EVOLUTION

COMPARATIVE METHODS FOR THE ANALYSIS OF CONTINUOUS VARIABLES: GEOMETRIC INTERPRETATIONS

F. James Rohlf

Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, New York 11794-5245
E-mail: rohlf@life.bio.sunysb.edu

Abstract.—This study is concerned with statistical methods used for the analysis of comparative data (in which observations are not expected to be independent because they are sampled across phylogenetically related species). The phylogenetically independent contrasts (PIC), phylogenetic generalized least-squares (PGLS), and phylogenetic autocorrelation (PA) methods are compared. Although the independent contrasts are not orthogonal, they are independent if the data conform to the Brownian motion model of evolution on which they are based. It is shown that uncentered correlations and regressions through the origin using the PIC method are identical to those obtained using PGLS with an intercept included in the model. The PIC method is a special case of PGLS. Corrected standard errors are given for estimates of the ancestral states based on the PGLS approach. The treatment of trees with hard polytomies is discussed and is shown to be an algorithmic rather than a statistical problem. Some of the relationships among the methods are shown graphically using the multivariate space in which variables are represented as vectors with respect to OTUs used as coordinate axes. The maximum-likelihood estimate of the autoregressive parameter, ρ, has not been computed correctly in previous studies (an appendix with MATLAB code provides a corrected algorithm). The importance of the eigenvalues and eigenvectors of the connection matrix, **W**, for the distribution of ρ is discussed. The PA method is shown to have several problems that limit its usefulness in comparative studies. Although the PA method is a generalized least-squares procedure, it cannot be made equivalent to the PGLS method using a phylogenetic model.

Key words.—Autoregressive models, comparative method, generalized least-squares, independent contrasts, maximum likelihood, phylogenetic autocorrelation, phylogeny, regression.

Received September 15, 2000.   Accepted June 7, 2000.

It is now generally recognized that specialized statistical methods are required to take into account the expected lack of independence of observations when samples are taken across phylogenetically related species. This has been demonstrated in simulation studies by Grafen (1989), Martins and Garland (1991), Purvis et al. (1994), Harvey and Rambaut (1998), and Garland and Díaz-Uriarte (1999). Special methods are needed because most of the familiar statistical methods require the fundamental assumption that the observations are independent (Sokal and Rohlf 1995). The use of standard nonparametric methods does not get around this problem because they also require independence of observations. Garland and Adolph (1994), Martins (1995), and Martins and Hansen (1996b) give good general introductions to this problem and survey the methods that have been proposed for its solution. Pagel (1999), Purvis and Webster (1999), and Harvey (2000) provide recent reviews. These techniques have been used in a very large number of studies. Felsenstein's (1985) method of phylogenetically independent contrasts (PIC) is the most commonly used technique (Garland et al. 1999). The phylogenetic generalized least-squares (PGLS)

and phylogenetic autocorrelation (PA) methods are compared below. Another method that is sometimes used is based on comparisons of selected pairs of species (Møller and Birkhead 1992). Maddison (2000) presents an algorithm for efficiently selecting pairs of species from a phylogeny. This method will not be investigated here because of the convincing list of problems and limitations that Maddison (2000) presents for this approach.

There have been conflicting suggestions about which of the available methods should be used and exactly how they should be applied. In part this is because the methods are complex, they all require rather unrealistic assumptions, and some involve operations that seem somewhat arbitrary. Pagel (1993) shows that regression using several approaches estimate the same parameter, however, they do not give the same results in practice because they differ in how the observations are weighted. Miles and Dunham (1996) point out that the differences among the methods are important because in some cases they may lead to dramatically different results.

Because of the complexity of the methods, there have been a number of studies using computer simulations to determine

some of their properties empirically (e.g., Díaz-Uriarte and Garland 1996; Martins 1999). Such experimental studies are very useful for investigating the robustness of methods with respect to violations of their assumptions. Many of the results of these studies have been complex and difficult to summarize. The approach here, however, will be to work out what can be done analytically assuming a simple Brownian motion model of evolution. One of the purposes of this note is to bring together some of the expected relationships among the methods. The present study will only consider relatively simple statistical analyses such as estimation of means, correlations, and linear regressions. More complicated statistical analyses can also be performed using these methods (e.g., Clobert et al. 1998), but more work is needed on models for multivariate applications.

The usefulness of interpreting these methods in terms of their multidimensional geometry is discussed. This is done in terms of spaces in which variables are represented as points with respect to the observations used as the coordinate axes. This space is dual to the more familiar multivariate space in which observations are represented as points plotted with respect to variables as coordinate axes. Carroll and Green (1997, pp. 268–270) refer to this as an ''alternative representation.'' Although it is less familiar, it is a standard technique commonly used in multivariate analysis. The use of the dual space often makes it easier to understand important properties of multivariate methods.

In this paper, the objects being studied will be referred to as operational taxonomic units (OTUs; Sneath and Sokal 1973) because they could be at any taxonomic level (although they usually are species in comparative studies). The ancestral (internal) nodes in the tree will be referred to as hypothetical taxonomic units (HTUs). In addition to the branching pattern, the phylogenetic tree must also include estimates of the branch lengths, which are assumed to be proportional to expected variance in the amount of evolutionary changes along each branch. They will also be proportional to the elapsed time if the rates of evolution are constant, as would be the case under a model of pure Brownian motion character evolution. For some models they may all be taken as equal in length. The methods described below all require a good estimate of the phylogeny for the organisms being studied. Hansen and Martins (1996) show how a phylogenetic tree can be used to construct the expected covariance matrix for the OTUs when a Brownian motion (random walk in continuous time) model of evolutionary divergence is assumed. The PIC and PGLS methods are based on this phylogenetic covariance matrix, Σ. The standard symbol for a parametric covariance matrix is used here (rather than $\mathbf{C}$ as in Garland and Ives [2000] or $\mathbf{V}$ as in Martins and Hansen [1997]) because the methods discussed here treat this matrix as if it was derived from the true phylogenetic tree. The diagonals of this matrix are the sums of the branch lengths between each OTU and the root (i.e., the height of each OTU on the phylogenetic tree). The off-diagonal elements give the covariances between pairs of OTUs (the height of their most recent common ancestor; Piazza and Cavalli-Sforza 1975; Felsenstein 1985). Note that this is the covariance matrix among observations (OTUs) whether the observations are for a single dependent variable or a vector of measurements for
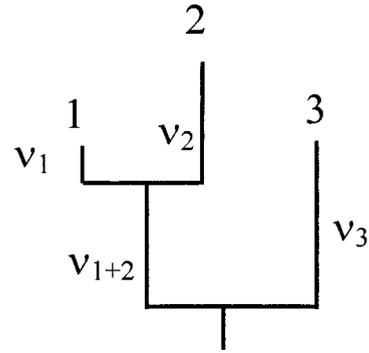

FIG. 1. Phylogenetic tree with three terminal taxa (OTUs). The quantities v_1, v_2, v_3, and v_{1+2} are the lengths of the branches supporting the taxa indicated by their subscripts. The lengths of these branches (1, 4, 6, 5, respectively) will be used in the numerical examples. No length is assigned to the branch supporting the entire tree (the stem below the root) because it does not influence the results of the methods discussed here.

several dependent variables. The covariance matrix for variables and allowances for different evolutionary rates in different variables are separate parameters that may need to be specified in a particular analysis. The Σ matrix can be constructed by inspection of the tree using an algorithm similar to that used to compute a matrix of cophenetic values (Sokal and Rohlf 1962). The simple three OTU phylogenetic tree given in Figure 1 is sufficient to illustrate many of the properties of the PIC and PGLS methods. Assuming the Brownian motion model, the phylogenetic covariance matrix for this tree is

$$\Sigma = \begin{pmatrix} v_1 + v_{1+2} & v_{1+2} & 0 \\ v_{1+2} & v_2 + v_{1+2} & 0 \\ 0 & 0 & v_3 \end{pmatrix}. \quad (1)$$

Martins and Hansen (1996a) call this the ''expected relationship matrix.'' The use of this matrix assumes that the phylogenetic covariance between pairs of species decreases linearly over time since their separation. Hansen and Martins (1996) describe several other evolutionary models and their expected covariance matrices, but these will not be considered here.

PHYLOGENETICALLY INDEPENDENT CONTRASTS

Felsenstein's (1985) PIC method is defined by an algorithm that creates contrasts for each internal node of a binary tree. The algorithm starts with the internal nodes near the tips of a tree and works down to the root node. Each contrast score measures the difference in a variable, Y, between the left and right branches at a node. As Grafen (1992) points out, the term ''contrast'' is often used ambiguously. In this paper I will distinguish between the vector of weights that define a contrast and the score or value for a contrast when applied to a variable. In the literature of the comparative method and in most software, the scores are usually called contrasts and the actual contrast vectors are not given. The PIC method is often simply called the independent contrasts method. That is not done here because the contrasts are only expected to be independent under a particular evolutionary model. The

algorithm is recursive because the computation of the scores at one level depends on branch lengths above it and the algorithm modifies these as it progresses down the tree. For a fully resolved (binary) tree there will be $n - 1$ contrasts for a tree with n OTUs.

The simple phylogenetic tree with three OTUs shown in Figure 1 is sufficient to illustrate the algorithm. There are three steps to be taken for each internal node. First, a contrast is constructed for OTUs 1 and 2. It is convenient to represent the contrast as a vector with elements $1, -1, 0$. This vector is standardized by dividing by $\sqrt{v_1 + v_2}$, the square root of the sum of the lengths of the two branches descending from this internal node. This is the expected standard deviation of the score on this contrast assuming a Brownian motion model of evolutionary change. This yields

$$\mathbf{c}_1 = \frac{1}{\sqrt{v_1 + v_2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}. \tag{2}$$

The score for this contrast is obtained by multiplying this vector by the vector of observations for the variable being studied, thus $s_1 = \mathbf{c}_1^t \mathbf{y}$, where the superscript t indicates the matrix transpose.

Second, a value for Y is estimated for this node, the ancestor of OTUs 1 and 2, as

$$Y_{1+2} = \frac{\dfrac{1}{v_1} Y_1 + \dfrac{1}{v_2} Y_2}{\dfrac{1}{v_1} + \dfrac{1}{v_2}}. \tag{3}$$

This is called a ''local'' estimate of the value of Y for this HTU because it is a weighted average of just the two descendant nodes and does not take into consideration the other nodes in the tree as does the squared-change parsimony method described below.

Third, length of the branch supporting this internal node is adjusted using the following relationship:

$$v'_{1+2} = v_{1+2} + \frac{1}{\dfrac{1}{v_1} + \dfrac{1}{v_2}}. \tag{4}$$

These three steps are then repeated for the next internal node in the tree. Logically, the second contrast is $(1, -1)$ to contrast HTU_{1+2} with OTU_3 but, as we have just seen, HTU_{1+2} is a weighted linear combination of OTUs 1 and 2. Thus, the contrast vector is actually

$$\left(\frac{\dfrac{1}{v_1}}{\dfrac{1}{v_1} + \dfrac{1}{v_2}}, \quad \frac{\dfrac{1}{v_2}}{\dfrac{1}{v_1} + \dfrac{1}{v_2}}, \quad -1 \right)^t$$

in terms of all three OTUs. As before, the contrast vector is standardized by dividing by the square root of the sum of the two branch lengths descending from it to yield

$$\mathbf{c}_2 = \frac{1}{\sqrt{v'_{1+2} + v_3}} \left(\frac{v_2}{v_1 + v_2}, \frac{v_1}{v_1 + v_2}, -1 \right)^t, \tag{5}$$

after simplification using the relationship $1/v_1 + 1/v_2 = (v_1 + v_2)/v_1 v_2$. Note that v'_{1+2} is the adjusted branch length from the previous step.

Although not part of Felsenstein's (1985) algorithm, one can estimate the value for the root node as

$$Y_{1+2+3} = \frac{\dfrac{1}{v'_{1+2}} Y_{1+2} + \dfrac{1}{v_3} Y_3}{\dfrac{1}{v'_{1+2}} + \dfrac{1}{v_3}} \tag{6}$$

(see Garland et al. 1993, 1999). The length of the branch supporting the root node (initially zero) becomes $[(1/v'_{1+2}) + (1/v_3)]^{-1}$, the expected variance of the estimate for the root. If there were more than three OTUs, the above steps would be repeated until the contrast is computed for the basal node of the tree.

