Evolution, 56(1), 2002, pp. 1–13

ADAPTIVE CONSTRAINTS AND THE PHYLOGENETIC COMPARATIVE METHOD: A COMPUTER SIMULATION TEST

EMÍLIA P. MARTINS,[1,2] JOSÉ ALEXANDRE F. DINIZ-FILHO,[3,4] AND ELIZABETH A. HOUSWORTH[5,6]
[1]*Department of Biology, Indiana University, Bloomington, Indiana 47408*
[2]*E-mail: emartins@indiana.edu*
[3]*Departamento de Biologia, Ciências Biológicas, Universidade Federal de Goiás, Cx. P. 131, 74.001-970, Goiânia, Goiás, Brasil*
[4]*E-mail: diniz@icb1.ufg.br*
[5]*Department of Biology and Department of Mathematics, University of Oregon, Eugene, Oregon 97403*
[6]*E-mail: eah@math.uoregon.edu*

Abstract.—Recently, the utility of modern phylogenetic comparative methods (PCMs) has been questioned because of the seemingly restrictive assumptions required by these methods. Although most comparative analyses involve traits thought to be undergoing natural or sexual selection, most PCMs require an assumption that the traits be evolving by less directed random processes, such as Brownian motion (BM). In this study, we use computer simulation to generate data under more realistic evolutionary scenarios and consider the statistical abilities of a variety of PCMs to estimate correlation coefficients from these data. We found that correlations estimated without taking phylogeny into account were often quite poor and never substantially better than those produced by the other tested methods. In contrast, most PCMs performed quite well even when their assumptions were violated. Felsenstein's independent contrasts (FIC) method gave the best performance in many cases, even when weak constraints had been acting throughout phenotypic evolution. When strong constraints acted in opposition to variance-generating (i.e., BM) forces, however, FIC correlation coefficients were biased in the direction of those BM forces. In most cases, all other PCMs tested (phylogenetic generalized least squares, phylogenetic mixed model, spatial autoregression, and phylogenetic eigenvector regression) yielded good statistical performance, regardless of the details of the evolutionary model used to generate the data. Actual parameter estimates given by different PCMs for each dataset, however, were occasionally very different from one another, suggesting that the choice among them should depend on the types of traits and evolutionary processes being considered.

Key words.—Adaptation, comparative method, computer simulation, generalized least squares, mixed model, phylogeny, spatial autoregression.

Received February 9, 2001. Accepted September 6, 2001.

Phylogenetic comparative methods (PCMs; Table 1) are now used widely to estimate parameters and to test hypotheses with interspecific data (Harvey and Pagel 1991; Martins and Hansen 1997). Recently, there has been some discussion about whether PCMs should be used as regularly as they have been, because some of these methods may not be appropriate for traits undergoing adaptive evolution (e.g., Westoby et al. 1995; Price 1997; Harvey and Rambaut 2000; Martins 2000). For example, Felsenstein's (1985) independent contrasts (FIC) method assumes that the data have resulted from an evolutionary process similar to Brownian motion (BM). Population geneticists have used BM mostly to describe traits undergoing random genetic drift (e.g., Felsenstein 1988), but most traits in comparative studies are of interest precisely because they are thought to have been subjected to natural or sexual selection (Martins 2000). Thus, FIC may not be appropriate for the analysis of these traits. Because models of character evolution in a PCM are usually translated into branch lengths on a phylogeny (see below), the problem has also been described as one of knowing the appropriate branch lengths for the phylogeny. Thus, some researchers have developed diagnostic techniques and other statistical tools to infer the correct branch lengths and to ensure that the PCMs are robust to violations of their assumptions (e.g., Gittleman and Kot 1990; Grafen 1989). Other researchers have developed more complex PCMs directly incorporating the effects of natural selection (e.g., Hansen 1997; Martins and Hansen 1997; Baum and Donoghue 2001; Orzack and Sober 2001). In this study, we use computer simulation to consider whether selection and adaptation pose a serious problem for existing PCMs. Specifically, we measure the impact of BM and realistic evolutionary alternatives on several PCMs, including FIC, Martins and Hansen's (1997) phylogenetic generalized least squares approach (PGLS), Lynch's (1991) phylogenetic mixed model (PMM; E. A. Housworth, E. P. Martins, and M. Lynch, unpubl. ms.), Cheverud et al.'s (1985) spatial autoregressive method (ARM), and Diniz-Filho et al.'s (1998) phylogenetic eigenvector regression (PVR).

All PCMs require at least some assumptions regarding the underlying evolutionary process. In most cases, assumptions are made in a two-step process (e.g., Martins and Hansen 1996). First, we obtain a reasonable phylogeny, and second, we assume that the characters of interest have evolved along that phylogeny in a particular way. The two assumptions are then used to infer how much we expect traits measured in phylogenetically related species to be similar based on shared ancestry. For example, we might start with a phylogeny developed using mtDNA data with branch lengths in units of relative time. We then assume that the trait has evolved along that phylogeny via a neutral, gradual process such that we expect the trait to increase or decrease randomly at each unit of time (e.g., BM). The result is that we expect taxa to be more phenotypically similar the more recently they diverged. Alternatively, we might start with the same phylogeny but make different evolutionary assumptions (e.g., that selection has been acting on the trait or that the organisms undergo change only at speciation events). There are numerous possibilities and it is difficult to choose among them.

© 2002 The Society for the Study of Evolution. All rights reserved.

TABLE 1. Acronyms used in this paper.

PCM	phylogenetic comparative method
TIPS	a nonphylogenetic approach; estimation of a Pearson product-moment correlation between two traits without incorporating any phylogenetic information
FIC	Felsenstein's (1985) independent contrasts method; equivalent to Grafen's (1989, 1991) standard regression and Martins and Hansen PGLS assuming a linear model
PGLS	phylogenetic generalized least-squares regression; extension of Felsenstein's (1985) and Grafen's (1989) methods in which a two-parameter, exponential weighting matrix (Martins and Hansen 1997) is applied to allow for flexibility in the underlying microevolutionary assumptions
PMM	Lynch's (1990) phylogenetic mixed model, using the restricted maximum-likelihood estimation procedure developed in E. A. Housworth, E. P. Martins, and M. Lynch (unpubl. ms.)
ARM	spatial autoregressive method
ARM1	as originally described by Cheverud et al. (1985)
ARM2	including a power parameter in the relationship matrix as suggested by Gittleman and Kot (1990)
PVR	phylogenetic eigenvector regression conducting a principal coordinate analysis of the relationship matrix and regressing the data on the eigenvectors of this matrix (Diniz-Filho et al. 1998)

Fortunately, for many phylogenetic analyses, all we really need is the final product—a description of how similar we expect phylogenetically related taxa to be (i.e., the phylogenetic dependence structure). How the traits reached that structure is evolutionarily interesting, but not a necessary assumption. Surprisingly, a wide variety of population genetic models of phenotypic evolution result in only two basic forms of phylogenetic dependence (Hansen and Martins 1996; Martins and Hansen 1997). Many models (e.g., random genetic drift or directional selection with a shifting optimum) result in phenotypic similarity being directly proportional to phylogenetic similarity, in what is usually termed a ''linear'' or ''clocklike'' model. Several other models (e.g., stabilizing selection, adaptation, primarily those involving some sort of constraint) result in phenotypic similarity decreasing more quickly, in an exponential or constrained fashion. Evolutionary constraints (e.g., selection) on phenotypes tend to lead to the loss of historical information (i.e., phylogenetic constraint, inertia, or heritability), making the information contained in a phylogeny less directly relevant to studies of traits measured in extant taxa. When the constraints are small, the exponential model approaches the linear model above. Thus, although we usually do not know whether linear or constrained models are more appropriate for any particular dataset, identifying the two allows us to bound the possibilities.

Existing PCMs (e.g., Table 1) differ primarily in the perspective that was used to develop them. Some methods were developed from an evolutionary perspective and rely directly on explicit assumptions regarding the evolutionary process. For example, FIC was derived directly from population genetic theory and requires an assumption that the traits of interest have evolved via the linear BM process outlined above. PGLS (Martins and Hansen 1997) expands the assumptions of FIC to allow for other evolutionary scenarios, but again requires that these scenarios be explicitly defined or estimated statistically. Similarly, PMM (Lynch 1991; E. A. Housworth, E. P. Martins, and M. Lynch, unpubl. ms.) is derived from quantitative genetics, and partitions phenotypic variation into phylogenetically heritable and nonheritable components. As described, it assumes that evolution occurs via BM followed by a burst of variation due to nonhistorical factors (e.g., environmental change accompanied by phenotypic plasticity) at the tips of the phylogeny. Another class of methods relies primarily on statistical assumptions. For example, ARM (Cheverud et al. 1985) requires an assumption that the traits are well described by an autoregressive model, PVR (Diniz-Filho et al. 1998) assumes that all the important variation in the relationship matrix is explained by only a few eigenvectors, and Grafen's (1989) independent contrast regression requires only that the error terms fit the usual regression requirements (i.e., normally distributed with a mean of zero and constant variance). None of these assumptions has been used previously in theoretical descriptions of phenotypic evolution, but all three methods provide reasonable statistical estimates with comparative data generated via a BM process (Grafen 1989; Martins 1996b; Diniz-Filho et al. 1998).

