Syst. Biol. 47(4):654–672, 1998

Effects of Branch Length Errors on the Performance of Phylogenetically Independent Contrasts

RAMÓN DÍAZ-URIARTE AND THEODORE GARLAND JR.

Department of Zoology, 430 Lincoln Drive, University of Wisconsin, Madison, Wisconsin 53706-1381, USA; E-mail: tgarland@macc.wisc.edu

Abstract.— We examined Type I error rates of Felsenstein's (1985; Am. Nat. 125:1–15) comparative method of phylogenetically independent contrasts when branch lengths are in error and the model of evolution is not Brownian motion. We used seven evolutionary models, six of which depart strongly from Brownian motion, to simulate the evolution of two continuously valued characters along two different phylogenies (15 and 49 species). First, we examined the performance of independent contrasts when branch lengths are distorted systematically, for example, by taking the square root of each branch segment. These distortions often caused inflated Type I error rates, but performance was almost always restored when branch length transformations were used. Next, we investigated effects of random errors in branch lengths. After the data were simulated, we added errors to the branch lengths and then used the altered phylogenies to estimate character correlations. Errors in the branches could be of two types: fixed, where branch lengths are either shortened or lengthened by a fixed fraction; or variable, where the error is a normal variate with mean zero and the variance is scaled to the length of the branch (so that expected error relative to branch length is constant for the whole tree). Thus, the error added is unrelated to the microevolutionary model. Without branch length checks and transformations, independent contrasts tended to yield extremely inflated and highly variable Type I error rates. Type I error rates were reduced, however, when branch lengths were checked and transformed as proposed by Garland et al. (1992; Syst. Biol. 41:18–32), and almost never exceeded twice the nominal P-value at $\alpha = 0.05$. Our results also indicate that, if branch length transformations are applied, then the appropriate degrees of freedom for testing the significance of a correlation coefficient should, in general, be reduced to account for estimation of the best branch length transformation. These results extend those reported in Díaz-Uriarte and Garland (1996; Syst. Biol. 45:27–47), and show that, even with errors in branch lengths and evolutionary models different from Brownian motion, independent contrasts are a robust method for testing hypotheses of correlated evolution. [Branch lengths; Brownian motion; continuous characters; independent contrasts; Ornstein–Uhlenbeck model; simulations; speciational model; Type I error.]

Felsenstein's (1985) method of phylogenetically independent contrasts (IC) is the most widely used method for analysis of comparative data (e.g., Miles and Dunham, 1993; Garland and Adolph, 1994; Martins and Hansen, 1996; Ricklefs and Starck, 1996; Price, 1997; Garland et al., in press). The method is conceptually simple and easy to apply and generally performs as well as, or better than, alternative methods (Grafen, 1989; Harvey and Pagel, 1991; Martins and Garland, 1991; Pagel, 1993; Purvis et al., 1994; Martins, 1996a; Martins and Hansen, 1996). Using computer simulations, we have shown previously that IC can exhibit broad-sense validity even under extreme deviations from a Brownian motion model of character evolution (Díaz-Uriarte and Garland, 1996). Nevertheless, our knowledge of the robustness of IC methods is still limited (Ricklefs and Starck, 1996; Price, 1997).

The four main assumptions of IC are (Díaz-Uriarte and Garland, 1996; Martins and Hansen, 1996): (1) a correct phylogenetic topology is available; (2) branch lengths of the phylogeny are available in units of (or proportional to) expected variance of character evolution; (3) character evolution can be modeled by a Brownian motion process; and (4) within-species variation is negligible or does not exist. If these assumptions are true, and the resulting IC meet the assumptions of the statistical test being applied, then IC yield nominal Type I error rates (low probability of rejecting the null hypothesis when it is true). When assumptions are violated, however, inflated Type I error rates are common (see Díaz-Uriarte and Garland [1996] and references therein), which results in the rejection of the null hypothesis more frequently than specified by the nominal P-value. Type I error rates are

not the only, or necessarily the best, criterion to judge the performance of a phylogenetically based statistical method, but they represent an important starting point, as comparative studies are frequently used to test hypotheses about correlated evolution (e.g., Miles and Dunham, 1993; papers in Martins, 1996b; Ricklefs and Starck, 1996; Price, 1997).

The importance of the first and last assumptions is not well understood. Several approaches have been suggested to deal with problems of topological uncertainty. For example, polytomies can be retained and degrees of freedom bounded (Purvis and Garland, 1993; Garland and Díaz-Uriarte, in press), analyses can be carried on the most plausible of the topologies (e.g., Bauwens and Díaz-Uriarte, 1997), or the analyses can be repeated on subsets of randomly generated topologies (Losos, 1994; Abouheif, 1998). Martins (1994; see also Martins and Hansen, 1997; Martins and Lamont, 1998) has suggested ways of incorporating information on within-species variation in the analysis of comparative data.

Misspecifications of branch lengths and of evolutionary models have the net effect of simultaneously violating the second and third assumptions of IC. As discussed previously (Díaz-Uriarte and Garland, 1996:28), however, even if the assumptions of IC are to some extent unrealistic, this does not preclude application of the method to real data. Rather, we need to know how violation of assumptions affects inference. Our previous work focused on the third assumption listed (Brownian motion), but we noted that the second and third assumptions are generally closely related (Díaz-Uriarte and Garland, 1996:28–29, 45). Here we investigate the performance of IC when the true evolutionary model is unknown and branch lengths of the phylogeny contain error. The errors in branch lengths investigated here are unrelated to the microevolutionary processes. The main types of errors examined in this work are random errors in branch lengths, which can only be avoided with infinite amounts of data (e.g., Swofford et al., 1996). Because these errors arise from multiple sources, such as dating of fossil ages, gaps in the fossil record, and cloning of DNA sequences, they are likely to affect virtually every comparative study. These errors are qualitatively different from errors in branch lengths caused by assuming the wrong evolutionary model (for example, assuming Brownian motion when the true model is Ornstein–Uhlenbeck; see also Díaz-Uriarte and Garland, 1996; Martins, 1994).

METHODS

Overview

We examined the performance of IC with branch length checks and transformations (ICblt and ICblte; see Table 2 and below) when branch lengths of the phylogeny are incorrectly specified. We examined effects of both systematic (i.e., all branch lengths were a constant, nonrandom function of the original ones used for simulation of data) and random branch length errors. To examine the effects of systematic errors, we simulated Brownian motion evolution on phylogenies where branch lengths were either raised to some power of the original ones or subjected to a rho transformation (Grafen, 1989) of the original branch lengths. We then used IC to estimate correlation coefficients between the two traits by using a phylogeny with the original (incorrect) branch lengths.

Our main focus was random errors in branch lengths. We simulated independent evolution of two traits along a phylogeny, using Brownian motion as well as several other evolutionary models. After the tip data were simulated, we added error (either a fixed amount or a variable, normally distributed amount) to the branch lengths of the phylogeny. The phylogenies with altered branch lengths were then used to estimate correlation coefficients by independent contrasts.

In both cases (systematic and random errors) our main focus was Type I error rates, which in the present case is the probability of rejecting the true null hypothesis of no correlation between the two traits. We also examined the bias of the estimated correlation coefficients. The methods of analysis for the systematic and random errors are equivalent.

The present work is an extension of Díaz-Uriarte and Garland (1996, hereafter abbreviated as D-U&G), and many of the methods used are similar to the ones in that paper. Therefore, we focus here on the methods and analyses that are particular to this paper, and refer the reader to our previous paper for more details on common features.

Models of Character Evolution

We simulated the evolution of two continuously valued characters along two different phylogenies, one for 49 species of Carnivora and ungulates (Garland et al., 1993: their Fig. 1), hereafter $Tree_{49}$, and another for 15 species of salamanders (Sessions and Larson, 1987; as used also by Martins and Garland, 1991; D-U&G; Martins, 1996a), hereafter $Tree_{15}$. To study the effects of systematic errors, we employed only a Brownian motion model of evolution. To examine effects of random errors, we used three basic evolutionary models: Brownian motion, Ornstein–Uhlenbeck, and speciational. Brownian motion and Ornstein–Uhlenbeck were used in D-U&G. The speciational model is equivalent to a Brownian motion model on a phylogeny with all branch lengths set equal to unity (see also Martins and Garland, 1991). This model differs from punctuated equilibrium, in which change occurs in only one of the daughter species, not both (see references in D-U&G and their page 43).

As in our previous work, for each of the three basic models we simulated character evolution with limits, using either the "Replace" or the "Truncate Change" algorithm. (See D-U&G for details; examples of the effects of the two limit algorithms on the distribution of tip data are shown in their Fig. 1.) In addition, we modeled pure Brownian motion without limits; this serves as a "baseline" for the effects of errors in branch lengths without the additional effects of violations in model specification.

