; Martins and Hansen 1997; Garland and Ives 2000; Rohlf 2001; Monte Carlo simulations: Martins and Garland 1991; Garland et al. 1993; Reynolds and Lee 1996). Moreover, as emphasized here (see also Freckleton et al. 2002), various types of branch length-transformations can allow one to conduct analyses on trees that vary continuously from a star with contemporaneous tips to the original candidate tree and on to a tree with branches transformed to imply even more phylogenetic signal than implied by the original candidate tree. The optimal branch-length transformation can be chosen objectively by such criteria as maxi-

mum likelihood (ML; best fit to data; e.g., Grafen 1989), thus allowing what amounts to a simultaneous test for phylogenetic signal (see also Lynch 1991; Mooers et al. 1999; Pagel 1999; Harvey and Rambaut 2000; Martins et al. 2002), and/or analyses can be conducted across a range of branch-length transformations as a sensitivity analysis (e.g., Butler et al. 2000).

Given the limitations of existing methods (see Discussion), our first purpose is to provide a new test for phylogenetic signal in continuous-valued characters, based on simple randomization procedures. This test allows us to address how common phylogenetic signal really is. Second, we provide a descriptive statistic, *K*, to gauge the amount of phylogenetic signal. Such a statistic is important because it allows comparisons of different traits across different trees and hence the possibility of discovering general patterns of relative evolutionary lability across trait types (cf. de Queiroz and Wimberger 1993; Gittleman et al. 1996a,b; Wimberger and de Queiroz 1996; Ackerly and Donoghue 1998; Ackerly and Reich 1999; Böhning-Gaese and Oberrath 1999; Morales 2000; Prinzing et al. 2001). Third, we implement two biologically motivated transformations of branch lengths, one of which corresponds to the Ornstein-Uhlenbeck (OU) process and the other to a new proposed model of accelerating versus decelerating rates of character evolution (ACDC). Maximum likelihood estimation of the OU or ACDC transformation parameter provides an alternative approach to testing for phylogenetic signal and can be used to guide the choice of branch lengths for a comparative analysis.

As described in Felsenstein's (1985) seminal paper, the branch lengths used in phylogenetically based statistical methods are assumed to be proportional to expected variance of character evolution for the trait(s) in question. Such branch lengths cannot be known directly, except perhaps in some laboratory organisms with known histories, generation times, mutation rates, and selective regimes. Therefore, reasonable approximations must be employed. Starter branch lengths may be real, such as estimated genetic distances or divergence times, or they may be arbitrary, as derived from a simple setting rule (e.g., Cheverud et al. 1985; Grafen 1989; Pagel 1992; all = 1). Real branch lengths obviously can enhance the biological relevance of statistical inferences, but arbitrary branch lengths do not necessarily negate biological interpretations (e.g., see Grafen 1989; Garland 1992; Clobert et al. 1998). Whatever starter branch lengths are employed, performance of phylogenetically based statistical methods, in terms of Type I error rate, can often be improved by transformation of the branch lengths (Grafen 1989; Garland et al. 1992; Díaz-Uriarte and Garland 1996, 1998; Martins et al. 2002). Transformation of branch lengths will also affect the interpretation of parameters estimated from comparative data, including the degree of phylogenetic signal and the form of the relationship between traits (e.g., see Gittleman and Kot 1990; Martins 1996; Garland et al. 1999). Therefore, we also review various ways of transforming branch lengths. Finally, we apply all of our proposed procedures to empirical examples taken from the literature. We ask how common significant phylogenetic signal really is, compare the amount of phylogenetic signal among different types of traits, and compare results of the randomization and branch-length-transformation tests for signal.

A Randomization Test for Phylogenetic Signal

A randomization procedure can be used to test whether a given set of comparative data exhibit a significant tendency for related species to resemble each other. This test makes no assumption about the model of evolution that produced the data observed at the tips (terminal nodes) of the phylogeny, except to the extent that (possibly arbitrary) branch lengths are used in the calculations and they imply something about how much variance in character evolution occurs between nodes (e.g., Felsenstein 1985; Díaz-Uriarte and Garland 1996, 1998).

The basic idea is to ask whether a given tree (topology and branch lengths) better fits a set of tip data as compared with the fit obtained when the data have been randomly permuted across the tips of the tree, thus destroying any phylogenetic signal that may have existed. The test can be implemented via a GLS approach (Grafen 1989; Martins and Hansen 1997; Garland et al. 1999; Pagel 1999; Butler et al. 2000; Garland and Ives 2000; Rohlf 2001), for which we have written the MatLab program PHYSIG.M (see Appendix 1; this and other programs are available at http://www.biology.ucr.edu/faculty/Garland/PHYSIG.html). Alternatively, it can be implemented via phylogenetically independent contrasts, for which existing modules of our Phenotypic Diversity Analysis Program (PDAP: http://www.biology.ucr.edu/faculty/Garland/PDAP.html) can be used.

With independent contrasts, the method works as follows. For a given set of real data, one computes standardized phylogenetically independent contrasts in the usual way (Felsenstein 1985; Garland et al. 1992; Purvis and Garland 1993). The variance of the contrasts is then computed as an index of how well the tree fits the data. If related species tend to have similar values for a given trait (i.e., they exhibit phylogenetic signal), then the variance of contrasts will tend to be low. To determine whether the observed phylogenetic signal is statistically significant, the variance of contrasts for the real data can be compared with the values obtained after the data have been permuted randomly across the tips of the tree without regard to phylogenetic relationships. If, say, 95% of the permuted datasets show variances of contrasts that are greater than that for the data in their correct positions, then we reject the null hypothesis of no phylogenetic signal. This would be appropriate for a one-tailed test. The test can also detect cases for which the resemblance of relatives is actually lower than expected (which we might term phylogenetic ''antisignal''). Here, the permuted data would have lower variances than for the data in their original, correct positions on the tree. The procedure for computing the test with PDAP modules (Garland et al. 1999; Garland and Ives 2000; Lapointe and Garland 2001) is described in its accompanying documentation (PDIMSTRW.DOC). In GLS mode, the mean squared error (MSE) is used as the test statistic, rather than the variance of contrasts, but the test is performed in exactly the same fashion as described above and the results will be identical (following from the results in Appendices 4 and 5).

Type I Error Rate and Power of the Randomization Test

To calculate the approximate Type I error rate of our proposed randomization test for phylogenetic signal, we simulated 1000 datasets under the Brownian motion assumption (PDSIMUL of Garland et al. 1993), using a star phylogeny (contemporaneous tips) with 32 species (one large polytomy with equal branch lengths). For these datasets, species have no relatives that are closer than any other, and hence no phylogenetic signal can exist. Therefore, under the null hypothesis and holding α at 0.05, 5% of the simulated datasets should appear to exhibit significant phylogenetic signal when analyzed on any given hierarchical phylogenetic tree. Thus, for each of the simulated datasets, we used PHYSIGER.M to conduct the randomization test as described above (but using only 100 permutations for each of the 1000 simulated datasets), using three different hierarchical trees: a symmetrical tree with contemporaneous tip heights, a ladder-shaped (pectinate) phylogeny with contemporaneous tip heights, and a ladder-shaped tree with all branch lengths set equal to one. In each case, we calculated how many of the 1000 simulated datasets appeared to show significant phylogenetic signal (i.e., 95 or more of the permuted datasets showed greater variance than for the data in their original positions on the tree), resulting in a one-tailed estimate for the Type I error rate. In future studies, it would be of interest to examine the effects of different evolutionary models on Type I error rates, as well as power to detect phylogenetic signal.

We expected the power (1 − Type II error rate) of the randomization test to depend on both tree size and shape. Therefore, we considered the same three tree shapes as described in the previous paragraph. For the symmetrical trees, we used sample sizes of eight, 16, 32, and 64 species. For the ladder-shaped trees, we used those sizes plus 12, 24, and 48. For each tree, we simulated 100 datasets under Brownian motion. For each of the 100 simulated datasets, we used PHYSIGER.M (with 100 permutations) to determine the proportion for which we rejected the null hypothesis (for $\alpha = 0.05$) of no phylogenetic signal. Given that the null hypothesis was indeed false in all cases (i.e., the simulations were produced on a hierarchical tree), this proportion is an estimate of the power of the test.

For very small trees, there are few distinct possible permutations of the data. For example, on a symmetrical tree with four tips, the data can only be permuted in, at most (assuming all values are unique), three ways that yield distinct values of MSE. Thus, it is simply not possible to get a statistically significant result (i.e., detect phylogenetic signal at $P < 0.05$) with a four-tipped tree. The number of possible distinct MSE categories depends on tree topology in addition to size and on the number of unique tip values. Therefore, when analyzing real data with small trees (e.g., $N < 7$), the user should check the histogram of MSE values for the permuted data that is provided by PHYSIG.M and make sure that it is possible to detect a signal at $P < 0.05$.

All else being equal, one would expect that character evolution along a strongly hierarchical tree would lead to a greater degree of phylogenetic signal than would be the case for evolution along a tree that was more starlike in shape (i.e., a rapid radiation involving multiple, nearly simultaneous speciation events, followed by a relatively long period of independent evolution leading to the tip species). To examine the effect of such variation in tree shape, we used a phylogeny of 49 Carnivora and ungulates as a starter tree (from fig. 1 of Garland et al. 1993). We then transformed that tree using the OU transformation (which can mimic stabilizing selection; see below), with parameter values of 1.005, 1.0, 0.95, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, and 0.2, the latter six values producing successively more starlike versions of the starter tree (Fig. 1 shows some examples). We then proceeded as before, simulating 100 datasets under the Brownian motion model for each of the seven trees, and conducting the randomization test on each dataset, with 100 permutations for each test.

Type I error rates for the randomization test were close or equal to the nominal value of 0.05 (0.048 for the symmetrical tree, 0.050 for the ladder-shaped tree, 0.051 for the ladder-shaped tree with all branch lengths equal to one). Power increases dramatically with sample size, with a value of 0.8 reached at approximately 17–20 species (Fig. 2). Power also varies somewhat with tree topology and/or branch lengths: both symmetrical trees and ladders with equal branch lengths tend to show slightly higher power at all sample sizes. Finally, as expected, transformation of the 49-species tree to be less hierarchical (Fig. 1) reduced power to detect phylogenetic signal (Fig. 3).

Empirical Examples of Testing for Phylogenetic Signal

For this initial application of our proposed randomization test for phylogenetic signal (and methods of branch-length transformation as discussed below), we did not do an exhaustive survey of published phylogenetically based comparative studies, but we did endeavor to include a wide range of organisms and trait types. We favored studies that have used PDAP (Garland et al. 1993, 1999; Garland and Ives 2000) simply because we were often able to obtain the original input files from the original authors, hence expediting our analyses and reducing the possibility of introducing errors when re-entering data. In some studies that presented data on multiple traits, some of the traits were redundant (e.g., multiple measures of body size), so we sometimes analyzed only a subset of the traits presented. For simplicity, we used the topology and branch lengths as reported in the original paper and did not try to either update phylogenetic information or transform branch lengths (see below).

In total, we analyzed 119 traits that were associated with 34 different phylogenies; sample size ranged from six to 254 (Appendix 6). Datasets included organisms that ranged widely in phylogenetic position and scope, including plants, vertebrates, mammals, birds, lizards, salamanders, fish, and *Drosophila*. For traits that varied closely with body size, we computed size-corrected values as follows. First, we log-transformed the trait as well as the measure of size (mass or snout-vent length). Second, we computed standardized phylogenetically independent contrasts for both traits. Third, we computed a least-squares linear regression through the origin for the contrasts and noted the slope, *b* (allometric exponent).

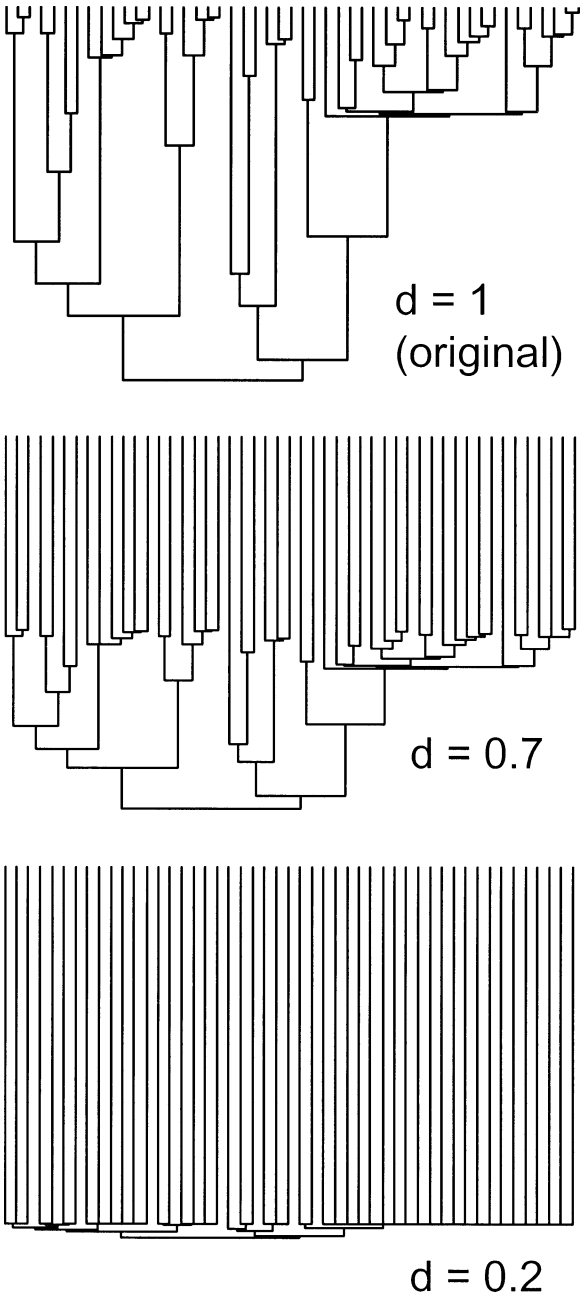

FIG. 1. Examples of the phylogenies used for simulations to illustrate the effect on statistical power of relative starness. Top is phylogeny with branch lengths proportional to estimated divergence times for 49 species of Carnivora and ungulates (fig. 1 of Garland et al. 1993). Middle is the same tree but transformed with a value of $d = 0.7$, where d is the restraining parameter in the Ornstein-Uhlenbeck transformation (see text and Appendix 2). Bottom is tree with $d = 0.2$. Simulations were conducted on trees with $d = 1.005$ (making it more hierarchical), 1.0 (original tree), 0.95, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, and 0.2 (see Fig. 3 for results).

Finally, we computed size-corrected values for the original trait (not contrasts) as: log[trait/(size^b)].

As expected from the power simulations, whether a trait showed significant ($P < 0.05$) phylogenetic signal was strongly related to the number of species included in the study: 92% (49 of 53) of the traits with sample sizes greater

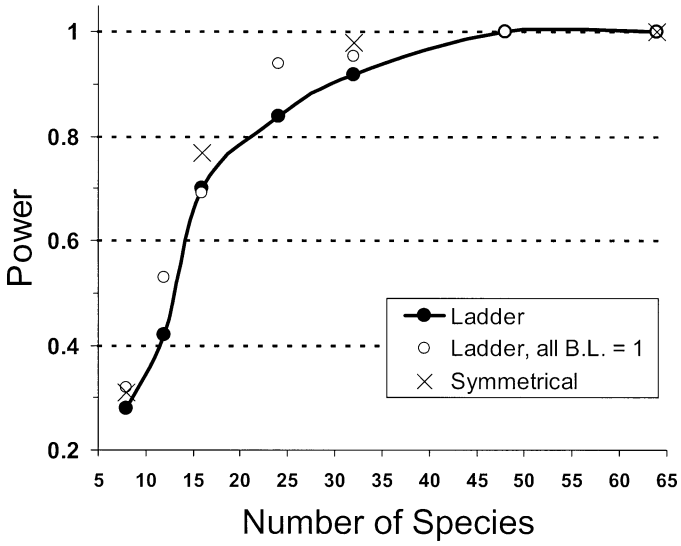

FIG. 2. Relationship between statistical power of the permutation test for detecting phylogenetic signal and sample size (number of species in the tree) as well as tree shape. Irrespective of tree shape, power is good (≥0.8) when sample size is greater than about 20 species. As a heuristic, a smoothed line is fitted to the data for the ladder-shaped tree.

than 20 showed signal, whereas signal was detected for only 41% (27 of 66) of the traits with N less than 20 (Appendix 6). For $N > 20$, the four traits that did not show significant phylogenetic signal were: log (male/female) body mass of 101 genera of bats ($P = 0.089$; Nee's arbitrary branch lengths, Hutcheon 2001), log mass-corrected maximal metabolic rate of 47 birds ($P = 0.260$; DNA hybridization branch lengths, Rezende et al. 2002), rhytidome allometric coefficient of 32 *Pinus* spp. ($P = 0.090$; all = 1, Jackson et al. 1999), and log

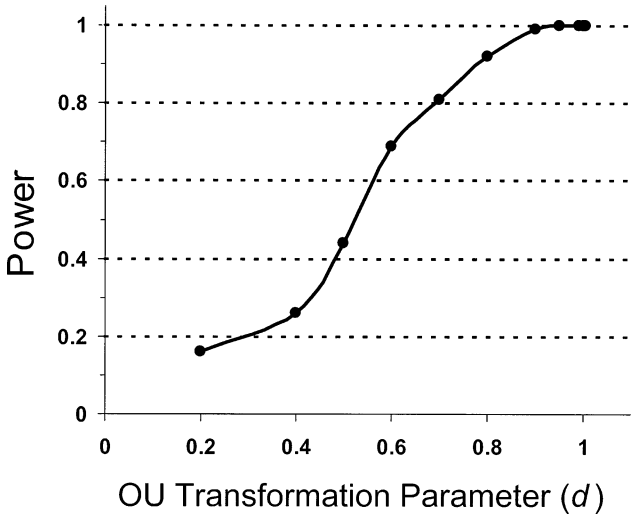

FIG. 3. Power of the permutation test in relation to tree starness. d is the restraining parameter in the Ornstein-Uhlenbeck transformation; $d = 1.005$ makes the tree more hierarchical, $d = 1$ represents the original tree, as shown in the top of Figure 1, and successively smaller values of d represent increasing starness. (A d-value of zero would yield a star phylogeny, i.e., a single polytomy with equal branch lengths.) As a heuristic, a smoothed line is fitted to the data.

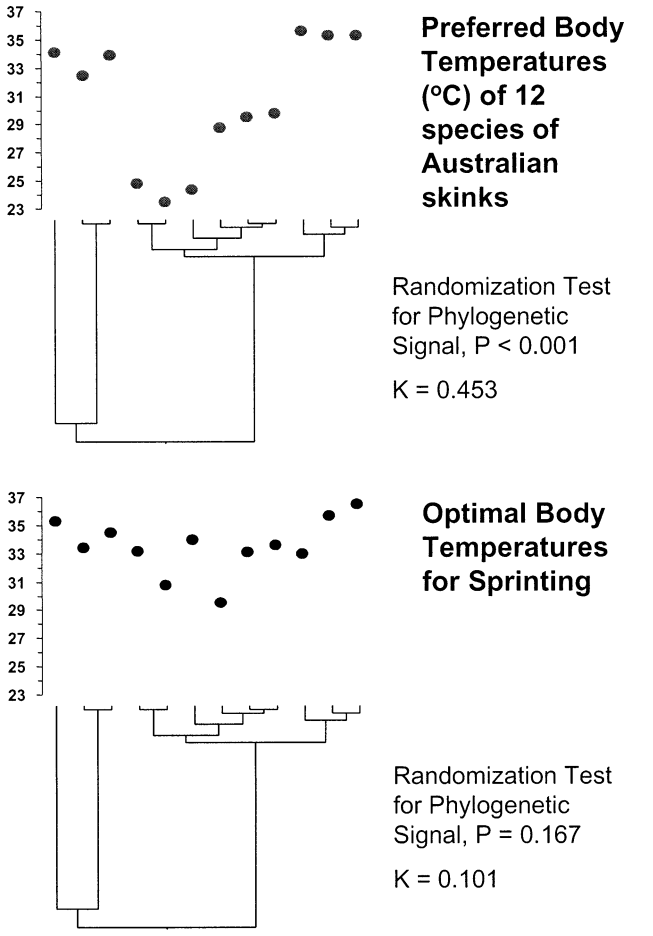

FIG. 4. Example in which significant phylogenetic signal is detected for one trait but not for another. (A) Preferred body temperature of some Australian skinks ($P < 0.001$, $K = 0.453$). (B) Optimal body temperature for sprinting ($P = 0.167$, $K = 0.101$). Data and tree are from Huey and Bennett (1987) and Garland et al. (1991), respectively.

of mass-corrected group size of 28 macropod marsupials ($P = 0.161$; all = 1, Fisher and Owens 2000). Importantly, irrespective of sample size, we found no case in which the permuted data showed significantly less variance ($P > 0.95$) than the real data, that is, no case in which the data showed what might be termed ''antisignal.''