Because the method is defined by a recursive algorithm, rather than an explicit set of equations, it is more difficult to appreciate some of its general properties. It is clear that the two contrasts shown above are not orthogonal. This can be shown by computing the cosine of the angle between the contrast vectors as

$$\cos(\theta_{12}) = \frac{\mathbf{c}_1^t \mathbf{c}_2}{\sqrt{\mathbf{c}_1^t \mathbf{c}_1 \mathbf{c}_2^t \mathbf{c}_2}}$$

$$= \frac{\dfrac{v_2}{v_1 + v_2} - \dfrac{v_1}{v_1 + v_2}}{\sqrt{2 \left[\left(\dfrac{v_2}{v_1 + v_2} \right)^2 + \left(\dfrac{v_1}{v_1 + v_2} \right)^2 + 1 \right]}}$$

$$= \frac{v_2 - v_1}{2\sqrt{v_1^2 + v_2^2 + v_1 v_2}}. \tag{7}$$

This will be zero if and only if the two branch lengths are equal. It can be positive or negative depending on the arbitrary choice of which branch is on the left or right. It takes on a maximum value if v_1 is equal to zero and v_2 is greater than zero. For example, if $v_1 = 0$ and $v_2 = 1$, then equation (7) yields $(1 - 0)/(2\sqrt{0 + 1 + 0}) = 1/2$. With more than three OTUs, the contrasts will not all be orthogonal even if all of the branch lengths are equal because of the adjustments to the branch lengths in the third step of the algorithm. Thus, despite the fact that these are usually called independent contrasts, the contrasts will not be independent unless the data are actually obtained from the given phylogeny (see the section Relationships between the PIC and PGLS Methods).

The contrast vectors are used to compute contrast scores for the variables of interest. The vector of contrast scores for the ith variable is computed as $\mathbf{s}_i = \mathbf{C}^t \mathbf{y}_i$, where $\mathbf{C}$ is an $n \times (n - 1)$ matrix with the columns corresponding to the contrast vectors. The uncentered correlations between these scores for different variables or the uncentered regression of the scores for one variable on the scores for another variable gives a relationship that takes into account the expected covariance among observations due to their shared phylogenetic history. The uncentered correlation between Y_1 and Y_2 is computed as
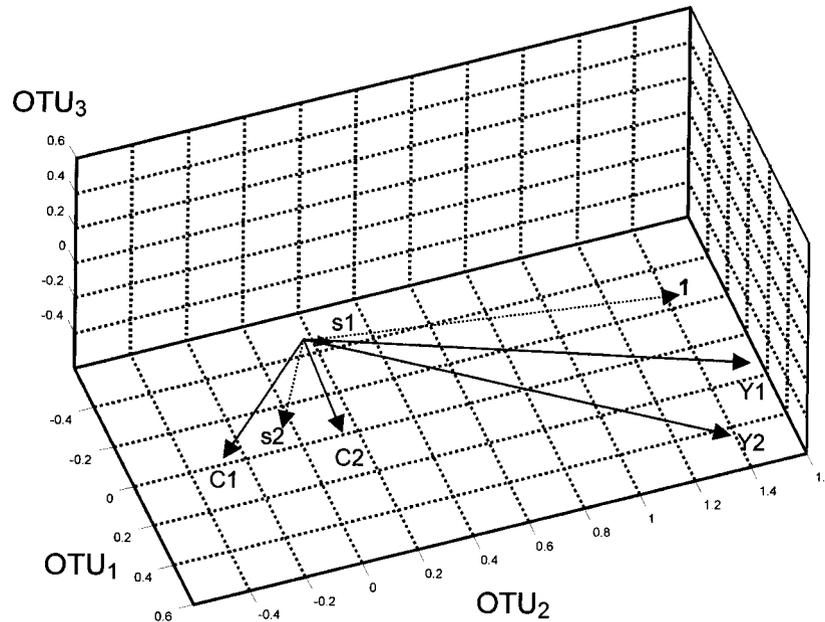
FIG. 2. Diagram in the dual space showing vectors plotted with the OTUs used as coordinate axes. Vectors $\mathbf{C}_1$ and $\mathbf{C}_2$ are the contrast vectors defining the contrast plane and $\mathbf{1}$ is the unit vector orthogonal to this plane. $\mathbf{Y}_1$ and $\mathbf{Y}_2$ correspond to two variables (with values equal to 1, 1.25, 0.5 and 1.5, 1, 0.75, respectively, in the artificial example). Their contrast score vectors ($\mathbf{s}_1$ and $\mathbf{s}_2$) are also shown. The lengths of the score vectors have been exaggerated by a factor of two to increase their visibility in this and subsequent figures (vector $\mathbf{s}_1$ is still very short).

$$r = \frac{\mathbf{s}_1^t \mathbf{s}_2}{\sqrt{\mathbf{s}_1^t \mathbf{s}_1 \mathbf{s}_2^t \mathbf{s}_2}}. \tag{8}$$

An uncentered correlation or a regression through the origin is used (e.g., Garland et al. 1992) because the left-right arrangement of branches on a tree is arbitrary. Changing the order of the branches at a node changes the signs of the corresponding contrast scores, which will usually change the mean for a variable, and this will lead to different results for the usual centered correlation and regression coefficients.

It is useful to interpret an uncentered correlation between two variables as the cosine of the angle between two vectors. Note that equations (7) and (8) are of the same form. This angle and the PIC contrast vectors themselves are in a space that is said to be dual to the usual multivariate space in which the observations are represented as points with respect to the variables used as coordinate axes. In the dual space, one visualizes variables as vectors from the origin with the endpoints plotted with respect (somewhat counter-intuitively) to the observations used as coordinate axes. The use of this dual space helps one understand the PIC method and its relationships to the PGLS approach discussed below.

Figure 2 is a plot in the dual space for our simple three-OTU example. The two contrast vectors, $\mathbf{C}_1$ and $\mathbf{C}_2$, have a cosine of 0.3273 and thus are separated by an angle of 70.9°. The set of contrast vectors span a hyperplane (an $(n - 1)$-dimensional subspace of an n-dimensional space, an ordinary plane in our example with just three OTUs). This hyperplane is orthogonal to the unit vector, $\mathbf{1}$, a column vector with all n elements equal to one. This figure also includes vectors, $\mathbf{y}_1$ and $\mathbf{y}_2$, representing two variables whose correlation or regression is to be investigated (the values of the variables are given in the caption). Their cosine is 0.9543 (correspond-

ing to an angle of 17.4°), but this is not the quantity of interest because it is an uncentered correlation and it does not take the expected covariance due to phylogeny into consideration. If the vectors were projected onto the hyperplane orthogonal to the unit vector (i.e., onto the plane defined by the contrasts) then the cosine of the angle between these vectors would be the ordinary product moment correlation between Y_1 and Y_2. This is because projecting a vector onto a hyperplane orthogonal to the unit vector removes the contribution of the mean from the vector (centers it). This cosine is the same as that of the angle between the vectors $\mathbf{y}_1 - \bar{\mathbf{y}}_1$ and $\mathbf{y}_2 - \bar{\mathbf{y}}_2$. In our example their cosine is 0.5, corresponding to an angle of 60°. This cosine is the usual centered product moment correlation that does not take into account the expected phylogenetic covariance among the OTUs.

The vectors of independent contrast scores, $\mathbf{s}_1$ and $\mathbf{s}_2$ for variables 1 and 2, lie in the hyperplane spanned by the contrast vectors. In our example, the uncentered correlation between these $(n - 1)$ dimensional vectors is 0.0926 (corresponding to an angle of 84.7°). This cosine is what is interpreted as the phylogenetically corrected correlation between the two variables. Because the score vectors are in the hyperplane orthogonal to the unit vector, this quantity is actually a centered correlation between the two variables (see below). The contrast scores can be represented in the n-dimensional dual space by plotting the products $\mathbf{C}\mathbf{s}_i$ as vectors (these correspond to the $\mathbf{s}_1$ and $\mathbf{s}_2$ vectors in Fig. 2). Geometrically, the contrast score vectors can be interpreted as inner products of a vector $\mathbf{y}$ with each contrast vector separately. This is not the same as the projections of a vector $\mathbf{y}$ onto the plane defined by the contrast vectors because the contrasts are not necessarily orthogonal. The computation of

the cosines between these vectors in the n-dimensional dual space is more complicated than in the $(n - 1)$-dimensional space of the contrasts because the score vectors are expressed as oblique coordinates relative to the contrast vectors. The cosine for a pair of variables is obtained by dividing the corresponding off-diagonal element of the $(\mathbf{Cs})'\mathbf{\Sigma Cs}$ matrix by the square roots of the diagonals of this matrix for the variables being compared ($\mathbf{\Sigma}$ is the phylogenetic covariance matrix defined above). In our three-OTU example, this will again yield a cosine of 0.0926. If one were interested in regressing the scores for Y_1 onto those for Y_2, the regression through the origin would yield a slope of 0.0617 for our example (geometrically, this is the projection of the $\mathbf{s}_1$ vector onto the $\mathbf{s}_2$ vector).

The algorithm as described above requires fully resolved phylogenetic trees, that is, just two descendants from each internal node. There have been various suggestions about how to generalize the method to allow trees with multifurcations. This issue will be discussed below after the PGLS method is described.

<h2 style="text-align:center">PHYLOGENETIC GENERALIZED LEAST-SQUARES</h2>

Grafen (1989) pointed out that when a phylogeny (including branch lengths) is known, one can use the standard technique of generalized least-squares analysis to take into account the expected covariance among OTUs. (He also proposed a new method to handle cases in which knowledge of the phylogeny is incomplete, i.e., soft polytomies where more than two branches descending from one or more internal nodes.) Martins and Hansen (1997) and Garland and Ives (2000) give good introductions to the use of the PGLS method in comparative studies. See Rao and Toutenburg (1999) for a general discussion of least-squares and generalized least-squares methods.

The PGLS regression model, $\mathbf{y} = \mathbf{Xb} + \boldsymbol{\epsilon}$, is the same as the usual least-squares regression model except for the structure of the error term. In this model, $\mathbf{y}$ is the $n \times 1$ vector of the dependent variable (or $n \times p$ for a multivariate generalized linear model), $\mathbf{X}$ is the $n \times q$ matrix of independent variables (the first column is usually the unit vector to include the y-intercept in the model, the other columns vary depending of the statistical design being employed), $\mathbf{b}$ is the $q \times 1$ vector of partial regression coefficients, and $\boldsymbol{\epsilon}$ is the error term. In an ordinary least-squares model, $\boldsymbol{\epsilon}$ is assumed to follow a normal distribution with a mean of zero and a variance of σ^2. However, in a PGLS model we are able to take into consideration the expected lack of independence of the observations due to the phylogenic history of the species. We assume here that $\boldsymbol{\epsilon}$ follows a multivariate normal distribution with a mean of $\mathbf{0}$ (a null vector of length n) and an $n \times n$ covariance matrix $\sigma^2\mathbf{\Sigma}$. The diagonal elements of the $\sigma^2\mathbf{\Sigma}$ matrix give the variances of each of the n OTUs and the off-diagonals contain the expected covariances between pairs of OTUs. The $\mathbf{\Sigma}$ matrix is, as described above, a function of the heights of the OTUs and the internal nodes in the phylogenetic tree. The symbol $\mathbf{\Sigma}$ is used because it is treated as the true covariance matrix rather than a sample estimate. The scalar σ^2 parameter is introduced to allow the covariances to be just proportional to the heights on the tree and to allow

for the units of a particular dependent variable. If measurement error or other sources of variance are known (and in suitable units), then they can be added to the $\sigma^2\mathbf{\Sigma}$ matrix, but that will not be considered here because it does not correspond to the model used by the PIC method.

The usual generalized least-square estimate (see section 4.2 of Rao and Toutenburg 1999) of the parameters is

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{y}. \tag{9}$$

Note that the σ^2 term canceled out. For our three-OTU example the intercept is 0.6698 and the slope is 0.0617. The covariance matrix of the regression coefficients is

$$\mathbf{S}_b = \sigma^2(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}. \tag{10}$$

Estimates of $\mathbf{y}$ are computed as

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}} = \mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{y} = \mathbf{Hy}, \tag{11}$$

where $\mathbf{H}$ is the generalization of the hat matrix of ordinary least-squares regression (e.g., Sokal and Rohlf 1995). The residuals from the model can be computed as

$$\mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{Xb} = \mathbf{y} - \mathbf{Hy} = (\mathbf{I} - \mathbf{H})\mathbf{y}. \tag{12}$$

The $n \times n$ covariance matrix of the vector of predicted observation is

$$\mathbf{S}_{\hat{\mathbf{y}}} = \sigma^2\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'. \tag{13}$$

These computations can be carried out using various software packages. The NTSYSpc software (Rohlf 2000) is convenient because the COPH module computes the $\mathbf{\Sigma}$ matrix from a tree (which can be in the nexus format) and the MUL-REGR module performs a generalized least-square (GLS) regression for one or more dependent variables and one or more independent variables. With clever construction of independent variables, a variety of types of analyses can be performed that will take the expected phylogenetic covariance into account.