Unlike our previous study, we did not model evolutionary trends (different starting and final means in the simulations). Our previous work showed that the imposition of a trend, in addition to limits to character evolution, did not have major effects on the performance of IC. Therefore, in all of the current simulations, the initial and expected final means were set to be the same (100).

In summary, for systematic errors, we used only Brownian motion without limits, with six types of error. For random errors, a total of seven models of evolution were used (Table 1). For each one of seven models (and on two phylogenies), we applied a total of six types of random errors (relative normal and fixed [see below], each with three levels). Therefore, a total of 84 basic cases were examined (7 models × 2 phylogenies × 6 types of random error).

Parameters of the Computer Simulations

As in our previous paper, we used the PDSIMUL program (Garland et al., 1993) to simulate bivariate evolution of continuous-valued characters along a specified phylogenetic tree. All parameters of the simulations (Initial Values, Variances-Tip, Final Means, Upper and Lower Limits, Adaptive Peak, and Decay Constant for Ornstein–Uhlenbeck models) were set as in D-U&G. Because we are concerned only with Type I error rates in the present study, all input correlations were set equal to 0. In other words, we were interested in how frequently the true null hypothesis (input correlation = 0) was rejected by the different methods, and how this compared with the nominal level of 0.05.

For every one of the 84 basic cases of random errors (model × phylogeny × type of error), we replicated six times the simulated data set of n = 1,000 (i.e., each data set consisted of 1,000 simulated evolutionary processes, each consisting of either 15 or 49 species at the tips of the phylogeny). The parameters of the six simulated data sets were identical within each model × phylogeny × type of error, except for the seed of the pseudorandom number generator, which was chosen from a table of random numbers. For each of the 12 cases of systematic error (phylogeny × error), we also used six replicates, each one with 1,000 simulated evolutionary processes.

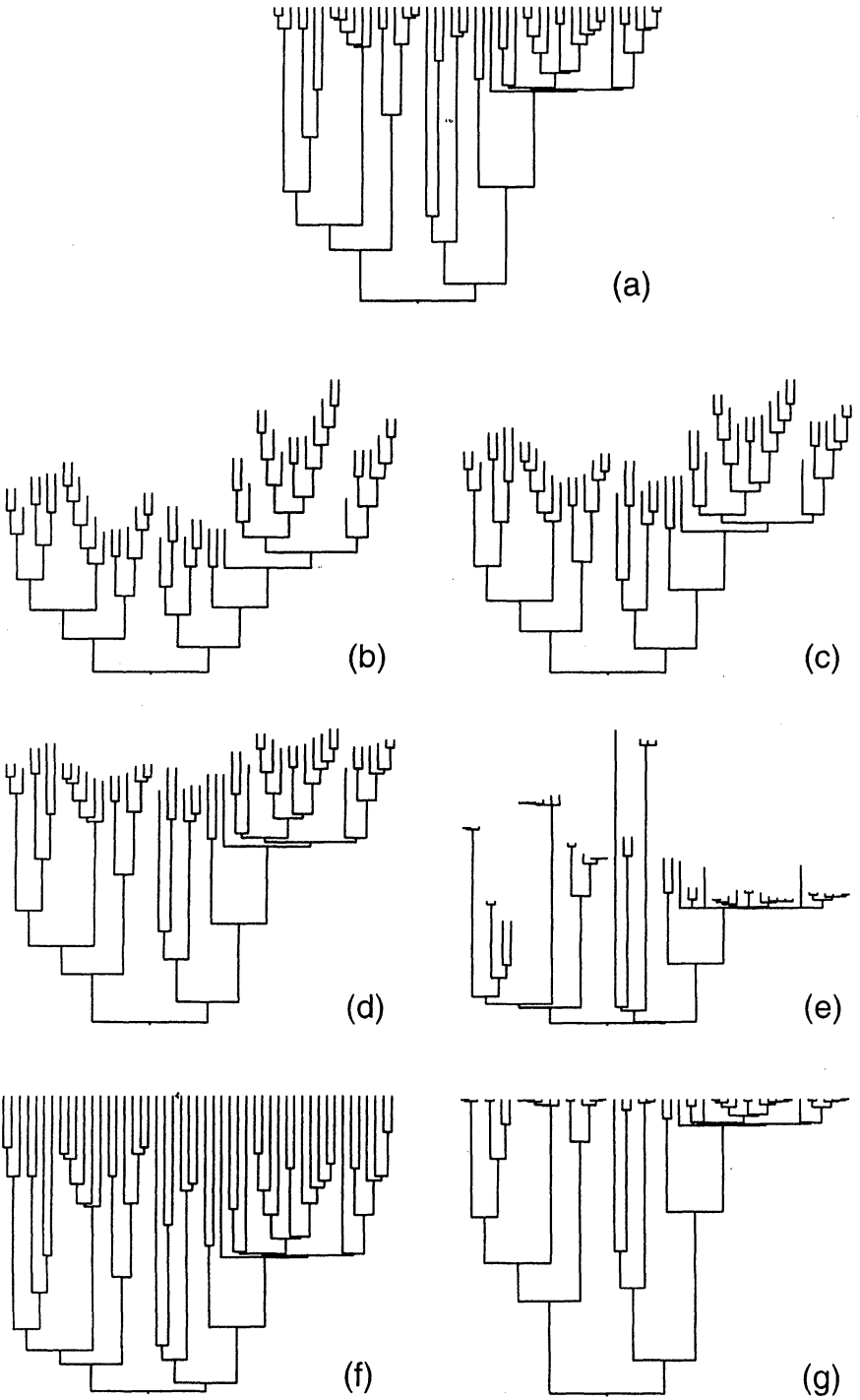

FIGURE 1. Systematic distortions of branch lengths on the 49 species phylogeny: (a) original branch lengths (as in Garland et al., 1993); (b–e) branch lengths raised to the (b) 0.25 power; (c) 0.5 power; (d) 0.75 power; (e) squared; (f) rho (Grafen, 1989) of 0.5; (g) rho of 2.

TABLE 1. Description of the seven models of evolution used. See text for details on parameters of simulations.

Abbreviation	Description
BM	Brownian motion
BMReplace	BM, limits with Replace algorithm
BMTruncate	BM, limits with Truncate algorithm
OUReplace	Ornstein–Uhlenbeck, limits with Replace algorithm
OUTruncate	Ornstein–Uhlenbeck, limits with Truncate algorithm
SReplace	Speciational BM, limits with Replace algorithm
STruncate	Speciational BM, limits with Truncate algorithm

Systematic Errors in Branch Lengths

Suppose that for a certain character the rate of evolution has not been proportional to time but instead to some transformation of time, such as the square root. If we had a phylogeny whose branch lengths are expressed in units of time and we knew that the rate of evolution was proportional to the square root of time (the "transformation that evolution used"), then we should use the square root of the branch lengths in the analyses. If we did not know the transformation that evolution used, then we would make an error by using the branch lengths in unmodified units of time, an error we call a systematic error.

We simulated evolution on phylogenies where branch lengths were a transformation of the time-units branches. We used two families of transformations: the power transformation, where the true branches are the time-units branches raised to some power, and Grafen's (1989) rho transformation. The rho transformation takes a power of the "height" of a node, not of the branch length. Rho values <1 compress the tree near the root, and expand it near the tips, whereas values >1 compress the tree near the tips, and expand it near the root. For the power transformations we used powers of 0.25, 0.5, 0.75, and 2. For rho, we used 0.5 and 2. These distortions are shown in Figure 1 for $Tree_{49}$. For each of the two phylogenies, we simulated independent evolution of two traits by using Brownian motion. The branches of these phylogenies had been transformed as specified above. We then analyzed the data by using the original (incorrect) branches in units of time.

Random Errors in Branch Lengths

We used the computer program PDERROR (see D-U&G) to implement random alterations of branch lengths. We used two types of random errors, fixed and relative normal. With "fixed" error, each branch of the phylogeny is either shortened or lengthened by a specified constant fraction. This fraction (e.g., 0.1, or 10%) is the same for all branches of the phylogeny. Therefore, every branch length is altered, and always by the same relative amount, but whether it is lengthened or shortened is random (with a probability of 0.5). We used three different values for the alteration of branch lengths: 0.1, 0.5, and 0.9. Use of 0.1 produces only small alterations of branch lengths (for example, a branch that originally measures 100 million years will become either 90 or 110 million years), whereas the 0.9 produces very large alterations (a branch that originally measures 100 million years will become either 10 or 190 million years). Effects of these three types of distortion are shown in Figure 2.