Figure 4 shows an example in which phylogenetic signal is detected for one trait but not another within a given study. Data on body temperatures of Australian skinks are plotted above their corresponding phylogeny (as used in a published comparative analysis by Garland et al. 1991). For preferred body temperature, close relatives tend to be similar: phylogenetic signal is detected ($P < 0.001$). For the optimal temperature for sprinting, however, relatives are not significantly more similar than if placed randomly on the tree ($P = 0.167$).

Quantifying Phylogenetic Signal

The randomization test described above provides a method whereby the degree of resemblance among relatives can be distinguished from random. Here, we derive a descriptive statistic that indicates the strength of phylogenetic signal, as compared with an analytical expectation based only on the tree structure (topology and branch lengths) and assuming Brownian motion character evolution.

A statistic to quantify phylogenetic signal begins with the ratio of the mean squared error of the tip data, measured from the phylogenetically correct mean (MSE_0), divided by the mean squared error of the data calculated using the variance-covariance matrix derived from the candidate tree (MSE). If the candidate tree accurately describes the variance-covariance pattern observed in the data, then the value of MSE will be relatively small, leading to a large value of MSE_0/MSE. Conversely, if there is little covariance within the tip data that is explained by the candidate tree, then MSE_0 will be relatively small, leading to a smaller value of MSE_0/MSE. Thus, high values of MSE_0/MSE imply more phylogenetic signal. MSE_0 is calculated simply as:

$$\mathrm{MSE}_0 = \frac{(\mathbf{X} - \hat{a})'(\mathbf{X} - \hat{a})}{n - 1}, \tag{1}$$

where â is the estimate of the phylogenetically correct mean (equal to the estimated trait value at the root node of the tree; e.g., see Garland et al. 1999) and **X** is the data vector containing n values. MSE is calculated as:

$$\mathrm{MSE} = \frac{(\mathbf{U} - \hat{a})'(\mathbf{U} - \hat{a})}{n - 1}, \tag{2}$$

where $\mathbf{U} = \mathbf{DX}$ is the transformed **X** vector obtained from the generalized least-squares procedure. The matrix **D** satisfies the equation: $\mathbf{DVD}' = \mathbf{I}$, where **V** is the variance-covariance matrix and **I** is the identity matrix (Garland et al. 1999).

The ratio MSE_0/MSE is calculated from the data, and although relatively large values imply more phylogenetic signal, values of MSE_0/MSE from different phylogenetic trees are not directly comparable. For such comparisons, however, it is possible to scale the value of MSE_0/MSE by that value that is predicted under the assumption of Brownian motion evolution along the specified tree. Specifically, the expected ratio under Brownian motion is:

$$\frac{\mathrm{MSE}_0}{\mathrm{MSE}} = \left(\frac{1}{n - 1}\right)\left(tr\mathbf{V} - \frac{n}{\Sigma\Sigma\mathbf{V}^{-1}}\right), \tag{3}$$

where $tr\mathbf{V}$ is the trace (sum of diagonal elements) of **V**, and $\Sigma\Sigma\mathbf{V}^{-1}$ is the sum of all elements of the inverse matrix of **V**. Note that it is not possible to extract separately the numerator and denominator of this expectation. The expected MSE_0/MSE is a useful descriptor of how hierarchical a given tree is, with larger values corresponding to greater degrees of hierarchy.

Figure 5 shows that, for the 35 trees represented in Appendix 6, the expected MSE_0/MSE varies widely and also shows a positive dependence on tree size. For several trees, the expected MSE_0/MSE is rather far above or below the general trend, and consideration of their branch lengths makes it apparent why. For example, the largest tree in our sample, with 254 species, is for birds (Reynolds and Lee 1996) and is based primarily on the DNA hybridization phylogeny of Sibley and Ahlquist (1990). Many bifurcations occur near the base (root), leading to many long branches (this

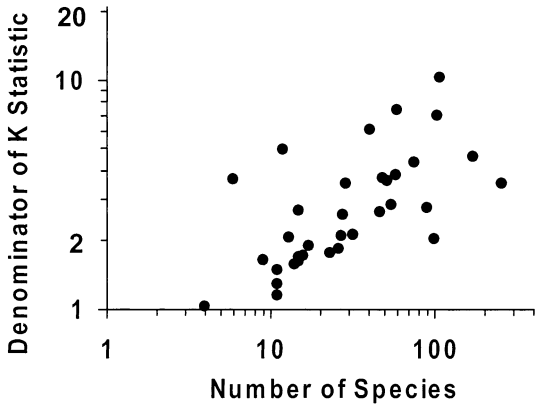

FIG. 5. Effect of tree size on expected amount of phylogenetic signal (expected MSE_0/MSE) for 35 phylogenetic trees used in published comparative studies. Note the strong positive effect of tree size and that some values deviate from the general trend because of unusual branch lengths (see text for discussion).

tree is shown in fig. 5A of Garland and Ives 2000) and hence relatively less expected phylogenetic resemblance ($MSE_0/MSE = 3.54$) than would be the case for a tree of equal size but heights of nodes more uniformly distributed between the root and the top of the tree. The same tree with Pagel's (1992) arbitrary branch lengths yields a value of 6.37, with Grafen's (1989) arbitrary branch lengths the value is 30.69, and with all branch lengths set equal to one the value is 6.29. Also consider two trees above the general trend. The tree with $N = 12$ (estimated divergence times, as shown in Fig. 4) has few long, internal branches and many short branches at or near the top of the tree, yielding an expected MSE_0/MSE of 4.95. With Grafen's arbitrary branch lengths the value is 2.00, with Pagel's arbitrary the value is 1.47, and for all branch lengths set equal to one the value is 1.50. Perry and Garland (2002; $N = 108$) used Grafen's arbitrary branch lengths, which tend to result in long branches at the base of the tree, yielding an expected MSE_0/MSE of 10.20. Expected MSE_0/MSE using Pagel's (1992) branch lengths is 4.60, more in line with the general trend (with all branch lengths equal to one, $MSE_0/MSE = 3.54$).

Given that the expected value of MSE_0/MSE depends on tree size and shape, it is natural to scale the observed MSE_0/MSE ratio by the expected ratio, thus allowing comparisons of traits regardless of tree characteristics. Therefore, we used the following statistic,

$$K = \text{observed}\frac{\text{MSE}_0}{\text{MSE}} \Big/ \text{expected}\frac{\text{MSE}_0}{\text{MSE}}. \quad (4)$$

A K less than one implies that relatives resemble each other less than expected under Brownian motion evolution along the candidate tree. This could be caused by either departure from Brownian motion evolution, such as adaptive evolution that is uncorrelated with the phylogeny (i.e., homoplasy), or ''measurement error'' in the broad sense (see Discussion). A K greater than one implies that close relatives are more similar than expected under Brownian motion evolution.

An alternative to K involves the ratio MSE*/MSE, where MSE* is calculated from the observed data on a star phylogeny with contemporaneous tips:

$$\text{MSE}^* = \frac{(\mathbf{X} - \bar{a})'(\mathbf{X} - \bar{a})}{n - 1}, \quad (5)$$

where $\bar{a}$ is the estimate of the mean of the raw data $\mathbf{X}$, rather than the phylogenetically correct mean $\hat{a}$ as used in the calculation of MSE_0. The ratio MSE*/MSE has the expected value under the assumption of Brownian motion of:

$$\frac{\text{MSE}^*}{\text{MSE}} = \left(\frac{1}{n - 1}\right)\left(tr\mathbf{V} - \frac{\Sigma\Sigma\mathbf{V}}{n}\right). \quad (6)$$

Dividing observed by expected values yields a statistic that we term K^*. Here, we only present results for K, which we feel has greater heuristic value. However, K and K^* are highly correlated, so results using the latter would be similar.

EMPIRICAL EXAMPLES OF QUANTIFYING PHYLOGENETIC SIGNAL

We considered the same set of traits and trees as above, plus one study involving only four species of birds (which is too few species for the randomization test). Hence, the total number of traits was 121, from 35 trees (Appendix 6). Considering all of these datasets, K ranged from 0.084 for the duration of head bobs in *Cyclura* lizards (a behavioral trait; Martins and Lamont 1998) to 4.02 for body mass of female macropod marsupials (Fisher and Owens 2000). As above in the analysis using the randomization test for signal, we used the topology and branch lengths as reported in the original paper and did not transform branch lengths. As shown in Figure 6, K does not vary in relation to tree size, but K is less than one for most traits, indicating a general tendency for less signal than expected under Brownian motion along the specified tree.

To compare the amount of phylogenetic signal across traits of different types, we categorized them as adult body size, life history, morphological, physiological, behavioral, or ecological. Categorizing traits is not always straightforward (see also Gittleman et al. 1996a; Stirling et al. 2002), and we view the following analyses primarily as a heuristic. We would encourage others to perform similar analyses with larger and more comprehensive datasets and alternative categorization schemes. Adult body size ($N = 24$) was treated as a separate category because it has sometimes been viewed as a morphological trait, yet is often highly correlated with various life-history and physiological traits (Calder 1984; Roff 1992; Gittleman et al. 1996a), and is the end result of the ontogenetic growth trajectory, which would usually be considered a life-history trait. Our morphological category ($N = 35$) included such traits as size-corrected limb length, bill length, brain size, testes mass, and sperm length, as well as individual leaf area and petiole length. Life-history traits ($N = 20$) included age at maturity, length of nestling period, adult mortality, clutch size, log male/female body size (an index of sexual dimorphism), seedling height, and seed size and viability. Physiological traits ($N = 21$) included metabolic rates, maximal sprint speed, stride frequency, critical thermal minimum and maximum, and enzyme activities. Behavioral traits ($N = 17$) included daily movement distance, prey size, display characteristics, and preferred body temperature. Only four traits were in our ecological category: seasonality of

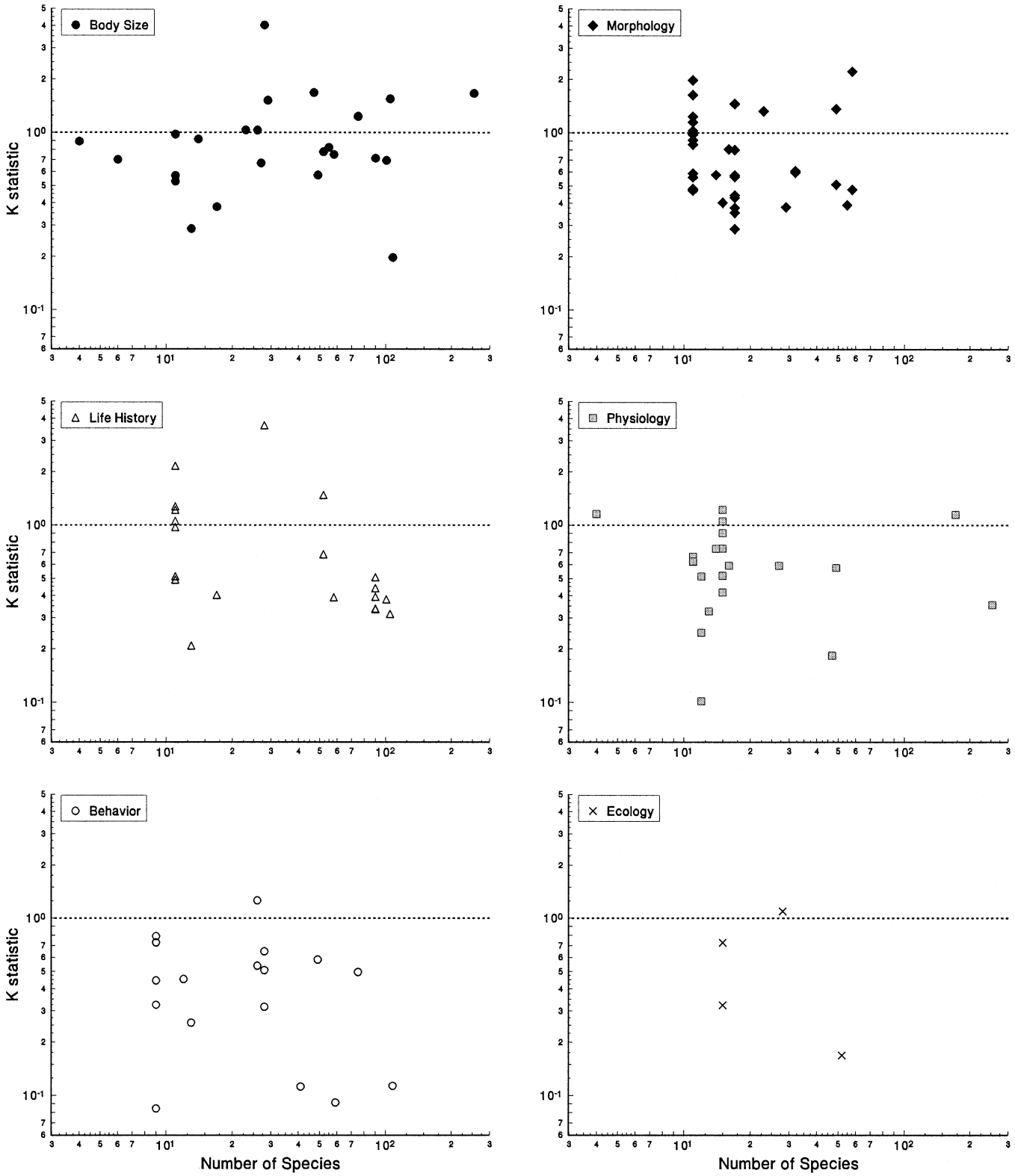

FIG. 6. Empirical relationship between tree size (number of species) and *K*, a statistic that indicates the amount of phylogenetic signal observed in a set of comparative data divided by the amount expected under Brownian motion character evolution along the specified tree topology and branch lengths. Most traits show less signal than expected, that is, *K* less than one. Analysis of covariance indicates that log *K* does not vary systematically with log tree size but does differ significantly among trait types (see text).

environments of 52 Carnivora (coefficient of variation of actual evapotranspiration averaged across year and across geographic range; $K = 0.17$, Ferguson and Lariviere 2002), rainfall in habitats of 28 macropods (corrected for correlation with body mass; $K = 1.10$, Fisher and Owens 2000), mean annual temperature in the waters of 15 *Fundulus* fish ($K = 0.32$, Pierce and Crawford 1997), latitude of ranges of 15 *Drosophila* ($K = 0.73$, Gibert and Huey 2001). (The datasets analyzed by Freckleton et al. [2002] include a large number of ecological traits.)

After exclusion of the four traits in the ecological category owing to small sample size, data were analyzed by a conventional one-way ANOVA, following log transformation to homogenize variances. The datasets included clearly do not meet the assumption of independence—some of them include the same taxa and even some of the same species—but we could not envision any simple way to deal with the situation. In any case, as noted above, we present the analyses primarily as a heuristic.

Trait categories differed significantly in mean $\log_{10}K$-value ($F = 4.97$, df = 4, 112, $P = 0.0010$). Levene's test indicated no significant difference in variance among categories (two-tailed $P = 0.266$). Based on Duncan's multiple range comparison ($P = 0.05$, and using harmonic mean cell sample size of 22.0575), behavioral traits showed significantly lower $\log_{10} K$ than did the other four types of traits. In addition, physiological traits showed lower $\log_{10} K$ than did body size. Mean values (95% CI) for $\log_{10} K$ were: body size −0.08 (−0.20, 0.03); morphology −0.15 (−0.23, −0.07); life history −0.20 (−0.35, −0.05); physiology −0.27 (−0.39, −0.14); and behavior −0.45 (−0.63, −0.27). To verify that $\log_{10} K$ was uncorrelated with tree size, we repeated the analysis as an ANCOVA and found that $\log_{10}$ tree size was not a significant covariate ($F = 1.88$, df = 1, 111, $P = 0.173$), whereas trait type remained highly significant ($F = 5.22$, df = 4, 111, $P = 0.001$).

We were concerned that the finding of lower *K*-values for behavioral traits could be misleading if most of the behavioral traits were from a small number of studies. That is not the case, however, as 10 studies included behavioral traits (see Appendix 6, Fig. 6). Five of the 17 behavioral traits were from a single study of nine lizards (Martins and Lamont 1998). One of these traits exhibited the lowest *K* of any trait that we examined ($K = 0.08$), but the other four traits exhibited *K*-values that were not unusual (range = 0.32–0.79; Appendix 6, Fig. 6). Therefore, we do not think that our results are biased by particular studies. Nevertheless, another way to compare traits is to consider only studies that include multiple types of traits. Given that the ANOVA indicated behavioral traits to have lower levels of phylogenetic signal as compared with other traits, it was of interest to examine those studies that included both behavioral and other traits. Unfortunately, this amounted to only seven of the 35 studies listed in Appendix 6. Averaging values within categories for the seven studies yielded: 108 lizards, behavior = 0.11, body size = 0.20; 75 antelope, behavior = 0.50, body size = 1.23; 49 Carnivora + ungulates, behavior = 0.58, body size = 0.57; 28 macropod marsupials, behavior = 0.49, mean of two other traits = 3.84 (one ecological trait excluded, as in overall ANOVA described above); 26 hystricognath rodents, mean of two behavior traits = 0.90, body size = 1.03; 27 swallows, behavior = 0.26, mean of three other traits = 0.27; and 12 skinks, behavior = 0.45, mean of three physiological traits = 0.29. Thus, for this subsample of studies, behavioral traits still tend to exhibit lower *K*, but a paired *t*-test of the logarithms of these seven values indicates no statistical difference ($t = 1.488$, two-tailed $P = 0.187$). We suspect that this lack of significance reflects mainly a lack of power caused by small sample size.

Probability of Detecting Signal Depends on Both N and K

Whether phylogenetic signal, as judged by the randomization test, is detected for a given trait would be predicted to depend on both sample size (number of species) and the amount of signal for that trait. Therefore, as a heuristic, we performed a multiple regression of log (*P*-value + 1) on log *N* and log *K*. The overall model was highly significant (multiple $r^2 = 0.40$, $F = 38.0$, df = 2,116, $P < 0.0001$) and both independent variables were significant negative predictors (log *N*: $F = 44.6$, $P < 0.0001$; log *K*: $F = 36.9$, $P < 0.0001$; the correlation between log *N* and log *K* was −0.074). (The log-transformed *P*-values exhibit a nonnormal and truncated distribution, because only 1000 permutations were used to generate them, but nonetheless residuals did not deviate excessively from normality.)

Statistical and Biological Approaches to Branch-Length Transformations

As noted in the introduction, branch lengths are an integral part of comparative analyses, including the proposed randomization test for phylogenetic signal and the descriptive *K*-statistic. Moreover, certain branch-length transformations can allow an alternative test of signal by comparing fits to data on a series that includes a star as one end of the continuum (e.g., Grafen 1989; next section). The choice of branch lengths for phylogenetically based statistical analysis involves several issues, including: (1) the type of starter branch lengths employed (e.g., whether they are arbitrary or based on data); (2) types of transformations to be considered; and (3) the inferences that may or may not be appropriately drawn from analyses that employ a given type of branch length and/or transformation. Ideally, branch lengths may be derived from data, such as genetic distances or estimates of divergence times derived from the fossil record. Many comparative analyses, however, have employed arbitrary branch lengths, such as setting all segments equal to one or aligning tip species at the top of the tree and setting the depths of internal nodes by an arbitrary algorithm (Grafen 1989; Pagel 1992; S. Nee cited in Purvis 1995). Typically, this is done because real branch lengths are unavailable, but some workers actually seem to prefer arbitrary branch lengths (especially all equal to one) as somehow making fewer assumptions about the data or analysis. Whatever the reason, approximately one-third of the studies that we have analyzed employed arbitrary branch lengths (13 of 35 trees, 55 of 121 traits).

