PHYLOGENIES AND THE COMPARATIVE METHOD: A GENERAL APPROACH TO INCORPORATING PHYLOGENETIC INFORMATION INTO THE ANALYSIS OF INTERSPECIFIC DATA

Emilia P. Martins1,* and Thomas F. Hansen2†

1Department of Biology, University of Oregon, Eugene, Oregon 97403; 2Department of Biology, University of Oregon, Eugene, Oregon 97403, and Department of Biology, Division of Zoology, University of Oslo, P.O. Box 1050 Blindern, N-0316 Oslo 3, Norway

Submitted January 18, 1996; Revised July 31, 1996; Accepted August 1, 1996

Abstract.—This article considers the statistical issues relevant to the comparative method in evolutionary biology. A generalized linear model (GLM) is presented for the analysis of comparative data, which can be used to address questions regarding the relationship between traits or between traits and environments, the rate of phenotypic evolution, the degree of phylogenetic effect, and the ancestral state of a character. Our approach thus emphasizes the similarity among evolutionary questions asked in comparative studies. We then discuss ways of specifying the sources of error involved in a comparative study (e.g., measurement error, error due to evolution along a phylogeny, error due to misspecification of a phylogeny) and show how the impact of these sources of error can be taken into account in a comparative analysis. In contrast to most existing phylogenetic comparative methods, our procedure offers substantial flexibility in the choice of microevolutionary assumptions underlying the statistical analysis, allowing researchers to choose assumptions that are most appropriate for their particular set of data and evolutionary question. In developing the approach, we also propose novel ways of incorporating within-species variation and/or measurement error into phylogenetic analyses, of estimating ancestral states, and of considering both continuous (quantitative) and categorical (qualitative or "state") characters in the same analysis.

In the last few years, there has been considerable debate regarding what can be inferred from the analysis of comparative or interspecific data and how such analyses should be conducted. Some researchers would like to determine whether two traits evolved in a correlated fashion; others estimate the degree of "phylogenetic inertia,"" the ancestral states of a character, or the rate of phenotypic evolution. Several statistical methods have been proposed to analyze comparative data in a phylogenetic context (see Harvey and Pagel 1991 or Martins and Hansen 1996b for reviews). These methods have been developed primarily to address the problem that comparative data are not statistically independent and therefore violate one of the most basic assumptions of most standard statistical procedures. The methods differ in scope, context, assumptions, and exact question being answered. They were also derived from a wide variety of sources and include statistical transformations of the data, algorithmic techniques, and computer ran-
domination procedures. Thus, it is difficult to know whether an estimate of ances-
tral states from one method is analogous to the states estimated by another
method or how an estimate of phylogenetic inertia differs from the estimate of an
evolutionary correlation. Similarly, there has been considerable debate regarding
whether comparative data can be used effectively to infer adaptation (Lauder et
al. 1993; Frumhoff and Reeve 1994; Leroi et al. 1994; Hansen 1996) and about
what might be appropriate assumptions for a comparative analysis (e.g., Nee et
al. 1996; Ridley and Grafen 1996).

In the current article, we develop a general approach for the analysis of compar-
ative data from an evolutionary perspective. This perspective emphasizes the
various components that are a part of any phylogenetic comparative analysis. By
combining these components creatively, biologists can design the most appro-
priate phylogenetic comparative method for their set of data and evolutionary
questions. The approach is explicitly statistical and differs from most previous
methods in that it allows for the incorporation of a variety of microevolutionary
assumptions, the consideration of both continuous and categorical variables, and
within-species variability or measurement error. It clarifies the critical issues and
assumptions that are a part of all comparative analyses and suggests how to make
informed choices among assumptions. When the chosen assumptions are correct,
it is also guaranteed to solve the problem of statistical nonindependence of com-
parative data. The approach shows how various evolutionary questions are re-
lated to one another and provides a starting place for the development of new
methods to address those questions and for a direct analytical comparison to
existing methods.

TRANSLATING EVOLUTIONARY QUESTIONS INTO STATISTICAL MODELS

The first step in a comparative study is to choose a hypothesis and to make
that hypothesis statistically explicit. A number of questions can be addressed
with comparative data and phylogenetic analyses (Martins 1995b; Martins and
Hansen 1996a). We concentrate on four possibilities, which may not be mutually
exclusive but which encompass many of the questions in the recent evolutionary
literature. Once a particular question has been chosen, it must be translated
into a statistical model describing the relationships among relevant characters,
environments, and other aspects of the comparative data before quantitative
analysis of the data can be conducted. In phylogenetic analyses, this involves specifica-
tion of the details of a statistical model. Such a model will usually include both
a mean structure relating the characters to one another and the structure of an
assumed error term \( Y = \mu + \epsilon \). The mean structure \( \mu \) is the expected value
of the response variable as a function of the predictor variables (e.g., \( \mu = \beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots \)). This mean structure is combined with an error term \( \epsilon \)
describing the impact of any factors that affect our ability to obtain accurate
parameter estimates. For example, most parametric statistical tests assume that
\( \epsilon \) is normally distributed with a mean of 0 and constant variance. We concentrate
first on the mean structure.
Fruit Size
Disperser Size

2.4 2.6 2.0 3.2 2.2 1.3 1.5 3.6 4.0
80 100 45 88 53 60 95 73 92

Fig. 1.—Fictitious example of a phylogeny and comparative data used to address questions involving the evolution of seed disperser size and fruit size. The original idea for this hypothesis is taken from Jordano (1995), but we have invented the data and phylogeny purely for illustrative purposes. Trait values are given above the location of each extant taxon. Numbers beside branches refer to their lengths. Lowercase letters are names for the internal nodes.