Theoretical consideration of different PCMs (e.g., Hansen 1996; E. A. Housworth, E. P. Martins, and M. Lynch, unpubl. ms.; Rohlf 2001) suggests that those with explicit evolutionary assumptions should be sensitive to violations of those assumptions and that those with statistical flexibility will be less sensitive to uncertainty in the underlying evolutionary model. We do not yet know, however, how the degree of assumption violation or statistical improvement corresponds to realistic evolutionary scenarios. For example, Díaz-Uriarte and Garland (1996) used computer simulation to show that FIC can perform poorly when its assumptions are violated. Their simulations, however, were based on algorithmic rather than evolutionary assumptions, and it is difficult to translate their simulation parameters into evolutionary terms (e.g., strength of selection). Thus, it is not clear whether FIC will reach the levels of poor performance illustrated in their study when applied to real data. Similarly, statistical methods such as ARM and PVR seem likely to perform better in a wider variety of situations, but may not perform better in the sorts of situations we regularly confront with comparative data.

In the current study, we use computer simulation to consider the statistical performance of a variety of PCMs under a variety of possible evolutionary conditions. Specifically, we apply each PCM to estimate a correlation coefficient describing the relationship between two traits. Although there are many other possibilities, correlation coefficients have broad practical appeal and have been used commonly in comparative studies and earlier simulation studies. We calculate these correlation coefficients using data generated under a model of gradual (linear) evolution and also under several possible models of constrained (exponential) evolution, thereby considering the performance of methods under the family of microevolutionary scenarios reviewed in Hansen and Martins (1996). Instead of focusing only on one method, we apply FIC (Felsenstein [1985] or equivalently, Grafen's [1989] standard regression or PGLS assuming a linear model), PGLS (Martins and Hansen [1997], estimating a sin-

gle constraint parameter), the PMM (Lynch 1991; E. A. Housworth, E. P. Martins, and M. Lynch, unpubl. ms.), ARM (Cheverud et al. [1985], as originally described, and with the Gittleman and Kot [1990] α parameter extension), and PVR (Diniz-Filho et al. 1998).

Methods

Generating the Data

Data were generated in Mathematica (Wolfram 1999) using an Ornstein-Uhlenbeck (OU) process. OU processes have been used to describe both neutral evolution and the evolution of phenotypes under various types of constraints (Lande 1976; Felsenstein 1988; Hansen and Martins 1996). It has been described as a ''rubber-band'' process in which the phenotype is held near a fixed optimum by a force, such that the pull toward the optimum is stronger as the phenotype drifts further away. When the pull toward the optimum is very small, the OU process approximates the BM model assumed by FIC and used in nearly all previous simulation studies (e.g., Martins and Garland 1991; Diniz-Filho 2001). There are several possible interpretations of the OU constraining forces, including, for example, most of the evolutionary models summarized in Hansen and Martins (1996).

Our simulation parameters can be described most easily in terms of the exponential covariance structure (Hansen and Martins 1996, eq. 5) that results from Lande's (1976) model of weak stabilizing selection and random genetic drift. Specifically, Lande's (1976) OU model involves two types of parameters. First, there is a set of BM parameters (in a matrix, **G**) indicating the increase in variance expected at each generation due to random and independent evolution along different branches of the phylogeny. For two traits, there is one of these variance-generating parameters for each trait ($\mathbf{G}[1, 1] = \sigma_X$ and $\mathbf{G}[2, 2] = \sigma_Y$) and a covariance ($\mathbf{G}[1, 2] = \mathbf{G}[2, 1] = \sigma_{XY}$) indicating, for example, genetic covariation between the traits. Second, there is a set of α parameters (in a matrix, **W**) that describe the magnitude of the constraint acting on phenotypic evolution (the width of the rubber band). Again, there might be separate constraining forces acting on each trait separately ($\mathbf{W}[1, 1] = \alpha_X$ and $\mathbf{W}[2, 2] = \alpha_Y$, forming an adaptive peak) and also a constraining force acting on the two simultaneously (in our case, $\mathbf{W}[1, 2] = \mathbf{W}[2, 1] = \alpha_{XY}$, e.g., describing an environmental force acting on both traits and forming an adaptive ridge). If the OU process has reached equilibrium, the true phenotypic correlation between two traits (estimated by most PCMs) is a function only of the evolutionary constraints, and can be calculated as: $-\alpha_{XY}/(\alpha_X\alpha_Y)^{0.5}$ (Hansen and Martins 1996). If the process has not reached equilibrium, the true phenotypic correlation may be somewhat different, but can also be calculated explicitly (Hansen and Martins 1996).

To generate realistic data under the Lande (1976) model, we must consider the actual values of σ, α, and the total length of the phylogeny. Building on Charlesworth (1984), the α parameters can be described as functions of the selective load—the percent of the population succumbing to mortality in each generation due to selection on the trait of interest. Specifically, we can translate between the α and σ parameters for a single trait and the selective load (L) using $\alpha_X = ([2L - L^2]/\sigma_X^2[1 - L]^2)^{0.5}$. There are several other interpretations of the parameters. For example, Hansen (1997) used an OU model to describe the evolution of adaptive phenotypes in a complex selective regime. In this case, the selective optimum shifts at defined points along the phylogeny and the α parameters describe the rate at which the phenotype responds to the new selective pressures. For simplicity, however, our discussion below focuses only on the Lande model.

We used a few preliminary runs to determine values of α at which the tested PCMs could not distinguish the results from BM data (minimum α) or from phylogenetically independent data (maximum α). In terms of the Lande model, for a phylogeny scaled to have total length of 50 million generations, we varied α parameters so that selective loads ranged across three orders of magnitude, starting with a minimum of 0.00000000000001% of the population dying off each generation due to selection on each trait individually and continuing up to 0.00000000001%. Because of the direct (albeit not linear) relationships between these parameters (see above for formula), our simulations are equivalent to considering a 1 million-generation phylogeny with a selective load ranging from 0.00000000003% to 0.00000002% or a 1000-generation phylogeny with selective load ranging from 0.00003% to 0.02%. Note, however, that the OU approximation proposed in the Lande (1976) model requires that selection be very weak and breaks down when α is large (e.g., in the 1000-generation example above).

We generated data on four phylogenies to consider the effects of phylogeny structure and relative sample size on PCM performance. All four structures were chosen as illustrations of the sorts of phylogenies typically employed in comparative analyses, and we make no attempt to evaluate the validity of the trees as hypotheses regarding true evolutionary relationships among taxa. We focused primarily on a 26-taxa hummingbird phylogeny (Bleiweiss et al. 1997; Fig. 1A). We then also applied a 13-taxa lizard phylogeny (Losos 1990; Fig. 1B) as an illustration of statistical performance with small sample sizes. Finally, we considered a 42-taxa carnivore phylogeny and a 50-taxa primate phylogeny to see the possible improvement of PCMs with larger sample sizes. Both of these latter phylogenies are parts of a composite supertree (constructed by Purvis 1995; Bininda-Emonds et al. 1999) containing South American taxa of interest to an earlier macroecological study (Diniz-Filho et al. 2000). Data generated on the 42-taxa phylogeny are expected to be particularly dependent, because this tree has many short branches near the tips. In contrast, the 50-taxa phylogeny leads to relatively independent data, because it includes several taxa evolving independently for long periods near the tips of the tree. We scaled branch lengths on all four phylogenies so that the total length from root to tips was identical.

Given an essentially infinite number of possible combinations of the six α and σ parameters on all four trees, we focused on a subset of 37 combinations (Table 2) that illustrate the performance of methods under different biologically realistic scenarios, and generated 1000 datasets for each phylogeny and parameter combination. In all cases, we set $\sigma_X = \sigma_Y = 1$, because these parameters cancel out of most of the calculations involved in estimating a correlation using the tested PCMs and thus had negligible impact on our re-