However starter branch lengths are obtained, they should be checked for statistical adequacy based on one or more

diagnostics (Grafen 1989; Garland et al. 1991, 1992; Purvis and Rambaut 1995; Díaz-Uriarte and Garland 1996, 1998; Reynolds and Lee 1996; Garland and Díaz-Uriarte 1999; Harvey and Rambaut 2000). If they fail these diagnostics, then transformations of the branch lengths (or of the tip data) may be employed as a remedial measure. Most commonly, the type of branch-length transformation used has been chosen arbitrarily and from a purely statistical perspective. Examples include log transformation, raising all branches to a power, and setting all branches equal to one (equivalent to raising branch lengths to a power of zero). In addition, branch lengths may be transformed proportional to the estimated relative amount of evolution (character variance) that has occurred above and below them on the tree, using Grafen's (1989) ρ (see also Pagel 1994, 1999). Branch lengths may also be transformed by extending only the terminal branches, which may be appropriate in the presence of measurement error in the tip data (T. Garland and A. R. Ives, unpubl. data). Finally, all branch segments between internal nodes may be collapsed to zero and branches leading to tip species (terminal taxa) set to be equal in length, thus enforcing a star topology on the tree. This can be viewed as one end of the continuum of possible transformations, one which is appropriate under the assumption that all traits evolved independently from a common ancestor, as is implicit when conventional statistical methods are applied to comparative data (e.g., see Grafen 1989; Purvis and Garland 1993; Garland et al. 1999). Note also that if the tips of the beginning phylogenetic tree are contemporaneous, then using Grafen's (1989) transformation with $\rho = 0$ also results in a star phylogeny.

Alternatively, if branch lengths are viewed as entities that may contain important biological information in their own right, then transformations of branch lengths according to some explicit model of evolution and estimation of the parameters in the model may provide useful information about underlying evolutionary processes (Martins 1994; Mooers et al. 1999; Martins et al. 2002). (Of course, this argument has greatest weight if the starter branch lengths are real rather than arbitrary.) To date, only one biologically motivated transformation has been suggested, the Ornstein-Uhlenbeck (OU) process, which has been proposed as a model of stabilizing selection (Felsenstein 1988; Garland et al. 1993; Hansen and Martins 1996; Martins et al. 2002). Testing for significance of the parameter associated with the OU model can thus be viewed as a test for stabilizing selection (Martins 1994). Very strong stabilizing selection can obliterate the effects of history such that phylogenetic signal disappears; accordingly, one end of the continuum of an OU transformation results in a star phylogeny. Similarly, as we show here, a transformation can be developed that models the acceleration or deceleration of Brownian motion evolution from the root to the tips of the tree (hereafter ACDC). Rapidly accelerating rates of evolution can also erase the effects of descent with modification, and again one end of the ACDC transformation results in a phylogeny that is virtually a star (very little hierarchical structure).

Even if branch lengths and transformations of them are arbitrary, biological inferences can sometimes still be made, provided traits are compared with respect to the same set(s) of branch lengths. For example, if multiple traits are considered, then trait-specific differences in the rate of evolution between two clades may be apparent (e.g., Clobert et al. 1998), although it may not be clear which clade is evolving relatively faster. Along these lines, Pagel (1999) and Freckleton et al. (2002) employ a transformation parameter (λ) that is not based on an explicit model of evolution, but they argue that various biological interpretations may nonetheless be possible when the estimate differs significantly from unity. It is also possible to estimate the length of each branch segment from the tip data themselves (Garland et al. 1992, 1999), and Mooers et al. (1999) used an ML approach that allowed them to draw biological inferences in comparisons of different types of traits.

Whatever type of starter branch lengths are used and whatever types of transformations one is willing to consider, the decision as to whether a transformation should actually be applied is not always clear-cut. The problem can be described as a continuum between two philosophies. At one end is the perspective that transformations should always be applied to satisfy assumptions of the ensuing statistical test as closely as possible. At the other end of the continuum is the idea that transformations should only be performed when absolutely necessary, for example, when diagnostics show statistically significant problems; minor departures from the assumptions of ensuing statistical tests can be ignored because most are robust to at least some departure from their assumptions (see discussion in Díaz-Uriarte and Garland 1996).

Transformation of Branch Lengths under Evolutionary Models

The OU model describes the motion of a species in the phenotypic space whereby the species moves randomly within the space, but is influenced by a central tendency such that large deviations from the central optimum receive a stronger force back toward the optimum (Felsenstein 1988; Garland et al. 1993; Martins 1994; Hansen and Martins 1996; Butler et al. 2000; Martins et al. 2002). We implement a simple version of the OU process, in which the covariance relationship among the characters can be represented by the formulae:

$$V\{X_i\} = \frac{1 - d^{2(\tau_{ij}+\tau_i)}}{1 - d^2}\sigma^2_\gamma \quad \text{and} \tag{7a}$$

$$\text{cov}\{X_i, X_j\} = d^{\tau_i+\tau_j}\frac{1 - d^{2\tau_{ij}}}{1 - d^2}\sigma^2_\gamma, \tag{7b}$$

where σ^2_γ denotes the rate of evolutionary divergence through time, and d governs the strength of the central tendency of the OU process, with $d = 1$ corresponding to Brownian motion evolution and lower values of d giving stronger stabilizing selection to a central mean for all species. The node-to-tip branch length for species i is denoted τ_i, and τ_{ij} denotes the shared branch length between tips i and j. The resulting trait values from this OU process follow a multivariate normal distribution with covariance matrix $\mathbf{V}$ given by equation (7a). The derivation of these expressions given in Appendix 2 is comparable to that presented in the appendix of Hansen (1997) and to simulations under PDSIMUL (Garland et al. 1993; see PDINSTRW.DOC).

The ACDC model is new. It describes evolution that either

increases (accelerates, AC) or decreases (decelerates, DC) in rate over time. The covariance relationship among the characters can be represented by the formulae:

$$V\{X_i\} = \frac{1 - g^{-(\tau_i + \tau_{ij})}}{1 - g^{-1}}\sigma_\gamma^2 \quad \text{and} \tag{8a}$$

$$\text{cov}\{X_i, X_j\} = \frac{1 - g^{-\tau_{ij}}}{1 - g^{-1}}\sigma_\gamma^2, \tag{8b}$$

where g is the overall rate of acceleration (g less than one) or deceleration (g greater than one), and other variables are as for the OU model (Appendix 3). The resulting distribution of trait values is multivariate normal with covariance matrix **V** given by by equation (8a).

The OU and ACDC parameters, d and g, can be estimated using ML in GLS mode (PHYSIGOU.M, PHYSIGACDC.M), and Appendix 4 demonstrates that the ML estimates are also the estimated generalized least-squares estimates. These parameters can also be estimated using independent contrasts, as described in Appendix 5.

Single-parameter transformations, such as OU and ACDC, can provide information about the degree of phylogenetic structure in the data, by direct inspection of the estimates of the OU and/or ACDC parameters. For example, the OU transformation has one parameter (d). When $d^* = 1$, the original candidate tree best fits the data (the asterisk refers to the ML estimate of the parameter). When d^* is between zero and one, the data show less structure than would be expected from the original candidate topology and branch lengths (e.g., see Fig. 1). For d^* greater than one, the data show more structure than expected, given the original tree (as is also true for the λ of Pagel 1999; Freckleton et al. 2002). When $d^* = 0$, the best-fitting tree is estimated to be a star phylogeny. The ACDC transformation also has one parameter (g), and interpretation is similar. When g^* is less than one, the transformed tree is more starlike than the original candidate. Unlike d, g can never equal zero; however, very small values of g (for estimation in MatLab, the smallest possible value of g is $2.225073858507201 \times 10^{-308}$) can yield matrices that are virtually the identity matrix (e.g., mathematically indistinguishable in terms of MSE), and limits on numerical precision in some programs may cause rounding or truncation that leads to an identity matrix. A g^* greater than one yields a tree that is more hierarchical than the original candidate tree.

A test of whether d is significantly different from zero provides information that is somewhat similar to our proposed randomization test for phylogenetic signal (see above), but different in an important way. If the best estimate of d for a given dataset is zero, then a star phylogeny better fits the data than does the original candidate tree or any OU transformations of it. The randomization test attempts to detect resemblance among relatives, but does not actually compare the candidate tree with a star. Nevertheless, results of the two procedures would be expected to be correlated (at least for trees with $N \geq 20$): if no phylogenetic signal is detected, then we would probably estimate an optimal d that is close to zero. The same would be true for g, the ACDC transformation parameter, except g can never equal exactly zero.

To examine the effect of tree size on the estimation of d and g, we simulated data (PDSIMUL) under Brownian motion along three sets of trees with four, eight, 16, and 32 tips: symmetrical with contemporaneous tip heights, ladder-shaped (pectinate) with contemporaneous tip heights, and ladder-shaped with all branch lengths set equal to one. We did 100 simulations for each of the 12 trees. We then calculated d and g for each simulated tree (using our program PHYSIGDG.M) and plotted the frequency distribution for each statistic at each sample size. As shown in Figure 7, sampling distributions for d and g were not normal and contain many values that are at (d) or near (g) zero, particularly at small sample sizes. Thus, d and g cannot be estimated reliably for trees of fewer than about 20 tips.

The highly nonnormal distributions for estimates of d and g are problematic for construction of traditional confidence intervals. Therefore, we devised a randomization procedure to determine whether the given set of real data could have produced a d- or g-value as much greater than zero as that observed, if there were actually no phylogenetic signal in the data (PHYSIGOU.M, PHYSIGACDC.M). First, we calculated the observed d (or g) for the original tree and data. Second, data were permuted equiprobably across the tips of the tree (1000 times), and then the ML estimate of the parameter d (or g) was calculated for each permuted dataset, along with the corresponding MSE. The number of cases for which MSE of the permuted data was greater than the value for the original data in their correct positions was tallied. If many cases (e.g., 95%) were found for which the MSEs for permuted data were greater than the MSE for the real data, then the hypothesis that d is zero could be rejected. Interpretation is similar for estimation of g (the ACDC parameter), except that, as noted above, g can never actually take on a value of zero.

Freckleton et al. (2002) recently presented an approach to testing for phylogenetic signal that is similar to that just described. Their analysis is based on Pagel's (1999) λ, which is a parameter multiplied to each of the off-diagonal elements of the variance-covariance matrix **V**. Although λ is not associated with an explicit model of evolution, it is nonetheless a straightforward way to decrement the candidate phylogenetic structure of the data by uniformly decrementing all covariances. Thus, when $\lambda = 0$, the tree has a single polytomy at the basal node for all species, whereas when $\lambda = 1$ the original candidate tree is recovered. Statistical tests for phylogenetic signal are performed under the null hypothesis that $\lambda = 0$, and tests for less signal than the candidate tree are performed under the null hypothesis that $\lambda = 1$. As we do for d and g of the OU and ACDC processes, Freckleton et al. (2002) estimate the best-fitting λ using ML. In contrast to our use of randomization tests for statistical inference (obtaining P-values), Freckleton et al. (2002) use a log-likelihood ratio test. We prefer randomization tests, because log-likelihood ratio tests are only asymptotically valid, and for small sample sizes (where, unfortunately, what is meant by small depends on the structure of the data, but is generally fewer than 30) log-likelihood ratio tests can give noticeably imprecise P-values.

We can also test the null hypothesis that d (or g) = 1 (i.e., that the original candidate tree adequately fits the data). For this, we devised another randomization procedure. For any

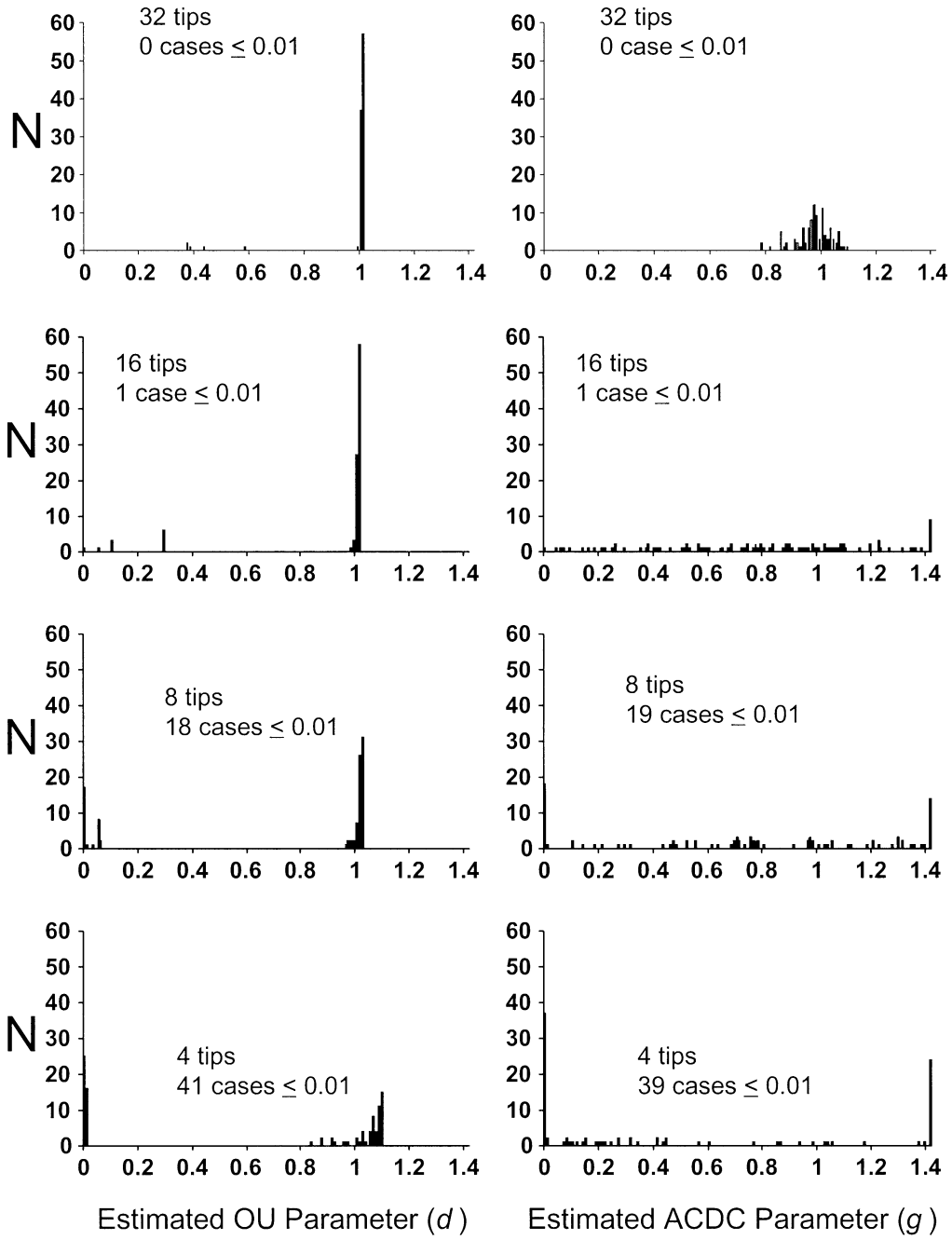

FIG. 7. Distribution of estimated Ornstein-Uhlenbeck (OU; *d*) or acceleration-deceleration (ACDC; *g*) branch-length transformation parameters in relation to tree size for a ladder-shaped (pectinate) tree with four, eight, 16, or 32 tips. Data were simulated under simple Brownian motion (PDSIMUL.EXE; see Garland et al. 1993). For all panels, $N = 100$, and values of *d* or *g* greater than 1.4 are placed into a single bin. Results were similar for ladder-shaped trees with all branch lengths set equal and for symmetrical trees with contemporaneous tips (results not shown). Estimation of *d* and *g* is unstable for smaller trees, so we recommend relying on these types of branch-length transformations only for trees with about 20 or more species.

value of *d*, the fit of the OU model can be judged by examining the MSE. To test the null hypothesis that $d = 1$, one can calculate the difference between the MSE when $d = 1$, denoted $\text{MSE}_{d=1}$, and the MSE under the ML estimate of *d*, denoted $\text{MSE}_{d=d\text{ML}}$. The difference $\text{MSE}_{d=1} - \text{MSE}_{d=d\text{ML}}$ measures the decrease in MSE (and hence the degree of fit) of the model when *d* equals the ML value compared to when $d = 1$. Statistical confidence in distinguishing the ML estimate of *d* from $d = 1$ requires estimating the probability of obtaining the observed difference $\text{MSE}_{d=1} - \text{MSE}_{d=d\text{ML}}$ under the null hypothesis that $d = 1$. If the observed value of the difference $\text{MSE}_{d=1} - \text{MSE}_{d=d\text{ML}}$ falls outside the 95% interval of values of $\text{MSE}_{d=1} - \text{MSE}_{d=d\text{ML}}$ obtained from permutation of the data under the assumption that $d = 1$, then *d* differs from one with 95% confidence. To create a permutation dataset of values of $\text{MSE}_{d=1} - \text{MSE}_{d=d\text{ML}}$ under

the assumption that $d = 1$, we first fit the model with $d = 1$. In GLS mode, this fitting produces residuals that, under the assumption that $d = 1$, are independent in the data space defined by transforming the data according to the covariance matrix **V** of the phylogenetic tree; these residuals are denoted $\boldsymbol{\alpha}$ in Garland and Ives (2000). We then permuted these residuals equiprobably and back-transformed the permuted values to create a new permutation dataset (Efron and Tibshirani 1993). The permutation dataset was analyzed in the same way as the true dataset: the value of the ML estimate of d was calculated, and the fit of the resulting model was compared to the case of $d = 1$ using the difference $\text{MSE}_{d=1} - \text{MSE}_{d=d\text{ML}}$. By repeating this procedure many times (e.g., 1000 permutation datasets), we obtained a collection of equiprobable values of $\text{MSE}_{d=1} - \text{MSE}_{d=d\text{ML}}$ under the assumption that $d = 1$. If the observed value of $\text{MSE}_{d=1} - \text{MSE}_{d=d\text{ML}}$ for the real data falls outside the 95% confidence interval of the permutation distribution, then we concluded that $d \neq 1$. Note that because $\text{MSE}_{d=1} - \text{MSE}_{d=d\text{ML}}$ will always be positive, the 95% confidence interval is one-tailed. The test for whether $g = 1$ under the ACDC transformation is conducted in an identical way. These calculations are performed in PHYOUH0d.m and PHYACDCH0g.m. Finally, note that this procedure can be used to test the statistical significance of any hypothesized value of d, not just the hypothesis $d = 1$, by setting d to the tested value before constructing the permutation datasets.

Empirical Examples of Branch-Length Transformation

As noted above, ML estimation of the OU transformation parameter, d, for a given set of data on a given candidate tree is equivalent to testing for the best fitting tree along the continuum from a star (with contemporaneous tips; $d = 0$) to the original tree ($d = 1$) and on to a tree that is even more hierarchical than the original ($d > 1$). For the 20 trees with more than 20 tips (see Appendix 6), the null hypothesis that d was equal to zero was rejected ($P < 0.05$) for 89% of the traits considered (47 of 53). Thus, a star phylogeny usually could be rejected in favor of a hierarchical tree.

For 10 cases, the estimate of d was significantly greater than zero but less than one, indicating that the best-fitting tree is neither the original candidate nor a star, but rather something intermediate in terms of amount of hierarchical structure. For 19 of 53 traits on trees with more than 20 tips, the ML estimate of d was greater than one, and it was significantly greater than one in three of these: plasma osmolarity of 172 vertebrates (starting branch lengths all = 1, $d = 1.027$, $P < 0.001$; Garland et al. 1997); log body mass of 75 antelope (divergence time branch lengths, $d = 1.011$, $P = 0.001$; Brashares et al. 2000); log mass-corrected bill length of 58 birds (all = 1, $d = 1.025$, $P = 0.049$; Székely et al. 2000). Thus, occasionally a tree more hierarchical than the original better fit the data.

Similar to estimation of the OU transformation, ML estimates of the ACDC parameter, g, were significantly greater than zero for 49 of 53 cases (two of the cases that were not greater than zero were the same as for d, but two were different). Thus, again a star phylogeny usually could be rejected in favor of a hierarchical tree. (For the trees with all branch lengths = 1, estimation of g was sometimes problematic [e.g., irregular likelihood surface]; hence, results for such trees should be interpreted with caution.) For 16 cases, the estimate of g was significantly greater than zero but less than one. Figure 8 shows an example for which both d and g were significantly different from both zero and unity, and for which the ACDC-transformed tree yielded the best fit to the data (lowest MSE). For 19 of the 53 cases, g was greater than one, and five of these were significantly so: plasma osmolarity of 172 vertebrates (Garland et al. 1997); log female body mass, log mass-corrected wing length, log mass-corrected bill length of 58 birds (Székely et al. 2000); log mass-corrected tail length of 23 anguid lizards (DNA sequence branch lengths; Wiens and Slingluff 2001).