Note that if $\mathbf{X} = \mathbf{1}$, then $\hat{\mathbf{y}} = (\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{y}$, this is $\bar{\mathbf{y}}_{\text{GLS}}$, the estimate of the grand mean. For our example, these are 0.7331 and 1.0254. These are identical to the estimates of the root given by the PIC method and it is also identical to the squared change parsimony estimate for the root node (Garland et al. 1999; Garland and Ives 2000). Thus, the $\mathbf{H}$ matrix gives the same matrix of weights for computing the root from the OTUs as in McArdle and Rodrigo (1994), assuming that the root is placed at an internal node and not on a separate branch as in some of Maddison's (1991) models. An advantage of the PGLS approach is that it also provides an estimate of the variance of a prediction. For example, the variance of the mean is

$$S^2_{\bar{\mathbf{y}}_{\text{GLS}}} = \sigma^2(\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1})^{-1}, \tag{14}$$

where $\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}$ is the sum of the elements of the inverse of the phylogenetic covariance matrix. This quantity is the same as the variance for the root obtained using the method of independent contrasts (see Garland et al. 1999). It is also identical to Schluter et al. (1997) maximum-likelihood estimate of the variance.

The vector of estimates for all of the HTUs is given by

$$\hat{\mathbf{y}}_H = \bar{\mathbf{y}}_{\text{GLS}} + \mathbf{\Sigma}_{\text{HO}}\mathbf{\Sigma}^{-1}(\mathbf{y} - \bar{\mathbf{y}}_{\text{GLS}}), \tag{15}$$

where $\mathbf{\Sigma}_{HO}$ is the $(n - 1) \times n$ matrix of covariances between the HTUs (rows) and the OTUs (columns). If an OTU is a descendant of a given HTU then the covariance is equal to the height of the HTU above the root and it is zero otherwise. Note that the row for the root HTU will contain only zeros. In our three OTU example this matrix is

$$\mathbf{\Sigma}_{HO} = \begin{pmatrix} 0 & 0 & 0 \\ v_{1+2} & v_{1+2} & 0 \end{pmatrix}. \tag{16}$$

The estimates obtained for the other HTUs are also identical to those obtained using squared-change parsimony (see Martins and Hansen 1997; Garland and Ives 2000). This procedure seems to be the simplest direct algorithm for estimating the HTUs and their variances—simpler than, for example, repeatedly rerooting the tree and using the independent contrasts method as suggested by Garland et al. (1999). It is also identical to but simpler than the maximum-likelihood estimation method of Schluter et al. (1997). Equation (15) for the estimate for the ith HTU can be expressed as a weighted sum of the Y values

$$\hat{\mathbf{y}}_{Hi} = \mathbf{w}_i\mathbf{y}, \tag{17}$$

where

$$\mathbf{w}_i = \{(\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1} + \mathbf{\Sigma}_{H_iO}[\mathbf{I} - \mathbf{\Sigma}^{-1}\mathbf{1}\mathbf{1}'(\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1})^{-1}]\}\mathbf{\Sigma}^{-1} \tag{18}$$

and $\mathbf{\Sigma}_{H_iO}$ is the ith row of $\mathbf{\Sigma}_{HO}$. Thus, the estimated Y for the ith HTU is a weighted average of the Y values of the OTUs.

An advantage of expressing the estimates of the HTUs as weighted sums is that one can then derived their covariance matrix using the standard relationship

$$\mathbf{S}_{wy} = \mathbf{w}\mathbf{S}\mathbf{w}^t. \tag{19}$$

The square roots of the diagonal elements give the standard errors of the estimates of the HTUs. Martins and Hansen (1997) give a different equation for the variances of the estimates for the HTUs (the expected variance in the HTUs after the effects of the OTUs have been held constant). The maximum-likelihood estimation method of Schluter et al. (1997) yields standard errors similar to those produced by Martins and Hansen's (1997) method. However, Martins's (1999) simulation study demonstrates that there is a serious problem with both of these approaches because they yield estimates of the standard errors that are usually much too small. Equation (19) yields the correct standard errors (confirmed by simulations, not shown). The discrepancies for the other two methods were for the standard errors of HTUs other than the root. Thus, one could also find the correct standard errors by repeatedly rerooting the tree, as suggested by Garland et al. (1999) and Garland and Ives (2000). Of course that procedure does not provide the full covariance matrix for the estimates. This is important because the estimates for different HTUs can be very highly correlated.

As noted by Garland and Ives (2000), the solution to the PGLS equations given above can be simplified by transforming the dependent and independent variables. Define

$$\mathbf{z} = \mathbf{\Lambda}^{-1/2}\mathbf{E}'\mathbf{y} \quad \text{and} \tag{20a}$$

$$\mathbf{U} = \mathbf{\Lambda}^{-1/2}\mathbf{E}'\mathbf{X}, \tag{20b}$$

where $\mathbf{\Sigma} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}'$ and $\mathbf{E}$ and $\mathbf{\Lambda}$ are the normalized eigen-

vectors and the eigenvalues of $\mathbf{\Sigma}$. Butler et al. (2000) describe a closely related approach using a Cholesky decomposition of $\mathbf{\Sigma}^{-1}$ (they also provide SAS/IML code). Piazza and Cavalli-Sforza (1975) describe some of the relationships between these eigenvectors and eigenvalues and the structure of the phylogenetic tree. The PGLS regression model then becomes

$$\mathbf{z} = \mathbf{U}\mathbf{b} + \delta, \tag{21}$$

where the error term, δ, has a covariance matrix of

$$\mathbf{\Sigma}_\delta = \sigma^2\mathbf{\Lambda}^{-1/2}\mathbf{E}'\mathbf{\Sigma}\mathbf{E}\mathbf{\Lambda}^{-1/2} = \sigma^2\mathbf{I}. \tag{22}$$

Thus, ordinary least-squares methods can be used on these transformed variables (although one needs to transform back to $\mathbf{X}$ and $\mathbf{Y}$ for prediction). Any standard least-squares analysis can be performed using $\mathbf{z}$ (such as ANOVA, ANCOVA, or regression). If $\mathbf{Y}$ has more than one column, then one can also perform standard multivariate analyses. For all of these analyses, one would normally include the term for the intercept (mean). The slope is $\hat{\mathbf{b}} = (\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{z}$, which yields results identical to those obtained using PGLS on the original $\mathbf{y}$ and $\mathbf{X}$. An advantage of the PGLS approach is that it can be extended to more complex statistical models. A simple extension is to allow constant for measurement error or other variation in a dependent variable within each OTU. This can be implemented by adding $\sigma_w^2\mathbf{I}_n$ to the phylogenetic covariance matrix. If this is done then the eigenvectors of the resultant covariance matrix will be unchanged but σ_w^2 will be added to each of the eigenvalues.

This transformation also makes it easier to visualize the PGLS equations in the dual space. This transformation corresponds to rotating the dual space to the principal axes of the phylogenetic covariance matrix, $\mathbf{\Sigma}$, and then scaling the space so that the length of the ith eigenvector is equal to $\lambda_i^{-1/2}$. Figure 3 shows a plot in the dual space as in Figure 2 but with the three unit length eigenvectors included. Figure 4 shows the same space after a rotation and a rescaling so that the eigenvectors are length $\lambda_i^{-1/2}$. The relationships among the other vectors are, of course, changed as a result. In this space, the transformed unit vector $(\mathbf{\Lambda}^{-1/2}\mathbf{E}'\mathbf{1})$ is no longer orthogonal to the plane defined by the contrasts. It now corresponds to the weights for computing $\bar{\mathbf{Y}}_{GLS}$. In this space the contrast vectors are orthogonal and the cosine of the angle between the projections of $\mathbf{y}_1$ and $\mathbf{y}_2$ onto the plane orthogonal to the transformed unit vector gives the desired correlation between Y_1 and Y_2 corrected for the covariance due to phylogeny.

PHYLOGENETIC AUTOREGRESSION

Cheverud and Dow (1985) and Cheverud et al. (1985) proposed the use of autocorrelation methods to take into account the expected correlation among OTUs due to their shared evolutionary history. They describe how this technique can be used to estimate the proportion of the variance in a variable that can be explained by phylogeny (sometimes called phylogenetic inertia). The approach is adapted from methods used successfully in the analysis of spatial patterns (for an introduction, see Upton and Fingleton 1985).
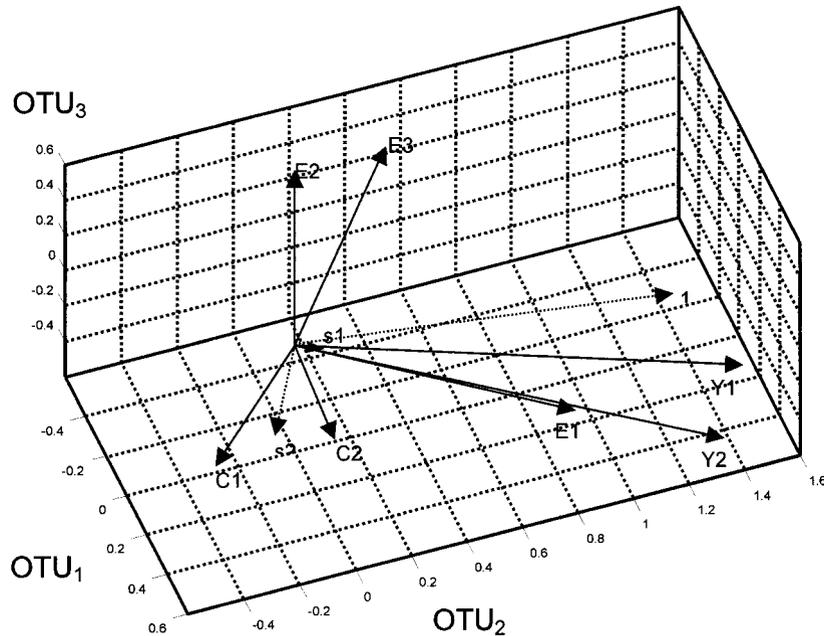
FIG. 3. Diagram in the dual space of Figure 2 with the addition of the three normalized eigenvectors ($\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$) of the phylogenetic covariance matrix. The vectors are as defined in Figure 2.

The model used (called the pure autocorrelation or, more properly, the pure autoregressive model) is

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \epsilon, \qquad (23)$$

where $\mathbf{y}$ is an observed data vector (usually standardized to have a mean of zero and a variance of unity), $\mathbf{W}$ is an $n \times$

FIG. 4. A diagram in the dual space of Figure 3 rotated to align the eigenvectors with the coordinate axes and then scaled so the eigenvectors are of length $\lambda_i^{-1/2}$. The vectors are as defined in Figures 2 and 3.

n connection matrix (each row of which sums to one, and ϵ is the error (assumed normal). The parameter ρ is sometimes called the autocorrelation coefficient but it is better called an autoregressive coefficient because it is not a correlation restricted to the interval -1 to 1. The vector $\hat{\mathbf{y}} = \rho\mathbf{W}\mathbf{y}$ is the estimate of a data vector $\mathbf{y}$. The ρ coefficient describes the relationship between the vectors $\mathbf{y}$ and $\mathbf{W}\mathbf{y}$. This model differs from that used in spatial analysis only in that the $\mathbf{W}$ matrix describes the strengths of the phylogenetic connections between pairs of OTUs in the phylogenetic tree rather than relationships between pairs of localities in space. To construct the $\mathbf{W}$ matrix, Cheverud and Dow (1985) and Cheverud et al. (1985) first created a distance matrix, $\mathbf{D}$, based on the phylogenetic tree. Cheverud et al. (1985) arbitrarily assigned a 1 for pairs of species in the same genus, a 2 for pairs of species in different genera but in the same family, a 3 for those in different families but in the same superfamily, etc. They used these values for levels because they did not have estimates of branch lengths. Their use of integer codes for distances in the tree does not mean that this technique assumes only rank order relationships. This distance matrix satisfies the ultrametric condition, which is more stringent than the triangle inequality (ultrametric distances satisfy the relationship $d_{ij} \leq \max\{d_{ik}, d_{jk}\}$ for all OTUs i, j, and k; see Rohlf and Sokal 1981). Each nonzero element of $\mathbf{D}$ was then replaced by its reciprocal (making what one might call an ultrametric similarity matrix). This matrix is of the form of a phylogenetic covariance matrix except that it has zeros down the principal diagonal. For each pair of OTUs, the entry can be interpreted as the height above the root of their most recent common ancestor. This matrix is then scaled so that each row of the resulting $\mathbf{W}$ matrix will sum to one. Note that if a single OTU joins the tree at the base (e.g., an outgroup), then the corresponding row will contain only zeroes
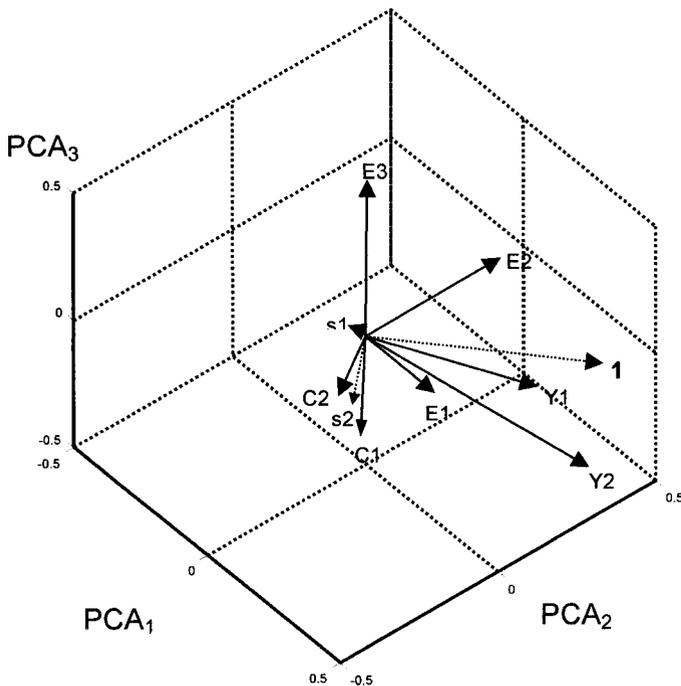
TABLE 1. Eigenvalues, λ, and eigenvectors based on the normalized connectivity matrix, **W,** from table 2 of Cheverud and Dow (1985). The top half gives the right eigenvectors and eigenvalues of **W.** The bottom half of the table gives the eigenvectors and eigenvalues of $\mathbf{W}' = (1/2\,\mathbf{N}\,(\mathbf{W} + \mathbf{W}')\mathbf{N}$, where $\mathbf{N} = \mathbf{I} - 1/n\,\mathbf{11}'$).