For "relative normal" errors we added an amount drawn from a normal distribution with mean 0 and a standard deviation equal to a specified relative normal error multiplied by branch length. In other words, the standard deviation of the distribution of errors is scaled to the length of each branch. Applying errors this way, branches of different lengths are all subject to the same relative error. For example, with a relative normal error of 0.05, the standard deviation of the distribution of errors will be 5 if the branch length measures 100 ($5 = 100 \times 0.05$) and 50 if the branch length measures 1,000. This in turn makes the distributions of the final branches (original branch + error) linearly comparable among branches of different lengths. For a branch length of 100, $\sim 95.44\%$ of the errors will be between -10 and 10, so 95.44% of the final branch lengths will be between 90 and 110 (remember that, in a normal distribution, 95.44% of the obser-

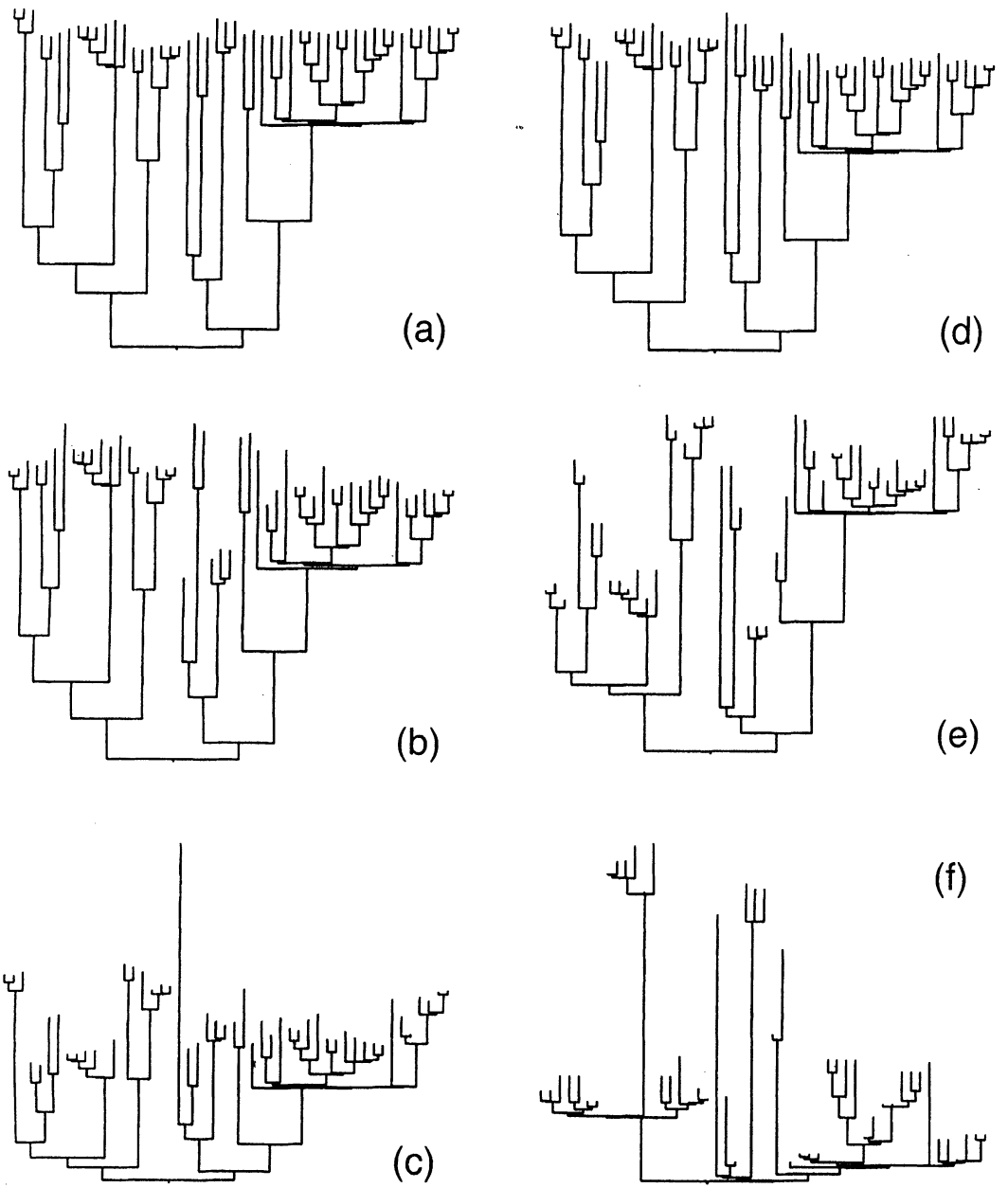

FIGURE 2. Examples of distortions created by adding random errors (see text) to the 49 species phylogeny: (a) 0.05 relative normal error; (b) 0.25 relative normal error; (c) 0.45 relative normal error; (d) 0.1 fixed error; (e) 0.5 fixed error; (f) 0.9 fixed error. See Figure 1a for the original phylogenetic tree without distorted branches.

vations are within two standard deviations of the mean).

We used three relative normal errors: 0.05, 0.25, and 0.45. With these errors, for a branch length of 100, 95.44% of the final branch lengths will be within 90 to 110, 50 to 150, and 10 to 190, respectively. See Figure 2 for examples.

In the variable error implementation, branch lengths can become negative; for example, with a relative error of 0.45, the probability of a branch length becoming negative is 0.0131. PDERROR prevented branch lengths from becoming negative by giving the affected branches the value of 1.1. (This is virtually equivalent to a branch length of 0, compared with average branch lengths of tens or hundreds of million years, as in $Tree_{15}$ and $Tree_{49}$.)

These implementations of errors allow for direct interpretation of the error levels, both in the "fixed" and the "relative normal error" types. The absolute error increases linearly with branch length (so the relative error remains constant). Other types of relationships (for example, multiplicative, where larger branches also have larger relative errors) are conceivable (*see* Discussion).

Analysis of Simulated Data Sets

The simulated data sets were analyzed with three versions of phylogenetically independent contrasts (Table 2) through use of PDERROR (see D-U&G for details). This program analyzes each simulated data set and can produce, among other output, the correlation coefficient estimated by two types of IC. The first one ("FL1G" of Martins and Garland, 1991) is the ordinary correlation (through the origin) of IC and is referred to as IC. The second one is ICblt (IC with *b*ranch *l*ength *t*ransformations). Here, branch lengths can be transformed before a correlation with IC is computed, as proposed by Garland et al. (1992). For each trait of each simulated data set, the program examines all transformations of branch lengths obtained by raising the branch lengths to powers ranging from 0 to 2 in intervals of 0.1, plus the log (base 10) of the branch lengths (note that a power of 0 yields all branch lengths equal to unity, and a power of 1 leaves branch lengths unchanged). To determine the best branch length transformation, the program computes for each trait and for each possible branch-length transform the Pearson correlation (not through the origin) between the absolute value of the standardized contrasts and their standard deviations, and then selects the transform that gives the smallest absolute value of the correlation coefficient. Once branch lengths have been transformed, standardized IC are computed as usual, and the correlation for the standardized IC of the two traits is also computed (through the origin). Because this procedure is performed independently for the two traits (in each simulated data set), the correlation coefficient between contrasts

TABLE 2. Methods and degrees of freedom (df) used in the analyses (n is number of species, either 15 or 49). The df were used to establish the conventional critical values for the correlation coefficients.

Method	Description	df
IC	Felsenstein's (1985) method of phylogenetically independent contrasts with no branch length transformations. "FL1G" of Martins and Garland (1991).	$n-2$
ICblt	Felsenstein's (1985) method after checking for adequate branch length standardization and using *b*ranch *l*ength *t*ransformations if appropriate, as indicated in Garland et al. (1992).	$n-2$
ICblt(df-4)	As above. ICblt of Díaz-Uriarte and Garland (1996).	$n-4$
ICblte	Same as for ICblt, but checking and *e*xcluding those simulated data sets in which the branch length transformations did not achieve an adequate standardization of contrasts. Data sets for which the correlation between absolute value of standardized contrasts and their standard deviations (square root of sum of branch lengths) was statistically significant for either one or both traits at the $P = 0.05$ level (for $n-3$ df) were regarded as not appropriately standardized (Garland et al., 1992) and were excluded.	$n-2$
ICblte(df-4)	As above. ICblte of Díaz-Uriarte and Garland (1996).	$n-4$

can be computed for two traits that have been standardized with different branch length transformations (for some examples with real data, see Garland et al., 1992; Garland, 1994).

The above procedure does not guarantee that an adequate transformation of branch lengths will be achieved. Therefore, as in D-U&G, we checked whether the transformed branch lengths yielded a Pearson correlation between the absolute values of standardized contrasts and their standard deviations that was not statistically significant at $P = 0.05$ for a two-tailed test with $n - 3$ df, where n = number of species, and with this correlation not forced through the origin (therefore, the critical values are 0.532 for $Tree_{15}$ and 0.285 for $Tree_{49}$). We then excluded those simulations for which the correlation remained statistically significant and recomputed Type I error rates. We call this ICblte (IC with *b*ranch *l*ength *t*ransformation checking and *e*xcluding some cases; Table 2). The results for ICblte may give the best indication of the "real-world" performance of IC(see D-U&G for more details). With a real data set, a practitioner presumably would not proceed with conventional tests if adequate standardization could not be achieved.