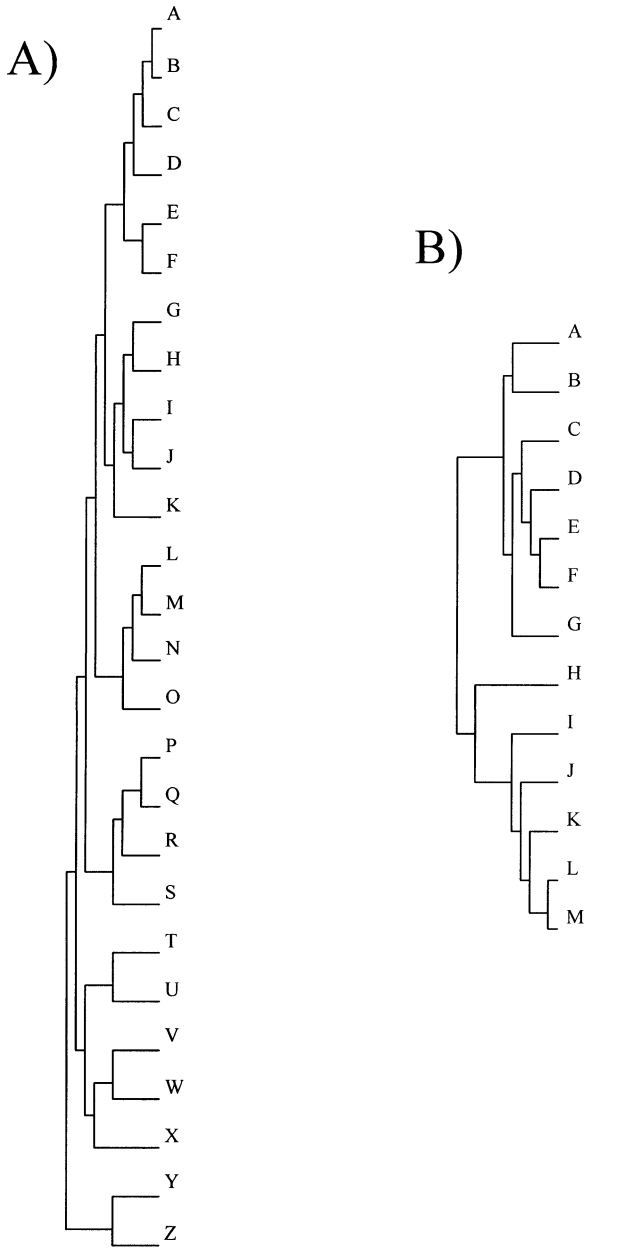

FIG. 1. Two of the four phylogenies used in this study. (A) A 26-taxa hummingbird phylogeny (Bleiweiss et al. 1997); (B) a 13-taxa lizard phylogeny (Losos 1990). Analyses were also conducted on a 42-taxa carnivore phylogeny and a 50-taxa primate phylogeny (Diniz-Filho et al. 2000).

sults. We began by focusing on the 26-taxa phylogeny and considering the impact of increasing α_X and α_Y separately, and of also including covariance restraining forces (α_{XY}) and variance-generating forces (σ_{XY}). We then confirmed the generality of our results by running a smaller set of simulations on each of the four phylogenies, varying the true phenotypic correlation between 0.0 and −0.9.

Phylogenetic Comparative Methods

Seven methods (Table 1) were used to calculate correlation coefficients between two traits for each dataset. We began by estimating a Pearson product-moment correlation coefficient between the raw trait values without incorporating the phylogeny (TIPS). Previous simulation studies considering data resulting from a BM process (e.g., Martins and Garland 1991) have shown that this nonphylogenetic approach performs poorly. Others, however, may recommend TIPS for data produced by processes other than BM.

We also applied three evolutionary PCMs, beginning with FIC, which gives correlation coefficients identical to those produced by Grafen's (1989) standard regression. FIC can also be considered to be a special case of either the PGLS (Martins and Hansen 1997) assuming that within-species variation is negligible and that phenotypic evolution is well described by BM, or the PMM (Lynch 1991; Housworth et al. in review) assuming that the phylogenetic heritability equals one. FIC involves transforming each trait into phylogenetically relevant contrasts based on a known phylogeny and then using those contrasts to calculate a correlation coefficient between the two traits. In applying this method, we assumed that the traits were evolving via a gradual or clocklike process (i.e., BM) such that the expected amount of change was proportional to the length of the branches on the phylogeny. FIC is known to perform quite well with BM-generated data, but can perform poorly with data generated under other models (Díaz-Uriarte and Garland 1996; E. A. Housworth, E. P. Martins, and M. Lynch, unpubl. ms.). We do not apply Díaz-Uriarte and Garland's (1996) algorithmic suggestion of transforming branch lengths used with FIC because this approach serves essentially the same statistical purpose as PGLS and PMM (see below), without offering an explicit evolutionary interpretation.

We also applied a form of PGLS (Grafen 1989; Martins and

TABLE 2. Simulation parameter values for the Ornstein-Uhlenbeck process considered in this study. Variance-generating parameters acting on each trait separately (σ_X and σ_Y) were set equal to one in all cases. Covariance between variance-generating forces is indicated as σ_{XY}, acting on both traits simultaneously. Constraining forces acting on each trait separately are indicated as α parameters (α_X and α_Y) and are given in terms of a phylogeny with total length of 50 million generations. Covariance constraining force is given as a correlation $\rho_\alpha = \alpha_{XY}/(\alpha_X\alpha_Y)^{0.5}$ True phenotypic correlations are given as ρ_{Eq} (true value expected at equilibrium). For the two cases in which equilibrium is not expected to have been reached in our study, we offer also the nonequilibrium true correlation (in parentheses). Finally, a_C is an estimate of the composite α parameter estimated by the PGLS method.

Tree	σ_{XY}	$\alpha_X = \alpha_Y$	ρ_α	ρ_{Eq}	a_C
all	0.0	0.00000000	0.0	0.0	0.0
all	0.0	0.00000005	0.0	0.0	2.5
all	0.0	0.00000050	0.0	0.0	25.0
all	0.0	0.00000050	0.5	−0.5	12.5
all	0.0	0.00000050	0.9	−0.9	2.5
all	−0.5	0.00000100	0.5	−0.5	37.5
all	0.5	0.00000100	0.5	−0.5	12.5
26	0.0	0.00000001	0.0	0.0	0.0
26	0.0	0.00000010	0.0	0.0	5.0
26	0.0	0.00000030	0.0	0.0	15.0
26	0.0	0.00000005	0.5	−0.5 (−0.47)	1.3
26	0.0	0.00000100	0.5	−0.5	25.0
26	0.0	0.00000005	0.9	−0.9 (−0.76)	0.3
26	0.0	0.00000100	0.9	−0.9	5.0
26	−0.9	0.00000100	0.5	−0.5	7.5
26	0.9	0.00000100	0.5	−0.5	2.5

Hansen 1997), estimating a correlation between two traits based on the results of a generalized least-squares (GLS) regression. As recommended in Martins and Hansen (1997), we set the elements of the PGLS weight matrix to $s_C \exp[-a_C t_{ij}]$, where t_{ij} is the phylogenetic distance separating taxa i and j, s_C is estimated directly from the regression, and a_C is estimated using a maximum-likelihood grid search. Although PGLS allows for the incorporation of within-species variation, herein we assumed that such variation was negligible to make this method comparable to the others. Although it is also possible to use PGLS to estimate all six parameters in the Lande model, it is exceedingly difficult to estimate that many parameters with the sample sizes commonly available for phylogenetic analysis, and doing so restricts the interpretation of the PGLS results to this specific model. Instead, the PGLS applied in the current study might be viewed as an extension of Felsenstein (1985) in which a single extra parameter (a_C) is used to describe the strength of an evolutionary constraint acting on the phenotype. We can use the multivariate covariance developed for this model in Hansen and Martins (1996, eq. 8) to relate the composite a_C approximately to the six parameters of the Lande model as: $a_C \cong \min(\lambda_i + \lambda_j)$, where λ_i are the eigenvalues of $\mathbf{QW}$, where $\mathbf{Q}$ is the matrix of σ parameters and $\mathbf{W}$ is the matrix of α parameters. Thus, although PGLS comes closest of all the tested PCMs to matching the simulation procedure, the match is perfect only when α_{XY} equals zero.

For a last evolutionary approach, we applied the PMM to the data (Lynch 1991; E. A. Housworth, E. P. Martins, and M. Lynch, unpubl. ms.). The PMM draws an analogy with quantitative genetics to partition data into an overall mean (μ), heritable (a), and nonheritable (e) components ($y = \mu + a + e$). With the PMM, we estimate the ancestral state at the root of the phylogeny (the grand mean), and partition remaining interspecific variation into phylogenetically heritable (passed on between taxa along the phylogeny) and nonheritable components (e.g., phenotypic plasticity). Herein, we estimated the total phenotypic correlation (ρ) of a bivariate form of the PMM using the new algorithm and REML estimators proposed in E. A. Housworth, E. P. Martins, and M. Lynch. (unpubl. ms.). When nonheritable components are estimated to be zero, this method is equivalent to FIC. The PMM applied herein requires estimation of six parameters, including two phylogenetic heritabilities, two total phenotypic variances, and two covariances (between additive and nonadditive components) to calculate a total phenotypic correlation between two traits. Although the PMM is based on explicit evolutionary assumptions, these do not directly match the assumptions of this simulation study. Nevertheless, the estimation of six parameters may give this method enough statistical flexibility to yield reasonable estimates of the phenotypic correlation despite the violation of its assumptions.

We then applied three more statistical methods, including two forms of spatial autoregression models. Although the assumptions underlying these methods have evolutionary implications, methods were developed from statistical rather than evolutionary principles, and the interpretation of these assumptions is not obvious (e.g., Martins and Hansen 1996). ARM1 is Cheverud et al.'s method (1985) using $\mathbf{y} = \rho\mathbf{Wy} + \boldsymbol{\varepsilon}$ to partition variation in each trait ($\mathbf{y}$) into phylogenetic ($\rho\mathbf{Wy}$) and specific ($\boldsymbol{\varepsilon}$) effects, where $\mathbf{W}$ is a divergence matrix describing the phylogeny (except that diagonals are set equal to zero) and ρ is a constant that is estimated as part of the procedure. Note that although Rohlf (2001) proposes an alternative procedure for estimating the autoregression coefficient (ρ, testing a wider range of possibilities), we apply herein only the procedure given with the original description of the method (Cheverud et al. 1985). Once the variation in each trait was partitioned, we estimated a Pearson-product moment correlation between the specific effects for the two traits. ARM2 is the same model but includes an extra power parameter to which each element of the $\mathbf{W}$ matrix is raised (Gittleman and Kot's [1990] α). Despite earlier evidence that there is little improvement due to ARM2 with small number of species (Martins 1996b), we test this version again because the parameter offers ARM some statistical flexibility, shaping the $\mathbf{W}$ relationship matrix to fit the data.