Group	Eigenvectors							
	1	2	3	4	5	6	7	8
A	0.5000	−0.8165	0.1890	−0.1233	0	0	0	0
K	0.5000	0.4082	0.1890	0.7607	0	0	0	0
L	0.5000	0.4082	0.1890	−0.6373	0	0	0	0
J	0.5000	−0.0000	−0.9449	0.0000	0	0	0	0
F	0	0	0	0	0.5000	−0.8165	0.1890	0.1233
I	0	0	0	0	0.5000	0.4082	0.1890	−0.7607
E	0	0	0	0	0.5000	0.4082	0.1890	0.6373
C	0	0	0	0	0.5000	−0.0000	−0.9449	0
λ	1.0000	−0.4000	−0.2000	−0.4000	1.0000	−0.4000	−0.2000	−0.4000
A	−0.3713	0.3536	−0.2041	−0.1696	0.7749	−0.7513	−0.4713	−0.2119
K	−0.3713	0.3536	−0.2041	−0.1696	−0.2407	0.1932	0.7398	0.4259
L	−0.3713	0.3536	−0.2041	−0.1696	−0.5343	0.5581	−0.2685	−0.2140
J	−0.2938	0.3536	0.6124	0.6432	−0.0000	0.0000	0.0000	0.0000
F	0.3713	0.3536	−0.2041	0.1696	−0.0976	0.1213	−0.3242	0.6732
I	0.3713	0.3536	−0.2041	0.1696	−0.0959	0.1192	0.1422	−0.1817
E	0.3713	0.3536	−0.2041	0.1696	0.1935	−0.2405	0.1820	−0.4916
C	0.2938	0.3536	0.6124	−0.6432	−0.0000	0.0000	0.0000	−0.0000
λ	1.0110	0.0000	−0.2000	−0.2110	−0.4000	−0.4000	−0.4000	−0.4000

and cannot be scaled to sum to unity. The choice of **D** and the computation of the **W** matrix implicitly specify an evolutionary model (Martins 1996). Before adjusting each row sum to one, the off-diagonals of **W** are the same as those of the phylogenetic covariance matrix. The PA method is just as dependent on having a correct phylogenetic tree (including branch lengths determining the heights of the HTUs), as are the PIC and PGLS methods. In spatial autoregression analysis various distances are sometimes used to take into account different models for the effective distances between localities. Some authors (e.g., Gittleman and Kot 1990; Miles and Dunham 1992) have suggested using path length distances (patristic distances) or the number of nodes separating pairs of species on the tree to define the **W** matrix. Although these distances are usually highly correlated with ultrametric distances, their interpretation is different. These distances are invariant to the choice of the root and they measure the amount of expected divergence between pairs of OTUs rather than the expected amount of independent evolution since divergence from a common ancestor. Note that when the OTUs are required to be at the same height in the tree, the patristic distances will be simply twice the ultrametric distances and the identical **W** matrix will be obtained.

The residuals, ϵ, from the model are estimated as

$$\mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \rho\mathbf{W})\mathbf{y}. \qquad (24)$$

These have been used, for example, to estimate the correlation between two variables with the effects of phylogeny removed.

While sample estimates, $\hat{\rho}$, are often within the interval -1 to 1, the limits are actually from $\lambda_{-\min}^{-1}$ to $\lambda_{+\min}^{-1}$, where $\lambda_{-\min}$ and $\lambda_{+\min}$ are the negative and positive eigenvalues of **W** that are closest to zero. For the tree given in Cheverud and Dow (1985), $\lambda_{-\min} = -0.2$ and $\lambda_{+\min} = 1$ (see Table 1) and thus the interval of possible values of $\hat{\rho}$ are from -5 to 1. A description of the algorithm used here to obtain the maximum likelihood estimates of ρ is given in the Appendix. The AUTOREGR module of NTSYSpc (Rohlf 2000) can be used to perform these computations. Quite different estimates of ρ were obtained ($-1.44, -0.76, -0.93, -0.51,$ and -1.84) from those values ($-0.86, -0.56, -0.68, -0.40,$ and $-0.97,$ respectively) reported by Cheverud and Dow (1985). The reasons for these discrepancies are given in the Appendix.

Gittleman and Kot (1990) used Moran's (1950) I statistic, which is computed as

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}. \qquad (25)$$

Because each row of **W** sums to unity, $S_0 = n$. Equation (25) can be written as

$$I = \frac{(\mathbf{y} - \bar{\mathbf{y}})^t \mathbf{W}(\mathbf{y} - \bar{\mathbf{y}})}{(\mathbf{y} - \bar{\mathbf{y}})^t(\mathbf{y} - \bar{\mathbf{y}})}. \qquad (26)$$

Gittleman and Kot (1990) suggest dividing this coefficient by, $I_{\max}$, their upper limit for $|I|$ (they express it as a ratio of standard deviations but the division by $[n - 1]$ in both the numerator and denominator cancel),

$$I_{\max} = \left(\frac{(\mathbf{y}'' - \bar{\mathbf{y}}'')^t(\mathbf{y}'' - \bar{\mathbf{y}}'')}{(\mathbf{y} - \bar{\mathbf{y}})^t(\mathbf{y} - \bar{\mathbf{y}})}\right)^{0.5}, \qquad (27)$$

where $\mathbf{y}'' = \mathbf{W}(\mathbf{y} - \bar{\mathbf{y}})$ and $\bar{\mathbf{y}}''$ is the mean of the $\mathbf{y}''$ values. Although $I_{\max}$ does not actually give the maximum value for I, division by it yields

$$I' = \frac{I}{I_{\max}} = \frac{(\mathbf{y} - \bar{\mathbf{y}})^t \mathbf{y}''}{\sqrt{(\mathbf{y} - \bar{\mathbf{y}})^t(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y}'' - \bar{\mathbf{y}}'')^t(\mathbf{y}'' - \bar{\mathbf{y}}'')}}. \qquad (28)$$

This is simply the ordinary product moment correlation between **y** and $\mathbf{y}''$ (having $\mathbf{y}''$ in the numerator, rather than $\mathbf{y}'' - \bar{\mathbf{y}}''$, does not matter because the results are identical). Note that the interpretations of the I and I' coefficients are very different. Moran's I provides information on spatial scale (taxonomic level) but it does not provide a measure of fit to

an autoregressive model. I', however, measures the degree of fit but provides no information on spatial scale.

de Jong et al. (1984) provide exact lower and upper bounds for I for a given connection matrix. For our case where the rows of $\mathbf{W}$ sum to unity, the limits are given by the smallest and largest eigenvalues of

$$\mathbf{W}' = \frac{1}{2}\mathbf{N}(\mathbf{W} + \mathbf{W}^t)\mathbf{N}, \qquad (29)$$

where $\mathbf{N} = (\mathbf{I} - (1/n)\mathbf{11}^t)$. However, if one of these extreme eigenvalues is zero, then the second smallest or largest eigenvalue is the lower or upper bound. The pre- and post-multiplication by $\mathbf{N}$ constrains the solution so that eigenvectors with nonzero eigenvalues will be in the plane orthogonal to the unit vector (i.e., their coefficients will sum to zero). These limits are not used to rescale I.

Phylogenetic Autocorrelograms

Gittleman and Kot (1990) used a sequence of $\mathbf{W}$ matrices corresponding to relationships at different taxonomic levels. They called the plot of these I' values against taxonomic level a phylogenetic correlogram. This diagram allows one to assess the amount of autocorrelation at different taxonomic levels in a phylogenetic tree. In their examples they found significant positive I' values at the genus and family levels, a significant negative value at the superfamily level, and no evidence of autocorrelation at the order level. At a particular taxonomic level the $\mathbf{W}$ matrix is block diagonal with the size of the ith block equal to the size of the ith group. The off-diagonal elements are equal to $1/(n_i - 1)$, where n_i is the size of the ith group. All other elements in the matrix are equal to zero.

It is interesting to note that the pure autoregressive model would fit perfectly (i.e., $\mathbf{y} - \hat{\mathbf{y}}$ would be a vector of zeros) if $\mathbf{y}$ happened to be proportional to one of the right eigenvectors of $\mathbf{W}$. The matrix $\mathbf{E}$ is a matrix of right eigenvectors of $\mathbf{W}$ if $\mathbf{E}\boldsymbol{\Lambda} = \mathbf{W}\mathbf{E}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues. The rows of $\mathbf{E}^{-1}$ are the left eigenvectors of $\mathbf{W}$ (this distinction is needed because $\mathbf{W}$ is not symmetric due to the row normalization). If $\mathbf{y}$ is the ith right eigenvector of $\mathbf{W}$, then $\lambda_i \mathbf{y} = \mathbf{W}\mathbf{y}$, where λ_i is the ith eigenvalue. Thus, the vectors $\mathbf{y}$ and $\mathbf{W}\mathbf{y}$ would be proportional and the maximum-likelihood estimate of ρ would be equal to λ_i^{-1}. This relationship between ρ_i and λ_i holds for all λ_i unless a constraint is applied. If $\mathbf{y}$ is proportional to one of the eigenvectors then the I' statistic will be ± 1, with the sign the same as the corresponding eigenvalue. Moran's I differs from ρ in that if $\mathbf{y} - \bar{\mathbf{y}}$ is proportional to the ith eigenvector of $\mathbf{W}'$ then $I = \lambda_i$, where λ_i is the ith eigenvalue of $\mathbf{W}'$. Of course it is unlikely that a data vector for a measured variable would have values *exactly* proportional to one of the eigenvectors. However, data vectors can be expressed as linear combination of the eigenvectors. If a data vector corresponds to a linear combination of the eigenvectors with negative eigenvalues, then it will tend to be negatively correlated with its product with $\mathbf{W}$ (this is an approximate relationship because if $\mathbf{y}$ is not identical to an eigenvector, then $\mathbf{W}\mathbf{y}$ will be closer than $\mathbf{y}$ to the eigenvector with the largest eigenvalue). Thus, an understanding of the properties of the eigenvalues and ei-

genvectors of $\mathbf{W}$ is important for interpreting an autoregression analysis. The importance of the eigenvalues has long been recognized in spatial autocorrelations analyses, for example, to simplify the computation of the maximum-likelihood estimate of ρ. The practical importance of examining the eigenvectors seems to only have been appreciated relatively recently (see Griffith 1996; Griffith and Fellows 1999). They are also useful for understanding the relationships of PA to the other methods described above.