In summary, for each simulated data set, we obtained three different distributions of correlation coefficients that can be used to estimate evolutionary correlations between two continuous-valued characters (see Table 2): IC (Felsenstein's [1985] method of phylogenetically independent contrasts, applied naively), ICblt (Felsenstein's method after checking and transforming branch lengths as suggested by Garland et al. [1991, 1992]), and ICblte (ICblt after excluding those simulations in which adequate standardization was not achieved).

Computing and Testing Type I Error Rates

To analyze the performance of each method with regard to hypothesis testing, we calculated the rates of Type I error (probability of rejecting the null hypothesis when it is true) in two different ways. In both cases, the question asked was whether a given method yields Type I error rates significantly different from those expected under standard normal theory when the true evolutionary correlation between traits is zero (on average). For further details, see D-U&G.

First, we compared the overall distribution of the 1,000 (or fewer for ICblte; see Table 3 and below) correlation coefficients for a given set of simulated data with the theoretical distribution of Pearson correlation coefficients under the null hypothesis. We determined the number of observed correlation coefficients whose value was between the critical values of 12 successive α levels. These observed numbers were compared with the numbers expected, based on a standard Pearson's r distribution (Table B.16 in Zar, 1984), by using a chi-square test with 11 df (for 12 intervals; see D-U&G for details). These chi-square tests are sensitive to any departures from the expected distribution of correlation coefficients (i.e., both inflation and deflation of Type I error rates, and bias in the distribution of correlation coefficients).

To obtain the cutting points of the intervals (i.e., the critical values of the correlation coefficients), we needed to define the number of degrees of freedom for the correlation coefficients. For IC, df $= n - 2$, where n is the number of species (we obtain $n - 1$ independent contrasts, and we lose 1 df by estimating the correlation—note that no intercept term exists; see Garland et al., 1992). For analyses in which branch lengths are transformed for both characters, the appropriate degrees of freedom are probably $n - 4$. The additional 2 df are subtracted because the data are used to estimate the "optimal" branch length transformation (see D-U&G; also Box and Draper, 1987; Reynolds and Lee, 1996). To examine empirically the most appropriate degrees of freedom, we determined the cutting points of the intervals for the correlation coefficients, using both $n - 4$ and $n - 2$ df Therefore, we have a total of five methods of analysis: IC, ICblt, ICblt(df-4), ICblte, and ICblte(df-4) (see Table 2).

Second, we computed the observed frequency of correlation coefficients (of 1,000 or fewer total) for a nominal α level of 0.05. In other words, we determined the number

TABLE 3. Mean number of simulations that were considered adequately standardized in the analyses with ICblte (see Table 2 and text for details); maximum possible is 1,000.

Systematic errors, type of branch length distortion	Number	
	$Tree_{15}$	$Tree_{49}$
** 0.25	985	985
** 0.5	990	998
** 0.75	996	1000
** 2	1000	958
Rho 0.5	972	973
Rho 2	999	822

	Random errors, $Tree_{15}$					
	Relative normal error			Fixed error		
Evolutionary model	0.05	0.025	0.45	0.1	0.5	0.9
BM	997	996	998	997	998	999
BMReplace	877	890	911	882	902	950
BMTruncate	966	969	969	966	976	987
OUReplace	861	871	893	867	887	948
OUTruncate	935	946	948	934	953	972
SReplace	834	843	870	839	866	934
STruncate	900	917	926	911	931	968

	Random errors, $Tree_{49}$					
	Relative normal error			Fixed error		
Evolutionary model	0.05	0.025	0.45	0.1	0.5	0.9
BM	1000	1000	1000	1000	1000	1000
BMReplace	996	998	996	998	996	994
BMTruncate	1000	1000	1000	1000	1000	1000
OUReplace	649	653	669	655	649	718
OUTruncate	968	960	958	965	959	956
SReplace	109	102	142	102	129	344
STruncate	624	614	657	628	641	777

of correlation coefficients exceeding the critical value for $\alpha = 0.05$ in a two-tailed test. As above, we used critical values for $n - 4$ and $n - 2$ df for the ICblt and ICblte methods. We then used a binomial test (Conover, 1980) to obtain the P-value of each observed frequency (six for each model × phylogeny combination). This binomial test computes the probability (P-value) of obtaining the same or larger number of correlation coefficients that are greater than the critical value (under the null hypothesis that, by chance, 50 out of 1,000 correlation coefficients should be larger than the critical value). This procedure specifically tests for inflated Type I error rates.

Absolute and Relative Performance of Methods

We evaluated both absolute and relative performance of the different methods in terms of Type I error rates as in D-U&G. Briefly, to evaluate absolute Type I error rates of the different methods, we computed combined probabilities (see Sokal and Rohlf, 1981:779–781) for the six replicate binomial tests and for the six replicate chi-square tests (Fig. 3 for systematic errors, Fig. 4 for random errors). The six P- values to be combined come from the six replicate computer-simulated data sets (each using a different random number seed) for each phylogeny × error combination (systematic error) or

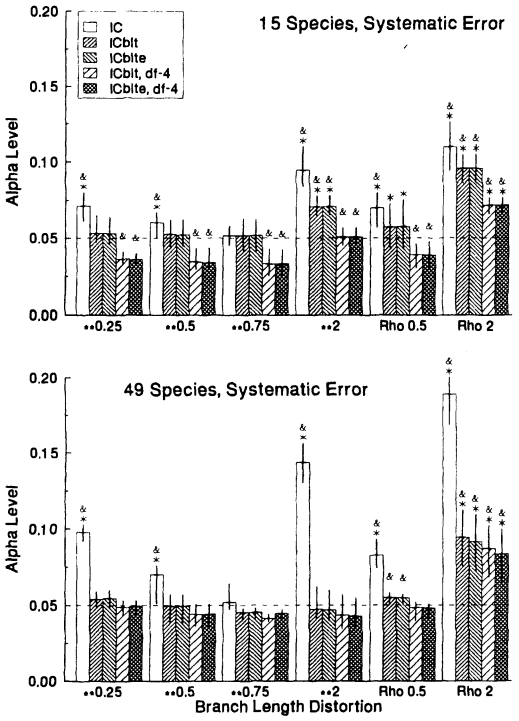

FIGURE 3. Type I error rates of phylogenetically IC for testing the significance of a bivariate evolutionary correlation, with systematic errors in branch lengths (see text). For each of the five methods of analysis (Table 2), we report the mean (bar) and range (thin vertical lines) of the actual alpha level for a nominal $\alpha = 0.05$, for the six replicate simulations. Asterisk: Combined P-value of the six binomial tests for inflated Type I error rates is significant at $P = 0.05$. Ampersand: Combined P-value of the six chi-square tests is significant at $P = 0.05$. Note that performance of ICblt and ICblte with rho-type systematic errors is worse than it might potentially be because the rho transformation (Grafen, 1989) was not examined when searching for the best possible branch length transformation (see text).

model × phylogeny × error combination (random error). The combined P-values in Figures 3 and 4 denoted with an asterisk test whether the P-values at $\alpha = 0.05$ are inflated. The combined P-values in Figures 3 and 4 shown with an ampersand indicate whether the overall distribution of correlation coefficients obtained with the different methods is significantly different from the expected overall distribution.

To assess relative performance of the methods, we compared them by using the six independent P-values of (1) the chi-square tests, which indicate deviations from the expected for overall distribution of correlation coefficients, and (2) the number of correlation coefficients exceeding the critical value at $\alpha = 0.05$ (Figs. 3, 4). The two ways of comparing methods generally yielded similar results. Statistical significance of differences between methods was tested with Wilcoxon signed-rank tests (Conover, 1980) by using two-tailed tests. (These are matched-pairs tests because the methods of analysis being compared use the same data set [see above, *Analysis of Simulated Data Sets*].) These tests were performed separately (1) for each of the model × phylogeny × error combinations (for random errors) or phylogeny × error (for systematic errors), with $n = 6$ for each test, and (2) for the data from all models combined for a given phylogeny ($n = 252$ for random errors, $n = 36$ for systematic errors). We performed four independent tests to examine (1) whether ICblt performs better than IC; (2) if the appropriate degrees of freedom are $n - 2$ or $n - 4$ (by comparing ICblt[df-4] with ICblt and ICblte[df-4] with ICblte; and (3) if ICblte (which we suggest is the method that will most closely reflect "real world" performance) differs from ICblt.

Statistical analyses of the data obtained from PDERROR were performed with SAS, S-Plus for Windows v. 3.3, and SPSS/PC+ v. 5.0. Statistical significance was judged at $P = 0.05$.