Agreement between Randomization and Branch-Length Transformation Approaches to Testing for Phylogenetic Signal

Over 90% of the traits with sample sizes greater than 20 (49 of 53) showed statistically significant ($P < 0.05$) phylogenetic signal based on our randomization test, and results of that test were in close accord with estimation of the optimal d parameter for the OU branch-length transformation. Specifically, in 46 of the 49 cases that showed significant signal, d was significantly greater than zero, implying that a star phylogeny did not adequately fit the data (Appendix 6). Only three cases showed a (small) discrepancy between the two test procedures. For the log(male/female) body mass of 101 genera of bats (Nee's arbitrary branch lengths; data and topology from Hutcheon 2001), signal was not detected ($P = 0.089$) but d was greater than zero ($P = 0.031$). For log(age at maturity) (corrected for snout-vent length) of 90 lizards (all branch lengths = 1; Clobert et al. 1998), signal was detected ($P = 0.001$) but d was not greater than zero ($P = 0.l61$). For log(female home range area) (corrected for body mass) of 28 macropod marsupials (all branch lengths = 1; Fisher and Owens 2000), signal was detected ($P = 0.033$) but d was not greater than zero ($P = 0.184$). Estimation of the optimal ACDC parameter, g, also agreed closely with the randomization test for phylogenetic signal, as again 46 of 49 cases showed g that were significantly greater than zero. Furthermore, as would be expected, nonsignificant results from the randomization test were generally associated with low (close to zero) values for the ML estimate of both d and g.

Given that the optimal value of d is chosen based on the best fit of the tree to the data, we may expect that in some cases phylogenetic signal will not be detected with the original candidate branch lengths but will be detected after transformation. For trees with more than 20 tips, two such cases (P switched from > 0.05 to < 0.05) occurred: log(male/female) body mass of 101 bat genera (Nee's arbitrary branch lengths; data from Hutcheon 2001); rhytidome allometric coefficient for 32 *Pinus* spp. (all branch lengths = 1; Jackson et al. 1999). A third case was close: log of mass-corrected group size of 28 macropods (P changed from 0.161 to 0.059; Fisher and Owens 2000). With the ACDC transformation, the same two traits showed significant signal after branch-length

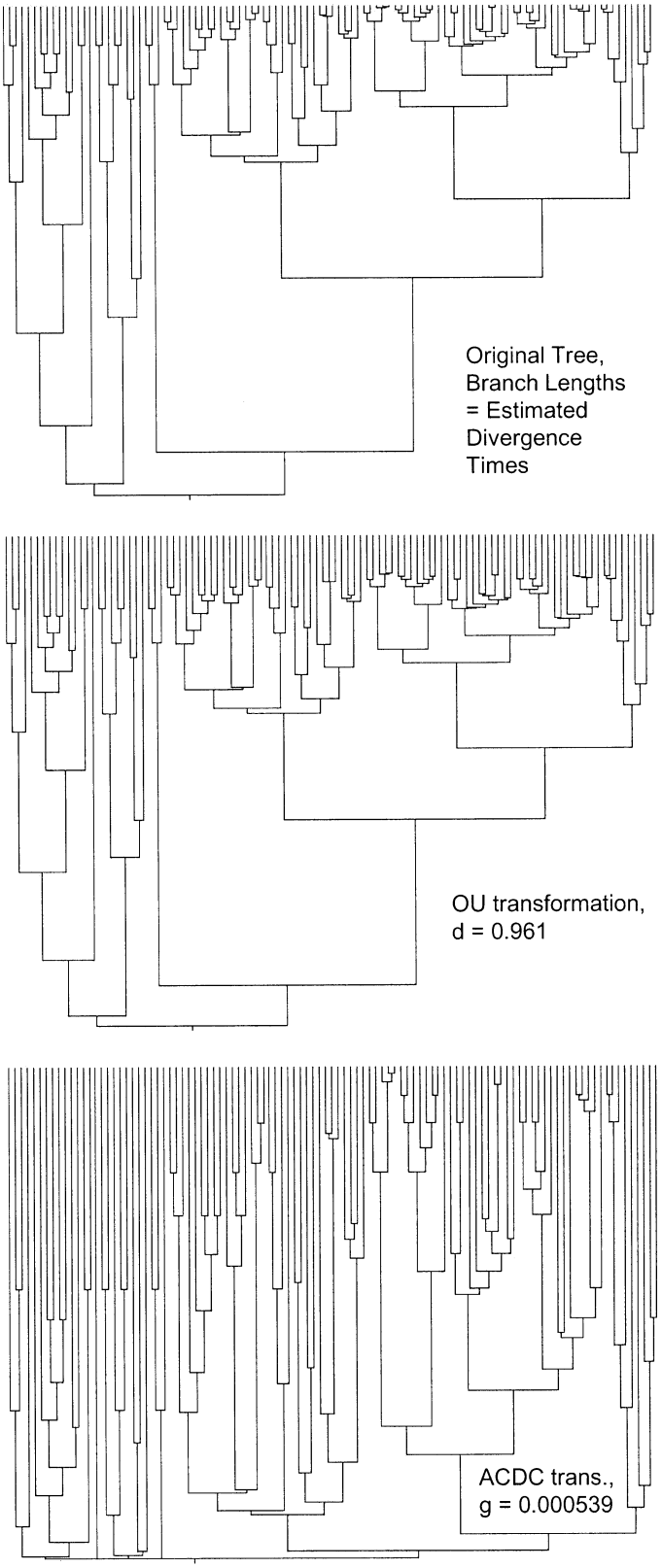

FIG. 8. Example of a real tree (estimated divergence times for 105 species of primates; from Smith and Cheverud 2002) transformed by the maximum likelihood estimate of the Ornstein-Uhlenbeck (OU) parameter (*d*) and of the acceleration-deceleration (ACDC) parameter (*g*) to obtain best fit to data on log(male/female) body mass (a measure of sexual dimorphism). Both parameters differ significantly from zero (both $P < 0.001$) and from unity ($P = 0.028$ and 0.002, respectively); hence, both transformed trees fit the data better than does either a star phylogeny (no hierarchical structure, MSE = 0.0544410) or the original candidate tree (MSE = 0.030422). Comparison of the MSEs indicates that ACDC-transformed tree (MSE = 0.0247463) fits the data better than does the OU-transformed tree (MSE = 0.0277609).

transformation, as did macropod group size (P changed to 0.036).

DISCUSSION

We have devised a simple randomization test for the presence of phylogenetic signal in continuous-valued traits. A survey of published datasets indicates that phylogenetic signal can be detected ($P < 0.05$) in most cases, especially for studies that involve 20 or more species (92%). This result demonstrates that phylogenetic signal is ubiquitous and provides empirical justification for the claim that estimates of mean values for species generally cannot be assumed to represent independent pieces of information in statistical analyses. We have also developed a descriptive statistic (K) that can be used to gauge the amount of phylogenetic signal relative to the amount expected for a character undergoing Brownian motion evolution along the specified topology and branch lengths. The literature survey indicates that on average most traits, other than body size, showed less signal than expected (i.e., $K < 1$). Analysis of variance and multiple-range comparison shows that, as a group, behavioral traits exhibit significantly lower signal than body size, morphological, life-history, or physiological traits (all corrected for body size) and that physiological traits show less signal than does body size. This result provides evidence consistent with the long-standing idea that behavior is relatively labile evolutionarily.

An alternative approach to testing for phylogenetic signal involves comparison of the fit of a series of trees whose branch lengths vary as dictated by a transformation parameter, one extreme of which yields a star phylogeny. Accordingly, we have also studied the behavior of two biologically motivated branch-length transformations, one based on the OU model of character evolution and the other based on a new model of accelerating/decelerating rates of evolution (ACDC). We have shown how to fit both models to data to obtain trees that better describe the data (lower MSE in GLS models or, equivalently, lower variance of phylogenetically independent contrasts). For trees with 20 or more species, tests of whether the OU and ACDC transformation parameters (*d* and *g*, respectively) are significantly greater than zero (which implies the presence of phylogenetic signal) are consistent with results of the randomization test: when the randomization test detects signal, the estimated *d* and *g* are also greater than zero (a ML estimate that equaled zero or near zero would imply that a star phylogeny best fit the data). These two transformations provide different biological perspectives from which the degree of phylogenetic signal in the data can be examined. Application to real datasets indicates that the ACDC model is at least as useful as the OU

model in terms of obtaining a set of branch lengths that provides a good fit to data. For trees with more than 20 tips, the estimated OU (25% of cases) and/or ACDC (40%) transformation parameter differed significantly from both one and zero, indicating that a tree less (or occasionally more) hierarchical than the original provided a better fit to the tip data and so could be preferred for comparative analyses.

Empirical Results: Presence of Phylogenetic Signal and Transformation of Branch Lengths

Regardless of sample size, our randomization test detected no traits that showed a significant tendency for relatives not to resemble each other. In other words, we found no cases in which relatives were less similar than if placed on the phylogeny at random (phylogenetic antisignal). Thus, for the examples that we considered, we found no evidence to suggest the presence of a process such as phylogenywide character displacement between close relatives. This is not to suggest that such a process did not occur between or among some close relatives, just that, if it did occur, its effects were swamped by the more general tendency for relatives to resemble each other.

The fit of any tree to data, regardless of whether its branch lengths have been transformed, can be compared by the MSE (under the constraint that the height of all trees have been scaled so that the determinant of the variance-covariance matrix equals unity; see Appendix 4). Two comparisons are instructive. First, in how many cases does a star phylogeny better fit the data than the original candidate tree? Second, in how many cases does a tree with branch lengths transformed according to the OU or ACDC process better fit the data than does a star phylogeny? Note that the transformed tree always must fit the data better than or as well as the original tree. Also, note that other types of branch-length transformations could be considered, such as raising the length of each branch segment to a power or taking the log (Garland et al. 1992; Díaz-Uriarte and Garland 1996, 1998; Reynolds and Lee 1996; Garland and Díaz-Uriarte 1999) or employing Pagel's λ (Pagel 1999; Freckleton et al. 2002).

For trees with more than 20 tips, a star phylogeny better fit the data than did the candidate tree in 12 of 53 cases (23%). For these trait-tree combinations, one might thus be tempted to employ conventional statistical analyses in a comparative study of, for example, character correlations. However, in all 12 of those cases, the OU- and/or the ACDC-transformed tree better fit the data (lower MSE) than did the star. Hence, for statistical analyses, the transformed tree would be a better choice. Comparing the transformations, in 25 cases the best fit was obtained by the OU model and in 28 by the ACDC model. Thus, the new ACDC model appears to be at least as useful as the OU model. For the 53 traits, ranks of the estimates of d and g were highly correlated (Spearman's $r = 0.824$, $P < 0.001$). (Note that the ACDC transformation is highly nonlinear: very small values can still imply substantial hierarchical structure.)

Finally, it is important to note that such models as OU or ACDC may provide improved fit to a set of tip data yet still be a poor representation of the real evolutionary processes that have been at work. Thus, inferences about the presence of OU- or ACDC-like processes should be made with caution.

Empirical Results: Amount of Phylogenetic Signal and Comparison of Trait Types

A K-value of one indicates that a trait shows exactly the amount of phylogenetic signal expected under Brownian motion evolution along the specified tree. For the 24 measures of adult body size, the mean $\log_{10} K$ was -0.08 with a 95% confidence interval that included zero (-0.20 to 0.03), or unity on the arithmetic scale. Thus, on average, body size exhibits neither more nor less phylogenetic signal than expected under a simple stochastic model of character evolution. This result is somewhat remarkable, given the various sources of error that will tend to reduce phylogenetic signal (see introduction and next paragraph). Most traits, however, exhibited K less than unity (see Fig. 6), and for all other trait types the 95% confidence interval excludes zero on the log scale, indicating that, as a group, they tend to exhibit less signal than expected (see Fig. 6).

Values of K less than unity may be caused by deviation from Brownian motion and/or measurement error, and variation in either factor could account for differences in K among individual traits or among types of traits. One obvious way that trait evolution may deviate from simple Brownian motion is if adaptation to a particular environmental factor occurs in some but not all members of a set of species. A simple hypothetical example of this is shown in figure 1 of Blomberg and Garland (2002). Measurement error can come from three sources: error in the measurement of the tip data, errors in branch lengths, or errors in tree topology. All three errors will generally make close relatives appear less similar than expected under Brownian motion, hence lowering K. We may expect the tip data to contain substantial measurement error in many of the datasets, given that values are often based on the measurement of few individuals, many studies include data taken by different investigators, and many of them were not based on common-garden studies, so genotype-environment interactions may have had unpredictable effects (see also Garland and Adolph 1991; Garland et al. 1992). The last of these represents an intractable problem for many comparative studies, especially ones that include a large range of body sizes and/or kinds of organisms (see also Garland 2001). For example, one simply cannot employ the same housing conditions nor feed the same food to a hummingbird and an ostrich. Moreover, many of the behavioral and ecological traits considered (e.g., home range size) exist only under natural conditions. Errors in branch lengths may also be large, especially given that approximately one-third of the studies that we have analyzed employed arbitrary branch lengths. Finally, topological errors are almost certain to exist in trees with relatively large numbers of species.

When analyzed on the original trees (without transformation of branch lengths), different types of traits showed significant variation in the amount of phylogenetic signal. Mean log MSE values ranked as body size $>$ morphology $>$ life history $>$ physiology $>$ behavior, and a multiple range comparison indicated that behavioral traits showed significantly lower MSE than did all other trait types, and physi-

ology had lower K than did body size. At least three nonexclusive hypotheses could explain this result. First, the studies included in our analysis may be a biased sample, unrepresentative of traits in general. As noted in the Results, however, the behavioral traits are from 10 studies, and those with particularly low K-values (on the order of 0.1) are from four different studies. Second, although we do not have direct information on pure measurement error, we would hypothesize that it is lower for body size and for other morphological traits (e.g., limb length) than for the other types of traits considered. At the opposite extreme, behavioral traits are notorious for being strongly subject to both immediate environmental effects and acclimation. As well, behavioral traits measured in the field (e.g., home range size) may show large seasonal or year-to-year variation.

Third, and most interesting, behavioral traits may be more evolutionary labile than other types of traits, as has often been suggested (e.g., Mayr 1963, 1982; Bush 1986; Huey and Bennett 1987; Plomin 1990; Gittleman et al. 1996a,b). Many examples are known in which behavioral traits seem to have evolved further than have some of the underlying morphological and physiological traits that would facilitate the behavior, such as in the foraging behavior of some anuran amphibians (Taigen and Pough 1985, p. 991) or the water ouzels (dippers), which dive but seem to have few morphological or physiological specializations that would facilitate diving ability (Futuyma 1986, p. 257). In other cases, changes in behavior may alter the selective environment of other traits and drive their evolution (Plotkin 1988; Wcislo 1989).

Gittleman et al. (1996a,b) used the true R^2 and phylogenetic autocorrelograms (Gittleman and Kot 1990; see next section), as well as estimates of evolutionary rates in darwins, to compare two morphological traits (brain size, body size), two life-history traits (gestation length, birth weight), and two behavioral traits (home range size, group size) across eight mammalian datasets ranging in size from 11 to 39 species. They found that the two behavioral traits (especially group size) generally showed the lowest R^2 and the weakest relationship with phylogeny in autocorrelograms, as well as higher estimated evolutionary rates. Morphological traits generally showed the highest R^2, and life-history traits were intermediate. Thus, their results are generally consistent with ours. (Note, however, that they did not correct the other traits for correlations with body size.) Given the mathematical and conceptual problems that have come to light with the autocorrelation method (Rohlf 2001), it would be of interest to reanalyze their datasets with our methods.

Different results emerge from comparisons among traits for both consistency indices and retention indices: behavioral characters showed no more homoplasy than did morphological characters (de Queiroz and Wimberger 1993; Wimberger and de Queiroz 1996). However, as noted by Wimberger and de Queiroz (1996), the behavioral traits that they considered were chosen by systematists specifically for potential utility in reconstructing phylogenetic relationships, and hence do not necessarily reflect properties of behavior and morphology in general. Moreover, they studied discrete rather than continuous-valued traits. Finally, studies of the heritability estimates for behavioral, life-history, physiological, and morphological traits show that behavioral and physiological traits have relatively low heritability, similar to life-history traits, and much lower than morphological traits (Mousseau and Roff 1987; Stirling et al. 2002). In principle, low heritabilities could reduce evolutionary lability, but the heritability estimates for behavioral traits are still generally high enough to allow rapid response to selection.

Other Approaches to Phylogenetic Signal

A number of previous techniques have been used, either explicitly or implicitly, to study phylogenetic signal in continuous-valued characters. Some of these have been used in attempts to test for phylogenetic inertia (see Blomberg and Garland 2002). Here, we briefly outline some of the more prominent methods. Several techniques, based on nested analysis of variance, have been used to examine the relative proportion of variance in a trait that can be partitioned among taxonomic levels (e.g., Stearns 1983; Harvey and Clutton-Brock 1985; Prinzing et al. 2001; see Harvey and Pagel 1991 for review). We do not consider these methods because taxonomic levels are arbitrary and the methods cannot accommodate full phylogenetic information (any topology and branch lengths).

Cheverud et al. (1985) proposed the first fully phylogenetic method that has been used to gauge the degree of phylogenetic burden or inertia in continuous-valued characters. They modified the tools of spatial autocorrelation and developed two statistics that might indicate the degree of phylogenetic signal, the autocorrelation coefficient (ρ) and the true R^2 (see Garland et al. 1999). Gittleman and Kot (1990) extended the method by allowing, in effect, for transformation of phylogenetic branch lengths; if this parameter is allowed to vary, then interpretation of ρ and the true R^2 would also be affected. Gittleman and Kot (1990) also proposed the use of Moran's I as another indicator of phylogenetic signal. Subsequently, Gittleman and coworkers have applied the phylogenetic autocorrelation method to a large number of comparative datasets (e.g., Geffen et al. 1996; Gittleman et al. 1996a,b; Gittleman and Van Valkenburgh 1997), as have others (e.g., Morales 2000; Smith and Cheverud 2002). Unfortunately, Rohlf (2001) has recently shown that the autocorrelation coefficient has not been estimated correctly in previous studies and, moreover, that the method has several problems that limit its usefulness in comparative studies. In particular, the interpretation of ρ is uncertain because it is not constrained between ± 1 and is sometimes considerably less than minus one. Also, both large positive and large negative values imply that most trait variation is at the base of the tree, whereas values closer to zero imply that variation is at the tips of the tree. Large values of ρ do not necessarily imply a better fit of the model to the data. Hence, significance tests of ρ are difficult to interpret. The use of the true R^2 as a measure of the variance explained by phylogeny is also problematic because a high R^2-value only signifies trait variation somewhere on the tree. Finally, the estimation of ρ does not correspond to the maximization of R^2 (Rohlf 2001).

The other major test for phylogenetic signal in comparative data (continuous-valued or categorical traits) is by Abouheif (1999), who uses a test for serial independence, originally developed by von Neumann et al. (1941) in a nonphyloge-

netic context. The von Neumann et al. test works by comparing the variation in the difference of successive observations to the sum of squares of the observations. It can be used to test for serial independence in any dataset, but it requires some modification for application to phylogenetic data. Specifically, Abouheif (1999) deals with the problem that any single tree topology can be represented in $2^{(n-1)}$ different ways, because branches can be rotated at nodes. Rotating branches at nodes results in a different ordering of the data in the dataset and changes the calculation of the statistic for serial independence. Abouheif's solution is to calculate *C* (the statistic for serial independence) on a large number of randomly generated trees that maintain the same relationships among species as the original tree topology. A mean *C*-value is then calculated. This process is then used in a permutation test to generate a null distribution of mean *C*-values to use in tests of significance. The main limitation of Abouheif's (1999) proposed test is that it does not incorporate branch length information. This is problematic because branch lengths provide important information about expected species' similarity that cannot be gained from the topology alone. In addition, he provides no analysis of Type I or Type II error rates.

A test similar to our randomization procedure, but for discrete characters, was proposed by Maddison and Slatkin (1991). They also randomized character values equiprobably across the tree, but instead of using the variance of phylogenetically independent contrasts (or the MSE in GLS mode) as a measure of fit of the tree to the data, they use the Fitch-Farris character optimization algorithm to calculate the minimum number of character changes necessary on the tree. A low number of character changes (low Fitch-Farris score), compared with the distribution of minimum character changes generated by permuting the species on the tree, implies that the traits show significant historical inertia. The Maddison-Slatkin method in effect assumes that all branch segments are of equal length.