The distance matrix, $\mathbf{D}$, used by Cheverud and Dow (1985) has very different properties than the distance matrices usually used in spatial autocorrelation analyses. In PA the distances satisfy the requirements of an ultrametric (Rohlf and Sokal 1981) rather than a Euclidean metric, as one expects for distances between points in a physical space. The eigenvalues and eigenvectors reflect the hierarchical nested relationships in the tree (for some notes, see Piazza and Cavalli-Sforza 1975). A principal coordinates analysis of this distance matrix will yield at least one zero eigenvalue corresponding to the fact that $n - 1$ rather than n dimensions are required to describe the structure of a tree. The other eigenvalues are usually positive, although that need not be true when arbitrary codes are assigned to the levels. However, what is of interest here are the eigenvalues and eigenvectors of $\mathbf{W}$, a row-normalized matrix of reciprocals of distances. The largest eigenvalue will be equal to one (a consequence of row normalization). The magnitudes of the other eigenvalues will depend upon the tree, but there will be many negative eigenvalues because the sum of the eigenvalues is zero. The pattern of the coefficients of the eigenvectors depends on the tree, but usually the groupings in the tree are apparent from the patterns of coefficients (as noticed by Piazza and Cavalli-Sforza [1975] for phylogenetic covariance matrices). Eigenvectors with large positive eigenvalues reflect structure at the base of the tree, eigenvectors with small positive eigenvalues represent structure near the base of the tree, and eigenvectors with large negative eigenvalues reflect structure near the tips of the tree. This is analogous to the relationship between large positive versus small and negative eigenvalues with large- and small-scale patterns found in spatial autocorrelation analysis (Griffith 1996. The relationship between ρ and the structure of the tree at different taxonomic levels is complicated by the inverse relationship between λ and $\hat{\rho}$. Small negative $\hat{\rho}$ values correspond to differences at furcations near the tips of the tree, larger negative $\hat{\rho}$ values correspond to differences at lower furcations until a furcation is reached at which the one obtains large positive $\hat{\rho}$ values. These decrease until one reaches the smallest positive $\hat{\rho}$ values at furcations at the base of the tree. This makes ρ an inconvenient index near taxonomic level. Note that $\hat{\rho}$, unlike an ordinary correlation, is not a measure of the strength of an association or the degree of fit.

The $\mathbf{W}$ matrices used to compute a phylogenetic autocorrelogram have a simple set of eigenvalues and eigenvectors. For a particular taxonomic level (height from the root) there will be as many eigenvalues equal to one as there are groups of size 2 or larger (with their eigenvectors corresponding to group sums). There will be as many zero eigenvalues as there are groups of size 1. The remaining eigenvalues will be equal to $-1/(n_i - 1)$ with each repeated $n_i - 1$ times (the orientation
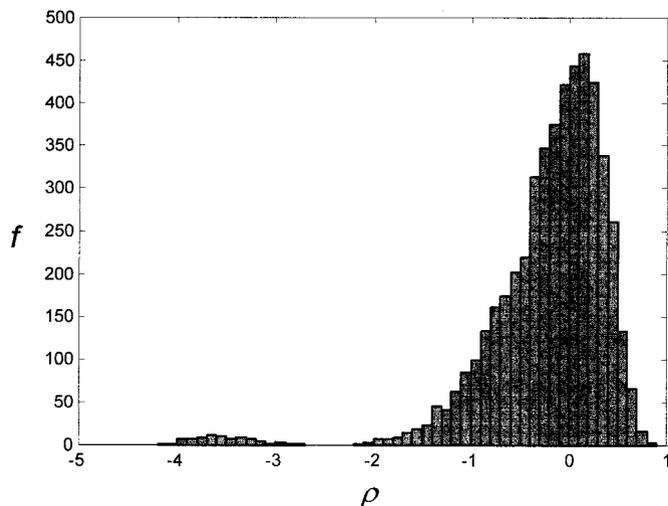
FIG. 5. Distribution of 5000 estimates of the autoregressive coefficient, ρ, for Cheverud and Dow's (1985) tree (see Fig. A1).

of their eigenvectors is arbitrary but taken together they span the space that defines each group, i.e., they correspond to contrasts among the members within each group). This means that at a particular taxonomic level, a variable that has values that are homogeneous within a group and different values in different groups will have a large positive ρ value, whereas a variable that varies within a group but not between groups will have a large negative value.

The example in Cheverud and Dow (1985) is convenient because they provide **D** and **W** matrices, **y** vectors for several variables, and estimated ρ values. The eight eigenvalues and right eigenvectors for their **W** matrix are given in the top half of Table 1. Note that only two of the eight eigenvalues are positive. Consequently, most of the eigenvectors will be negatively correlated with their product with **W**. Because the majority of eigenvalues are negative, most data vectors should have negative ρ values when this **W** matrix is used. To test this, a sampling experiment was performed using 5000 vectors constructed using a normally distributed random deviate for each of their eight elements. The resulting distribution of the maximum-likelihood estimate, $\hat{\rho}$, is shown in Figure 5. It is strongly skewed with about 80% of the estimates less than zero. The distribution for I' is similar but constrained to be within the interval -1 to 1. The distribution of I is constrained to fall within the interval from -0.4000 to 1.0110 because of the range of eigenvalues of the **W**' (see the bottom half of Table 1). Only one of the estimates for ρ for the five variables studied by Cheverud and Dow (1985) was significantly different from zero at the 5% level (and then only if one does not adjust for the fact that one is testing five coefficients). Overall, the ρ values are not much different from what one might expect by chance; thus, their conclusions are not well supported by the results of their analyses. However, their data do support their conclusion that recently split groups are more different than those that split earlier. If one standardizes the five variables and computes Euclidean distances between the samples from the eight populations, one finds a negative relationship between this matrix and a matrix of ultrametric distances based on their tree. The nor-

malized Mantel statistic (Sokal 1979) is -0.33493 and its probability is 0.0180 (based on a 999 random permutations). The form of the distribution in Figure 5 is very different from those reported by Martins (1996) because the data vectors in her study were created by a simulation of a particular evolutionary process rather than a random sample of independent observations.

The interpretation of significance tests for ρ and for I is complicated by the fact that larger values (either positive or negative) do not necessarily imply a better-fitting autoregressive model. In the example given above, $\rho = -2.5$ or -5.0 correspond to perfect fits and intermediate values of ρ correspond to a worse fit. Moran's I has similar complications.

Similar considerations apply if a sequence of **W** matrices is used to construct an autocorrelogram giving the strength of autocorrelation at different taxonomic levels. Gittleman et al. (1996) give several examples and their interpretations of the patterns they found. For a distance class corresponding to a taxonomic level h (starting with $h = 1$ for the root), the elements of **D** are set equal to one for all pairs of OTUs that are within the same taxon at level h, but belong to different taxa at level $h + 1$ and are zero otherwise. For example, at the family level an element is set equal to one if a pair of species belongs to different genera within the same family and is set to zero otherwise. The eigenvalues for the row normalized matrix **W** will be equal to one for eigenvectors whose nonzero elements are constant for one or more groups at level h, -1 for eigenvectors that contrast groups at level $h + 1$, and zero for all others. For levels beyond the first there will be multiples of these eigenvalues, and thus the eigenvectors will only determine subspaces of contrasts of the types described above (and therefore linear combinations of the eigenvectors with the same eigenvalue must also be considered). For the I coefficient, the appropriate eigenvalues and eigenvalues are those computed from the **W**' matrix defined at a particular taxonomic level, h. The eigenvalues of this **W**' matrix will be one for eigenvectors that contrast groups at level h, -1 for eigenvectors that contrast taxa at level $h + 1$, and zero otherwise.

These results suggest a simpler interpretation of phylogenetic correlograms such as given in Gittleman and Kot (1990). For simplicity, assume the tree is a perfectly balanced binary tree, the number of OTUs, n, is a power of 2, and that a variable y has been standardized. At the taxonomic highest level (the branch from the root), Moran's I is $[n/(n-1)]\bar{y}_1\bar{y}_2$, where $\bar{y}_1$ is the mean for the OTUs in the left branch and $\bar{y}_2$ is the mean for those in the right branch. The means will be equal but opposite in sign because the data have been standardized. Thus I must be negative (unless one were studying a variable with no variation). At the next level I will be equal to $\{n/[2(n-1)]\}(\bar{y}_1\bar{y}_2 + \bar{y}_3\bar{y}_4)$, where $\bar{y}_1$ and $\bar{y}_2$ are the means for the two branches within the left branch from the root and $\bar{y}_3$ and $\bar{y}_4$ are the means for the two branches within the right branch from the root. This quantity will be positive if most of the differences are between the left and right branches of the tree rather than among the subgroups within the left and right branches. It will be negative if the differences are among the subgroups rather than between the left and right branches. Similar relationships hold for the other levels in the tree.

Interpretation of the resulting correlogram is complicated by the fact than a strong negative I at a level h will tend to lead to strong positive I values at lower levels. For example, a variable equal to $+1$ for all OTUs in the left branch and -1 for all OTUs in the right branch will yield a value of -1 for Moran's coefficient at the root level and a value of $+1$ at all other levels in the tree. Thus, there are dependencies among the correlations at different levels. Gittleman and Luh (1994) suggest that the PA method should not be used if a significantly negative value is obtained in a phylogenetic correlogram. It seems difficult to justify this rule. It would seem more reasonable to insist on a consistent trend with negative I values at the base of the tree and positive ones near the tips as would be expected under the Brownian motion model for evolution. The results given above suggest that one should take advantage of the simple nested structure of a phylogenetic tree rather than using methods designed for dealing with spatial patterns.

The geometric visualization of PA in the dual space is similar to that used for the PIC and PGLS methods. The **y** vector is the projection of a variable onto the plane orthogonal to the unit vector. Its length will be $\sqrt{n-1}$ if it has been standardized. The premultiplication of **y** by **W** is a shear (it rotates **y** out of the plane orthogonal to the unit vector and then stretches it in the direction of the ith eigenvector by a factor of $\sqrt{\lambda_i}$). A standardization of this transformed vector projects **y**′ back into the plane orthogonal to the unit vector and scales its length to $\sqrt{n-1}$. The I' coefficient is the cosine of the angle between the **y** and **y**″ vectors in the plane orthogonal to the unit vector. Moran's I coefficient is the length of the projection of the **y**″ vector onto the $\mathbf{y} - \bar{y}$ vector. The autoregressive coefficient is more difficult to visualize because it is estimated from the maximum of a likelihood function. However, the residual vector is in the plane orthogonal to the unit vector. The correlations between the residuals of different variables can be quite different from those the phylogenetically corrected correlations obtained using the PIC or PGLS methods. The reason for this is given below.

Variance Due to Phylogeny

After the PA model was fit, Cheverud et al. (1985, p. 1340) used it to decompose the total variation of a variable into that part that could be said to be due to phylogeny ($\hat{y} = \rho \mathbf{W}y$) and that part due to residual specific variation ($y - \bar{y}$). Although Cheverud et al. state that these are by definition independent, the estimates are not independent. In fact they can be highly correlated (for the GM1 variable of Cheverud and Dow [1985] the correlation is 0.572). They would be uncorrelated if ρ were estimated using least-squares rather than maximum likelihood. They suggest using $\hat{\rho}^2$ as a standardized measure of the covariance between the total and the phylogenetic variation, that is, as the proportion of the total variation explained by phylogeny. However, this is not appropriate here because the maximum-likelihood estimate of ρ is not constrained to be between -1 and $+1$ (it is equal to -1.4415 for their GM1 variable). It is also not appropriate because ρ is a measure of spatial scale and thus $\hat{\rho}^2$ is not related to the proportion of variance explained. They also

suggest using the relationship $R^2 = 1 - (s_e^2/s^2)$ as another way to estimate the proportion of the variance explained by phylogeny, where s_e^2 is the maximum-likelihood estimate of the error variance and s^2 is the variance for the variable of interest. For their GM1 variable, $s_e^2 = 0.0098$, $s^2 = 0.0393$, and thus $R^2 = 0.7515$. There are a couple of problems with using this formula. R^2 is usually defined in terms of sums of squares rather than variances (this is a problem for small n because s_e^2 has one less degree of freedom than s^2, taking this into account would yield $R^2 = 0.7870$) and the estimation of ρ does not correspond to the maximization of R^2. Unlike the case of ordinary linear regression models, this R^2 does not give the same result as the direct computation of r_{yy}^2 (0.8094 for GM1). Note that if least-squares were used to estimate ρ (yielding $\hat{\rho} = -2.18$ for GM1) rather than maximum likelihood, then these two formulas would give identical results ($r_{yy}^2 = R^2 = 0.8094$ for GM1) and $\hat{y}$ and $y - \bar{y}$ would be uncorrelated. Of course, there are several ways (different values of ρ) in which variables can fit the PA model. This seems to make the estimate for ρ and the proportion of variance explained less interesting. A very high percent explained simply means that most of the variation in a variable is at *some* furcation in the tree (e.g., small values for descendants of one branch and large values for descendents of the other branch). The value of ρ simply indicates the level of this furcation on the tree.