Bias of Correlation Coefficients

We examined bias of correlation coefficients by using the methods in Martins and Garland (1991). In the present case, bias would mean that the expected value of the estimated correlation coefficients is different from 0 (the true correlation). For each combination of phylogeny × method × error (for random errors) or phylogeny × error (for systematic errors), we grouped the six simulated data sets, which gave a total of 6,000 estimated correlation coefficients (for ICblte the total sample size was generally smaller, sometimes as low as 620; see Table

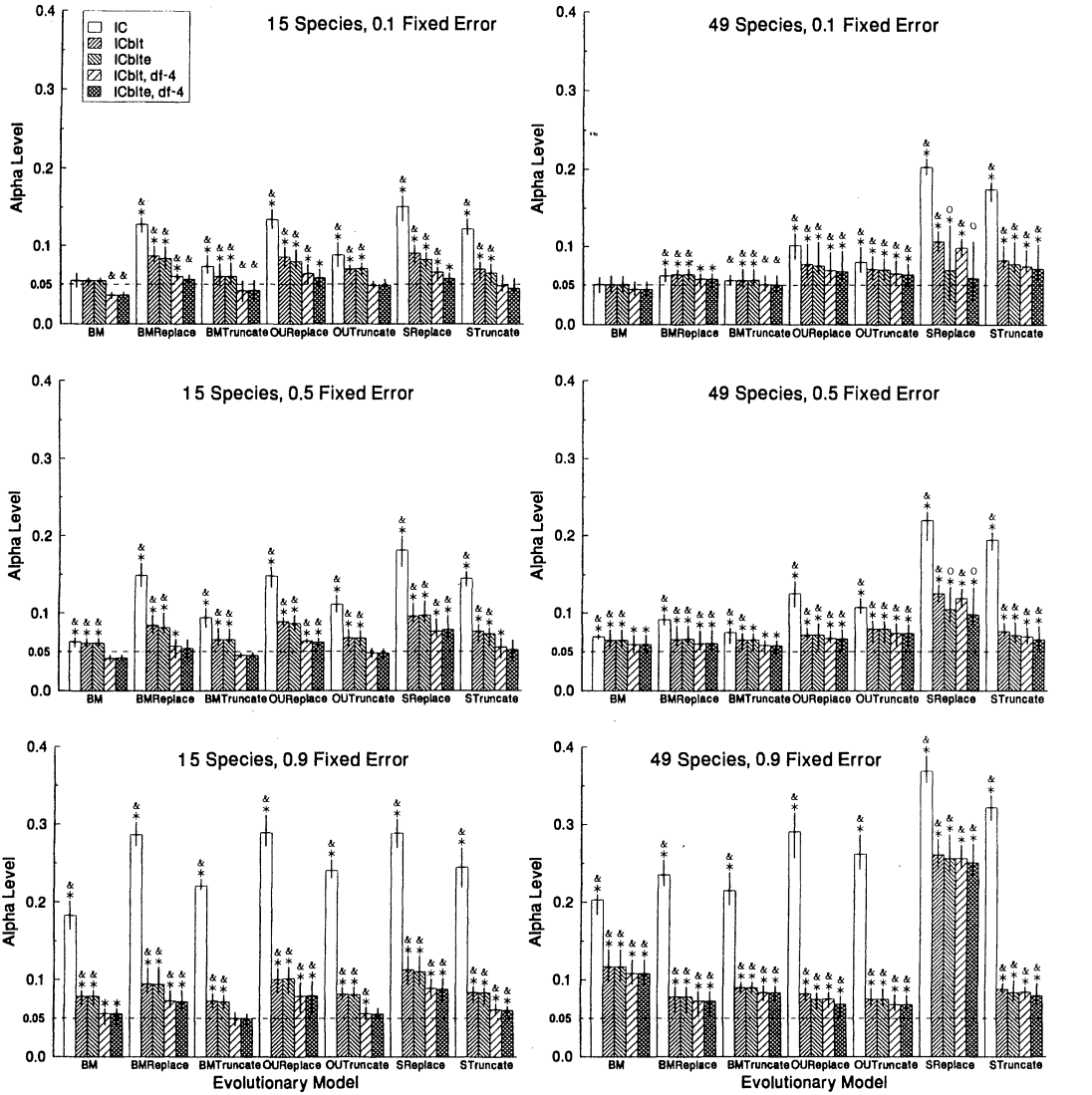

FIGURE 4. Type I error rates of phylogenetically IC for testing the significance of a bivariate evolutionary correlation, with random errors in branch lengths and misspecifications of the evolutionary model (see text). For each of the five methods of analysis (Table 2), we report the mean (bar) and range (thin vertical lines) of the actual alpha level for a nominal $\alpha = 0.05$ for the six replicate simulations. Asterisk: Combined P-value of the six binomial tests for inflated Type I error rates is significant at $P = 0.05$. Ampersand: Combined P-value of the six chi-square tests is significant at $P = 0.05$. With the 49 species phylogeny and SReplace model, the number of simulations that were adequately standardized was frequently too small to perform a chi-square (denoted by O; see Table 3).

3). We then computed a 95% confidence interval (C.I.) for the mean of these correlation coefficients. If no bias exists, then the 95% C.I. should include 0. We also performed a sign test, to determine whether the numbers of positive and negative correlation coefficients were equal. The 95% C.I. and sign tests gave very similar results.

We also compared IC, ICblt, and ICblte for differences in bias. The values of the correlation coefficients for ICblt and ICblte are exactly the same (ICblte is just ICblt with

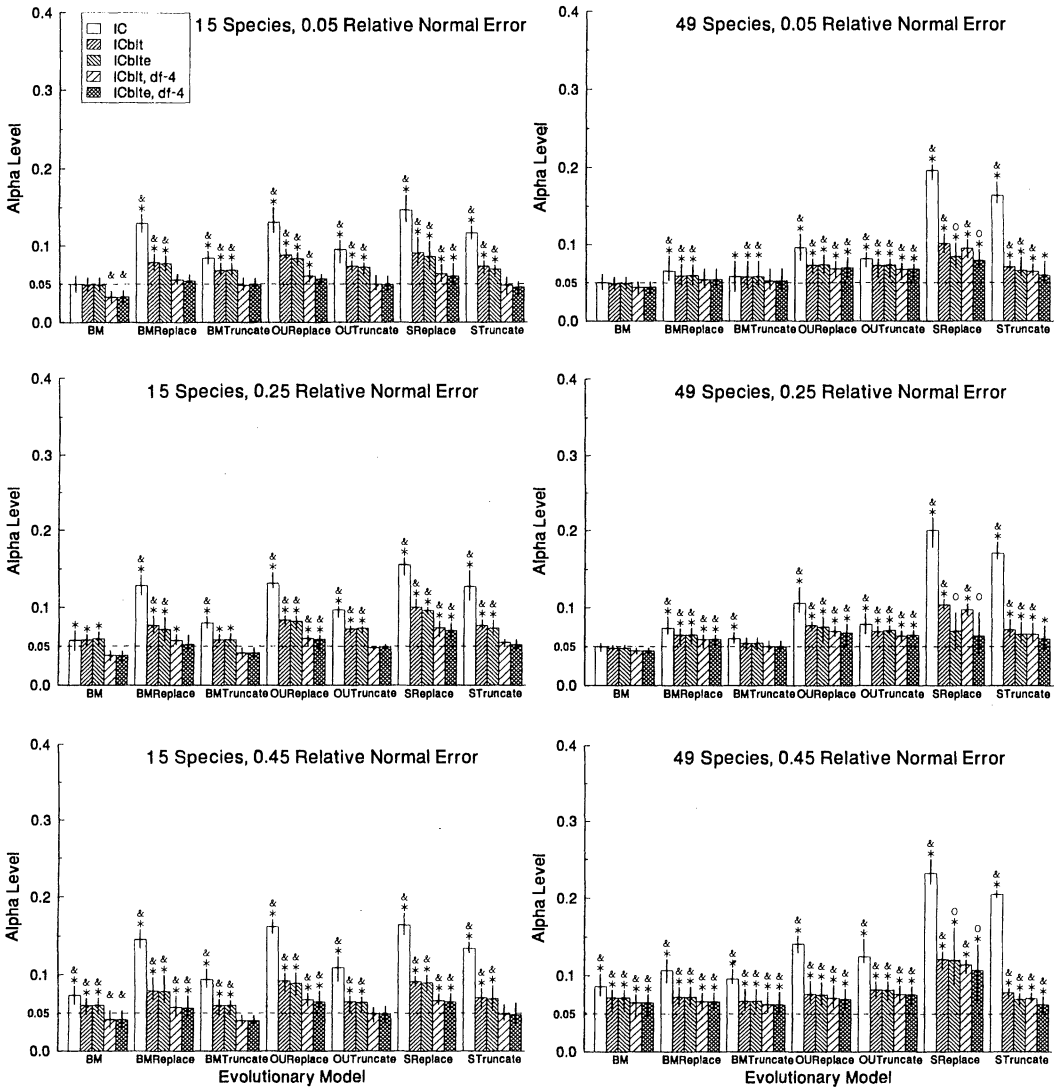

FIGURE 4. Continued.

some values excluded); therefore, the only meaningful comparisons in terms of bias are IC vs. ICblt and IC vs. ICblte. We carried out Wilcoxon signed-rank tests (Conover, 1980) for each combination of phylogeny (49 vs. 15) and type of error (random [fixed and variable] and systematic). Thus, we performed four tests (49 and relative normal, 49 and fixed, 15 and relative normal, 15 and fixed), each on 126,000 (or fewer for ICblte) correlation coefficients. For the systematic errors, we analyzed the data for the 15 and 49 phylogenies separately, each on 36,000 correlation coefficients.