The quantitative convergence index (QVI) of Ackerly and Donoghue (1998) can also be interpreted as an indicator of phylogenetic signal. This statistic is based on linear parsimony algorithms for continuous-valued characters. It is equivalent to one minus the retention index (Farris 1989) and measures the degree to which sister taxa are similar (QVI = 0) or different (i.e., QVI = 1 when the most distant taxa are most similar, and thus convergent evolution is maximized). Randomization methods are used to provide null distributions for testing the statistical significance of the QVI (see also Ackerly and Reich 1999; Prinzing et al. 2001).

Another method that partitions phylogenetic effects among different phylogenetic levels was developed by Legendre et al. (1994; for an application see Böhning-Gaese and Oberrath 1999). This method relies on the comparison of a matrix of trait dissimilarities with a matrix of phylogenetic distances. It can be used to compare different classes of traits, as in the present paper, but it is not based on any explicit model of evolution, and it must make implicit assumptions depending on the way matrices are constructed (Martins and Hansen 1996).

Jablonski (1987) analyzed the cross-species heritabilities (which might be viewed as a measure of phylogenetic signal) of geographical range in fossil mollusks. However, his analysis is unusual in that a detailed fossil record allowed identification of ancestor (parent)-descendent (offspring) relationships among species. These conditions are not normally met with other comparative datasets. Ashton (2001) proposed a similar test for phylogenetic conservatism, based on the correlation observed in a bivariate scatterplot of traits values for sister taxa (with ancestral values computed as simple averages of the sister taxa). This test did not use information on phylogenetic branch lengths, however, and its statistical properties are unknown.

Grafen (1989), who first proposed GLS methods for comparative data, also included estimation of a parameter ρ (not the same as the like-named autocorrelation coefficient in Gittleman and Kot 1990) that served to stretch/compress the internal nodes of a phylogenetic tree, much like the OU and ACDC transformations. He did so from the presumed starting point that only arbitrary branch lengths would be available and that the topology would contain soft polytomies. He implemented a ML method for estimating an optimal branch-length transformation, simultaneously with estimating the form of the relationship between two or more traits in a GLS model. If, and only if, the starting branch lengths have contemporaneous tips, then a ρ of zero will yield a star phylogeny. Thus, estimation of ρ could be interpreted as providing a test for phylogenetic signal. However, the branch-length modifications implied by values of ρ that differed from unity were not based on a model of evolution, and Grafen viewed it simply as a nuisance parameter. Pagel (1999) and Freckleton et al. (2002) employed a similar transformation parameter (λ), also not based on a model of evolution, but they suggest it can be used to differentiate among various ''modes'' of evolution. One might conjecture that nonbiologically motivated branch-length transformations would tend to fit real data less well than do their counterparts that are based on plausible biological models. However, it should also be noted that many real datasets show clear deviations from such simple models as OU and ACDC (e.g., absolute physical limits to character evolution, unequal rates of evolution across lineages [Garland 1992; Garland and Ives 2000], lineage-specific effects [sensu Arnold 1994]) and many studies use arbitrary branch lengths as a starting point, so it is possible that these biologically based transformations would fare little better than, for example, Grafen's (1989) ρ or direct exponential or logarithmic transformation of each branch segment (Garland et al. 1992; Díaz-Uriarte and Garland 1996, 1998; Reynolds and Lee 1996; Garland and Díaz-Uriarte 1999). This is an empirical question that could be examined by application of both types of transformations to a large number of real and simulated datasets (cf. Mooers et al. 1999).

It would, of course, be possible to apply transformations that represent even more complex models of evolution than OU and ACDC. Transformations with multiple parameters (e.g., the ''free'' model of Mooers et al. 1999) can be problematic, as the position of the tree in the multiparameter space must be considered and the meaning of the parameters can become uncertain (as in the phylogenetic autocorrelation method, see above). For example, the free model requires one parameter for each branch length (edge) in the tree (2*N*

− 2 parameters for N species in the tree). It is not clear how to simultaneously interpret all of the parameters in the model. Furthermore, estimating each branch length from the data can provide a description of the best set of branch lengths given the trait data, but does not incorporate any external information on branch lengths for the given tree, such as may be available from molecular or fossil data, although the information in the tree topology is used. However, intermediate transformations—those that affect only a subset of the branch lengths on the tree—may be very useful. One ad hoc empirical example is by Garland and Ives (2000), who describe differing rates of evolution in passerine and nonpasserine birds. They scaled the height of the branches in the passerine subclade separately from the rest of the tree in an attempt to enforce the equivalence of contrast variances across the tree as a whole. Such rescaling can be viewed as an intermediate transformation between single-parameter transformations, such as our implementation of OU or ACDC and the free model of Mooers et al. (1999), in which the number of parameters is equal to the number of branches in the tree because the rescaling transformation has a parameter corresponding to the different values of the contrast variances for passerines and nonpasserines.

Finally, the mixed-effects model of Lynch (1991) can be used to gauge phylogenetic signal. Analogous to quantitative-genetic analyses with pedigrees, this method attempts to decompose the total phenotypic variance among species (tips on a phylogeny) into components associated with heritable versus nonheritable effects. The method is still under development and seems to be quite demanding of sample size (Martins et al. 2002; E. A. Housworth and M. Lynch, pers. comm.). Discussion of its potential is premature and beyond the scope of the present paper (see also Blomberg and Garland 2002).

Recommendations for Analysis of Comparative Data

We recommend presentation of the following along with any univariate analysis of comparative data: (1) the tip data analyzed; (2) the actual phylogeny (topology and branch lengths) used for analyses; (3) the expected MSE_0/MSE as an index of how hierarchical the tree is (e.g., see Fig. 5); (4) the observed MSE_0/MSE and the K-statistic, the latter to allow direct comparisons among trait-tree combinations; (5) the MSE of the trait(s) on the tree used for analyses and on a star phylogeny; (6) results of the randomization test for phylogenetic signal; and (7) results of the traditional diagnostic test for adequacy of branch lengths (Pearson correlation [not through origin] of absolute values of standardized contrasts versus their standard deviations; Garland et al. 1991, 1992), along with comments about any notable outlier contrasts that may heavily influence the distribution. In addition, one should check for clade differences in this diagnostic plot that may indicate heterogeneity in rate of evolution, and hence the need for more complex branch-length transformations or statistical analytical models (e.g., see Garland 1992; Garland and Ives 2000).

We view all of the foregoing as essential. Additional important information depends on what types of branch-length transformations, if any, have been pursued. Whatever strategy is followed, a description of the procedures used to arrive at a final set of branch lengths for analysis should be presented. If optimal transformations are estimated, such as under the OU and ACDC models, then one should present the estimated transformation parameter, the MSE for the trait(s) on the transformed tree, and the randomization tests for whether d or g differs significantly from both zero and unity. In addition, one should generally subtract degrees of freedom during any subsequent comparative analyses (e.g., independent contrasts tests for a correlation between two traits) to account for estimation of the additional parameters (e.g., see Díaz-Uriarte and Garland 1996, 1998; Garland and Díaz-Uriarte 1999).

The randomization test for phylogenetic signal in combination with tests of whether d or g are different from zero allow one to choose objectively the branch lengths for a comparative analysis, such as by the method of phylogenetically independent contrasts, which is equivalent to GLS models (see also Freckleton et al. 2002). One possible outcome is that a star phylogeny fits the data as well as or better than other sets of branches. For traits with more than 20 species ($N = 53$), our empirical survey revealed only one such case. For the log of mass-corrected maximal metabolic rates of 47 species of birds (Rezende et al. 2002), estimates of both the OU and ACDC transformation parameter were close to zero (1×10^{-8} and 1×10^{-30}, respectively), yielding a tree that was mathematically indistinguishable from a star for the OU transform, but that had some small bifurcations remaining for the ACDC transform. In addition, no significant phylogenetic signal was detected by the randomization test on either the original tree ($P = 0.260$) or the ACDC-transformed tree ($P = 0.065$). Thus, for this trait, a star phylogeny would be a justifiable choice for statistical analyses. In agreement with our finding of no phylogenetic signal for maximal metabolic rate, Rezende et al. (2002) found no significant difference in a phylogenetic ANCOVA comparing passerines with nonpasserines. For these same birds, however, log(body mass) showed strong signal (randomization $P < 0.001$) and was best fit by the original tree (estimate of $d = 0.99$, $g = 2.3$; neither significantly different from unity).

Independent contrast analyses of multivariate or multivariable relationships can be performed with different sets of branch lengths, or different transformations thereof, for different traits (e.g., Bonine and Garland 1999). Because independent contrasts are a special case of GLS computations (Rohlf 2001), the latter may also employ different branches for different traits, although this is often computationally more awkward than when using independent contrasts (Garland and Ives 2000). An alternative is to use a single set of branch lengths for analyses that involve multiple traits (e.g., bivariate correlation, principal components analysis, Clobert et al. 1998; multiple regression, see Grafen 1989). The most appropriate strategy will depend on the assumptions that one is willing to make about evolution of the traits in the analysis as well as the assumptions inherent to the statistical model (e.g., independence and normality in a bivariate relationship, in multivariate space, for residuals from a regression). As noted elsewhere, multiple-regression type analyses often include methodological nuisance variables (such as calculation method for home range area, length of study; e.g., Wolf et al. 1998; Perry and Garland 2002) that are clearly nonphy-

logenetic and so may best be analyzed as if on a star phylogeny, which can be done easily with independent contrasts.

For trees with fewer than 20 tips, our simulations indicate relatively low power to detect phylogenetic signal by the proposed randomization test (Fig. 2) and also unreliable estimation of the OU and ACDC branch-length transformation parameters (Fig. 7), which can also serve as a test for signal. Consistent with these simulation results, our empirical survey found that a star better fit the data, as compared with the hierarchical tree, for 38 of 68 traits in studies with fewer than 20 species. Following transformation, the best-fitting tree was: star, 17; OU, 24; ACDC, 27 (the star was credited with best fit if the OU or ACDC parameter was estimated as zero). The fact that a star phylogeny sometimes better fits the data for trees with relatively few species may be attributable to the lower possible amount of phylogenetic signal on smaller trees (Fig. 5) and/or the difficulty in ML estimation for smaller trees (Fig. 7). For analysis of real data with small sample sizes, a finding that a star phylogeny better fits the data than do various hierarchical trees should not be taken as a directive to apply and favor the results of conventional statistical analyses. However, one would be well advised to apply analyses with several sets of branch lengths, whether arbitrary or based on biological models, and view the exercise as a sensitivity analysis (e.g., see fig. 2 in Garland et al. 1999; Butler et al. 2000).

Conclusions and Future Directions

Our empirical survey indicates that phylogenetic signal is pervasive in cross-species datasets, even for some behavioral traits that are expected to be evolutionarily malleable (e.g., highly adaptive) and/or quite subject to nongenetic environmental effects, such as home range size and group size (see also Prinzing et al. [2001] on niche position of plants). This result reinforces the importance of phylogenetically based statistical methods for analyses of comparative datasets. Our results also emphasize the importance of exploring branch-length transformations, whether biologically motivated or not. An important area for future research will be determining, for a wide range of evolutionary models, which types of transformations are most effective and robust with respect to the performance of subsequent statistical analyses, such as testing for correlated character evolution (e.g., Grafen 1989; Martins and Garland 1991; Pagel 1994, 1999; Purvis et al. 1994; Díaz-Uriarte and Garland 1996, 1998; Price 1997; Harvey and Rambaut 1998, 2000; Garland and Díaz-Uriarte 1999; Martins et al. 2002). Another important topic is how the various diagnostics that have been suggested previously for choosing branch-length transformations from a purely statistical perspective (Grafen 1989; Garland et al. 1991, 1992; Díaz-Uriarte and Garland 1996, 1998; Reynolds and Lee 1996; Garland and Díaz-Uriarte 1999; Harvey and Rambaut 2000) relate to the alternate optimality criterion of minimizing the MSE, as used here and by others who have implemented GLS approaches. Either kind of optimality criterion could be used to choose from within families of either statistically or biologically motivated branch-length transformations, but this has not yet been pursued. Our preliminary analyses, both analytical and empirical (results not shown), indicate that the traditional diagnostic test for adequacy of branch lengths (Pearson correlation [not through origin] of absolute values of standardized contrasts versus their standard deviations: Garland et al. 1991, 1992) often suggests the same tree as does the criterion of minimum MSE.

Although the OU and ACDC transformations may cover a fairly broad range of what might be termed ''well-behaved'' evolutionary models (see also Garland et al. 1993; Hansen and Martins 1996; Butler et al. 2000; Martins et al. 2002), they are clearly too simple for many real traits, such as those exhibiting strong lineage-specific effects (e.g., plasma osmolarity of vertebrates, Garland et al. 1997), clade differences in rates of evolution (Garland 1992; Barbosa 1993; Clobert et al. 1998; Barbosa and Moreno 1999; Garland and Ives 2000), or absolute limits to evolution (see Garland et al. 1993). Such cases will often exhibit problems of estimation that may be detected by inspection of a plot of the MSE versus the parameter value, as is provided in our MatLab programs (see also comments in Grafen 1989). When an irregular likelihood surface is noted, or when a more complicated evolutionary model is suspected, our randomization test is probably much more robust for simply detecting the presence of phylogenetic signal as compared with testing whether d or g differs significantly from zero. Our K-statistic is a useful descriptor of the amount of phylogenetic signal for any trait on any tree, regardless of how complicated its evolution may have been, but values much larger or smaller than unity do not offer any insight as to how evolution may have deviated from Brownian motion (even if we assume that measurement errors are negligible). Another way to detect complicated trait evolution is by inspection of, for example, bivariate scatterplots of traits versus body mass, both for raw data and for phylogenetically independent contrasts, including raw-data plots that are coded by major lineage (to detect grade shifts) and/or that have phylogenetically correct allometric lines superimposed (see also Garland et al. 1993; Ackerly and Donoghue 1998; Ackerly 1999; Ackerly and Reich 1999; Garland and Ives 2000; Nunn and Barton 2000; Garland 2001). Insight to adaptive radiation may also be gained by comparing the amount of signal (or the rate of evolution: Garland 1992; Garland and Ives 2000) present in different subclades. For example, a low amount of signal or high rate of evolution suggests that the clade in question may possess a key innovation. Of course, a single such clade represents only an anecdote, and the attribution of key innovations is exceedingly difficult if not logically impossible in the absence of replicate lineages that show both the putative key innovation and a high rate of evolution (e.g., see Losos and Miles 2002).

The tools employed in this paper should be applied to a much wider range of comparative datasets. Our preliminary results are encouraging in that they reveal differences in the amount of phylogenetic signal (as indicated by our K-statistic) among different types of traits. Consistent with some previous work (Gittleman et al. 1996a,b), we find that behavioral traits tend to show less phylogenetic signal than do other types of traits. Another topic of interest would be to explore the effects of using various types of real and/or arbitrary branch lengths, plus branch-length transformations, on comparisons of the amount of signal across trait types

(see also Mooers et al. 1999). For example, if a consistent set of branch lengths were imposed for all trees and traits, with or without the use of branch-length transformations, would we still find differences in phylogenetic signal across trait types?

Neither the descriptive statistic (*K*) nor the tests for phylogenetic signal (randomization; estimates of *d* and *g* vs. zero) attempt to take into account effects of adaptation (e.g., see fig. 1 in Blomberg and Garland 2002). In comparative studies, the usual way of inferring adaptation is by correlating character variation with continuous-valued or categorical descriptors of environmental factors that are thought to indicate variation in selective regime. Thus, it will be important to develop methods that can reliably estimate correlations of characters with environmental variables while simultaneously estimating the degree of phylogenetic signal. Such methods can be problematic in practice because variation in selective regime may itself be confounded with phylogenetic position; that is, the environmental predictor variables may themselves exhibit phylogenetic signal. Such confounding can lead to problems of both estimation and interpretation, similar to quantitative-genetic studies in which genotypes and environments are confounded. Also important for such attempts will be explicit incorporation of information on measurement error of various types. Such a comprehensive and practical method does not yet exist, but steps are being made in that direction (e.g., see Gittleman and Kot 1990, p. 231; Lynch 1991; Butler et al. 2000; Cornillon et al. 2000; Baum and Donoghue 2001; Housworth and Martins 2001; Orzack and Sober 2001; Martins et al. 2002; Y. Desdevises, P. Legendre, L. Azouzi, and S. Morand, pers. comm.). The effect of such simultaneous-inference procedures on the apparent amount of phylogenetic signal in traits is difficult to predict.

Acknowledgments

This work was supported by National Science Foundation grant DEB-0196384 to TG and ARI. SPB was supported in part by a Michael Guyer postdoctoral fellowship from the University of Wisconsin–Madison Department of Zoology. We thank all of our colleagues who provided data files for analyses. We also thank B. J. Crespi, A. de Queiroz, C. L. Nunn, and R. J. Smith for comments on the manuscript.

Literature Cited

Abouheif, E. 1999. A method for testing 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. Pp. 391–413 *in* M. C. Press, J. D. Scholes, and M. G. Braker, eds. Physiological plant ecology. The 39th symposium of the British Ecological Society held at the University of York, 7–9 September 1998. Blackwell Science, Oxford, U.K.

———. 2000. Taxon sampling, correlated evolution, and independent contrasts. Evolution 54:1480–1492.

Ackerly, D. D., and M. J. Donoghue. 1998. Leaf size, sapling allometry, and Corner's rules: phylogeny and correlated evolution in maples (*Acer*). Am. Nat. 152:767–791.

Ackerly, D. D., and P. B. Reich. 1999. Convergence and correlations among leaf size and function in seed plants: a comparative test using independent contrasts. Am. J. Bot. 86:1272–1281.

Antonovics, J., and P. H. van Tienderen. 1991. Ontoecogenophyloconstraints? The chaos of constraint terminology. Trends Ecol. Evol. 6:166–168.

Arnold, E. N. 1994. Investigating the origins of performance advantage: adaptation, exaptation and lineage effects. Pp. 123–168 *in* P. Eggleton and R. I. Vane-Wright, eds. Phylogenetics and ecology. Linnean Society Symposium Series no. 17. Academic Press, London.

Ashton, K. G. 2001. Body size variation among mainland populations of the western rattlesnake (*Crotalus viridis*). Evolution 55:2523–2533.

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

Barbosa, A., and E. Moreno. 1999. Evolution of foraging strategies in shorebirds: an ecomorphological approach. Auk 116:712–725.

Baum, D. A., and M. J. Donoghue. 2001. A likelihood framework for the phylogenetic analysis of adaptation. Pp. 24–44 *in* S. Orzack and E. Sober, eds. Adaptationism and optimality. Cambridge Univ. Press, Cambridge, U.K.

Bjorklund, M. 1997. Are 'comparative methods' always necessary? Oikos 80:607–612.

Blomberg, S. P., and T. Garland Jr. 2002. Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. J. Evol. Biol. 15:899–910.

Böhning-Gaese, B., and R. Oberrath. 1999. Phylogenetic effects on morphological, life-history, behavioural and ecological traits of birds. Evol. Ecol. Res. 1:347–364.

Bonine, K. E., and T. Garland Jr. 1999. Sprint performance of phrynosomatid lizards, measured on a high-speed treadmill, correlates with hindlimb length. J. Zool. (Lond.) 248:255–265.

Bonine, K. E., T. T. Gleeson, and T. Garland Jr. 2001. Comparative analysis of fiber-type composition in the iliofibularis muscle of phrynosomatid lizards (Sauria). J. Morphol. 250:265–280.

Brashares, J., T. Garland Jr., and P. Arcese. 2000. Phylogenetic analysis of coadaptation in behavior, diet, and body size in the African antelope. Behav. Ecol. 11:452–463.

Bush, G. L. 1986. Evolutionary behavior genetics. Pp. 1–5 *in* M. D. Huettel, ed. Evolutionary genetics of invertebrate behavior, progress and prospects. Plenum Press, New York.

Butler, M. A., T. W. Schoener, and J. B. Losos. 2000. The relationship between sexual size dimorphism and habitat use in Greater Antillean *Anolis* lizards. Evolution 54:259–272.

Calder, W. A., III. 1984. Size, function, and life history. Harvard Univ. Press, Cambridge, MA.

Chevalier, C. D. 1991. Aspects of thermoregulation and energetics in the Procyonidae (Mammalia: Carnivora). Ph.D. diss., University of California, Irvine, CA.

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. Anthropol. 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.

Cornillon, P.-A., D. Pontier, and M.-J. Rochet. 2000. Autoregressive models for estimating phylogenetic and environmental effects: accounting for within-species variations. J. Theor. Biol. 202:247–256.

Cruz-Neto, A. P., T. Garland Jr., and A. S. Abe. 2001. Diet, phylogeny, and basal metabolic rate in phyllostomid bats. Zoology 104:49–58.