Diniz-Filho et al. (1998) proposed a new method related to PA to estimate the proportion of the variance in a variable that can be attributed to common ancestry (they call it phylogenetic inertia) and to remove the effect of phylogeny when evaluating correlations among variables. As in PA, they start with a matrix, **D**, of ultrametric distances based on a phylogenetic tree. They use a principal coordinates analysis of **D** to obtain a normalized eigenvector matrix **E** and a diagonal matrix, **Λ**, of eigenvalues. The last eigenvalue will be equal to zero and its corresponding eigenvector will be equal to $(1/\sqrt{n})\mathbf{I}$. Thus, the other eigenvectors span a hyperplane orthogonal to the unit vector. They then use a multiple regression analysis to predict one or more dependent variables using $\mathbf{X} = \mathbf{E}\Lambda^{1/2}$ as a matrix of independent variables. If all but the last eigenvector is retained, the least-squares prediction equation will be

$$\hat{\mathbf{y}} = \mathbf{XB} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$
$$= \mathbf{E}\Lambda^{1/2}(\Lambda^{1/2}\mathbf{E}'\mathbf{E}\Lambda^{1/2})^{-1}\Lambda^{1/2}\mathbf{E}'\mathbf{y} = \mathbf{EE}'\mathbf{y} = \mathbf{Ny}$$
$$= \mathbf{y} - \bar{\mathbf{y}}, \tag{30}$$

where **N** is as defined earlier. Thus, the fit will always be perfect, $R^2 = 1$, and there will be no residual variation in which to investigate correlations with other variables. However, Diniz-Filho et al. (1998) only retained the first few eigenvectors—those for which the corresponding λ_i exceed the values expected from the broken stick model. The fit is then no longer perfect and the R^2 values simply measure the proportion of the variance attributable to average differences among the nodes corresponding to the retained eigenvectors (Diniz-Filho et al. 1998, table 3). Leaving out any eigenvector with $\lambda_i > 0$ corresponds to ignoring some part of the structure of the tree near the tips and thus not taking the entire phylogeny into account. By ignoring eigenvectors with small

eigenvalues, this approach confounds variation due to divergence among species that diverge near the tips of the tree with variation within the branches connecting each species to its more recent common ancestor. Thus, this only estimates part of the variation in a variable that can be accounted for by phylogeny. A more direct approach might be to compute the sum of squared differences between the values for each OTU and the squared-change parsimony estimates of the ancestral states for their most recent common ancestors. This would allow one to compute the proportion of the total variation that could be attributable to the structure of the tree versus variation within the terminal branches. This corresponds to the usual separating of explained and unexplained sums of squares in regression. It would, of course, be difficult to compare the results from different studies because they would be sensitive to different decisions about which species should be included in a study.

COMPARISONS AMONG METHODS

The Problem of Multifurcations

The algorithm for the independent contrasts method has the limitation that it is only defined for binary trees. There have been several suggestions for ways to get around this problem. First, one needs to distinguish between hard polytomies, where one estimates that the actual phylogeny contains a branch where more than two branches emerged nearly simultaneously, and soft polytomies, where the actual phylogeny is assumed to be bifurcating but the available data do not allow one to resolve one or more branches. There have been several papers (e.g., Grafen 1989; Losos 1994; Martins 1996; Garland and Díaz-Uriarte 1999) concerned with adjustments to deal with soft polytomies. They are beyond the scope of the present paper. Because the application of the PGLS approach is not restricted to binary trees, it can serve as a guide to what the correct adjustments should be for hard polytomies. Felsenstein (1985) suggested introducing zero length branches in the tree to arbitrarily resolve the multifurcations. Pagel (1992) was concerned about the arbitrary introduction of zero length branches and recommended ways to decide how to order the branches. He suggested using only one contrast per internal node no matter how many branches are involved. This agrees with Grafen's (1989, 1992) radiation principle. Although the procedure of introducing zero-length branches to arbitrarily resolve a tree seems ad hoc, it is actually a clever algorithmic trick because the different ways of introducing zero-length branches have no effect on the implied phylogenetic covariance matrix and thus have no effect on the estimates produced by the PIC and PGLS methods. Purvis and Garland (1993) point this out for the PIC method. Misunderstanding this point has led to the recommendation that tests should be made with one fewer degrees of freedom than the number of nodes in the tree (e.g., Pagel 1992). Taken to an extreme, this would imply that one would have zero degrees of freedom for a sample of independent observations because the corresponding phylogenetic tree would be a bush.

The effect of hard polytomies on the PGLS method can be seen in the eigenvector decomposition of the $\mathbf{\Sigma}$ matrix. If the OTUs are all at the same height in the tree, then a single

multifurcation with f branches will result in $f - 1$ identical eigenvalues. In such cases the orientation of the corresponding eigenvectors is not uniquely defined (any rotation of them would be just as effective in fitting the $\mathbf{\Sigma}$ matrix). The presence of multifurcations does *not* result in eigenvalues equal to zero, that is, a reduction in dimensionality. Purvis and Garland (1993) also point out that there should be no loss of degrees of freedom due the presence of multifurcations. If the OTUs are not all at the same height, then the presence of multifurcations need not lead to identical eigenvalues and the eigenvectors may be uniquely defined. Similarly, multifurcations are not a problem using the PA approach because one only needs to be able to define the connection matrix. Thus, the problem introduced by hard polytomies is the additional complexity of their implementation in some computer software (having to add small constants to some of the branch lengths to use certain software is a rather inelegant solution).

In the PIC method, the different ways of introducing zero length branches results in contrasts representing different choices of sets of vectors to span the same space. Different choices will still satisfy the important relationship $\mathbf{C}'\mathbf{\Sigma}\mathbf{C} = \mathbf{I}_{n-1}$, and thus they are statistically equivalent. If there is a single multifurcation with f branches this subspace is $(f - 1)$-dimensional and is orthogonal to a vector with ones for members of this set and zero otherwise. Although the contrasts differ for the different ways zero-length branches can be introduced, the cosine of the angle between the contrast score vectors remains the same (different choices correspond to different rotations of the dual space). If only one contrast is used to represent a node with a multifurcations, then the set of contrasts will no longer fully account for the phylogenetic covariance matrix (i.e., it corresponds to a projection into a lower dimensional space which, of course, loses information). Using fewer than $n - 1$ degrees of freedom reduces power and thus would lower the expected performance of the PIC method. Such adjustments were reviewed by Purvis and Garland (1993) in the context of soft and firm polytomies.

The Problem of the Intercept

Because the PIC procedure uses regression through the origin, a y-intercept is not estimated. Although this lack of an intercept is appropriate for contrasts, it would be useful to be able to predict y from x in their original units. Grafen (1992) says that such a ''re-introduction of a constant is a statistical nonsense.'' However, this is not really a problem. The PGLS approach provides estimates of both the slope and the y-intercept. As demonstrated below and pointed out by Garland and Ives (2000), the PIC regression line has the same slope as the PGLS regression line with an intercept included. The problem is simply that the PIC procedure does not provide an estimate of this parameter. Garland and Ives (2000) show that given the slope based on PIC, the PGLS estimate of the y-intercept can be obtained simply by forcing the regression line through the point $\bar{\mathbf{x}}_{GLS}$, $\bar{\mathbf{y}}_{GLS}$ (the estimates of the root for the x and y variables). The standard relationship, $a = \bar{\mathbf{y}}_{GLS} - b\bar{\mathbf{x}}_{GLS}$, can then be used to compute the intercept (Garland and Ives 2000). Garland et al. (1993) and Williams

(1996) provide examples of the use of this approach. Standard errors and confidence limits can be computed using equation (13) for $S_{\hat{y}}$ (simply set $x = 0$).

Relationships between the PIC and PGLS Methods

Because the independent contrast scores are computed as $\mathbf{C}'\mathbf{y}$ and $\mathbf{Y}$ is assumed to have a covariance matrix of $\boldsymbol{\Sigma}$, the covariance matrix for the contrast scores will be equal to $\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C}$. Felsenstein (1985) states that, under the Brownian motion model for evolution, the PIC scores are independent and each has a variance of unity. This implies that $\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C} = \mathbf{I}_{n-1}$. Although awkward to prove because the PIC method is just an algorithm, empirically one can show that this important relationship holds. Thus, even though the contrasts are not orthogonal, the contrast scores will be independent if the observations have a covariance matrix proportional to $\boldsymbol{\Sigma}$.

Let $\mathbf{T} = [\mathbf{YX}]$ be a combined matrix of the observations for the p Y variables and the q X variables (the unit vector is not included in $\mathbf{X}$). The PGLS estimate of the sums of squares and cross-products matrix is

$$\mathbf{A} = (\mathbf{T} - \bar{\mathbf{T}})'\boldsymbol{\Sigma}^{-1}(\mathbf{T} - \bar{\mathbf{T}}), \tag{31}$$

where $\bar{\mathbf{T}}$ is a vector containing the estimates of the Y and X variables. Expressed in terms of independent contrasts this matrix is

$$
\begin{aligned}
\mathbf{A} &= (\mathbf{C}'\mathbf{T} - \mathbf{C}'\bar{\mathbf{T}})'(\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C})^{-1}(\mathbf{C}'\mathbf{T} - \mathbf{C}'\bar{\mathbf{T}}) \\
&= (\mathbf{T} - \bar{\mathbf{T}})'\mathbf{C}(\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C})^{-1}\mathbf{C}'(\mathbf{T} - \bar{\mathbf{T}}) \\
&= (\mathbf{T} - \bar{\mathbf{T}})'\mathbf{C}\mathbf{C}'(\mathbf{T} - \bar{\mathbf{T}}) \\
&= (\mathbf{C}'\mathbf{T})'\mathbf{C}'\mathbf{T}. \tag{32}
\end{aligned}
$$

The third line of this equation uses the relationship $\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C} = \mathbf{I}_{n-1}$. The last step uses the fact that $\mathbf{C}$ and $\bar{\mathbf{T}}$ are orthogonal. Thus, the sums of squares matrix, expressed as deviations from the means (estimates for the root), is the same as for contrasts expressed as deviations from the origin. Note that the arbitrariness of trying to compute a mean contrast is avoided. From the $\mathbf{A}$ matrix one can compute correlations between any pair of variables or the regression of any variable on one or more of the others.

The residual sums of squares for Y is

$$\mathbf{A}_{Y.X} = \mathbf{A}_Y - \mathbf{A}_{YX}\mathbf{A}_X^{-1}\mathbf{A}_{XY}, \tag{33}$$

where $\mathbf{A}_Y$ is the upper left $p \times p$ block of $\mathbf{A}$ (the total sums of squares matrix for Y), $\mathbf{A}_{YX}$ is the lower left $p \times q$ block of $\mathbf{A}$, etc. When divided by its degrees of freedom, $n - 1 - q$, this estimates the stochastic variance of the model. Garland et al. (1999) interpret this variance as a measure of the rate of character evolution, but this is complicated by the fact that the branch lengths may be in rather arbitrary units. It would be of most interest if the branch lengths were all in units proportional to time. Calling this variance a rate does not seem appropriate in the context of a Brownian motion model of evolution (see Bookstein 1987). Garland (1992) also discusses these issues.

These results demonstrate another important relationship. What the change of variables in equation (20) does is to project a vector, y, onto the n-dimensional space spanned by the eigenvectors of the inverse of the phylogenetic covariance

matrix. Because $\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C} = \mathbf{I}_{n-1}$ this is almost the same as projecting a variable on to the $n - 1$ PIC vectors. The difference in the number of dimensions is because the space spanned by the eigenvectors includes the direction corresponding to the mean (the direction corresponding to the projection of the unit vector into this space). If that direction were removed (either by multiplication by the $\mathbf{N}$ matrix of equation [29] or by using deviations from the mean), then the space defined by these eigenvectors will differ from that of the PIC vectors by only a rigid rotation. Thus, statistical analyses of the projections and the PIC scores will yield identical results. Although the computation of the PIC scores is straight-forward, multiplication by $\boldsymbol{\Lambda}^{-1/2}\mathbf{E}^t$ can be done by standard software and automatically takes into account any multifurcations in the phylogenetic tree.

Relationships between the PA and PGLS Methods

Comparison of PA and PGLS methods is complicated due to ambiguity of what distances are used to define the $\mathbf{W}$ matrix used in PA. Gittleman and Kot (1990) refer to matrices of patristic distances (sums of branch lengths along paths connecting pairs of species), but seem to actually use ultrametric similarities (using the height above the root of the most recent common ancestor of each pair of species). Although not explicitly stated, Martins (1996) also uses ultrametric similarities. Diniz-Filho et al. (1998) say that the distance used is usually the percentage of divergence between pairs of species (which would also be a patristic distance), but they also use ultrametric similarities from an estimated phylogeny in their examples. In the present paper the $\mathbf{W}$ matrix is constructed by first entering the height above the root of the most recent common ancestor of each pair of species as in the computation of the phylogenetic covariance matrix (see above).