Phylogenetic eigenvector regression (PVR; Diniz-Filho et al. 1998; Diniz-Filho 2001) is similar to the ARM methods and to PMM in that it also partitions variation into phylogenetic and specific components. In this case, however, principal coordinate analysis is used to extract the most relevant eigenvectors of the phylogenetic distance matrix. The interspecific data ($\mathbf{y}$) are then regressed on these eigenvectors ($\mathbf{X}$, rather than $\mathbf{Wy}$ of ARM), estimating a variable number of regression coefficients ($\boldsymbol{\beta}$), and resulting in a set of residual or specific values that are independent of the phylogeny. The method provides considerably more statistical flexibility than ARM in that it may involve estimation of several regression parameters ($\boldsymbol{\beta}$) as opposed to a single ρ. (See Diniz-Filho et al. [1999] for a discussion of how to interpret these regression parameters.) Specifically, applying broken-stick criteria (as recommended in Diniz-Filho et al. 1998), we estimated two regression parameters for the 13-taxa tree, four for the 26-taxa tree, three for the 42-taxa tree, and five for the 50-taxa phylogeny. Although Diniz-Filho et al. (1998; Diniz-Filho 2001) only propose PVR as an alternative to ARM in estimating phylogenetic inertia, correlation between the residuals (specific values) for two traits might also be used to consider the relationship between traits, in much the same way as in ARM. Earlier simulation studies suggested that both ARM and PVR methods perform reasonably well under BM evolution, but that PVR may provide better estimates of inertia in some cases (Diniz-Filho et al. 1998).

PVR calculations were conducted using a Basic program (Coelho and Diniz-Filho 2000). FIC and ARM methods were calculated using COMPARE (Martins 2001). Calculations for PGLS and PMM for this study were conducted in SAS (SAS Institute 1990), but are also now available in COMPARE (Martins 2001).

Comparing Phylogenetic Comparative Methods

For each analysis of 1000 datasets using a particular PCM, we calculated several measures of method performance. First, to determine parameter estimation abilities, we compared the mean correlation coefficient for each method (e.g., FIC_i, where i refers to a particular dataset) to the true phenotypic correlation for that run (e.g., θ_i; calculated directly from the parameters used to generate the data: $-\alpha_{XY}/[\alpha_X\alpha_Y]^{0.5}$). The

TABLE 3. Pearson product-moment correlation coefficients between results for each PCM on each run of the simulation and the results of other tested methods. Upper triangle in each matrix refers to results for data simulated under a Brownian motion model, when FIC is expected to give the best performance. Values below the diagonal are results for data simulated under an Ornstein-Uhlenbeck model with strong evolutionary constraints ($\sigma_X = \sigma_Y = 1$, $\sigma_{XY} = 0$, $\alpha_X = \alpha_Y = 0.0000005$, $\alpha_{XY} = 0$), when TIPS is expected to give the best performance. Thus, boldface values are correlations between method results and the best results for each dataset. See Table 1 for method acronyms.

13 taxa	TIPS	FIC	PGLS	PMM	ARM1	PVR	26 taxa	TIPS	FIC	PGLS	PMM	ARM1	PVR
TIPS		**0.73**	0.88	0.91	0.80	0.45	TIPS		**0.74**	0.85	0.85	0.75	0.62
FIC	**0.87**		**0.89**	**0.81**	**0.85**	**0.84**	FIC	**0.65**		**0.95**	**0.82**	**0.80**	**0.80**
PGLS	**0.98**	0.90		0.92	0.88	0.68	PGLS	**0.97**	0.74		0.91	0.85	0.83
PMM	**0.96**	0.76	0.94		0.87	0.61	PMM	**0.93**	0.46	0.87		0.85	0.82
ARM1	**0.88**	0.71	0.87	0.86		0.65	ARM1	**0.93**	0.58	0.91	0.88		0.70
PVR	**0.92**	0.92	0.94	0.85	0.77		PVR	**0.90**	0.72	0.91	0.78	0.83	
42 taxa	TIPS	FIC	PGLS	PMM	ARM1	PVR	50 taxa	TIPS	FIC	PGLS	PMM	ARM1	PVR
TIPS		**0.36**	0.53	0.30	0.34	0.26	TIPS		**0.55**	0.65	0.68	0.70	0.53
FIC	**0.52**		**0.92**	**0.56**	**0.73**	**0.59**	FIC	**0.78**		**0.95**	**0.85**	**0.76**	**0.76**
PGLS	**0.82**	0.90		0.57	0.76	0.61	PGLS	**0.99**	0.81		0.89	0.80	0.78
PMM	**0.92**	0.38	0.70		0.50	0.42	PMM	**0.96**	0.68	0.94		0.80	0.77
ARM1	**0.95**	0.44	0.74	0.86		0.86	ARM1	**0.97**	0.73	0.96	0.92		0.63
PVR	**0.96**	0.53	0.80	0.86	0.92		PVR	**0.95**	0.78	0.95	0.91	0.92	

average difference between these two is our estimate of the bias for each method (e.g., bias = Σ [FIC$_i$ − θ_i]/1000, where the sum is over the 1000 runs for that particular combination of evolutionary parameters). We also calculated the mean squared error (MSE = Σ [FIC$_i$ − θ_i]^2/1000). The MSE is a more general measure that incorporates both the bias and an estimate of the accuracy or sharpness of the parameter estimate. The square root of the MSE gives us the root mean squared error (RMSE), a general measure of statistical performance of the method (with large values indicating that the estimator gives unreliable estimates). Note that both the bias and the RMSE are on the scale of the trait such that a RMSE of 0.2 for a normally distributed parameter estimate could be used to generate a rough confidence interval of ±0.4 (=1.96 × RMSE) about the result. When bias was small and the true phenotypic correlation was zero, we also sometimes rephrase this measure as Type I error rate (in this case, the probability of erroneously deciding that the phenotypic correlation is not zero when testing against standard tables) for comparison with other simulation studies. Note, however, that PMM hypothesis tests should probably be conducted against simulated null distributions rather than against standard tables (E. A. Housworth, E. P. Martins, and M. Lynch, unpubl. ms.). Finally, we use Pearson product-moment correlations to consider similarities among method results for each run of the simulation.

RESULTS

In general, method performance was quite good. For example, in most cases, correlations between PCM results and the best possible results (i.e., results in each case for the best of the PCMs tested) were greater than 80%, with all PCMs giving very good estimates of the evolutionary relationship between two traits (Table 3). Bias was negligible for all methods when using data generated under BM ($\sigma_{XY} = \alpha_X = \alpha_Y = \alpha_{XY} = 0$) and also in most cases when evolutionary constraints had been acting during phenotypic evolution (see below for exceptions). In those simulations for which the true phenotypic correlation was zero, Type I error rates for hypothesis tests of whether a correlation coefficient is significantly different from zero were also reasonable. For example, the Type I error at an expected *P*-value of 0.05 was almost always less than 0.10 (e.g., Fig. 2).

As expected, TIPS and FIC were the most extreme methods, yielding the best and the worst estimates of correlation coefficients depending on whether their assumptions were met (Figs. 2, 3; Table 3). FIC yielded the best performance (lowest RMSE) for BM data (Figs. 2A; white bars of Fig. 3), and performed well even with some simple evolutionary constraints (Figs. 2B–D). Increasing constraint caused the data to become less variable and more independent of the phylogeny, leading eventually to rather poor performance of FIC (e.g., Figs. 2E, F; black bars of Fig. 3). Data resulting from evolution under strong constraints (Figs. 2F; black bars of Fig. 3) were well analyzed by TIPS, which also gave reasonable performance with intermediate levels of constraint (Figs. 2B–D). TIPS performed quite poorly, however, with weak evolutionary constraints (Figs. 2A, B; white bars of Fig. 3). Although these general results were true for simulations on all four phylogenies, performance for TIPS and FIC was more extreme on the highly dependent 42-taxa tree (Fig. 3B; Table 3).

In general, the observed decrease in TIPS performance was greater than the observed decrease in FIC performance when their respective assumptions were violated (Fig. 3; Table 3). TIPS RMSE values for BM data were as high as 2.5 times greater than the optimal (Fig. 3B), such that the true 95% confidence interval for a correlation coefficient obtained using this approach could be as large as ±0.8. FIC RMSE was also occasionally large, but never exceeded 1.8 times the optimal, even when strong evolutionary constraints were acting along the 42-taxa tree (Fig. 3B). Unfortunately, bias was responsible for much of the increase in FIC RMSE when this PCM was applied to non-BM data. Although all methods were biased when restraining forces were very small (e.g., Fig. 4A), this bias usually disappeared as α_X, α_Y, and α_{XY} increased (e.g., Fig. 4B, C). FIC often retained a substantial bias, even when the restraining forces were large (e.g., Fig.