RESULTS

Overview of Results

Results for systematic branch length errors are summarized in Figure 3. When the distortion in branch lengths is a power of the

original branch lengths, Type I error rates are generally inflated. However, transformations usually yield correct Type I error rates, although the appropriate number of degrees of freedom varies with type of errors. When the systematic error is caused by a rho transformation, branch length transformations with $n-4$ df yield the best results, although in some cases (e.g., rho = 2) these results show significant deviations from the overall correct distribution of correlation coefficients. Note that performance of ICblt and ICblte with rho-type systematic errors is worse than it could potentially be, because the rho transformation was not included when searching for the best possible branch length transformation. Nevertheless, the real P-value at $\alpha = 0.05$ is always < 0.1.

Results for random branch length errors are shown in Figures 4 and 5. Results vary widely among models, phylogeny, and type of error. In all cases, except simple Brownian motion, the appropriate degrees of freedom for testing significance of the correlation coefficient are $n-4$, not $n-2$. For many models of evolution and under most of the error levels, all methods produced inflated Type I error rates, but when branch lengths are checked and transformed, the real Type I error rate at $\alpha = 0.05$ is generally < 0.1. As with systematic errors, even in the worst case, using branch length transformation and standardization with $n-4$ df to determine critical values, would almost always yield only slightly inflated Type I error rates. In comparison, using independent contrasts without any type of branch length transformation can produce highly inflated Type I error rates (see Fig. 5).

With respect to bias, for both systematic and random errors, the three methods produced unbiased estimates in almost all cases, as judged from the 95% CI not overlapping zero (sign tests produced comparable results). In the few cases where bias existed, the absolute amount of bias was very small (mean absolute value of the estimated correlation coefficient was < 0.015). Comparisons among the three methods, using Wilcoxon signed-rank tests, did not reveal any significant differences for either systematic or random errors.

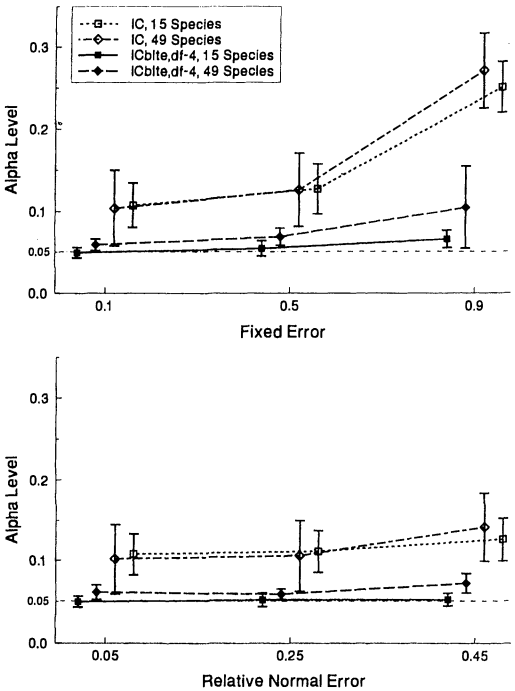

FIGURE 5. Summary results for the simulations with random branch length errors. Shown are means of the actual alpha levels (at a nominal $\alpha = 0.05$) for the seven models of evolution (Table 1) at each level of branch length error, ±2 standard errors (i.e., an approximate 95% C.I.). The standard errors were computed by using as individual data the mean of each of the models by error combinations (so the standard error is based on seven data points). These error bars, shown mainly for heuristic purposes, should not be used to compare IC vs. ICblte at the same error level (as described in the text, the appropriate comparison is a paired comparison, as was done with the Wilcoxon tests). The improved performance achieved by branch length transformations is apparent in every case.

Figure 6a shows the distribution of the branch length transformations employed by ICblt for some examples of systematic errors. As expected, the mode of these distributions is the same as the actual exponent used to produce the systematic distortion (note that for systematic distortions of 0.25 and 0.75 the closest exponents were, respectively, 0.2 or 0.3 and 0.7 or 0.8, as the exponents examined by the PDERROR program increased in intervals of 0.1). For relative normal errors (Fig. 6b), these distributions are slightly asymmetric, but the

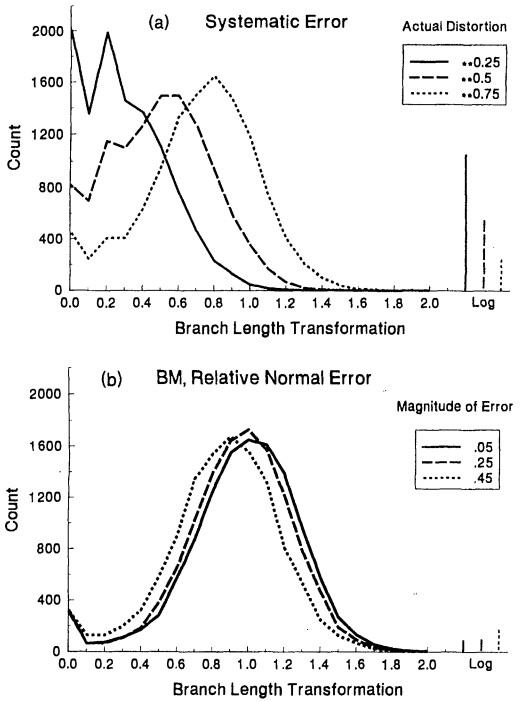

FIGURE 6. Distribution of branch length transformation exponents used by ICblt method for $Tree_{49}$ with a Brownian motion model of character evolution and (a) systematic or (b) random errors added to branches. Note that with systematic errors (a) the modal transformation used by the method, based on the diagnostic of Garland et al. (1992), corresponds quite well to the actual distortion employed. For random errors (b), no systematic distortion exists and, as should be the case, the modal transformation is close to no transformation. For examples of trees with errors as in (b), see Figure 2a, 2b, and 2c. Figures 6a and b use all 12,000 data points from the actual simulations (we used data from both traits, in each of the six replicates of 1,000 simulated data sets).

mode of branch length transformations is 1 or 0.9. Therefore, the transformation used was often no change at all in the branch lengths. Again, this is as expected, because no "global" distortion exists.

Systematic Errors

Absolute performance.—At $\alpha = 0.05$, the combined P-values in Figure 3 show that Type I error rates were inflated in 10 out of 12 instances for IC. Performance improved with branch length transformations; for ICblt and ICblte, both with $n-2$ df, Type I error rates were inflated in only 4 of 12 cases. ICblt and ICblte with $n-4$ df showed inflation of Type I error rates in only two cases (rho = 2, for both the 15 and the 49 species phylogenies). Regarding the overall distribution of correlation coefficients (Fig. 3), results were generally very similar, except that ICblt and ICblte performed relatively poorly with $n-4$: correlation coefficients tended to concentrate in the center of the distribution, so the number of correlation coefficients in the tails of the distribution was lower than it should have been. In other words, these methods were too conservative; for example, the number of correlation coefficients exceeding the critical value at $\alpha = 0.05$ is ~35—instead of the expected 50—for powers of 0.25, 0.5, and 0.75 on $Tree_{15}$. This effect is more evident for 15 species than for 49, because 13 vs. 11 df makes a large difference in the critical values (0.514 vs. 0.553), whereas 47 vs. 45 df does not (0.282 vs. 0.288).

The performance of each method depended on both phylogeny and type of error. Interactions among phylogeny, error, and method of analysis were significant as demonstrated by ANOVAs. We used the split-plot (Snedecor and Cochran, 1989; Yandell, 1997) model:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} + \gamma_l + (\alpha\gamma)_{il} + (\beta\delta)_{jl} + (\alpha\beta\gamma)_{ijl} + \delta_{ijkl}$$

where Y is the arcsine of square root of the P-value from the chi-square, or the log of the number of correlation coefficients larger than the critical value at $\alpha = 0.05$; α is the effect of phylogeny, β is the effect of type of error, and δ is the effect of method of analysis; ε and δ correspond to the "whole plot" and "subplot" errors, respectively. Second-order interactions, error × method and method × phylogeny, were significant in both cases. The error × phylogeny interaction was significant for number of correlation coefficients larger than the critical value at $\alpha = 0.05$, but only marginally significant for the chi-square P-value. The third-order interaction (error × method × phylogeny) was significant in both cases at $P < 0.001$. Plots of residuals vs. treatment and residuals vs. predicted values did not reveal heteroscedastic-

ity or errors in the specification of the model for the log of the number of correlation coefficients larger than the critical value at $\alpha = 0.05$. Residuals were ill-behaved for the chi-square *P*-values, so we repeated the analyses using rank-transformed data (Conover, 1980); these analyses yielded similar results.