Because of the form of equation (24), Purvis et al. (1994) and Purvis and Webster (1999) say that PA uses variation in the specific component after the component due to phylogeny is discarded whereas PIC (and hence PGLS) views all variation as potentially due to adaptation. Actually, both methods analyze the total variation relative to a matrix that standardizes the data to take into account the phylogenetic relationships among the OTUs—they simply use different matrices, compare equations (12) and (24). Martins and Hansen (1997), in contrast, interpret PA as a method with a simple error structure but a complex mean structure to account for the lack of independence among the OTUs. However, an autoregressive model can be expressed as a generalized least-squares model (Griffith 1988). Rearranging equation (23) yields $\boldsymbol{\epsilon} = (\mathbf{I} - \rho\mathbf{W})\mathbf{y}$ and thus the autoregressive model can be written as

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\epsilon}, \tag{34}$$

where $\boldsymbol{\epsilon}$ is assumed to be multivariate normal with a co-variance matrix of $\sigma^2\mathbf{I}_n$. This model assumes that the mean of $\mathbf{y}$ is zero. A more general model includes a term for the mean,

$$\mathbf{y} = \mu\mathbf{1} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\epsilon}. \tag{35}$$

Once ρ is estimated using maximum likelihood (see the Ap-

pendix), the mean and variance of Y are estimated as in a PGLS analysis by substituting $(\mathbf{I} - \rho\mathbf{W})^t(\mathbf{I} - \rho\mathbf{W})$ for $\mathbf{\Sigma}^{-1}$ in the PGLS equations. The usual use of deviations from the ordinary unweighted mean in a PA analysis is not correct because the expected covariance among the observations is not taken into account (see next section).

The model is easily extended to include the regression

$$\mathbf{y} = \mathbf{Xb} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\epsilon}. \qquad (36)$$

In the PGLS regression model, $\boldsymbol{\epsilon}$ is assumed to have a covariance matrix equal to $\boldsymbol{\sigma}^2\mathbf{\Sigma}$. In the simultaneous autoregressive model of equation (36), $(\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\epsilon}$ is the error term with a covariance matrix of $\boldsymbol{\sigma}^2[(\mathbf{I} - \rho\mathbf{W})^t(\mathbf{I} - \rho\mathbf{W})]^{-1}$. Again, the autoregressive parameter, ρ, must be estimated using maximum likelihood, but the other parameters can be estimated as in a PGLS analysis. For example, based on equation (9) we can estimate the regression parameters as

$$\hat{\mathbf{b}} = [\mathbf{X}^t(\mathbf{I} - \rho\mathbf{W})^t(\mathbf{I} - \rho\mathbf{W})\mathbf{X}]^{-1}\mathbf{X}^t(\mathbf{I} - \rho\mathbf{W})^t(\mathbf{I} - \rho\mathbf{W})\mathbf{y} \quad (37)$$

for the simultaneous autoregressive model in equation (35).

Martins and Hansen (1996b) make the point that the $\mathbf{W}$ matrix used by Cheverud and Dow (1985) and by Gittleman and Kot (1990) is not compatible with any known microevolutionary process (i.e., its use is not equivalent to a $\mathbf{\Sigma}$ matrix based on a microevolutionary process). The regression coefficients based on PA will not be the same as in a PGLS analysis unless $[(\mathbf{I} - \rho\mathbf{W})^t(\mathbf{I} - \rho\mathbf{W})]^{-1}$ is equal to $\mathbf{\Sigma}$. This is not possible if one requires the $\mathbf{W}$ matrix to be a connection matrix with zeros down its principal diagonal and rows that sum to unity. Thus, PA cannot correspond to PGLS based on any evolutionary model. If we ignore the requirements that there must be zeros down the diagonal and rows that sum to one, then there are an infinite number of possible solutions (corresponding to the different ways in which the $\mathbf{\Sigma}$ matrix can be factored). An obvious choice is to set $\mathbf{W}$ equal to $(1/\rho)(\mathbf{I} - \mathbf{\Lambda}^{-1/2}\mathbf{E}^t)$, where $\mathbf{\Lambda}$ and $\mathbf{E}$ are the matrices of eigenvalues and eigenvectors of $\mathbf{\Sigma}$. This would give the same fit to the data as in PGLS and thus yield the same phylogenetically adjusted regressions or correlations. One can verify this by the fact that this choice makes $(\mathbf{I} - \rho\mathbf{W})$ equal to $\mathbf{\Lambda}^{-1/2}\mathbf{E}^t\mathbf{y}$ as in equation (20). But this choice no longer corresponds to PA as defined by Cheverud and Dow (1985). The estimate, $\hat{\mathbf{y}} = \rho\mathbf{Wy}$, should also not be expected to be a very effective estimate for OTUs in a phylogeny. In the PA model the prediction for the ith OTU is $\rho \sum_{j=1}^{n} w_{ij}y_j$. In a simple phylogenetic model, a more reasonable estimate for an OTU would be an estimate of the state of its most recent common ancestor (e.g., the ith row of $\mathbf{W}$ should correspond to the weights needed to give the squared-change parsimony estimate of the most recent common ancestor of the ith OTU. The $\mathbf{W}$ matrix is not constructed in that way. In PA the weights for the ith OTU are simply the heights of the most recent common ancestors of each OTU and the ith OTU (except that the weight for the ith OTU itself is taken as zero).

Phylogenetic independent contrasts.—Despite their name, the contrasts are independent only if the data are generated by the simplest Brownian motion continuous random-walk model of evolution. Because PIC is expressed just as an algorithm, it is difficult to generalize it to allow for more complex evolutionary models. By interpreting the usual formulas in terms of the dual space, it is shown that the usual regressions and correlations computed using independent contrasts are through the GLS estimate of the mean not the origin, as commonly stated.

Phylogenetic generalized least-squares.—This standard statistical method can directly take into account the expected pattern of covariation among OTUs expected on the basis of a given phylogeny. The PIC method represents a particular algorithm for implementing the PGLS method on a continuous Brownian motion type random-walk model of evolution. If an expected covariance matrix among OTUs can be constructed for some alternative model, then the PGLS method is immediately extended. A formula that gives the correct standard errors for the squared-change parsimony estimates of the ancestral states is presented (unfortunately, the resulting standard errors are usually much larger than previous estimates based on both PGLS and on maximum likelihood).

Phylogenetic autoregression.—The maximum-likelihood estimates of the autoregressive parameter, ρ, have not been computed correctly (the Appendix explains the problem and provides a corrected algorithm). A more fundamental problem is that this method cannot be made equivalent to the PIC and PGLS methods for any known evolutionary model and thus it is difficult to interpret the results of PA biologically. A further complication is that, for a given tree, multiple perfect fits to the autoregressive model are possible corresponding to various values of ρ (not just $\rho = 1$). Problems with methods for estimating phylogenetic inertia are also demonstrated. It was concluded that the PA method is not very useful in comparative studies.

Although this paper only deals with the simplest evolutionary model and only for the unrealistic case where the phylogeny (including branch lengths) is known and, I hope that the results given here will be useful in clearing up some common misconceptions about some of the methods used in comparative studies. The field is still young, and many methodological papers have made conflicting suggestions. The use of comparative methods is important and should become routine. An advantage of the PGLS method is that it is a standard statistical approach and is easy to apply to a wide variety of statistical designs (e.g., regression, ANOVA, ANCOVA, MANOVA) without users having to invent special ad hoc approaches to handle a particular set of data. From the results presented here, it seems prudent to concentrate future effort on extending the PGLS approach to handle more realistic cases.

DISCUSSION

This study has led to a number of new results for the three major approaches used in comparative studies.

Literature Cited

Bookstein, F. L. 1987. Random walk and the existence of evolutionary rates. Paleobiology 11:258–271.

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

Carroll, J. D., and P. E. Green. 1997. Mathematical tools for applied multivariate analysis. Academic Press, New York.

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

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

Cliff, A. D., and J. K. Ord. 1981. Spatial processes: models and applications. Pion Ltd., London.

Clobert, J., T. Garland Jr.,and R. Barbault. 1998. The evolution of demographic tactics in lizards: a test of some hypotheses concerning life history evolution. J. Evol. Biol. 11:329–364.

deJong, P., C. Sprenger, and F. Vanveen. 1984. On extreme values of Morans-I and Gearys-C. Geogr. Anal. 16:17–24.

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.

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

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

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

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. 1999. Polytomies and phylogenetically independent contrasts: examination of the bounded degrees of freedom approach. Syst. Biol. 48:547–558.

Garland, T., Jr.,and A. R. Ives. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. Am. Nat. 155:346–364.

Garland, T., Jr., 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., P. E. Midford, and A. R. Ives. 1999. An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral values. Am. Zool. 39: 374–388.

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

Gittleman, J. L., and H. K. Luh. 1994. Phylogeny, evolutionary models and comparative methods: A simulation study. Pp. 103–122 *in* P. Eggleton and D. Vane-Wright, eds. Phylogenetics and ecology. Academic Press, London.

Gittleman, J. L., C. G. Anderson, M. Kot, and H. K. Luh. 1996. Comparative tests of evolutionary lability and rates using molecular phylogenies. Pp. 289–307 *in* P. H. Harvey, A. J. L. Brown, J. M. Smith, and S. Nee, eds. New uses for new phylogenies. Oxford Univ. Press, New York.

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

———. 1992. The uniqueness of the phylogenetic regression. J. Theor. Biology 156:405–423.

Griffith, D. A. 1988. Advanced spatial statistics. Kluwer Academic Publishers, Dordrecht, The Netherlands.

———. 1996. Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. Can. Geogr. 40:351–367.

Griffith, D. A., and P. L. Fellows. 1999. Pixels and eigenvectors: classification of LANDSAT TM imagery using spectral and locational information. Pp. 309–317 *in* K. Lowell and A. Jaton, eds. Spatial accuracy assessment: land information uncertainty in natural resources. Ann Arbor Press, Chelsea, MI.

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. 2000. Why and how phylogenetic relationships should be incorporated into studies of scaling. Pp. 253–265 *in* J. H. Brown and G. B. West, eds. Scaling in biology. Oxford Univ. Press, New York.

Harvey, P. H., and A. Rambaut. 1998. Phylogenetic extinction rates and comparative methodology. Proc. R. Soc. Lond. B Biol. Sci. 265:1691–1696.

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

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

———. 2000. Testing character correlation using pairwise comparisons on a phylogeny. J. Theor. Biol. 202:195–204.

Martins, E. P. 1995. Phylogenies and comparative data, microevolutionary perspective. Philos. Trans. R. Soc. Lond. B Biol. Sci. 349:85–91.

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

———. 1999. Estimation of ancestral states of continuous characters: a computer simulation study. Syst. Biol. 48:642–650.

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. 1996a. A microevolutionary link between phylogenies and comparative data. Pp. 273–288 *in* P. H. Harvey, A. J. L. Brown, J. M. Smith, and S. Nee, eds. New uses for new phylogenies. Oxford Univ. Press, New York.

———. 1996b. The statistical analysis of interspecific data: a review and evaluation of phylogenetic comparative methods. Pp. 22–75 *in* E. P. Martins, ed. Phylogenies and the comparative method in animal behavior. Oxford Univ. Press, Oxford.

———. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149:646–667.

MathWorks. 2000. Using MATLAB. MathWorks, Natick, MA.

McArdle, B., and A. G. Rodrigo. 1994. Estimating the ancestral states of a continuous-valued character using squared-change parsimony: An analytical solution. Syst. Biol. 43:573–578.

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

Miles, D. B., and A. E. Dunham. 1996. The paradox of the phylogeny: character displacement of analyses of body size in island *Anolis*. Evolution 50:594–603.

Møller, A. P., and T. R. Birkhead. 1992. A pairwise comparative method as illustrated by copulation frequency in birds. Am. Nat. 139:644–656.

Moran, P. A. P. 1950. Notes on continuous stochastic phenomena. Biometrika 37:17–23.

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

———. 1993. Seeking the evolutionary regression coefficient: an

analysis of what comparative methods measure. J. Theor. Biol. 164:191–205.

———. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Piazza, A., and L. L. Cavalli-Sforza. 1975. Spectral analysis of patterned covariance matrices and evolutionary relationships. Pp. 76–105 *in* G. F. Estabrook, ed. Proceedings of the eighth international conference on numerical taxonomy. W. H. Freeman, San Francisco, CA.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1989. Numerical recipes in Pascal. Cambridge Univ. Press, New York.

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

Purvis, A., and A. J. Webster. 1999. Phylogenetically independent comparisions and primate phylogeny. Pp. 44–68 *in* P. C. Lee, ed. Comparative primate socioecology. Cambridge Univ. Press, Cambridge, U.K.

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.

Rao, C. R., and H. Toutenburg. 1999. Linear models: least squares and alternatives, 2nd ed. Springer, New York.

Rohlf, F. J. 2000. NTSYS-pc: numerical taxonomy and multivariate analysis system. Exeter Software, Setauket, NY.

Rohlf, F. J., and R. R. Sokal. 1981. Comparing numerical taxonomic studies. Syst. Zool. 30:459–490.

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

Sneath, P. H. A., and R. R. Sokal. 1973. Numerical taxonomy: the principles and practice of numerical classification. W. H. Freeman, San Francisco, CA.

Sokal, R. R. 1979. Testing statistical significance of geographic variation patterns. Syst. Zool. 28:227–232.

Sokal, R. R., and F. J. Rohlf. 1962. The comparison of dendrograms by objective methods. Taxon 11:33–40.

———. 1995. Biometry. The principals and practice of statistics in biological research, 3rd ed. W. H. Freeman, San Francisco, CA.

Upton, G., and B. Fingleton. 1985. Spatial data analysis by example. John Wiley and Sons, New York.