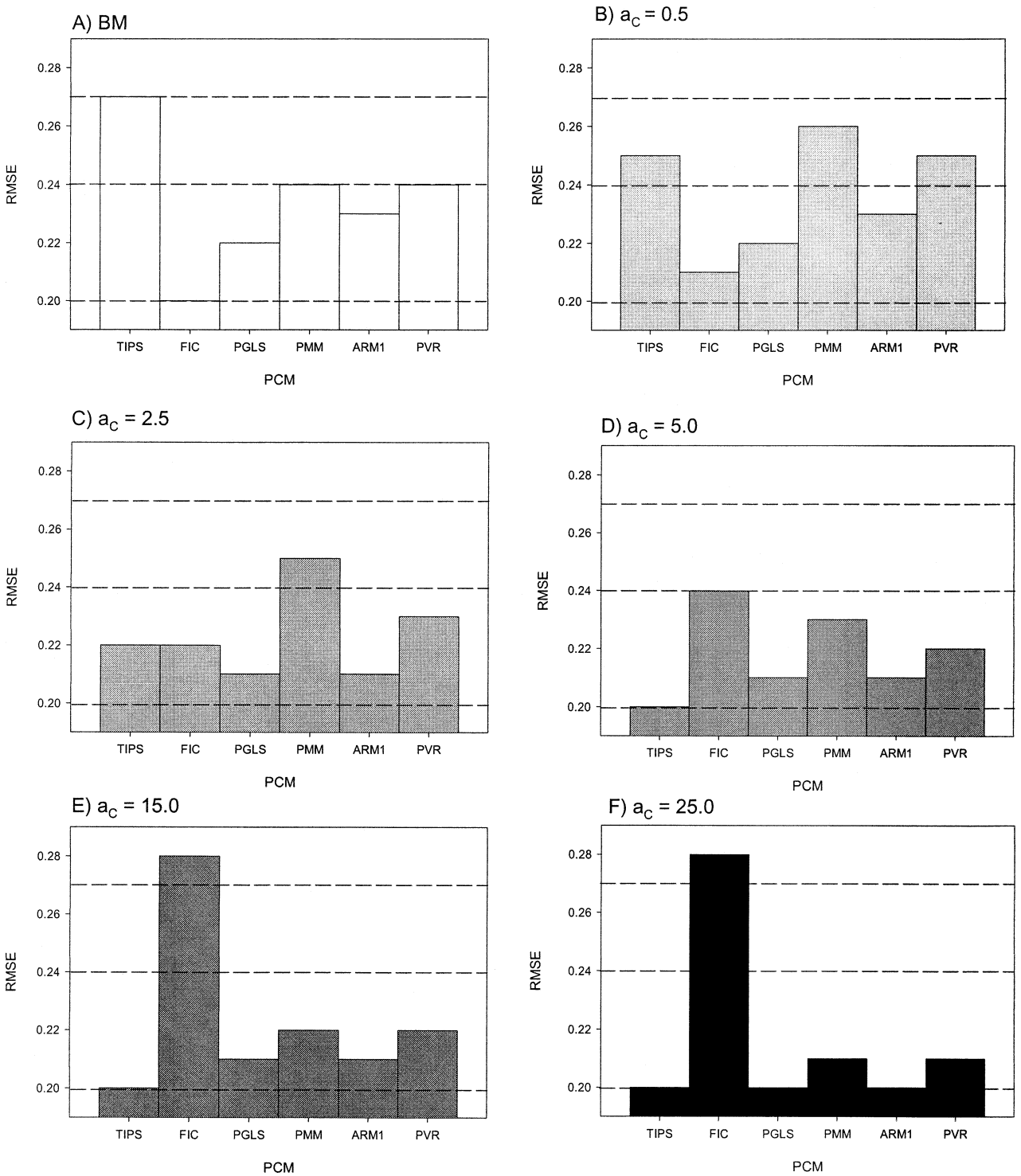

FIG. 2. Root mean squared error (RMSE) for each phylogenetic comparative method on the 26-taxa phylogeny when applied to data generated under (A) a Brownian motion model of phenotypic evolution, and an Ornstein-Uhlenbeck model of constrained phenotypic evolution (B–F, darker bars correspond to larger constraints). See Table 1 for method acronyms. In all cases, $\sigma_X = \sigma_Y = 1$, $\sigma_{XY} = 0$, $\rho_\alpha = 0.0$, and the total phenotypic correlation = 0.0. Constraints acting on the two traits separately were increased from $\alpha_X = \alpha_Y = 0.00000001$ to 0.0000005 (see Table 2 for actual values). Dashed horizontal lines translate RMSE values into Type I error at a true *P*-value of 0.05 for this tree and for these models of evolution and assuming that bias is negligible. RMSE of 0.20 indicates a method for which the Type I error at a true *P*-value of 0.05 would actually equal 0.05. RMSE of 0.24 indicates a method for which the Type I error at a true *P*-value of 0.05 equals 0.10. RMSE of 0.27 indicates a method for which the Type I error at a true *P*-value of 0.05 equals 0.15.

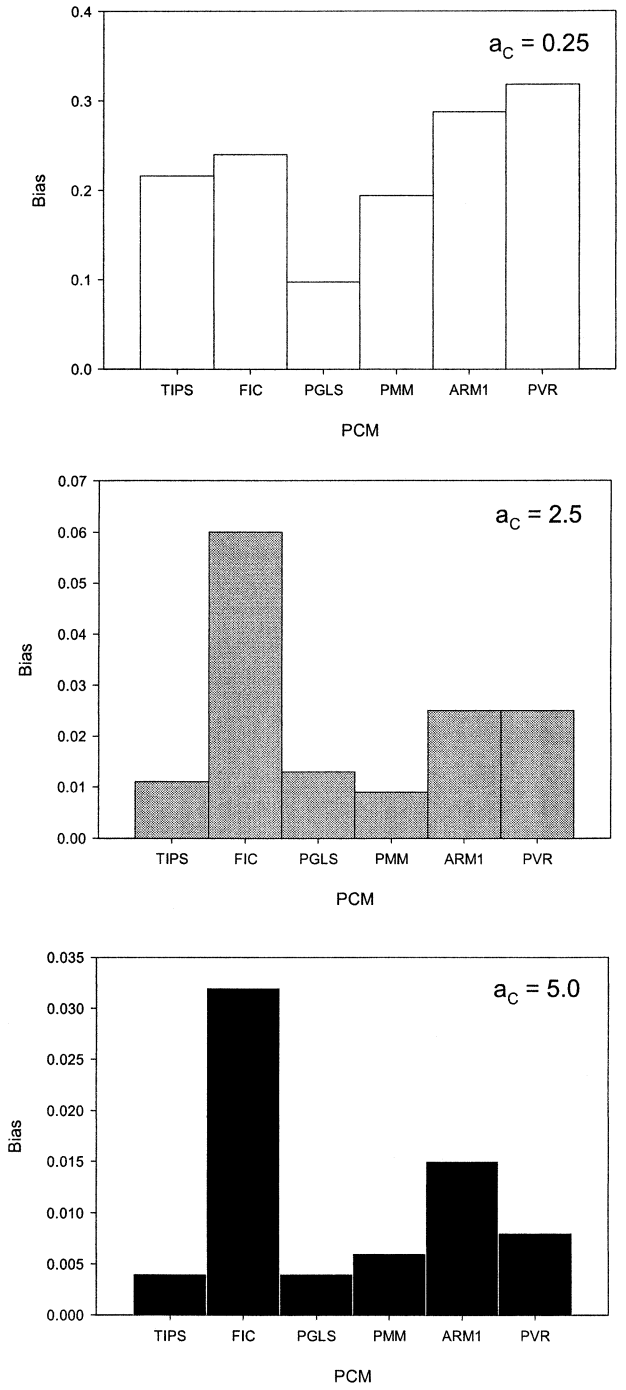

FIG. 3. Root mean squared deviation (RMSE) for each phylogenetic comparative method on when data were generated on the 26-taxa phylogeny under complex versions of the Ornstein-Uhlenbeck model of phenotypic evolution. See Table 1 for method acronyms. In all cases, $\sigma_X = \sigma_Y = 1$, $\sigma_{XY} = 0$, $\rho_\alpha = 0.9$, and the total phenotypic correlation $= -0.9$. Differences across graphs are in the absolute magnitude of the constraining forces (or equivalently in the length of time in which they act). (A) $\alpha_X = \alpha_Y = 0.00000005$; (B) $\alpha_X = \alpha_Y = 0.0000005$; (C) $\alpha_X = \alpha_Y = 0.000001$.

4B, C). (Note that ARM, but not PVR, also retained a slight bias even when the restraining forces were large; Fig. 4 B, C.) FIC was also usually biased in the direction of σ_{XY} and thus had much greater RMSE when σ_{XY} acted in opposition to α_{XY}. (Fig. 5). Variation in the σ parameters had very little impact on other PCMs.