Derrickson, E. M., and R. E. Ricklefs. 1988. Taxon-dependent diversification of life-history traits and the perception of phylogenetic constraints. Funct. Ecol. 2:417–423.

de Queiroz, A., and P. H. Wimberger. 1993. The usefulness of behavior for phylogeny estimation: levels of homoplasy in behavioral and morphological characters. Evolution 47:46–60.

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.

———. 1998. Effects of branch length errors on the performance of phylogenetically independent contrasts. Syst. Biol. 47: 654–672.
Ebensperger, L. A., and H. Cofre. 2001. On the evolution of group-living in the New World cursorial hystricognath rodents. Behav. Ecol. 12:227–236.
Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap: monographs on statistics and applied probability. Chapman and Hall, New York.
Farris, J. S. 1989. The retention index and homoplasy excess. Syst. Zool. 18:374–385.
Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.
———. 1988. Phylogenies and quantitative characters. Annu. Rev. Ecol. Syst. 19:445–471.
Ferguson, S. H., and S. Lariviere. 2002. Can comparing life histories help conserve carnivores? Anim. Cons. 5:1–12.
Fisher, D. O., and I. P. F. Owens. 2000. Female home range size and the evolution of social organization in macropod marsupials. J. Anim. Ecol. 69:1083–1098.
Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. Am. Nat. 160:712–726.
Futuyma, D. J. 1986. Evolutionary biology. 2d ed. Sinauer Associates, Sunderland, MA.
Garland, T., Jr. 1992. Rate tests for phenotypic evolution using phylogenetically independent contrasts. Am. Nat. 140:509–519.
———. 2001. Phylogenetic comparison and artificial selection: two approaches in evolutionary physiology. Pp 107–132 *in* R. C. Roach, P. D. Wagner, and P. H. Hackett, eds. Hypoxia: from genes to the bedside. Advances in experimental biology and medicine. Vol. 502. Kluwer/Plenum Academic, New York.
Garland, T., Jr., and S. C. Adolph. 1991. Physiological differentiation of vertebrate populations. Annu. Rev. Ecol. Syst. 22: 193–228.
———. 1994. Why not to do two-species comparative studies: limitations on inferring adaptation. Physiol. Zool. 67:797–828.
Garland, T., Jr., and R. Díaz-Uriarte. 1999. Polytomies and phylogenetically independent contrasts: an examination of the bounded degrees of freedom approach. Syst. Biol. 48:547–558.
Garland, T., Jr., and A. R. Ives. 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., and C. M. Janis. 1993. Does metatarsal/femur ratio predict maximal running speed in cursorial mammals? J. Zool. 229:133–151.
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., 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., A. W. Dickerman, C. M. Janis, and J. A. Jones. 1993. Phylogenetic analysis of covariance by computer simulation. Syst. Biol. 42:265–292.
Garland, T., Jr., K. L. M. Martin, and R. Díaz-Uriarte. 1997. Reconstructing ancestral trait values using squared-change parsimony: plasma osmolarity at the origin of amniotes. Pp. 425–501 *in* S. S. Sumida and K. L. M. Martin, eds. Amniote origins: completing the transition to land. Academic Press, San Diego, CA.
Garland, T., Jr., P. E. Midford, and A. R. Ives. 1999. An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral values. Am. Zool. 39: 374–388.
Geffen, E., M. E. Gompper, J. L. Gittleman, H.-K. Luh, D. W. Macdonald, and R. K. Wayne. 1996. Size, life-history traits, and social organization in the Canidae: a reevaluation. Am. Nat. 147: 140–160.
Gibert, P., and R. B. Huey. 2001. Chill-coma temperature in *Drosophila*. Physiol. Biochem. Zool. 74:429–434.
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., and H.-K. Luh. 1992. On comparing comparative methods. Annu. Rev. Ecol. Syst. 23:383–404.
———. 1994. Phylogeny, evolutionary models and comparative methods: a simulation study. Pp. 103–122 *in* P. Eggleton and D. Vane-Wright, eds. Phylogenetics and ecology. Academic Press, London.
Gittleman, J. L., and B. Van Valkenburgh. 1997. Sexual dimorphism in the canines and skulls of carnivores: effects of size, phylogeny and behavioural ecology. J. Zool. (Lond.) 242:97–117.
Gittleman, J. L., C. G. Anderson, M. Kot, and H.-K. Luh. 1996a. Phylogenetic lability and rates of evolution: a comparison of behavioral, morphological and life history traits. Pp. 166–205 *in* E. P. Martins, ed. Phylogenies and the comparative method in animal behavior. Oxford Univ. Press, Oxford, U.K.
———. 1996b. Comparative tests of evolutionary lability and rates using molecular phylogenies. Pp. 289–307 *in* P. H. Harvey, A. J. Leigh Brown, J. Maynard Smith, and S. Nee, eds. New uses for new phylogenies. Oxford Univ. Press, Oxford, U.K.
Grafen, A. 1989. The phylogenetic regression. Proc. R. Soc. Lond. B 326:119–156.
Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351.
Hansen, T. F., and E. P. Martins. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50: 1404–1417.
Harcourt, A. H., P. H. Harvey, S. G. Larson, and R. V. Short. 1981. Testis weight, body weight and breeding system in primates. Nature 293:55–57.
Harris, M. A., and K. Steudel. 1997. Ecological correlates of hindlimb length in the Carnivora. J. Zool. (Lond.) 241:381–408.
Harvey, P. H., and T. H. Clutton-Brock. 1985. Life history variation in primates. Evolution 39:559–581.
Harvey, P. H., and A. H. Harcourt. 1984. Sperm competition, testes size, and breeding system in primates. Pp. 589–600 *in* R. L. Smith, ed. Sperm competition and the evolution of animal mating systems. Academic Press, New York.
Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford Univ. Press, Oxford, U.K.
Harvey, P. H., and A. Rambaut. 1998. Phylogenetic extinction rates and comparative methodology. Proc. R. Soc. Lond. B 265: 1691–1696.
———. 2000. Comparative analyses for adaptive radiation. Proc. R. Soc. Lond. B 355:1599–1605.
Hillis, D. M., and J. P. Huelsenbeck. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. J. Hered. 83: 189–195.
Housworth, E. A., and E. P. Martins. 2001. Random sampling of constrained phylogenies: conducting phylogenetic analyses when the phylogeny is partially known. Syst. Biol. 50:628–639.
Huey, R. B., and A. F. Bennett. 1987. Phylogenetic studies of coadaptation: preferred temperatures versus optimal performance temperatures of lizards. Evolution 41:1098–1115.
Hutcheon, J. M. 2001. Molecular systematics of the Chiroptera: a phylogenetic analysis using DNA-DNA hybridization and the comparative method. Ph.D. diss., University of Wisconsin, Madison, WI.
Irschick, D. J., L. J. Vitt, P. A. Zani, and J. B. Losos. 1997. A comparison of evolutionary radiations in mainland and Caribbean *Anolis* lizards. Evolution 78:2191–2203.
Iwaniuk, A. N., and J. E. Nelson. 2001. A comparative analysis of relative brain size in waterfowl (Anseriformes). Brain Behav. Evol. 57:87–97.
Jablonski, D. 1987. Heritability at the species level: analysis of geographic ranges of Cretaceous mollusks. Science 238: 360–362.
Jackson, J. F., D. C. Adams, and U. B. Jackson. 1999. Allometry of constitutive defense: a model and a comparative test with tree bark and fire regime. Am. Nat. 153:614–632.
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.
Lapointe, F.-J., and T. Garland Jr. 2001. A generalized permutation model for the analysis of cross-species data. J. Classif. 18: 109–127.
Legendre, P., F.-J. Lapointe, and P. Casgrain. 1994. Modeling brain evolution from behavior: a permutational regression approach. Evolution 48:1487–1499.
Losos, J. B. 1990. Ecomorphology, performance capability, and scaling of West Indian *Anolis* lizards: an evolutionary analysis. Ecol. Monogr. 60:369–388.
———. 1999. Uncertainty in the reconstruction of ancestral character states and limitations on the use of phylogenetic comparative methods. Anim. Behav. 58:1319–1324.
———. 2000. Ecological character displacement and the study of adaptation. Proc. Natl. Acad. Sci. USA 97:5693–5695.
Losos, J. B., and D. B. Miles. 2002. Testing the hypothesis that a clade has adaptively radiated: iguanid lizard clades as a case study. Am. Nat. 160:147–157.
Lynch, M. 1991. Methods for analysis of comparative data in evolutionary biology. Evolution 45:1065–1080.
Maddison, W. P., and M. Slatkin. 1991. Null models for the number of evolutionary steps in a character on a phylogenetic tree. Evolution 45:1184–1197.
Martins, E. P. 1994. Estimating the rate of phenotypic evolution from comparative data. Am. Nat. 144:193–209.
———. 1996. 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. 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 J. Lamont. 1998. Estimating ancestral states of a communicative display: a comparative study of *Cyclura* rock iguanas. Anim. Behav. 55:1685–1706.
Martins, E. P., J. A. F. Diniz-Filho, and E. A. Housworth. 2002. Adaptive constraints and the phylogenetic comparative method: a computer simulation test. Evolution 56:1–13.
Mayr, E. 1963. Animal species and evolution. The Belknap Press of Harvard Univ. Press, Cambridge, MA.
———. 1982. The growth of biological thought: diversity, evolution, and inheritance. Belknap Press, Cambridge, MA.
Mazer, S. J. 1998. Alternative approaches to the analysis of comparative data: compare and contrast. Am. J. Bot. 85:1194–1199.
McKitrick, M. C. 1993. Phylogenetic constraint in evolutionary theory: Has it any explanatory power? Annu. Rev. Ecol. Syst. 24:307–330.
Møller, A. P., S. Merino, C. R. Brown, and R. J. Robertson. 2001. Immune defense and host sociality: a comparative study of swallows and martins. Am. Nat. 158:136–145.
Mooers, A. Ø., S. M. Vamosi, and D. Schluter. 1999. Using phylogenies to test macroevolutionary hypotheses of trait evolution in cranes (Gruinae). Am. Nat. 154:249–259.
Morales, E. 2000. Estimating phylogenetic inertia in *Tithonia* (Asteracea): a comparative approach. Evolution 54:475–484.
Mousseau, T. A., and D. A. Roff. 1987. Natural selection and the heritability of fitness components. Heredity 59:181–197.
Nunn, C. L., and R. A. Barton. 2000. Allometric slopes and independent contrasts: a comparative test of Kleiber's Law in primate ranging patterns. Am. Nat. 156:519–533.
Orzack, S. H., and E. Sober. 2001. Adaptation, phylogenetic inertia, and the method of controlled comparisons. Pp. 45–63 *in* S. H. Orzack and E. Sober, eds. Adaptationism and optimality. Cambridge Univ. Press, Cambridge, U.K.
Pagel, M. D. 1992. A method for the analysis of comparative data. J. Theor. Biol. 156:431–442.
———. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. R. Soc. Lond. B 255:37–45.
———. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.
Perry, G., and T. Garland Jr. 2002. Lizard home ranges revisited: effects of sex, body size, diet, habitat, and phylogeny. Ecology 83:1870–1885.
Pierce, V. A., and D. L. Crawford. 1997. Phylogenetic analysis of glycolytic enzyme expression. Science 276:256–259.
Pitnick, S. 1996. Investment in testes and the cost of making long sperm in *Drosophila* species. Am. Nat. 148:57–80.
Plomin, R. 1990. The role of inheritance in behavior. Science 248: 183–188.
Plotkin, H. C., ed. 1988. The role of behavior in evolution. MIT Press, Cambridge, MA.
Price, T. 1997. Correlated evolution and independent contrasts. Philos. Trans. R. Soc. Lond. B 352:519–529.
Prinzing, A., W. Durka, S. Klotz, and R. Brandl. 2001. The niche of higher plants: evidence for phylogenetic conservatism. Proc. R. Soc. London Ser. B 268:2383–2389.
Purvis, A. 1995. A composite estimate of primate phylogeny. Philos. Trans. R. Soc. Lond. B 348:405–421.
Purvis, A., and T. Garland Jr. 1993. Polytomies in comparative analyses of continuous characters. Syst. Biol. 42:569–575.
Purvis, A., and A. Rambaut. 1995. Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analysing comparative data. Computer Appl. Biosci. 11: 247–251.
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.
Reeve, H. K., and P. W. Sherman. 2001. Optimality and phylogeny: a critique of current thought. Pp. 45–63 *in* S. H. Orzack and E. Sober, eds. Adaptationism and optimality. Cambridge Univ. Press, Cambridge, U.K.
Reynolds, P. S., and R. M. Lee III. 1996. Phylogenetic analysis of avian energetics: passerines and nonpasserines do not differ. Am. Nat. 147:735–759.
Rezende, E. L., D. L. Swanson, F. F. Novoa, and F. Bozinovic. 2002. Passerines versus nonpasserines: so far, no statistical differences in the scaling of avian energetics. J. Exp. Biol. 205: 101–107.
Ridley, M. 1983. The explanation of organic diversity: the comparative method and adaptations for mating. Clarendon, Oxford, U.K.
Roff, D. A. 1992. The evolution of life histories: theory and analysis. Routledge, Chapman and Hall, New York.
Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. Evolution 55: 2143–2160.
Schluter, D., T. Price, A. Ø. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51: 1699–1711.
Schwenk, K., and G. P. Wagner. 2001. Function and the evolution of phenotypic stability: connecting pattern to process. Am. Zool. 41:552–563.
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.
Sibley, C. G., and J. E. Ahlquist. 1990. Phylogeny and classification of birds: a study in molecular evolution. Yale Univ. Press, New Haven, CT.
Smith, R. J., and J. M. Cheverud. 2002. Scaling of sexual size dimorphism in body mass: a phylogenetic analysis of Rensch's rule in primates. Int. J. Primatol. 23:1095–1135.
Stearns, S. C. 1983. The influence of size and phylogeny on patterns of covariation among life-history traits in mammals. Oikos 41: 173–187.
Stirling, D. G., D. Réale, and D. A. Roff. 2002. Selection, structure and the heritability of behaviour. J. Evol. Biol. 15:277–289.
Swiderski, D. L. 2001. The role of phylogenies in comparative biology: an introduction to the symposium. Am. Zool. 41: 485–486.
Székely, T., J. D. Reynolds, and J. Figuerola. 2000. Sexual size

dimorphism in shorebirds, gulls and alcids: the influence of sexual and natural selection. Evolution 54:1404–1413.
Taigen, T. L., and F. H. Pough. 1985. Metabolic correlates of anuran behavior. Am. Zool. 25:987–997.
Tobalske, B. W., and K. P. Dial. 2000. Effects of body size on takeoff flight performance in the Phasianidae (Aves). J. Exp. Biol. 203:3319–3332.
von Neumann, J., R. H. Kent, H. R. Bellinson, and B. I. Hart. 1941. The mean square successive difference. Ann. Math. Stat. 12: 153–162.
Wagner, G. P., and K. Schwenk. 2000. Evolutionarily stable configurations: functional integration and the evolution of phenotypic stability. Evol. Biol. 31:155–217.
Wake, D. 1991. Homoplasy: the result of natural selection, or evidence of design limitations?. Am. Nat. 138:543–567.
Wcislo, W. T. 1989. Behavioral environments and evolutionary change. Annu. Rev. Ecol. Syst. 20:137–169.
Westoby, M., M. Leishman, and J. Lord. 1995. On misinterpreting the phylogenetic correction. J. Ecol. 83:531–534.
Wiens, J. J., and J. L. Slingluff. 2001. How lizards turn into snakes: a phylogenetic analysis of body-form evolution in anguid lizards. Evolution 55:2303–2318.
Wimberger, P. H., and A. de Queiroz. 1996. Comparing behavioral and morphological characters as indicators of phylogeny. Pp. 206–233 *in* E. P. Martins, ed. Phylogenies and the comparative method in animal behavior. Oxford Univ. Press, New York.
Wolf, C. M., T. Garland Jr., and B. Griffith. 1998. Predictors of avian and mammalian translocation success: reanalysis with phylogenetically independent contrasts. Biol. Conserv. 86:243–255.

Corresponding Editor: B. Crespi

APPENDIX 1

List of computer programs in MatLab (and one in Xlisp-Stat). These are available on request from TG.

PHYSIG.DOC	Documentation as Microsoft Word file.
PHYSIG.M	Test for phylogenetic signal by randomization test; computes *K* and other MSE statistics. Allows transformation of tree prior to analyses, by user-specified value of OU or ACDC parameter. Uses ACDC.m, ACDCDetNE1.m, mse0ratio.m, mseStarratio.m, OU.m, OUDetNE1.m, PHYSIGfunct.m, and scalebydet.m, which must be in the same folder as PHYSIG.M.
PHYSIGOU.M	Maximum likelihood estimate of OU transformation parameter (*d*), randomization test for whether *d* differs significantly from zero, MSE statistics. Uses GetMSEOUfunct.m, mseratio.m, PHYSIGfunct.m, PHYSIGOUfunct.m, scalebydet.m, and tolerance.m.
PHYSIGACDC.M	Maximum likelihood estimate of ACDC parameter (*g*), randomization test for whether *g* differs significantly from zero, MSE statistics. Uses GetMSEACDCfunct.m, mseratio.m, PHYSIGACDCfunct.m, PHYSIGfunct.m, scalebydet.m, and tolerance.m.
PHYOUH0d.M	Tests null hypothesis that d = user-specified value (typically 1), via a randomization test. Uses GetMSEOUfunct.m, OU.m, PHYSIGfunct.m, PHYSIGOUfunct.m, and tolerance.m.
PHYACDCH0g.M	Tests null hypothesis that g = a user-specified value (typical 1), via a randomization test. Uses ACDC.m, GetMSEACDCH0gfunct.m, PHYSIGACDCfunct.m, and PHYSIGfunct.m.
PHYSIGER.M	Analyzes simulated data from PDSIMUL.EXE of the PDAP package to calculate Type I or Type II error rates of the randomization test that is implemented by PHYSIG.M. Uses PHYSIGfunct.m.
PHYSIGDG.M	Analyzes simulated data from PDSIMUL.EXE to estimate *d* and *g*, as well as computation of MSE statistics. Uses mseratio.m, PHYSIGACDCfunct.m, PHYSIGfunct.m, and PHYSIGOUfunct.m.
REGRESSION.M	Computes GLS (multiple) regressions.
TREECONV.LSP	Converts phylogenetic variance-covariance matrix to bracket format phylogenetic tree and vice versa. Particularly useful for visualizing results of OU and ACDC transformations (e.g., see Figs. 1, 8).

Appendix 2

Ornstein-Uhlenbeck Process

Suppose the evolution of a continuous trait x follows the discrete-time OU process

$$x(t) = dx(t - 1) + \gamma(t), \tag{A1}$$

where the time step 1 is arbitrary but is small compared to the time between speciation events. The parameter d measures the central tendency of the OU process, with a random walk occurring when $d = 1$ and $d = 0$ corresponding to a very strong attraction to the mean of zero. The random variable $\gamma(t)$ has mean zero and variance σ^2_γ, and represents evolutionary change per unit time. Starting at $x(0)$, the value of x after t time steps is

$$\begin{aligned} x(t) &= dx(t - 1) + \gamma(t) \\ &= d[dx(t - 2) + \gamma(t - 1)] + \gamma(t) \\ &= d^t x(0) + d^{t-1}\gamma(1) + d^{t-2}\gamma(2) + \cdots + d\gamma(t - 1) + \gamma(t). \end{aligned} \tag{A2}$$

Because $\gamma(t)$ are all independent and identically distributed with variance σ^2_γ, the variance in $x(t)$ is

$$V\{x(t)\} = d^{2t}V\{x(0)\} + \frac{1 - d^{2t}}{1 - d^2}\sigma^2_\gamma. \tag{A3}$$

This equation can be used to find the variance-covariance patterns for an OU process. Let the node-to-tip branch length for species i be τ_i, and let τ_{ij} denote the shared branch length between tips i and j. Because the base-to-tip branch length for species i is $\tau_{ij} + \tau_i$, the variance in X_i is

$$\begin{aligned} V\{X_i\} &= V\{x(\tau_{ij} + \tau_i)\} \\ &= d^{2(\tau_{ij}+\tau_i)}V\{x(0)\} + \frac{1 - d^{2(\tau_{ij}+\tau_i)}}{1 - d^2}\sigma^2_\gamma \\ &= \frac{1 - d^{2(\tau_{ij}+\tau_i)}}{1 - d^2}\sigma^2_\gamma. \end{aligned} \tag{A4}$$

Here, we have set the variance of the trait at the base of the tree, $V\{x(0)\}$, equal to zero, because the basal ancestral species is the fixed starting point of the phylogeny.