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

Corresponding Editor: T. Garland, Jr.

APPENDIX

Maximum-Likelihood Estimation of the Phylogenetic Autoregressive Coefficient, ρ

This appendix describes a maximum-likelihood procedure to estimate the autoregressive parameter, ρ, in the model

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \boldsymbol{\epsilon}, \qquad (A1)$$

where **y** is a vector containing the n observed values for some variable, **W** is an $n \times n$ connection matrix (with zeros in the principal diagonal and each row summing to one), and $\boldsymbol{\epsilon}$ is the error (assumed normal with a mean of zero and a variance σ^2). The maximum-likelihood estimate, $\hat{\rho}$, is the value of ρ that minimizes

$$L = \ln(\mathbf{y}'\mathbf{y} - 2\rho\mathbf{y}'\mathbf{W}\mathbf{y} + \rho^2\mathbf{y}''\mathbf{y}') - \frac{1}{n}\ln|\mathbf{I} - \rho\mathbf{W}|^2, \quad (A2)$$

where $\mathbf{y}' = \mathbf{W}\mathbf{y}$ (Cliff and Ord 1981). Note that the determinant is squared rather than multiplying its log by two as is done by most authors. This distinction is necessary because the determinant will be negative when ρ > 1. In practice, the determinant is computed using the relationship $|\mathbf{I} - \rho\mathbf{W}| = \Pi_{i=1}^{n} (1 - \rho\lambda_i)$ because it requires much less computation when searching for the value of ρ that minimizes L. Whereas the data vector **y** can be standardized (expressed as standard deviations from the mean), $\mathbf{y}'$ must *not* be standardized
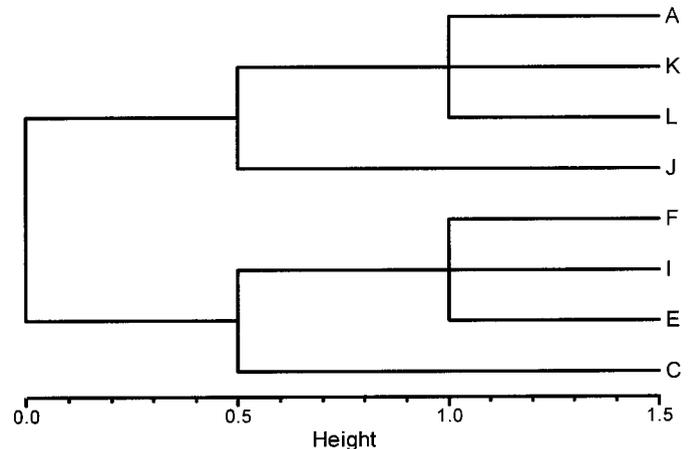

FIG. A1. Phylogenetic tree redrawn from Cheverud and Dow (1985). The height of each furcation is as used in the computations.

as was done by Cheverud and Dow (1985), Cheverud et al. (1985, p. 1341), and Gittleman and Kot (1990). This is because it would imply a rescaling of the **W** matrix in the first term of equation (A2) without a corresponding adjustment to the second term (D. Griffith, pers. comm.). Such standardization destroys the relationship between ρ and the eigenvalues and the right eigenvectors of **W** described in this paper. Cheverud et al. (1985, p. 1340) state that having each row of **W** sum to one constrains the estimate to $|\rho| < 1$. Actually, this just ensures that the largest eigenvalue will be equal to one. Griffith (1996) emphasizes the importance of the constraint $|\rho| < 1$ to ensure stationarity and invertibility between autoregressive and moving average models but this does not seem relevant for phylogenetic models.

Because larger values of ρ do not imply a better fit to the model, the actual residual sums of squares needs to be computed to assess how well the model fits the data. Note that, unlike conventional regression, these estimated residuals, $y - \hat{y}$, need not have a mean of zero.

The data from Cheverud and Dow (1985) serve as a convenient computational example. Figure A1 shows their phylogenetic tree. The $n = 8$ values for their GM1 variable are −0.12, 0.36, −0.1, 0.04, −0.15, 0.29, −0.11, and −0.06 (listed in the same order as in the phylogenetic tree). Table A1 gives both their matrix of distances between the eight groups and their row normalized connection matrix **W**. The upper half of Table 1 gives the right eigenvectors (the vectors, **v**, that satisfy the equation $\lambda\mathbf{v} = \mathbf{W}\mathbf{v}$) and their eigenvalues, λ.

The solid line in Figure A2 is a plot of L as a function of ρ for these data. Clearly, locating the value of ρ that minimizes L is complicated by the complexity of the likelihood function. Cheverud and Dow (1985) and Cheverud et al. (1985) used a direct search for the value of ρ that minimizes L but just over the interval −1 to 1 because they standardized the $\mathbf{y}'$ variable. Although estimates of ρ are often within that interval, the actual range of possible estimates is from $\lambda_{-\min}^{-1}$ to $\lambda_{+\min}^{-1}$, where $\lambda_{-\min}^{-1}$ and $\lambda_{+\min}^{-1}$ are the negative and positive eigenvalues that are closest to zero. For our example, $\lambda_{-\min} = -0.2$ and $\lambda_{+\min} = 1$ (see Table 1). Thus, the range of possible values of ρ for the given tree is from −5 to 1.

The search for the value of ρ that minimizes L is made more difficult by the fact that L is infinite when ρ is equal to the reciprocals of any of the eigenvalues. The likelihood function "blows up" because data with ρ values equal to the reciprocals of the eigenvalues will perfectly fit the model given in equation (A1). Figure A2 also shows the separate contributions of the two terms on the right-hand side of equation (A2). The first term (plotted as a dashed line) is well behaved. The second term (shown as a dotted line) is a function of the log of a determinant and goes to infinity when ρ is equal to the reciprocals of any of the eigenvalues because the determinant goes to zero in those cases. Note that this second term does not depend upon the observed data, **y**. Once one computes the

TABLE A1. Distance matrix and its corresponding normalized connection matrix from table 2 of Cheverud and Dow (1985).

	Distance matrix, **D**								Normalized connection matrix, **W**							
	A	K	L	J	F	I	E	C	A	K	L	J	F	I	E	C
A	0	1	1	2	0	0	0	0	0	0.4	0.4	0.2	0	0	0	0
K	1	0	1	2	0	0	0	0	0.4	0	0.4	0.2	0	0	0	0
L	1	1	0	2	0	0	0	0	0.4	0.4	0	0.2	0	0	0	0
J	2	2	2	0	0	0	0	0	0.333	0.333	0.333	0	0	0	0	0
F	0	0	0	0	0	1	1	2	0	0	0	0	0	0.4	0.4	0.2
I	0	0	0	0	1	0	1	2	0	0	0	0	0.4	0	0.4	0.2
E	0	0	0	0	1	1	0	2	0	0	0	0	0.4	0.4	0	0.2
C	0	0	0	0	2	2	2	0	0	0	0	0	0.333	0.333	0.333	0

eigenvalues of **W**, one can use standard methods to search separately in the intervals between pairs of the λ_i^{-1} values. The Golden section search method, as described in Press et al. (1989), was found to work well. The listing below gives a function written in the MATLAB language (MathWorks 2000) that carries out these operations. The function is called with two arguments: **y**, a column vector of n observed values, and **W**, an $n \times n$ connection matrix. The maximum-likelihood estimate of ρ is returned. The code assumes that the row sums of **W** are greater than zero. To save space, code for various validity checking is not included. This method is also available in the NTSYSpc software (Rohlf 2000). Note that the code is provided here to make the corrected algorithm more explicit, not to encourage the use of the PA method in comparative studies.

Using this approach a value of $\hat{\rho} = -1.4415$ was obtained for the GM1 variable. Cheverud and Dow (1985) reported a value of -0.86. Unfortunately, software implementing the PA method (and hence all studies based upon their use) has used the incorrect algorithm for estimating ρ.

MATLAB function autoregr and its support routines is as follows:

```
function [rho] = autoregr (y,W)
% Estimate auto regressive coeff (rho) using maximum likelihood
% Uses Golden Search between pairs of eigenvalues
% to use:
% [rho] = autoregr (y,W)
% solution must be in range 1/Ln to 1/Lp,
% where Ln and Lp are negative and positive eigenvalues
% closest to zero of the normalized W matrix.
```
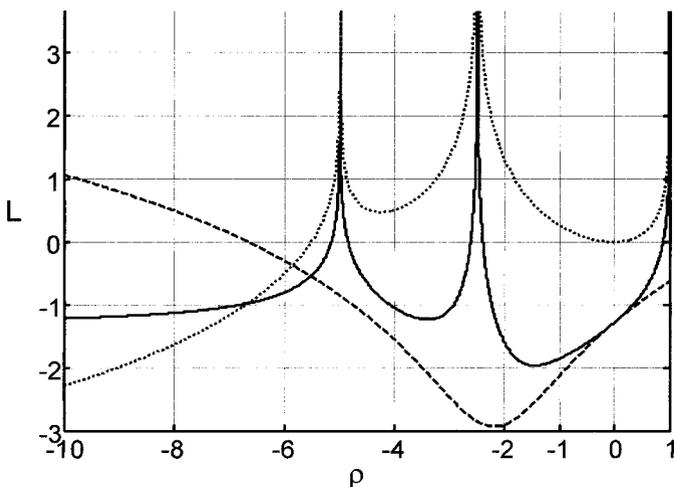

FIG. A2. Maximum-likelihood estimation of the autoregressive coefficient, ρ. The function to be minimized is given as a solid line and its two components are given as dashed and dotted lines. See text for explanation. The three peaks correspond to the three eigenvalues of Σ, $\lambda_i = 1, -0.2, -0.4$, and $\rho_i = 1/\lambda_i$. The global minimum is at $\rho = -1.44$.

```
% rows of W should not be constant.
   W = W - diag (diag(W)); % ensure diagonal is zero
   D = inv (diag(sum(W')));
   Wnorm = D*W; % row normalize W matrix
   [n,m] = size(y);
   y = y - ones (n,1)*mean(y); % deviations from mean
   Wy = Wnorm*y;

   Wlam = eig(Wnorm); % eigenvalues of W
   % find distinct eigenvalues
   sorted = sort(Wlam);
   clear Distinct;
   Distinct(1) = -inf;
   Distinct(2) = sorted(1);
   nDistinct = 2;
   for ii = 2:n
     if sorted(ii)-Distinct(nDistinct)>1e-6 % tolerance
        nDistinct = nDistinct+1;
        Distinct(nDistinct) = sorted(ii);
     end;
   end;
   % search for minimum of LL
   LL = inf;
   for ii = 2:nDistinct-1 % search between pairs of roots
     % [to constrain do not use positive roots < 1]
     ax = 1/Distinct(ii);
     cx = 1/Distinct(ii+1);
     [rhoi,LLi] = GoldenSearch(y,Wy,Wlam,ax,cx);
     if LLi<LL
        LL = LLi; rho = rhoi;
     end;
   end;
return

function [LL] = likelihoodrho(n,y,Wy,Wlam,rhohat)
% computes likelihood of rho
% assumes: logDet = log(det(eye(n) -rhohat*W))
   DetProd = 1;
   for j = 1:n
     prod = 1-rhohat*Wlam(j);
     DetProd = DetProd*prod;
   end;
   absValDet = abs(DetProd); % [abs to allow rho > 1]
   if absValDet==0
     logDet = -inf;
   else
     logDet = log(absValDet);
   end;
   LL = log(y'*y-2*rhohat*y'*Wy+rhohat*rhohat*Wy'*Wy)-logDet*2/n;
return

function [rhohat,likelihood] = GoldenSearch(y,Wy,Wlam,ax,cx)
% Golden section search over the interval ax to cx
% returns rhohat and likelihood value.
   r = 0.61803399; tol = 0.0001; % may wish to alter tol
   [n,m] = size(y);
   x0 = ax; x3 = cx; bx = (ax+cx)/2;
   if abs(cx-bx)>abs(bx-ax)
```

```
  x1 = bx; x2 = bx+(1-r)*(cx(??)-bx);          x3 = x2; x2 = x1;
else                                           x1 = r*x2+(1-r)*x0;
  x2 = bx; x1 = bx-(1-r)*(bx-ax);              f2 = f1; f1 = likelihoodrho(n,y,Wy,Wlam,x1);
end;                                         end;
f1 = likelihoodrho(n,y,Wy,Wlam,x1);        end;
f2 = likelihoodrho(n,y,Wy,Wlam,x2);        if f1 < f2
while abs(x3-x0) > tol*(abs(x1)+abs(x2))      likelihood = f1; xmin = x1;
  if f2<f1                                   else
    x0 = x1; x1 = x2;                          likelihood = f2; xmin = x2;
    x2 = r*x1+(1-r)*x3;                      end;
    f1 = f2; f2 = likelihoodrho(n,y,Wy,Wlam,x2);   rhohat = xmin;
  else                                     return
```