Other methods were intermediate, yielding slightly lower RMSE with less variable data and giving generally unbiased and accurate results in most cases (Figs. 2–5; Table 3). As expected given the match in assumptions with the simulation procedure, PGLS performed well, providing good estimates of the single a_C parameter even with the 13-taxa phylogeny. Performance improved, as expected, with increasing sample size, because the method was better able to estimate the a_C parameter. For example, with BM data and 13 taxa, RMSE for PGLS was a little larger than that for FIC (0.35 vs. 0.30, respectively). Still with as few as 26 taxa, RMSE for PGLS was already similar to that for FIC (0.22 vs. 0.20), and values of RMSE for the two methods were nearly identical for the 42- and 50-taxa phylogenies. Correlation coefficients between results for PGLS and FIC when applied to BM data were always at least 90%. In contrast to FIC, however, even with strong evolutionary constraints, PGLS bias was negligible and performance was usually also near the best of all methods tested (Table 3). In terms of Type I error rates, PGLS error was a little high on the 13-taxon phylogeny (e.g., 0.10 for a true α of 0.05), but not higher than 0.08 for other trees. Removal of an extra degree of freedom for estimation of the a_C parameter (not done here) should be sufficient for this PCM to produce accurate hypothesis tests.

PMM, ARM, and PVR methods were intermediate, being not as effective as FIC and PGLS, but giving roughly reasonable results in most cases (Figs. 2–5; Table 3). Results for ARM1 and ARM2 were virtually identical (correlation between these was usually greater than 0.98), and are thus presented only for ARM1, which provided slightly better results in all cases. ARM2 performed especially poorly with BM data generated along the 42-taxa phylogeny. PVR results were also similar to ARM in terms of RMSE, usually yielding less bias but more variability than did ARM methods. Actual values, however, often varied considerably from run to run, with ARM and PVR sometimes showing as little as 60% similarity (Table 3).

PMM and PVR performed slightly worse than other PCMs with simple evolutionary constraints (e.g., acting on the traits independently; Fig. 2), perhaps due in part to difficulties in estimating so many parameters (six and four, respectively, for 26-taxa phylogeny) using data from a relatively small number of taxa. Type I error rates with BM data were sometimes a little higher than acceptable (e.g., 0.21 when it should have been 0.05 for PVR on the 42-taxa tree), but were never as bad as TIPS (0.46 when it should have been 0.05 for the same tree). Note also that Type I error rates were calculated by comparing PCM results to standard tables. The randomization tests recommended for PMM in E. A. Housworth, E. P. Martins, and M. Lynch (unpubl.ms.) would always yield accurate Type I error rates. PMM performed especially well with complex evolutionary constraints, sometimes yielding lower RMSE than PGLS, despite the mismatch in assump-

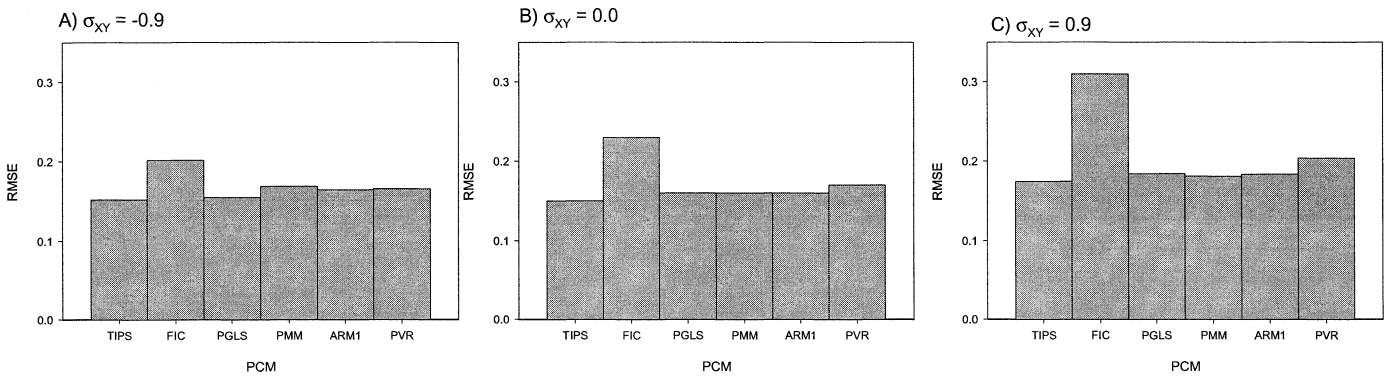

FIG. 4. Root mean squared deviation (RMSE) for each phylogenetic comparative method on the 26-taxa phylogeny when the bivariate variance generating parameter (σ_{XY}) was varied. See Table 1 for method acronyms. In all three cases, $\sigma_X = \sigma_Y = 1$, $\alpha_X = \alpha_Y = 0.000001$, $\rho_\alpha = 0.5$, and the total phenotypic correlation is -0.5. (A) $\sigma_{XY} = -0.9$; (B) $\sigma_{XY} = 0.0$; (C) $\sigma_{XY} = 0.9$.

tions (e.g., black bars in Fig. 3B, D; lower half of matrices in Table 3).

DISCUSSION

Our results are refreshingly optimistic. Most of the phylogenetic comparative methods (PCMs) tested did well in most situations (and better than a nonphylogenetic approach, TIPS) even when strong evolutionary constraints had been acting on the traits. PCMs were usually unbiased, estimates of the error were reasonable, and parameter estimates from different methods analyzing the same dataset were usually roughly comparable ($r > 0.7$). For example, when testing the hypothesis that a correlation coefficient was different from zero, although Type I error at an expected *P*-value of 0.05 was as high as 0.46 for TIPS, it was almost always less than 0.10 for all PCMs tested. The exception was FIC, which was robust to minor violations of its assumptions, but yielded biased estimates when its assumptions were seriously violated. All of the other PCMs, even those whose assumptions were seriously violated by the data used in this simulation study (e.g., PMM; Lynch 1991; E. A. Housworth, E. P. Martins, and M. Lynch, unpubl. ms.), performed similarly and well in most cases.

Phylogeny Does Matter

As with other simulation studies (e.g., Martins and Garland 1991), our results again confirm that how and whether phylogeny is incorporated can make a difference to the results of a comparative analysis, and that a nonphylogenetic analysis can often lead to very poor estimates. For example, RMSE for nonphylogenetic estimates of correlation coefficients (TIPS) in this study was sometimes as high as 0.36, such that instead of getting a 95% confidence interval, we would have had a 64% confidence interval. Similarly, the relationship between TIPS results and the best PCM estimates was sometimes as low as 36%. Only when evolutionary constraints were strong was TIPS a reasonable choice. Even then, TIPS performance was never more than slightly better (5% lower RMSE) than that of the PCMs, except FIC. Although TIPS may be useful when phylogenetic information is unavailable or unreliable, existing randomization tests (e.g., Martins and Housworth 2001) and Bayesian approaches (e.g., Huelsenbeck et al. 2000; Huelsenbeck and Bollback 2001) are likely to be more effective. Because PCMs also provide a number of other advantages (e.g., extra information about the evolutionary process provided by parameters such as α and h^2), there seems to be little to lose and much to gain by incorporating available phylogenetic information.

As suggested by some authors (e.g., Price 1997; Harvey and Rambaut 2000), taking phylogeny into account did not always solve the problem if the wrong method was chosen for a particular situation. For example, as in other studies (Díaz-Uriarte and Garland 1996; Price 1997; Harvey and Rambaut 2000; E. A. Housworth, E. P. Martins, and M. Lynch, unpubl. ms.), our results show that FIC can perform poorly when its assumptions are seriously violated. Specifically, FIC tended to overestimate the importance of variance-generating (i.e., BM) forces. When evolutionary constraints acted in opposition to the BM forces, FIC yielded correlation coefficients that were biased in the direction of the BM forces. Fortunately, the problem is easy to solve. All of the other tested PCMs (both evolutionary and statistical) provided enough flexibility to yield good statistical performance, even when their own assumptions were seriously violated. In our study, even a single extra parameter (e.g., as in PGLS) was usually sufficient to eliminate the FIC bias, and having too many extra parameters or parameters of the wrong type (e.g., as in PMM or PVR) did not seem to lead to further statistical problems.