Relative performance.—Branch length transformations always decreased Type I error rates; IC never had lower Type I error rates than ICblt or ICblte. In terms of the number of correlation coefficients larger than the critical value at $\alpha = 0.05$ (Fig. 3), ICblt(df-4) showed significantly lower Type I error rates than IC or ICblt for every single one of the six possible types of errors, both for $Tree_{15}$ and $Tree_{49}$. ICblte(df-4) showed no significant difference from ICblt(df-4) in $Tree_{15}$; for $Tree_{49}$, three of those differences were not significant, but in one ICblte did better (rho = 2) and in two ICblt did better. When data were combined, ICblte(df-4) and ICblt(df-4) showed no significant differences for either $Tree_{15}$ or $Tree_{49}$. Finally, Type I error rates of ICblte(df-4) were smaller than those of ICblte in every case for $Tree_{15}$, and in five cases for $Tree_{49}$. When all data were combined, ICblte(df-4) was significantly better than ICblte for both phylogenies.

Results for the overall distribution of correlation coefficients were generally similar. Nevertheless, we can see differences in the results when comparing Type I error rates and when comparing overall distributions of correlation coefficients (Fig. 3). ICblt and ICblte, both with $n - 4$ df, have lower Type I error rates than the $n - 2$ df alternatives: With $n - 4$ df, the number of correlation coefficients exceeding the critical levels is lower than expected (see preceding *Absolute performance*).

Random Errors

Absolute performance.—Figure 4 shows that, for a nominal $\alpha = 0.05$, both IC and ICblt produced inflated Type I error rates in 38 of 42 cases for $Tree_{49}$ and 40 cases for $Tree_{15}$. For ICblte, there were 37 cases of inflated Type I error rates in $Tree_{49}$, 42 in $Tree_{15}$. Performance improved substantially with ICblt(df-4) and ICblte(df-4), particularly for $Tree_{15}$; for ICblt(df-4) inflation occurred in 35 cases for $Tree_{49}$ and in 21 cases $Tree_{15}$. ICblte(df-4) did slightly better, with inflation in 33 and 15 cases in $Tree_{49}$ and $Tree_{15}$, respectively. Results were almost identical with respect to the overall distribution of correlation coefficients (Fig. 4). In summary, branch length transformation with $n-4$ df considerably improved performance, and Type I error rates at the nominal $\alpha = 0.05$ were always < 0.1, except for the "Speciational Replace" model (under several types of errors) and the Brownian motion model with 0.9 fixed error, both for $Tree_{49}$.

For each method, performance depended on phylogeny, model of evolution, and type of error. Interactions among phylogeny, evolutionary model, error, and method of analysis were significant, as demonstrated by ANOVAs. We used the split-plot model:

$$\begin{aligned} Y_{ijklm} = {} & \mu + \alpha_i + \beta_j + \omega_k + (\alpha\beta)_{ij} + (\alpha\omega)_{ik} \\ & + (\beta\omega)_{jk} + (\alpha\beta\omega)_{ijk} + \varepsilon_{ijkl} + \gamma_m \\ & + (\alpha\gamma)_{im} + (\beta\gamma)_{jm} + (\omega\gamma)_{km} \\ & + (\alpha\beta\gamma)_{ijm} + (\alpha\omega\gamma)_{ikm} + (\beta\omega\gamma)_{jkm} \\ & + (\alpha\beta\omega\gamma)_{ijkm} + \delta_{ijklm} \end{aligned}$$

where everything is as in the model for systematic errors (see above), except for the addition of ω, the effect of model of evolution. All second-, third-, and fourth-order interactions were significant for both log of number of correlation coefficients larger than the critical value at $\alpha = 0.05$, and for the arcsine of the square root of the *P*-value of chi-square. (Plots of residuals vs. treatment and residuals vs. predicted values were well behaved for the log of the number of correlation coefficients larger than the critical value at $\alpha = 0.05$. Residuals were ill-behaved for the *P*-values of chi-square, so we repeated the analyses using rank transformed data, and obtained similar results.)

Relative performance.—Branch length transformations always decreased Type I error rates. ICblt(df-4) had significantly lower Type I error rates than IC and ICblt in every one of the 42 combinations of error × evolutionary model for $Tree_{15}$. For $Tree_{49}$, ICblt(df-4) always had lower Type I error rates than ICblt; ICblt(df-4) had lower Type

I error rates than IC in 38 cases, and differences were not significant in four cases. When we combined all data, ICblt(df-4) did significantly better than IC and ICblt on both phylogenies. Type I error rates of ICblte(df-4) were significantly smaller than those of ICblte in every case for $Tree_{15}$ and in 38 cases for $Tree_{49}$. Finally, ICblte(df-4) had lower Type I error rates than ICblt(df-4) in 11 and 4 cases for $Tree_{15}$ and $Tree_{49}$, respectively (in the rest of the cases, differences were not significant). When all data were combined, ICblte(df-4) performed significantly better than ICblte and ICblt(df-4) for both $Tree_{15}$ and $Tree_{49}$ species phylogenies. Results were generally similar for the overall distribution of correlation coefficients, except for one case on $Tree_{15}$ (Brownian motion), where ICblt performed better than ICblt(df-4).

ICblte improves the performance of ICblt by excluding simulations in which standardization was clearly inadequate. Nevertheless, the improvement in performance of ICblte over ICblt is not closely related to the number of simulations excluded (see Table 3 and Fig. 4; for example, Speciational Replace in $Tree_{49}$).

Discussion

We used computer simulations to study the statistical performance of Felsenstein's (1985) phylogenetically based statistical method, IC, when errors exist in the branch lengths of the phylogeny and, in most cases, when the characters do not evolve by Brownian motion. We considered the case of testing a bivariate evolutionary correlation. As noted earlier, the main assumptions of IC are that (1) a correct phylogenetic topology is available; (2) branch lengths of the phylogeny are measured in units proportional to expected variance of character evolution; (3) character evolution can be described as a Brownian motion model; and (4) within-species variation is absent or negligible (D-U&G; Martins and Hansen, 1996). Our previous work focused on the third assumption, but we noted that the second and third assumptions are generally closely related (D-U&G, pp. 28–29, 45). In the present work, misspecifications of branch lengths and models had the net effect of simultaneously violating the second and third assumptions. We found that both systematic and random branch length errors often lead to inflated Type I error rates when independent contrasts are applied naively.

We also tested whether the use of branch length checks and transformations can improve the performance of IC (i.e., reduce the inflation of Type I error rates). Garland et al. (1992) proposed examining the relationship between the absolute value of the standardized contrasts and their standard deviations. The existence of patterns in these plots indicates that contrasts are not appropriately standardized. As a remedial measure, Garland et al. (1992) suggested transforming the branch lengths until appropriate standardization was achieved. Previously (D-U&G), we showed that these branch length checks and transformations can indeed improve the performance of IC when the assumption of Brownian motion is violated. Here, we report that transformations also help in the face of branch length errors.

Summary of Results

We first examined the utility of branch length transformations when the distortion of branch lengths is systematic (but the evolutionary model is simple Brownian motion; see Fig. 1 for examples of distortions). These analyses specifically examined the second assumption listed above, without violation of others. Independent contrasts with branch length transformations performed very well, particularly for the distortions that involved raising the branch lengths to a power (Fig. 3). Performance was less good for rho distortions (Figs. 1f and 1g), but this result was expected because we did not allow the algorithm for finding the best branch length transformation to use the rho family of transformations (Grafen, 1989; see D-U&G:45–46). Another kind of systematic distortion can occur if clades differ in rate of evolution (Garland, 1992). We did not specifically study such cases, but they too can be detected by examination of the diagnostics we used. Moreover, they can be addressed

statistically by differential transformation of branch lengths in different clades (Garland and Ives, in review). Therefore, systematic misspecifications of branch lengths (for example, the rate of evolution was proportional to the square root of time, but the branch lengths of the working phylogeny were in units of absolute divergence times) do not represent a major problem for the performance of independent contrasts when testing hypotheses about correlated evolution.

Next, we examined the performance of IC when phylogenetic branch lengths contain random errors (and character evolution is not Brownian motion). This is a different problem from above, because there is no global "true transformation" to find, and two of the above assumptions (2 and 3) are violated. Note that the errors added to the branch lengths are unrelated to the microevolutionary model. These types of errors must be widespread, as they arise from multiple sources and can be avoided only by using an infinite amount of data. Again, our results consistently show that when a simple diagnostic is used to check for adequate standardization, IC can show broad-sense validity. Figure 5 summarizes our results and shows that (1) the application of IC without branch length checks can lead to highly inflated Type I error rates; (2) branch length transformations and checks can considerably improve the performance of IC; (3) with branch length transformation, even for large branch length errors, and over a variety of evolutionary models, IC are reasonably robust, showing Type I error rates that, on average, are always lower than twice the nominal level at $\alpha = 0.05$.