To calculate covariances between trait values for species at the tips of the phylogeny, let X_{ij} denote the value of a trait at the first common ancestor of species at tips i and j, and let ΔX_i and ΔX_j denote the changes in trait value between this node and tips i and j. Then the covariance between the trait values at tips i and j, X_i and X_j, is

$$\begin{aligned} \mathrm{cov}\{X_i, X_j\} &= E\{(X_{ij} + \Delta X_i)(X_{ij} + \Delta X_j)\} \\ &= V\{X_{ij}\} + E\{\Delta X_i X_{ij}\} + E\{\Delta X_j X_{ij}\} \\ &\quad + E\{\Delta X_i \Delta X_j\}. \end{aligned} \tag{A5}$$

In the case of Brownian motion evolution, the change in trait value from the value at the common node to the value at the tip is independent of the value at the common node; that is, the covariance between X_{ij} and ΔX_i is zero. Also under Brownian motion evolution, ΔX_i and ΔX_j are independent. Therefore, in the above expression $E\{\Delta X_i X_{ij}\} = E\{\Delta X_j X_{ij}\} = E\{\Delta X_i \Delta X_j\} = 0$. However, when evolution follows an OU process, these random variables are not independent. To illustrate this, suppose X_{ij} is a relatively large value, so the common ancestor of species i and j is far from the OU optimum of zero. Then the subsequent evolution of both species from this value is more likely to move toward the optimal value. Thus, X_{ij} and ΔX_i will be negatively correlated, whereas ΔX_i and ΔX_j will be positively correlated.

To calculate $E\{\Delta X_i X_{ij}\}$, from by equation (A2) ΔX_i can be written

$$\begin{aligned} \Delta X_i &= -X_{ij} + X_i \\ &= -X_{ij} + d^t X_{ij} + d^{t-1}\gamma(\tau_{ij} + 1) + \cdots + \gamma(\tau_{ij} + \tau_i). \end{aligned} \tag{A6}$$

Therefore,

$$E\{\Delta X_i X_{ij}\} = (d^{\tau_i} - 1)V\{X_{ij}\}. \tag{A7}$$

It follows in a similar manner that

$$E\{\Delta X_i \Delta X_j\} = (d^{\tau_i+\tau_j} - d^{\tau_i} - d^{\tau_j} + 1)V\{X_{ij}\}. \tag{A8}$$

Combining by equations (A4), by (A5), by (A7), and by (A8) yields

$$\mathrm{cov}\{X_i, X_j\} = d^{\tau_i+\tau_j}\frac{1 - d^{2\tau_{ij}}}{1 - d^2}\sigma^2_\gamma. \tag{A9}$$

The simulation of an OU process by PDSIMUL (Garland et al. 1993) is qualitatively similar to the process described above and leads to the same covariance structure of tip trait values. Furthermore, this derivation is similar to that in Hansen and Martins (1996), although they assume that at the time of speciation the OU process resets, so that the optimal trait value is the value of the common ancestor of the two lineages (see the appendix in Hansen 1997; Butler et al. 2000). In contrast, we assume that the optimal trait value of the OU process is the same for all species in the phylogeny.

Appendix 3

ACDC Process

Suppose the evolution of a continuous trait x follows the discrete-time variable-rate process

$$x(t) = x(t - 1) + \gamma(t), \tag{A10}$$

where the time step 1 is arbitrary but is small compared to the time between speciation. The variance of $\gamma(t)$ (i.e., the rate of evolution) changes through time, with

$$V\{\gamma(t)\} = \sigma^2_\gamma\, g^{-t}. \tag{A11}$$

Here, σ^2_γ is the variance of $\gamma(0)$ at the base of the phylogenetic tree, and g measures the rate at which evolution changes through time. If g is less than one, then evolution becomes more rapid with time, whereas g greater than one implies a slowing of evolution. Note that we assume the change in the rate of evolution changes with the absolute time since the initial phylogenetic split at the basal node. Therefore, at any point in time the rate of evolutionary change is the same for every lineage.

Starting at $x(0)$, the value of x after t time steps is

$$x(t) = x(0) + \sum_{s=1}^{t} \gamma(s). \tag{A12}$$

Because $\gamma(s)$ are all independent with variances given by by equation (A11), the variance in $x(t)$ is

$$\begin{aligned} V\{x(t)\} &= V\{x(0)\} + \sum_{s=1}^{t} \sigma^2_\gamma g^{-s} \\ &= V\{x(0)\} + \sigma^2_\gamma \frac{1 - g^{-t}}{1 - g^{-1}}. \end{aligned} \tag{A13}$$

This equation can be used to find the variance-covariance patterns for the ACDC evolutionary process. Let the node-to-tip branch length for species i be τ_i, and let τ_{ij} denote the shared branch length between tips i and j. Because the base-to-tip branch length for species i is $\tau_{ij} + \tau_i$, the variance in X_i is

$$V\{X_i\} = \frac{1 - g^{-(\tau_i+\tau_{ij})}}{1 - g^{-1}}\sigma^2_\gamma. \tag{A14}$$

Here, we have set the variance of the trait at the base of the tree, $V\{x(0)\}$, equal to zero, because the basal ancestral species is the fixed starting point of the phylogeny. To calculate covariances between trait values for species at the tips of the phylogeny, let X_{ij} denote the value of a trait at the first common ancestor of species at tips i and j, and let ΔX_i and ΔX_j denote the changes in trait value between this node and tips i and j. In contrast to the OU process, X_{ij}, ΔX_i, and ΔX_j are all independent. Therefore, from equation (A14),

$$\mathrm{cov}\{X_i, X_j\} = V\{X_{ij}\} = \frac{1 - g^{-\tau_{ij}}}{1 - g^{-1}}\sigma^2_\gamma. \tag{A15}$$

Appendix 4

Maximum Likelihood Estimation

The procedure we use for estimating a parameter θ in a branch-length transform (i.e., *d* in the OU and *g* in the ACDC transforms) is based on ML. Let $\mathbf{V}(\theta) = \sigma^2(\theta)\mathbf{C}(\theta)$ be the covariance matrix of the phylogenetic tree given by a transform with parameter θ, where the matrix $\mathbf{C}(\theta)$ governs the covariance structure of $\mathbf{V}(\theta)$, and $\sigma^2(\theta)$ scales the rate of evolutionary divergence of traits. For any value θ, the estimate of a trait *x* at the basal node, $\hat{a}(\theta)$, that is, the phylogenetically based sample mean, is

$$\hat{a}(\theta) = [\mathbf{1}'\ \mathbf{C}(\theta)^{-1}\mathbf{1}]^{-1}\ [\mathbf{1}'\ \mathbf{C}(\theta)^{-1}\ \mathbf{X}], \tag{A16}$$

where **1** denotes the vectors of ones of length *N*, where *N* is the number of tips on the phylogenetic tree. The ML estimate of θ is the value that minimizes the negative log-likelihood function

$$-\log L = \frac{N}{2}\log(2\pi) + \frac{1}{2}\log[|\mathbf{C}(\theta)|] + \frac{1}{2}[\mathbf{X} - \hat{a}(\theta)]\mathbf{C}(\theta)^{-1}[\mathbf{X} - \hat{a}(\theta)]. \tag{A17}$$

The corresponding estimate for $\sigma^2(\theta)$ is given by the MSE (Garland and Ives 2000):

$$\hat{\sigma}^2(\theta) = [\mathbf{X} - \hat{a}(\theta)]'\ \mathbf{C}(\theta)^{-1}\ [\mathbf{X} - \hat{a}(\theta)]/(N - 1). \tag{A18}$$

An instructive alternative formula for calculating the ML estimate of θ involves rescaling the matrix $\mathbf{C}(\theta)$ by its determinant to give

$$\tilde{\mathbf{C}}(\theta) = \frac{\mathbf{C}(\theta)}{\det[\mathbf{C}(\theta)]}. \tag{A19}$$

For this rescaling, the negative log-likelihood function is minimized by minimizing

$$\mathrm{MSE}(\theta) = (\mathbf{X} - \hat{a})'\ \tilde{\mathbf{C}}(\theta)^{-1}\ (\mathbf{X} - \hat{a})/(N - 1). \tag{A20}$$

Thus, the ML estimate of θ gives the minimum MSE for the suite of matrices $\tilde{\mathbf{C}}(\theta)$ that have determinant equaling one. by Equation (A20) is known as the estimated generalized least-squares (EGLS) estimator of θ (Judge et al. 1985). Thus, the ML and EGLS estimates of θ are the same.

Appendix 5

Estimation with Phylogenetically Independent Contrasts

In addition to estimating a parameter θ in a branch-length transform (i.e., *d* in the OU and *g* in the ACDC transforms) using ML (Appendix 4), it is possible to use the independent contrasts (IC) approach. Here we show that the ML estimate of θ is the value that minimizes the variances of the standardized contrasts and the phylogenetically based mean in IC, subject to a constraint placed on the average value of the branch lengths.

For a phylogenetic tree with *N* tip species, let $\mathbf{G}(\theta)$ be the $(N - 1) \times N$ matrix such that $\Delta\mathbf{X}(\theta) = \mathbf{G}(\theta)\mathbf{X}$ are standardized contrasts. Matrix $\mathbf{G}(\theta)$ is the same as the matrix $\mathbf{C}$ derived in Rohlf (2001) and should not to be confused with the definition of $\mathbf{C}$ we are using as giving the covariance structure of matrix $\mathbf{V}$, that is, $\mathbf{V}(\theta) = \sigma^2(\theta)\mathbf{C}(\theta)$. By construction, the standardized contrasts are independent and have variance equal to one. Therefore, $\Delta\mathbf{X}(\theta)\Delta\mathbf{X}'(\theta) = \mathbf{I}_{N-1}$, the $(N - 1) \times (N - 1)$ identity matrix. Following Rohlf (2001):

$$\Delta\mathbf{X}(\theta)\Delta\mathbf{X}'(\theta) = \mathbf{G}(\theta)\mathbf{X}[\mathbf{G}(\theta)\mathbf{X}]' = \mathbf{G}(\theta)\mathbf{X}\mathbf{X}'\mathbf{G}'(\theta) = \mathbf{G}(\theta)\mathbf{C}(\theta)\mathbf{G}'(\theta). \tag{A21}$$

This shows that $\mathbf{G}(\theta)\mathbf{C}(\theta)\mathbf{G}'(\theta) = \mathbf{I}_{N-1}$. With this we can write equation (A18) as

$$\begin{aligned}\hat{\sigma}^2(\theta) &= [\mathbf{X} - \hat{a}(\theta)]'\mathbf{C}(\theta)^{-1}[\mathbf{X} - \hat{a}(\theta)]/(N - 1)\\ &= [\mathbf{X} - \hat{a}(\theta)]'\mathbf{G}'(\theta)\mathbf{G}(\theta)[\mathbf{X} - \hat{a}(\theta)]/(N - 1)\\ &= \Delta\mathbf{X}'(\theta)\Delta\mathbf{X}(\theta)/(N - 1).\end{aligned} \tag{A22}$$

From Appendix 4, the estimated general least-squares estimator of θ is obtained by minimizing by equation (A17) subject to the constraint that $\det[\mathbf{C}(\theta)]$ is constant. It is possible to show that

$$\det[\mathbf{C}(\theta)]^{-1} = \left[\frac{\nu_1'(\theta)\nu_{1'}'(\theta)}{\nu_1'(\theta) + \nu_{1'}'(\theta)}\right]\prod_{i=1}^{N-1}[\nu_i'(\theta) + \nu_{i'}'(\theta)], \tag{A23}$$

where $\nu_1'(\theta)$ and $\nu_{1'}'(\theta)$ are the corrected branch lengths above the basal node, and $\nu_i'(\theta)$ and $\nu_{i'}'(\theta)$ are the corrected branch lengths above node *i* (i.e., $\nu_i'(\theta) + \nu_{i'}'(\theta)$ is the sum of branch lengths separating two species whose most recent common ancestor is at node *i*). From by equation (A20), the EGLS estimator of θ is

$$\begin{aligned}\mathrm{MSE}(\theta) &= [\mathbf{X}(\theta) - \hat{a}(\theta)]\tilde{\mathbf{C}}(\theta)^{-1}\frac{[\mathbf{X}(\theta) - \hat{a}(\theta)]}{(N - 1)}\\ &= \frac{1}{\det[\mathbf{C}(\theta)]^{-1}}\frac{\{[\mathbf{X}(\theta) - \hat{a}(\theta)]'\mathbf{C}(\theta)^{-1}[\mathbf{X}(\theta) - \hat{a}(\theta)]\}}{(N - 1)}\\ &= \frac{1}{N - 1}\left\{\frac{\Delta\mathbf{X}(\theta)}{\det[\mathbf{C}(\theta)]^{-1/2}}\right\}'\left\{\frac{\Delta\mathbf{X}(\theta)}{\det[\mathbf{C}(\theta)]^{-1/2}}\right\}.\end{aligned} \tag{A24}$$

Thus, the EGLS (and hence the ML) estimate of θ is the value that minimizes the mean sum of squares of the independent contrasts divided by the square root of the quantity

$$\left(\frac{\nu_1'\nu_{1'}'}{\nu_1' + \nu_{1'}'}\right)\prod_{i=1}^{N-1}(\nu_i' + \nu_{i'}'). \tag{A25}$$

APPENDIX 6

Descriptive statistics and tests for phylogenetic signal for 121 traits from 35 phylogenies. Trait type: S, body size; P, physiology; B, behavior; L, life history; M, morphology; E, ecology. *P* signal is from randomization test; *d* and *g* are, respectively, estimated Ornstein-Uhlenbeck model and acceleration-deceleration (ACDC) model transformation parameters. *P*-values of exactly 0 actually indicate $P < 0.001$; a indicates not reliable because of sensitivity to starting values or other numerical problems; b indicates not relevant because $d = 0$ or *g* is virtually 0 or because they are > 1. SVL, snout-vent length. FOG, fast oxidative glycolytic; FG, fast glycolytic.

Organism	Trait	Branch lengths	Tree no.	Trait type	*N*	*P* signal	*d*	*P* $d = 0$	*P* $d = 1$
Birds	log body mass	DNA hyb	1	S	254	0	0.996	0	0.525
Birds	log basal metabolic rate (mass corrected)	DNA hyb	1	P	254	0	0.835	0	0
Vertebrates	plasma osmolarity	all = 1	2	P	172	0	1.027	0	0
Lizards	log SVL	Grafen	3	S	108	0	0.967	0	0.005
Lizards	log home range (SVL corrected)	Grafen	3	B	108	0	0.878	0	0
Primates	log female body mass	time	4	S	105	0	1.001	0	0.625
Primates	log(male/female body mass)	time	4	L	105	0	0.961	0	0.028
Bats	log female body mass	Nee	5	S	101	0	0.925	0	0.08
Bats	log (male/female body mass)	Nee	5	L	101	0.089	0.304	0.031	0
Lizards	log SVL	all = 1	6	S	90	0	1.008	0	0.428
Lizards	log age at maturity (SVL corrected)	all = 1	6	L	90	0.001	0.015	0.161	0.017
Lizards	log clutch size (SVL corrected)	all = 1	6	L	90	0	1.037	0	0.175
Lizards	log brood frequency (SVL corrected)	all = 1	6	L	90	0	0.900	0.003	0.205
Lizards	log mortality (SVL corrected)	all = 1	6	L	90	0	1.001	0.001	0.223
Lizards	log fecundity (SVL corrected)	all = 1	6	L	90	0.008	0.024	0.014	0
Antelope	log body mass	time	7	S	75	0	1.011	0	0.001
Antelope	log group size (mass corrected)	time	7	B	75	0	0.955	0	0.016
Carnivora	log prey size (mass corrected)	time	8	B	59	0.02	0.856	0.01	0.001
Shore birds	log female body mass	all = 1	9	S	58	0	1.018	0	0.179
Shore birds	log (male/female body mass)	all = 1	9	L	58	0	0.878	0.004	0.121
Shore birds	log female wing length (mass corrected)	all = 1	9	M	58	0	1.020	0	0.145
Shore birds	log female bill length (mass corrected)	all = 1	9	M	58	0	1.025	0	0.049
Anseriformes	log brain size (mass corrected)	Pagel	10	M	55	0	0.969	0.017	0.581
Anseriformes	log body mass	Pagel	10	S	55	0	0.829	0.017	0.065
Carnivora	log female body mass	time	11	S	52	0	0.980	0	0.053
Carnivora	log litter size	time	11	L	52	0	0.972	0	0.07
Carnivora	log gestation length (mass corrected)	time	11	L	52	0	0.994	0	0.533
Carnivora	seasonality	time	11	E	52	0.027	0.686	0.001	0
Carnivora, ungulates	log body mass	time	12	S	49	0	0.991	0	0.635
Carnivora, ungulates	log home range (mass corrected)	time	12	B	49	0	0.827	0	0.135
Carnivora, ungulates	log sprint speed	time	12	P	49	0	1.006	0	0.456
Carnivora, ungulates	metatarsal/femur ratio	time	12	M	49	0	1.005	0	0.711
Carnivora, ungulates	log hindleg length (mass corrected)	time	12	M	49	0	0.956	0	0.011
Birds	log body mass	DNA hyb	13	S	47	0	0.994	0	0.095
Birds	log maximum metabolic rate (mass corrected)	DNA hyb	13	P	47	0.26	1.0×10^{-8}	0.402	0
Carnivora	log daily movement distance (mass corrected)	time	14	B	41	0.02	0.242	0.074	0
Pinus	predicted allometric rank	all = 1	15	M	32	0.008	0.782	0.005	0.266
Pinus	rhytidome allometric coefficient	all = 1	15	M	32	0.09	0.616	0.072	0.11
Primates	log body mass	time	16	S	29	0	1.002	0	0.79
Primates	log testis mass (mass corrected)	time	16	M	29	0	0.955	0	0.11
Macropod marsupials	log female body mass	all = 1	17	S	28	0	1.033	0	0.439
Macropod marsupials	log (male/female body mass) (mass corrected)	all = 1	17	L	28	0	1.033	0	0.407
Macropod marsupials	log female home range (mass corrected)	all = 1	17	B	28	0.033	0.684	0.184	0.773
Macropod marsupials	log density (mass corrected)	all = 1	17	B	28	0.005	1.014	0.015	0.757
Macropod marsupials	log rainfall (mass corrected)	all = 1	17	E	28	0.009	0.895	0.014	0.466
Macropod marsupials	log group size (mass corrected)	all = 1	17	B	28	0.161	0.617	0.064	0.037
Bats	log body mass	Pagel	18	S	27	0	0.975	0.004	0.28
Bats	log basal metabolic rate (mass corrected)	Pagel	18	P	27	0	0.867	0.002	
Hystricognath rodents	log body mass	Pagel	19	S	26	0	1.033	0	0.51
Hystricognath rodents	log group size	Pagel	19	B	26	0.015	0.744	0.022	0.112
Hystricognath rodents	plant cover (not transformed)	Pagel	19	B	26	0	1.046	0	0.194
Anguidae	log SVL	DNA seq	20	S	23	0	1.043	0	0.333

APPENDIX 6. Extended.