FIC, as originally described, relies on an assumption that phenotypic evolution is well described by BM—a powerful model that has been used to describe a variety of types of phenotypic evolution, including evolution under directional selection and random genetic drift (Felsenstein 1988; Hansen and Martins 1996). Our results confirm earlier suggestion (Price 1997; Díaz-Uriarte and Garland 1996; Harvey and Rambaut 2000; E. A. Housworth, E. P. Martins, and M. Lynch, unpubl. ms.) that caution should be used in applying FIC when its BM assumptions are seriously violated. When the violations were less serious, however, we often found that FIC still gave the best performance of all PCMs tested. For example, using the Lande (1976) model, when less than about 0.000000001% of the population suffered mortality each generation throughout a 1 million–generation phylogeny due to selective forces acting on the traits of interest (or about

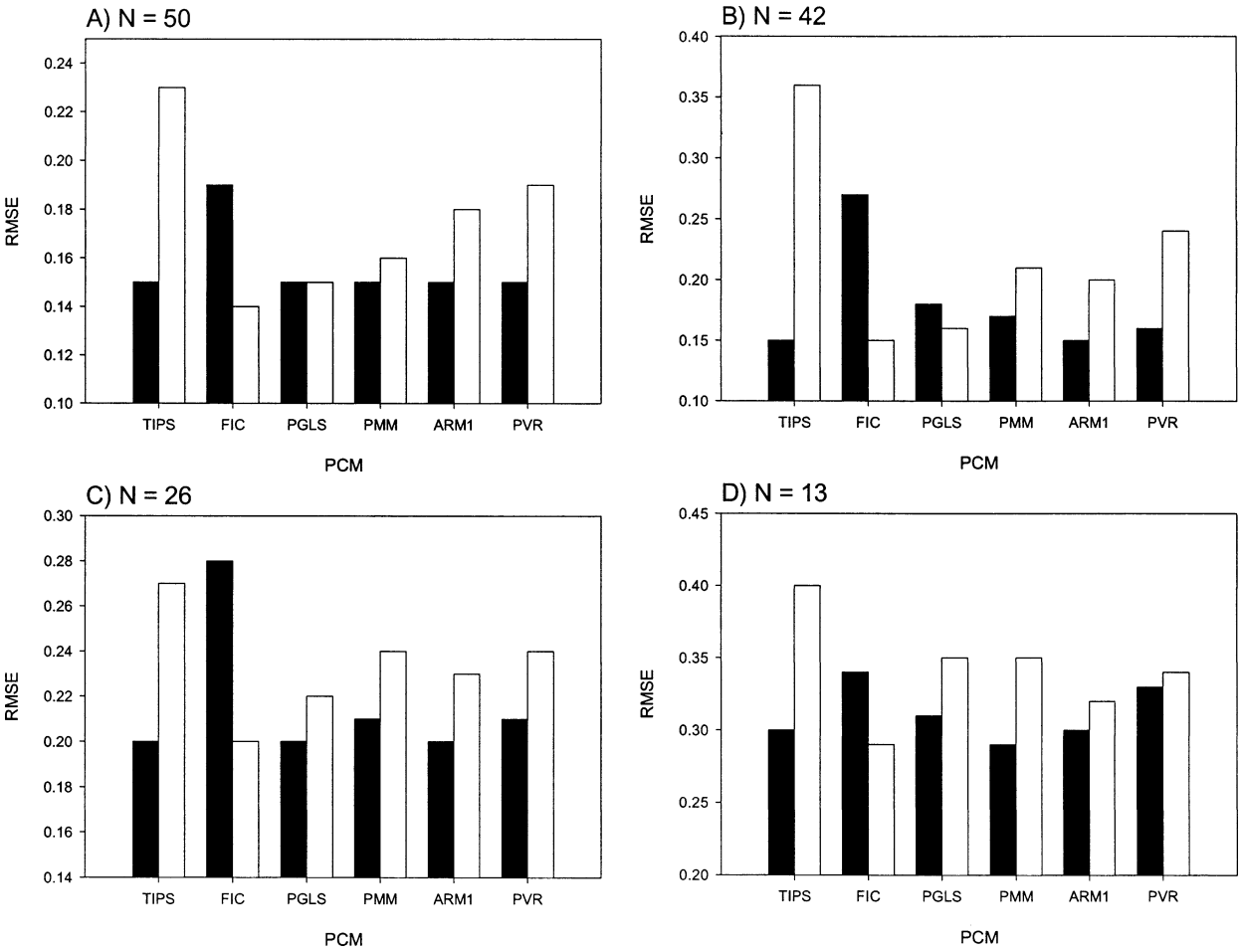

FIG. 5. Root mean squared error (RMSE) for each phylogenetic comparative method when data were generated under a Brownian motion model (white bars; $\sigma_X = \sigma_Y = 1$, $\sigma_{XY} = 0$, $\alpha_X = \alpha_Y = \alpha_{XY} = 0$) and an Ornstein-Uhlenbeck model of phenotypic evolution with strong evolutionary constraints (black bars; $\sigma_X = \sigma_Y = 1$, $\sigma_{XY} = 0$, $\alpha_X = \alpha_Y = 0.0000005$, $\alpha_{XY} = 0$). See Table 1 for method acronyms. Data were generated along (A) 50-taxa phylogeny; (B) 42-taxa phylogeny; (C) 26-taxa phylogeny (Fig. 1A); (D) 13-taxa phylogeny (Fig. 1B).

0.000000000001% for 50 million generations, or 0.001% for 1000 generations), FIC was still the best choice. Other PCMs lose a little accuracy in estimating extra, apparently unnecessary, parameters. In the most extreme case, with BM data generated along the 42-taxa highly dependent phylogeny, correlations between FIC and other PCM results fell as low as 60%. Only with stronger selective pressures and/or longer periods of time, did other PCMs become more effective.

Unfortunately, it is difficult to know when FIC assumption violations are sufficiently serious to warrant caution. Proposed diagnostic techniques are of limited utility because they are not directly linked to parameter estimates (e.g., correlation coefficients). Even when a rates test (e.g., Martins 1994) or a residual plot of independent contrasts versus their standard deviations (Díaz-Uriarte and Garland 1996) suggests that a BM model does not fit the data well, FIC (as originally described, without branch length transformations or other corrections) may still give very good statistical performance and be preferred because of its straightforward interpretation. One possibility would be to apply FIC always in combination with at least one other PCM, interpreting the results more cautiously if the results for the two PCMs differ dramatically.

Choosing among Methods

In most cases, the tested PCMs (except FIC) gave results that were similar to each other, with correlations between methods exceeding 80%. Nevertheless, with small sample sizes or highly dependent data, the results for different PCMs on a single dataset could differ dramatically from each other, with correlations among results from different PCMs reaching as low as 0.4 (Table 3). Thus, although any PCM (except FIC) can be used generally to get reasonable parameter estimates and hypothesis tests, they may occasionally yield dramatically different results with any single dataset. Some of these differences may be due to limitations in how we applied the PCMs. For example, allowing negative a_C-values in PGLS, negative phylogenetic heritabilities in the PMM, or the wider range of ρ-values for the ARM recommended by Rohlf (2001) may decrease the differences among these methods. These PCMs also differ in the number and type of extra statistical parameters that are estimated as well as in their evolutionary interpretations.

Surprisingly enough, two parameters (as in PGLS and ARM) were often sufficient to describe the dependence in-

troduced by the six-parameter OU model underlying the data in this study. The simplified form of the PGLS (adding a single a_C parameter to the FIC model) performed well even with small sample sizes (e.g., 13 taxa) and highly dependent data (e.g., from evolution along the 42-taxa phylogeny). Type I error rates were only slightly elevated and could easily be corrected by subtracting an extra degree of freedom for estimation of the composite a_C parameter. Although PGLS clearly gave the best performance in the current study, this is likely due to the clear match between PGLS and the evolutionary model used to generate the data in this study.

ARM methods also performed well, but could give results with a single dataset that were different from those produced by PGLS. As in Martins (1996b), the Gittleman and Kot (1990) power parameter did not improve the statistical performance of ARM1 with the phylogenies and sample sizes applied in this study. In most cases, results for ARM1 and ARM2 were nearly identical, and when they differed, ARM1 gave better estimates. Unlike Martins (1996b), we found that ARM statistical performance was reasonably good even for the 13-taxa phylogeny, suggesting that the pathologies described for a 15-taxa phylogeny in Martins (1996b) were due to something more than the small number of taxa (e.g., phylogeny shape).

As expected, the two PCMs that estimated the most parameters (PVR and PMM) performed less well than other PCMs with data generated under simple models of evolution (e.g., BM). By choosing only two to five of the eigenvectors to describe each relationship matrix, the PVR throws out information regarding all but the deepest parts of each phylogeny. This method is thus robust to minor errors in tree structure and may be particularly useful when the phylogeny and/or model of phenotypic evolution are poorly unknown. Although the evolutionary models underlying the simulated data were different from that underlying the PMM, this method offers a great deal of statistical flexibility and is specifically designed to partition variation due to different and possibly opposing forces. Again, in our study, performance of both PVR and PMM was roughly comparable to ARM (RMSE was sometimes higher and sometimes lower), even though results for each dataset were sometimes rather different from those produced by ARM or by each other (r as low as 42%). Much as in Diniz-Filho et al. (1998), PVR performed well even with small phylogenies. PMM performed especially well when strong selective constraints acted in opposition to variance-generating forces.