With IC, the Speciational Brownian motion model produced greater inflation of Type I error rates than other models, in particular with $Tree_{49}$, analogous to what was reported for Punctuated Equilibrium in D-U&G. There (D-U&G:43) we suggested that the poor performance of independent contrasts under Punctuated Equilibrium was caused mainly by the way daughter nodes are algebraically related to the ancestor node. The present data suggest, on the contrary, that the extreme violation of assumptions about evolutionary rates might be largely responsible for the poor performance of IC under both Speciational Brownian motion and Punctuated Equilibrium. Although both Punctuated Equilibrium and Speciational Brownian motion imply that branch lengths are all equal to each other, the algebraic relation between ancestors and descendants is similar in Speciational Brownian motion, Brownian motion, and Ornstein–Uhlenbeck models (and different from Punctuated Equilibrium).

Even if Punctuated Equilibrium or Speciational Brownian motion were the true models of evolution, with real data IC should be able to perform better than shown by the present simulations. Most empirical studies include only a sample of species from any given clade. Such sampling will result in missing speciation events in the branches that connect ancestors and descendants, with more distantly related species being connected by branches where more speciation events are missing. Therefore, in the observed data, the amount of change will tend to be more continuously distributed, and longer branches will show more change than shorter ones (which resembles Brownian motion more than speciational change or punctuated equilibrium). Future studies should consider the effects of sampling on the performance of comparative methods.

Ours is the first explicit study of the effects of random errors in branch lengths on the performance of phylogenetically based statistical methods. We used two very different ways of introducing random errors of branch length, but both yield qualitatively similar results. In our implementation of both kinds of errors, the absolute error increased linearly with branch length (so the relative error remains constant). Effects of alternative mechanisms of adding branch lengths error, and their biological meaning, warrant further study.

Use of Additional Information to Fit More Complex Models

This and our previous paper (D-U&G) demonstrate that branch length transformations are useful remedial measures for the application of phylogenetically inde-

pendent contrasts. In particular, they allow some recovery of statistical performance (reduction in Type I error rates) when branch lengths contain errors or when character evolution is not Brownian motion. In a general statistical sense, the use of branch length (or character) transformations is equivalent to any use of a richer model to better fit the data. Fitting a richer model to data involves estimation of more parameters. The values of these additional parameters are obtained by examination of the data themselves, which is the reason for subtracting additional degrees of freedom in final hypothesis testing. The additional parameters estimated are generally regarded as nuisance parameters (Grafen, 1989), although they can be used to address certain questions, such as those involving rates of evolution (e.g., see Garland, 1992; Martins, 1994; Martins and Hansen, 1997; Pagel, 1998; discussion in Garland et al., in press).

In this context, it is easy to see that branch length transformations per se do not lead to difficulties in the interpretation of results. Some authors (e.g., Martins and Hansen, 1996, 1997; see also Wenzel and Carpenter, 1994) have criticized the use of branch length transformations, arguing that they can lead to problems of interpretation: After use of branch length transformations, the correlation coefficient can no longer be interpreted as one of the parameters of a particular model of evolutionary change, which would preclude an "evolutionary interpretation" of the results. However, the evolutionary interpretation requires correct identification of the model of evolutionary change. This is unlikely to be possible (but see Martins, 1994) for the vast majority of comparative studies (see also Leroi et al., 1994). Additionally, these criticisms do not indicate what to do when model assumptions are violated, and our data clearly indicate that it is inappropriate to use untransformed branch lengths when assumptions of the model are violated. Therefore, we do not attempt to interpret the correlation coefficient as a particular parameter in some specific evolutionary model. Nevertheless, a correlation significantly different from zero means that changes in one trait have not been independent of changes in the other. This is a more modest interpretation, but it is probably much more robust. And for many users of independent contrasts, the simple test of independence will be enough.

ACKNOWLEDGMENTS

We thank P. E. Midford for computer programming; E. V. Nordheim, and R. J. Chappell for statistical advice; E. V. Nordheim, R. J. Chappell, and J. A. W. Kirsch for discussion; and three reviewers for comments. This work was supported by NSF grants IBN-9157268 (PYI), DEB-9220872, and DEB-9509343 to T.G.

REFERENCES

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

BAUWENS, D., AND R. DÍAZ-URIARTE. 1997. Covariation of life-history traits in lacertid lizards: A comparative study. Am. Nat. 149:91–111.

BOX, G. E. P., AND N. R. DRAPER. 1987. Empirical model building and response surfaces. John Wiley & Sons, New York.

CONOVER, W. J. 1980. Practical nonparametric statistics, 2nd edition. John Wiley & Sons, New York.

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.

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

GARLAND, T., JR. 1992. Rate tests for phenotypic evolution using phylogenetically independent contrasts. Am. Nat. 140:509–519.

GARLAND, T., JR. 1994. Phylogenetic analyses of lizard endurance capacity in relation to body size and body temperature. Pages 237–259 (+ references) *in* Lizard ecology: Historical and experimental perspectives (L. J. Vitt and E. R. Pianka, eds.). Princeton Univ. Press, Princeton.

GARLAND, T., JR., AND S. C. ADOLPH. 1994. Why not to do two-species comparative studies: Limitations on inferring adaptation. Physiol. Zool. 67:797–828.

GARLAND, T., JR., AND R. DÍAZ-URIARTE. In press. Polytomies and independent contrasts: An examination of the bounded degrees of freedom approach. Syst. Biol. 48.

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., 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., AND A. R. IVES. Preliminary data. Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods.

GARLAND, T. JR., P. E. MIDFORD, AND A. R. IVES. In press. An introduction to phylogenetically based statistical

methods, with a new method for confidence intervals on ancestral values. Am. Zool.

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

HARVEY, P. H., AND M. D. PAGEL. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford.

LEROI, A. M., M. R. ROSE, AND G. V. LAUDER. 1994. What does the comparative method reveal about adaptation? Am. Nat. 143: 381–402.

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

MARTINS, E. P. 1994. Estimating the rate of phenotypic evolution from comparative data. Am. Nat. 144:193–209.

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

MARTINS, E. P. (ed.). 1996b. Phylogenies and the comparative method in animal behavior. Oxford University Press, Oxford.

MARTINS, E. P., AND T. GARLAND JR. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: A simulation study. Evolution 45:534–557.

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

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

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

MILES, D. B., AND A. E. DUNHAM. 1993. Historical perspectives in ecology and evolutionary biology: The use of phylogenetic comparative analyses. Ann. Rev. Ecol. Syst. 24:587–619.

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

PAGEL, M. 1998. Inferring evolutionary processes from phylogenies. Zool. Scr. 26:331–348.

PRICE, T. 1997. Correlated evolution and independent contrasts. Philos. Trans. R. Soc. Lond. B 352:519–529.

PURVIS, A., AND T. GARLAND JR. 1993. Polytomies in comparative analyses of continuous characters. Syst. Biol. 42:569–575.

PURVIS, A., J. L. GITTLEMAN, AND H.-K. LUH. 1994. Truth or consequences: Effects of phylogenetic accuracy on two comparative methods. J. Theor. Biol. 167:293–300.

REYNOLDS, P. S., AND R. M. LEE III. 1996. Phylogenetic analysis of avian energetics: Passerines and nonpasserines do not differ. Am. Nat. 147:735–759.

RICKLEFS, R. E., AND J. M. STARCK. 1996. Applications of phylogenetically independent contrasts: A mixed progress report. Oikos 77:167–172..

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.

SNEDECOR, G. W., AND W. G. COCHRAN. 1989. Statistical methods, 8th edition. Iowa State University Press, Ames, Iowa.

SOKAL, R. R., AND F. J. ROHLF. 1981. Biometry, 2nd edition. W. H. Freeman, New York.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, AND D. M. HILLIS. 1996. Phylogenetic inference. Pages 407–514 *in* Molecular systematics, 2nd edition (D. M. Hillis, C. Moritz and B. K. Mable, eds.). Sinauer Associates, Sunderland, Massachusetts.

WENZEL, J. W., AND J. M. CARPENTER. 1994. Comparing methods: Adaptive traits and tests of adaptation. Pages 79–101 *in* Phylogenetics and ecology (P. Eggleton and R. I. Vane-Wright, eds.). Linnean Society Symposium Series 17. Academic Press, London.

YANDELL, B. S. 1997. Practical data analysis for designed experiments. Chapman and Hall, New York.

ZAR, J. H. 1984. Biostatistical analysis, 2nd edition. Prentice-Hall, Englewood Cliffs, New Jersey.

Received 10 July 1997; accepted 25 May 1998
Associate Editor: D. Cannatella