g	P $g = 0$	P $g = 1$	Expected MSE_0/ MSE	Observed MSE_0/ MSE	K	MSE star	MSE candidate	MSE OU	MSE ACDC	Source
1.449	0	0.7	3.54	5.89	1.66	0.8165	0.1592	0.1585	0.1588	Reynolds and Lee 1996
0.001	0	0	3.54	1.26	0.36	0.0174	0.0184	0.0123	0.0141	Reynolds and Lee 1996
590.0	0	0.015	4.62	5.29	1.14	48695	19056	16307	10044	Garland et al. 1997
7.6×10^{-6}	0	0.011	10.2	2.02	0.20	1.8495	0.9470	0.7000	0.7643	Perry and Garland 2002
1.4×10^{-9}	0	0	10.2	1.16	0.11	0.8043	0.6976	0.4809	0.4976	Perry and Garland 2002
3.271	0	0.415	6.96	10.79	1.549	2.358	0.3074	0.3042	0.3029	Smith and Cheverud 2002
0.001	0	0.002	6.96	2.19	0.315	0.0544	0.0304	0.0278	0.0247	Smith and Cheverud 2002
0.141	0	0	2.04	1.42	0.69	0.2231	0.1579	0.1519	0.1524	Hutcheon 2001
7.0×10^{-5}	0.039	0	2.04	0.78	0.38	0.0041	0.0055	0.0041	0.0040	Hutcheon 2001
2.541	0	0.278	2.77	1.98	0.71	0.0589	0.0322	0.0322	0.0316	Clobert et al. 1998
0.000	b	b	2.77	0.93	0.34	0.0573	0.0619	0.0571	0.0573	Clobert et al. 1998
5.563	0	0.07	2.77	1.40	0.51	0.0927	0.0664	0.0646	0.0635	Clobert et al. 1998
1.193	0	0.97	2.77	1.09	0.39	0.0973	0.0935	0.0907	0.0934	Clobert et al. 1998
1.433	0	0.7	2.77	1.22	0.44	0.0555	0.0480	0.0480	0.0479	Clobert et al. 1998
0.000	b	b	2.77	0.94	0.34	0.1064	0.1222	0.1024	0.1064	Clobert et al. 1998
0.934	0	0.94	4.37	5.39	1.23	0.3134	0.0641	0.0559	0.0641	Brashares et al. 2000
0.019	0	0.09	4.37	2.17	0.50	0.1646	0.0790	0.0682	0.0722	Brashares et al. 2000
2.8×10^{-20}	0.024	0	7.34	0.67	0.09	0.6602	1.1602	0.6285	0.6254	Harris and Steudel 1997
9.154	0	0.001	3.85	2.89	0.75	0.1891	0.1262	0.1216	0.1127	Székely et al. 2000
0.285	0	0.017	3.85	1.51	0.39	0.0053	0.0048	0.0044	0.0045	Székely et al. 2000
17.94	0	0	3.85	1.84	0.48	0.0068	0.0037	0.0035	0.0030	Székely et al. 2000
14.87	0	0	3.85	8.55	2.22	0.0375	0.0235	0.0216	0.0191	Székely et al. 2000
0.002	0	0	2.86	1.12	0.39	0.0032	0.0029	0.0027	0.0024	Iwaniuk and Nelson 2001
0.151	0	0.325	2.86	2.35	0.82	0.1290	0.0599	0.0591	0.0575	Iwaniuk and Nelson 2001
0.091	0	0.23	3.65	2.84	0.78	1.1062	0.4053	0.3626	0.3830	Ferguson and Lariviere 2002
0.056	0	0.185	3.65	2.50	0.68	0.0848	0.0342	0.0256	0.0314	Ferguson and Lariviere 2002
1.052	0	0.983	3.65	5.39	1.48	0.0722	0.0134	0.0129	0.0134	Ferguson and Lariviere 2002
2.3×10^{-19}	0.037	0	3.65	0.62	0.17	0.0855	0.1473	0.0754	0.0810	Ferguson and Lariviere 2002
0.327	0	0.37	3.72	2.13	0.57	0.5040	0.2415	0.2397	0.2372	Garland et al. 1993
0.008	0	0.015	3.72	2.17	0.58	0.8480	0.6490	0.6086	0.5618	Garland et al. 1993
0.200	0	0.283	3.72	2.14	0.57	0.0288	0.0151	0.0148	0.0147	Garland and Janis 1993
0.659	0	0.842	3.72	5.08	1.37	0.0989	0.0230	0.0227	0.0230	Garland and Janis 1993
0.054	0	0.054	3.72	1.89	0.51	18.03	10.95	9.45	9.96	Garland and Janis 1993
2.324	0	0.585	2.67	4.48	1.68	0.3104	0.1198	0.1064	0.1181	Rezende et al. 2002
1.0×10^{-30}	0.16	0	2.67	0.49	0.18	0.0039	0.0080	0.0039	0.0038	Rezende et al. 2002
2.2×10^{-12}	0.038	0	6.07	0.68	0.11	0.1582	0.2323	0.1556	0.1453	Harris and Steudel 1997
0.011	0	0.02	2.11	1.26	0.60	4.1573	3.9951	3.3533	2.9175	Jackson et al. 1999
0.029	0.027	0.01	2.11	1.29	0.61	0.0136	0.0164	0.0131	0.0130	Jackson et al. 1999
2.125	0	0.547	3.52	5.35	1.52	0.3723	0.1052	0.1049	0.1036	Harcourt et al. 1981; Harvey and Harcourt 1984; Purvis 1995
0.001	0	0.002	3.52	1.34	0.38	0.1254	0.1052	0.0860	0.0714	Harcourt et al. 1981; Harvey and Harcourt 1984; Purvis 1995
2.931	0	0.392	2.59	10.44	4.03	0.1894	0.0676	0.0650	0.0650	Fisher and Owens 2000
1.898	0	0.645	2.59	9.47	3.65	0.2265	0.0794	0.0765	0.0783	Fisher and Owens 2000
0.133	0.039	0.545	2.59	1.32	0.51	0.3825	0.3984	0.3767	0.3677	Fisher and Owens 2000
2.050	0.022	0.491	2.59	1.68	0.65	0.2400	0.2182	0.2176	0.2148	Fisher and Owens 2000
0.420	0.053	0.418	2.59	2.84	1.10	0.1071	0.0981	0.0954	0.0958	Fisher and Owens 2000
0.012	0.046	0.058	2.59	0.82	0.32	0.0544	0.0668	0.0513	0.0451	Fisher and Owens 2000
0.061	0	0.165	2.10	1.41	0.67	0.0805	0.0572	0.0567	0.0532	Cruz-Neto et al. 2001
0.017	0.002	0.095	2.10	1.24	0.59	0.0097	0.0086	0.0081	0.0076	Cruz-Neto et al. 2001
1.200	0	0.934	1.85	1.91	1.03	0.4819	0.2790	0.2731	0.2789	Ebensperger and Cofre 2001
0.007	0.01	0.045	1.85	1.00	0.54	0.1791	0.1824	0.1592	0.1476	Ebensperger and Cofre 2001
2.902	0	0.605	1.85	2.33	1.26	4.4435	1.9136	1.8005	1.8870	Ebensperger and Cofre 2001
0.452	0	0.26	1.75	1.81	1.03	0.0379	0.0210	0.0203	0.0208	Wiens and Slingluff 2001

APPENDIX 6. Continued.

Organism	Trait	Branch lengths	Tree no.	Trait type	*N*	P signal	*d*	*P* *d* = 0	*P* *d* = 1
Anguidae	log tail length (mass corrected)	DNA seq	20	M	23	0	1.034	0	0.51
Maples	mature height	DNA seq	21	S	17	0.691	0.000	b	
Maples	seed size	DNA seq	21	L	17	0.275	0.231	0.129	
Maples	leaf + petiole length	DNA seq	21	M	17	0.217	0.450	0.021	
Maples	inflorescence + peduncle length	DNA seq	21	M	17	0.123	0.668	0.027	
Maples	petiole length	DNA seq	21	M	17	0.915	0.000	b	
Maples	twig-cross-sectional area	DNA seq	21	M	17	0.192	0.334	0.05	
Maples	inflorescence length	DNA seq	21	M	17	0.005	0.753	0.003	
Maples	leaf pairs per shoot	DNA seq	21	M	17	0.041	1.060	a	
Maples	individual leaf area	DNA seq	21	M	17	0.559	0.204	0.105	
Maples	shoot leaf area	DNA seq	21	M	17	0.379	0.140	0.11	
Maples	bifurcation angle	DNA seq	21	M	17	0.001	1.005	0	
Maples	leader dominance	DNA seq	21	M	17	0	0.923	0.001	
Shore birds	femur length	time	22	M	16	0.002	1.013	0.016	
Shore birds	stride frequency	time	22	P	16	0.032	1.028	0.061	
Drosophila	shill-coma temperature	DNA seq	23	P	15	0.457	0.000	b	
Drosophila	latitude	DNA seq	23	E	15	0.023	0.877	0.037	
Drosophila	wing length	DNA seq	23	M	15	0.409	0.000	b	
Fundulus fish	mean annual temperature	speciations	24	E	15	0.258	0.029	0.395	
Fundulus fish	*Gapdh* enzyme activity	speciations	24	P	15	0.006	1.057	0	
Fundulus fish	*Tpi* enzyme activity	speciations	24	P	15	0	1.045	0	
Salamanders	*C*-value	time	25	P	15	0	1.030	0	
Salamanders	regenerative growth rate	time	25	P	15	0.092	0.000	b	
Salamanders	differentiation rate	time	25	P	15	0.006	0.995	0.001	
Anolis	log body mass	time	26	S	14	0.049	0.840	0.063	
Anolis	log hindleg length (mass corrected)	time	26	M	14	0.164	0.000	b	
Anolis	log sprint speed (mass corrected)	time	26	P	14	0.121	0.000	b	
Swallows	log mean colony size	Grafen	27	B	13	0.372	0.000	b	
Swallows	nestling period	Grafen	27	L	13	0.726	0.000	b	
Swallows	log body mass	Grafen	27	S	13	0.346	0.000	b	
Swallows	wing web swelling nestlings	Grafen	27	P	13	0.208	0.499	0.094	
Skinks	preferred body temperature	time	28	B	12	0	1.005	0.001	
Skinks	optimal temperature for sprinting	time	28	P	12	0.167	0.000	b	
Skinks	critical thermal minimum	time	28	P	12	0.005	0.964	0.021	
Skinks	critical thermal maximum	time	28	P	12	0	1.008	0	
Finches	wing length	time	29	M	11	0.325	0.000	b	
Drosophilia	dry body mass	time	30	S	11	0.085	0.903	0.168	
Drosophila	relative testis mass	time	30	M	11	0.027	1.109	0.031	
Drosophila	sperm length	time	30	M	11	0.135	1.089	0.211	
Drosophila	sperm number	time	30	P	11	0.463	0.000	b	
Phrynosomatidae	proportion of FOG fibers	Pagel	31	M	11	0.002	1.092	0	
Phrynosomatidae	proportion of FG fibers	Pagel	31	M	11	0	1.130	0	
Phrynosomatidae	log forelimb span (mass corrected)	Pagel	31	M	11	0.362	0.000	b	
Phrynosomatidae	log thigh muscle cross-sectional area (mass corrected)	Pagel	31	M	11	0.012	0.954	0.118	
Phrynosomatidae	log hindlimb span (mass corrected)	Pagel	31	M	11	0.044	0.864	0.079	
Phrynosomatidae	log body mass	Pagel	31	S	11	0.316	0.000	b	
Phrynosomatidae	log SVL (mass corrected)	Pagel	31	M	11	0.017	0.905	0.064	
Tithonia	seed size	all = 1	32	L	11	0.011	0.827	0.022	
Tithonia	flower size	all = 1	32	M	11	0.328	0.000	b	
Tithonia	leaf size	all = 1	32	M	11	0.595	0.236	0.287	
Tithonia	head size	all = 1	32	M	11	0.607	0.163	0.337	
Tithonia	flowers per head	all = 1	32	M	11	0.015	0.803	0.017	
Tithonia	seeds per head	all = 1	32	L	11	0.115	0.000	b	
Tithonia	seedling height	all = 1	32	L	11	0.022	1.131	0.031	
Tithonia	growth rate	all = 1	32	P	11	0.15	1.121	0.205	
Tithonia	germination time	all = 1	32	P	11	0.507	0.000	b	
Tithonia	establishment	all = 1	32	L	11	0.664	0.149	0.079	
Tithonia	viability	all = 1	32	L	11	0.018	0.880	0.041	
Tithonia	germination	all = 1	32	L	11	0.741	0.162	0.316	
Tithonia	resource allocation	all = 1	32	L	11	0	1.129	0	
Tithonia	adult height	all = 1	32	S	11	0.652	0.000	b	
Cyclura	total display duration	DNA seq	33	B	9	0.419	0.000	b	
Cyclura	head bob duration	DNA seq	33	B	9	0.925	0.000	b	
Cyclura	pause duration	DNA seq	33	B	9	0.253	0.000	b	
Cyclura	number of bouts	DNA seq	33	B	9	0.098	1.017	0.112	
Cyclura	number of bobs	DNA seq	33	B	9	0.036	0.998	0.061	
Procyonidae	log body mass	time	34	S	6	0.057	0.998	0.073	
Phasianidae	log body mass	time	35	S	4		0.511	0.358	
Phasianidae	log wingbeat frequency (mass corrected)	time	35	P	4		0.912	0.189	

APPENDIX 6. Continued, extended.

g	P $g = 0$	P $g = 1$	Expected MSE_0/ MSE	Observed MSE_0/ MSE	K	MSE star	MSE candidate	MSE OU	MSE ACDC	Source
128.1	0	0	1.75	2.32	1.33	0.0327	0.0153	0.0146	0.0121	Wiens and Slingluff 2001
0.877	0.819		1.89	0.72	0.38	0.2167	0.3501	0.2167	0.3499	Ackerly and Donoghue 1998
0.059	0.091		1.89	0.76	0.40	0.5657	0.7462	0.5411	0.5630	Ackerly and Donoghue 1998
0.070	0.066		1.89	0.81	0.43	0.1520	0.1929	0.1269	0.1415	Ackerly and Donoghue 1998
0.012	0.015		1.89	1.06	0.56	0.1231	0.1428	0.1023	0.0889	Ackerly and Donoghue 1998
1.0×10^{-30}	0.131		1.89	0.54	0.29	0.3120	0.5805	0.3120	0.5270	Ackerly and Donoghue 1998
1.0×10^{-30}	0.118		1.89	0.84	0.44	0.5724	0.7001	0.5120	0.5724	Ackerly and Donoghue 1998
0.152	0.002		1.89	1.51	0.80	0.3482	0.2792	0.2106	0.2330	Ackerly and Donoghue 1998
33.877	0.169		1.89	1.09	0.57	0.5639	0.5826	0.5149	0.4225	Ackerly and Donoghue 1998
1.0×10^{-30}	0.078		1.89	0.67	0.35	0.5788	0.8886	0.5501	0.5788	Ackerly and Donoghue 1998
1.0×10^{-30}	0.196		1.89	0.71	0.38	0.8997	1.2623	0.8637	0.8997	Ackerly and Donoghue 1998
2.247	0		1.89	2.75	1.45	152.85	72.79	72.78	70.93	Ackerly and Donoghue 1998
0.950	0.001		1.89	2.76	1.46	0.0181	0.0085	0.0082	0.0085	Ackerly and Donoghue 1998
0.021	0.002		1.71	1.38	0.81	73.29	56.30	55.58	52.37	Barbosa and Moreno 1999
0.006	0.074		1.71	1.01	0.59	0.2523	0.2490	0.2382	0.2264	Barbosa and Moreno 1999
0.000	a		1.61	0.67	0.42	1.8108	2.9523	1.8108	1.8108	Gibert and Huey 2001
0.145	0.163		1.61	1.17	0.73	297.29	282.65	269.69	268.66	Gibert and Huey 2001
0.000	a		1.61	0.65	0.40	571.85	887.01	571.85	571.85	Gibert and Huey 2001
0.000	a		2.71	0.88	0.32	30.94	43.13	30.40	30.94	Pierce and Crawford 1997
2.965	0.057		2.71	2.00	0.74	52.48	34.26	18.89	33.87	Pierce and Crawford 1997
8.536	0.006		2.71	2.85	1.05	11986	4436	3892	4045	Pierce and Crawford 1997
4.285	0		1.68	2.05	1.22	98.361	48.403	41.853	47.594	Sessions and Larson 1987
1.1×10^{-4}	0.063		1.68	0.87	0.52	2.7641	3.2169	2.7641	2.4550	Sessions and Larson 1987
0.027	0.006		1.68	1.52	0.90	2.9627	2.3101	2.3079	2.1449	Sessions and Larson 1987
0.028	0.091		1.58	1.45	0.92	0.2028	0.2041	0.1931	0.1912	Losos 1990
0.003	0.116		1.58	0.91	0.58	0.0016	0.0017	0.0016	0.0014	Losos 1990
0.001	0.155		1.58	1.16	0.74	0.0014	0.0016	0.0014	0.0014	Losos 1990
2.6×10^{-18}	0.492		2.05	0.53	0.26	1.0016	1.8975	1.0016	1.0049	Møller et al. 2001
1.0×10^{-30}	0.466		2.05	0.43	0.21	0.0016	0.0039	0.0016	0.0016	Møller et al. 2001
1.0×10^{-30}	0.377		2.05	0.59	0.29	0.0119	0.0208	0.0119	0.0119	Møller et al. 2001
1.0×10^{-30}	0.437		2.05	0.67	0.33	0.0575	0.0879	0.0560	0.0575	Møller et al. 2001
0.007	0.001		4.95	2.24	0.45	20.318	10.184	9.446	8.667	Garland et al. 1991
8.8×10^{-18}	0.109		4.95	0.50	0.10	3.8282	8.3739	3.8282	3.5663	Garland et al. 1991
1.1×10^{-5}	0.008		4.95	1.22	0.25	10.752	10.041	9.123	6.775	Garland et al. 1991
0.006	0		4.95	2.55	0.51	5.7239	2.3777	1.6605	2.0410	Garland et al. 1991
1.4×10^{-21}	0.434		1.48	0.71	0.48	18.840	27.784	18.840	18.839	Schluter et al. 1997
0.184	0.127		1.14	1.12	0.98	21177	20210	19929	19807	Pitnick 1996
1.381	0.027		1.14	1.31	1.15	8.7720	6.7291	3.5356	6.7244	Pitnick 1996
0.077	0.175		1.14	1.04	0.91	276.5	269.2	247.8	264.2	Pitnick 1996
1.4×10^{-21}	0.51		1.14	0.76	0.66	957654	1263129	957654	957661	Pitnick 1996
41.18	0		1.30	2.12	1.63	0.0190	0.0090	0.0078	0.0068	Bonine et al. 2001
3353	0		1.30	2.57	1.98	0.0255	0.0099	0.0041	0.0035	Bonine et al. 2001
0.001	0.999		1.30	0.73	0.56	0.0008	0.0011	0.0008	0.0008	Bonine et al. 2001
0.686	0.01		1.30	1.60	1.23	0.0208	0.0130	0.0128	0.0130	Bonine et al. 2001
0.117	0.035		1.30	1.11	0.86	0.0041	0.0037	0.0035	0.0035	Bonine et al. 2001
0.001	0.999		1.30	0.74	0.57	0.1527	0.2079	0.1527	0.1551	Bonine et al. 2001
0.290	0.031		1.30	1.27	0.98	0.0010	0.0008	0.0008	0.0008	Bonine et al. 2001
0.315	0.032		1.48	1.80	1.22	1.5708	1.1589	1.0524	1.0938	Morales 2000
4.620	0.717		1.48	0.87	0.59	3.0719	3.9850	3.0719	3.7020	Morales 2000
0.003	0.304		1.48	0.71	0.48	1443	2092	1393	1194	Morales 2000
0.007	0.435		1.48	0.70	0.47	0.4981	0.7189	0.4846	0.4366	Morales 2000
0.101	0.027		1.48	1.52	1.03	614.8	506.7	477.2	434.7	Morales 2000
0.707	0.366		1.48	1.56	1.05	385.7	412.3	385.7	411.2	Morales 2000
0.000	0.057		1.48	1.44	0.97	16.22	13.65	11.85	8.12	Morales 2000
7.738	0.229		1.48	0.92	0.62	0.0212	0.0234	0.0206	0.0214	Morales 2000
0.219	0.681		1.48	0.92	0.63	30.88	4.04	30.88	41.83	Morales 2000
0.000	0.099		1.48	0.76	0.51	337.1	522.8	330.9	125.3	Morales 2000
0.001	0.004		1.48	1.88	1.27	919.2	660.0	636.2	475.4	Morales 2000
0.000	0.123		1.48	0.73	0.49	589.8	938.5	577.4	301.3	Morales 2000
1.263	0.001		1.48	3.18	2.15	1293.9	487.7	317.7	487.5	Morales 2000
13.60	0.737		1.48	0.78	0.53	1.6085	2.4524	1.6085	2.0733	Morales 2000
0.003	0.977		1.64	0.53	0.32	0.9251	1.9218	0.9251	1.4589	Martins and Lamont 1998
0.001	0.999		1.64	0.14	0.08	0.0826	0.5956	0.0826	0.3992	Martins and Lamont 1998
0.266	0.939		1.64	0.73	0.44	0.0500	0.0684	0.0500	0.0668	Martins and Lamont 1998
0.017	0.037		1.64	1.20	0.73	0.6622	0.6394	0.6161	0.5308	Martins and Lamont 1998
0.000	0.065		1.64	1.30	0.79	4.9866	3.8667	3.8661	3.0159	Martins and Lamont 1998
0.004	0.064		3.68	2.59	0.70	0.1013	0.0462	0.0295	0.0323	Chevalier 1991; Garland and Adolph 1994
0.012	0.521		1.02	0.91	0.89	0.3611	0.4068	0.3550	0.3587	Tobalske and Dial 2000
0.634	0.158		1.02	1.18	1.15	0.0017	0.0015	0.0015	0.0015	Tobalske and Dial 2000