Overall, it was difficult to choose among methods (other than FIC and TIPS) in most situations, based purely on their abilities to estimate correlation coefficients from data generated in this study. Instead, given the substantial differences among PCMs in terms of their interpretations, researchers may want to choose among PCMs based on the types of characters and questions involved in a particular study. For example, researchers with relatively small datasets may strike a compromise between statistical and evolutionary flexibility by choosing PGLS and thereby avoiding the serious problems possible with TIPS and FIC, while also being realistic about the number of parameters that can be estimated from their data. Statistically inclined researchers may prefer the flexibility of PVR, using model-fitting procedures to determine the best number of eigenvectors to include and then interpreting those chosen as evidence of exactly where the phylogenetic effect lies in a particular phylogeny and dataset (e.g., Diniz-Filho 2001). Researchers with data from large numbers of taxa available may prefer the extra evolutionary insight offered by the PMM. Again, possibly the best solution is to apply a combination of methods and model-fitting procedures. For example, if an initial run shows that the PGLS a_C is close to zero or the PMM phylogenetic heritability (h^2) is close to one, a researcher might apply FIC to get the best estimate of a phenotypic correlation between characters. Differences among results obtained from different PCMs can be used to bound the possibilities.

Incorrect Phylogenetic Information

The question of microevolutionary model considered in this study might also be framed as one of incorrect phylogenetic information. As mentioned above, the microevolutionary model is usually described as a set of branch lengths on the phylogeny, translating branch lengths in units of time into units of expected amount of phenotypic change. Thus, in changing the underlying evolutionary model from BM to strong selection, we also modify the branch lengths on the phylogeny. A method (e.g., FIC) that assumes a BM model will thus be using incorrect branch length information if applied to data generated under strong selection. Some errors or uncertainties in the phylogenetic topology can also be described with simple changes in branch lengths. For example, a polytomy might be described as sets of bifurcations separated by very tiny branch lengths.

Díaz-Uriarte and Garland (1996, 1998; see also Harvey and Rambaut 2000) show that log-transformation of branch lengths can improve the performance of FIC when confronted with incorrect phylogenetic information (branch lengths or model of evolutionary change), resulting, for example, in Type I error rates that almost never exceeded twice the expected value (0.10 when $\alpha = 0.05$). As mentioned above, this level of performance is roughly comparable to that determined for the PCMs (except FIC) in our study, confirming our conclusion that most PCMs are roughly equivalent in terms of their abilities to solve the problem of phylogenetic dependence. Unfortunately, diagnostic tests such as plotting contrasts against their standard deviations (or parameter estimates such as the PGLS α and PMM h^2) are not perfect, and researchers applying these tests will sometimes erroneously conclude that transformation is necessary and obtain poor estimates of evolutionary relationships. With PGLS and PMM, a researcher can at least use the explicit evolutionary interpretations of model parameters (e.g., α and h^2) to inform the choice between what can be widely divergent results obtained by different methods with the same data (e.g., Table 3). Essentially, these answer Harvey and Rambaut's (2000) call for maximum-likelihood transformations that are directly linked to specific evolutionary models. In contrast, the Díaz-Uriarte and Garland (1996, 1998) branch length transformations do not have an explicit evolutionary interpretation, limiting the utility of FIC by essentially turning it into a statistical rather than an evolutionary approach.

Another suggestion for dealing with uncertain phyloge-

netic information is to use randomization procedures to generate large numbers of branch lengths on a known phylogeny (or large numbers of phylogenies; Losos 1994; Martins 1996a; Housworth and Martins 2001). Variation among results obtained using each of those possible phylogenies can then be incorporated into the final confidence interval (Martins 1996a). Ideally, this procedure would be used in conjunction with any of the PCMs to address the problem of uncertainty in the phylogeny. Several of the PCMs tested in the current study also consider a range of possible branch length combinations, choosing the one that fits the data best. The actual set of possible branch length combinations considered differs substantially among PCMs and between the PCMs and the randomization procedure described above. For example, PGLS considers only those sets of branch lengths that follow a particular exponential evolutionary model and PMM considers only those branch length combinations that follow the mixed model. A randomization procedure could be used in conjunction with either of these two (or with other PCMs) to consider yet a third type of branch length combination.

Additional research would be needed to determine how the above sets of branch length combinations relate to each other and those, for example, specified by DNA sequence data (e.g., as in Huelsenbeck et al. 2000; Huelsenbeck and Bollback 2001). Finally, Both PGLS and PMM can be extended to include measures of within-species variation or measurement error. Future research might thus also include the importance of within-species variation on PCM abilities. Although these factors seem unlikely to have much of an impact on the actual results of most individual comparative analyses, further development of phenotypic evolution and comparative method theory may also improve our ability to apply phylogeny reconstruction methods to increasingly divergent types of data.

Acknowledgments

We thank J. Huelsenbeck and two anonymous reviewers for useful comments. Our work was funded by U.S. National Science Foundation grants DEB-9720641 to EPM and DMS-0075143 to EAH. JAFD-F was supported by the Brazilian Conselho Nacional de Desenvolvimento Científico and Tecnológico, proc. no. 300762/94-1.

Literature Cited

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

Bininda-Emonds, O. R. P., J. L. Gittleman, and A. Purvis. 1999. Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). Biol. Rev. 74:143–175.

Bleiweiss, R., J. A. W. Kirsch, and J. C. Matheus. 1997. DNA hybridization evidence of the principal lineages of hummingbirds (Aves: Trochilidae). Mol. Biol. Evol. 14:325–343.

Charlesworth, B. 1984. The cost of phenotypic evolution. Paleobiology 10:319–327.

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.

Coelho, A. S. G., and J. A. F. Diniz-Filho. 2000. PVR, version 4.0. Computer programs for conducting phylogenetic eigenvector regression of comparative data. Distributed by the author via http://compare.bio.indiana.edu/ PVR.html

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. Effect of branch length errors on the performance of phylogenetically independent contrasts. Syst. Biol. 47: 654–672.

Diniz-Filho, J. A. F. 2001. Phylogenetic autocorrelation under distinct evolutionary processes. Evolution 55:1104–1109.

Diniz-Filho, J. A. F., C. E. R. Sant'Ana, and L. M. Bini. 1998. An eigenvector method for estimating phylogenetic inertia. Evolution 52:1247–1262.

Diniz-Filho, J. A. F., S. Fuchs, and M. C. Arias. 1999. Phylogeographical autocorrelation of phenotypic evolution in honey bees (*Apis mellifera* L.). Heredity 83:671–680.

Diniz-Filho J. A. F., A. S. G. Coelho, and C. E. R. Sant'Ana. 2000. Statistical inference of correlated evolution between macroecological variables using phylogenetic eigenvector regression. Ecol. Austral 10:27–36.

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

———. 1988. Phylogenies and quantitative characters. Annu. Rev. Ecol. Syst. 192:445–471.

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

Grafen, A. 1989. The phylogenetic regression. Philos. Trans. R. Soc. B 326:157–199.

Hansen, T. F. 1996. Adaptation, phylogeny and the comparative method. Ph.D. diss., Department of Biology, University of Oslo, Norway.

———. 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.

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. 2000. Comparative analyses for adaptive radiations. Philos. Trans. R. Soc. B 355:1599–1605.

Housworth, E. A., and E. P. Martins. 2001. Conducting phylogenetic analyses when the phylogeny is partially known: random sampling of constrained phylogenies. Syst. Biol. 50:628–639.

Huelsenbeck, J. P., and J. P. Bollbeck. 2001. Empirical and hierarchial Bayesian estimation of ancestral states. Syst. Biol. 50: 351–366.

Huelsenbeck, J. P., B. Rannala, and J. P. Masly. 2000. Accommodating phylogenetic uncertainty in evolutionary studies. Science 288:2349–2350.

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334.

Losos, J. B. 1990. Ecomorphology, performance capability and scaling of West Indian *Anolis* lizards: an evolutionary analysis. Ecol. Monogr. 65:369–388.

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

Lynch, M. 1991. Methods for the analysis of comparative data in evolutionary biology. Evolution 45:1065–1080.

Martins, E. P. 1994. Estimating rates of character change from comparative data. Am. Nat. 144:193–209.

———. 1996a. Conducting phylogenetic comparative analyses when the phylogeny is not known. Evolution 50:12–22.

———. 1996b. Phylogenies, spatial autoregression, and the comparative method: a computer simulation test. Evolution 50: 1750–1765.

———. 2000. Adaptation and the comparative method. Trends Ecol. Evol. 15:295–299.

———. 2001. COMPARE: statistical analysis of comparative data.

Computer programs distributed by the author via http://compare.bio.indiana.edu

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. A microevolutionary link between phylogenies and comparative data. Pp. 283–288 *in* P. Harvey, J. Maynard Smith, and A. Leigh-Brown, eds. New uses for new phylogenies. Oxford Univ. Press, Oxford, U.K.

———. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into analysis of interspecific data. Am. Nat. 149:646–667 (erratum *in* Am. Nat. 153:448).

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.

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

Purvis, A. 1995. A composite estimate of primate phylogeny. Philos. Trans. R. Soc. B 348:405–421.

Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: Geometric interpretations. Evolution 50: 2143–2160.

SAS Institute. 1990. SAS. Ver. 6.1. SAS Institute, Inc., Cary, NC.

Westoby, M., M. R. Leishman, J. M. Lord. 1995. On misinterpreting the ''phylogenetic correction.'' J. Ecol. 83:531–534.

Wolfram, S. 1999. The mathematica book. Cambridge Univ. Press, Cambridge, U.K.

Corresponding Editor: J. Huelsenbeck