Relationships between Traits or between Traits and Environments

The most common questions asked in modern comparative studies concern relationships between two or more variables (figs. 1, 2). Researchers may be interested in uncovering universal patterns among widely diverse taxa without making explicit evolutionary interpretations, in considering the constraints imposed on the evolution of one trait by the evolution of another, or in determining whether a trait is an adaptation to its environment. They might ask, for example, whether the fleshy fruits of angiosperm plants evolved in response to the type of animals that disperse their seeds (Jordano 1995; figs. 1, 2) or whether the evolution of warning coloration in insect larvae led to the evolution of gregarious social behavior (Sillén-Tullberg 1988, 1993).

The usual approach to these sorts of questions is to consider whether variation in a trait can be predicted by measures of other traits in the same taxon. For a single taxon, we can use the standard linear regression formula:

\[ B_i = \beta_0 + C_1 \beta_1 + C_2 \beta_2 + \ldots + \varepsilon_i, \]

(1a)

where \( B_i \) is a response variable in taxon \( i \) and the \( C \)'s are predictor variables. The regression coefficients, \( \beta_0, \beta_1, \) and \( \beta_2, \) describe the relationships between \( B \) and \( C_0, C_1, C_2, \) respectively, and \( \varepsilon_i \) is an error term. In matrix terms,
Fig. 2.—Scatterplot depicting a comparative study in which the question involves the relationship between fruit size and the size of the animal that disperses the seeds. We might use such a plot to consider, for example, Is there a universal relationship between fruit phenotype and disperser size? Have evolutionary changes in seed disperser driven the evolution of fruit phenotype? Have fruits adapted to the types of seed dispersers that are available? The line depicts the relationship between fruit size ($Y$) and seed disperser size ($X$) calculated using a simple form of equation (1) ($Y = \beta_0 + \beta_1 X + \epsilon$). Data and phylogeny are like those depicted in figure 1. Evolution is assumed to occur under a Brownian motion model, such that the error variance matrix $V = \{(13,9,4,4,2,0,0,0,0), (9,13,4,4,2,0,0,0,0), (4,4,13,6,2,0,0,0,0), (4,4,6,13,2,0,0,0,0), (2,2,2,2,13,0,0,0,0), (0,0,0,0,0,13,8,4,4), (0,0,0,0,0,8,13,4,4), (0,0,0,0,0,4,4,13,9), (0,0,0,0,0,4,4,9,13)\}$. Estimates of regression parameters and their standard errors (calculated using eq. [7]) are $\hat{\beta}_0 = 1.30 \pm 3.941$ and $\hat{\beta}_1 = 0.02 \pm 0.048$.

$$B = C\beta + \epsilon,$$  \hspace{1cm} (1b)

where $B$ and $C$ are matrices describing measures or functions of measures of extant species, $\beta$ is a vector of regression coefficients, and $\epsilon$ is vector of error terms.

The matrices $B$ and $C$ may consist of phenotypic measurements of extant taxa, measures of the environment, or even functions of these measured variables. For example, we might consider whether fruit size variation among angiosperms ($B$) can be explained by variation in disperser type ($C$; see figs. 1 and 2 for a worked example). In other cases, we might be interested in the relationship between variables that were not directly measured in extant taxa (e.g., usually some function of the measures of extant taxa). For example, a model of adaptation has been developed (Hansen 1996) involving the relationship between a character ($B$) and the history of the environment ($C$) in which that character is found. In this model, $C_{1i}$ is a function of the time from the root of the tree that taxon $i$ has
spent in state 1 of a particular environment. Similarly, many existing phylogenetic comparative methods (e.g., Cheverud et al. 1985; Felsenstein 1985; Grafen 1989; Maddison 1991) can be described as transformations of measured variables into C and B before application in a standard regression model. For example, Felsenstein’s (1985) method transforms measured species phenotypes into standardized independent contrasts (C and B) by assuming that phenotypic evolution can be described using a Brownian motion model and applying an algorithm involving the phylogeny. These contrasts can then be used to infer the relationship between traits using models such as equation (1).

From a statistical perspective, phylogenetic information can be incorporated either through the mean structure or through the error term of equation (1). For example, Cheverud et al. (1985) propose an autoregressive model that is a form of equation (1) with a complex mean structure. These authors propose partitioning each variable into "phylogenetic" and "specific" effects and the subsequent use of "specific" effects in standard statistical procedures (with simple error structures) to look at the relationship among variables. As with Felsenstein’s (1985) method, we might view this method as a version of equation (1) in which the phylogeny is incorporated into the mean structure by specification of B and C as complex functions of the measured traits. Alternatively, Grafen (1989) proposes the use of a simple mean structure in which both B and C are variables measured directly from extant taxa. In Grafen’s method, the phylogeny is translated into a complex error structure that is then incorporated into the analyses using generalized least squares (GLS) regression techniques. Our own approach here might be viewed as an extension of Grafen’s method in which we show how error structures can be more precisely determined for a particular set of data, phylogenies, and evolutionary question and incorporated using GLS procedures.

Note that methods with different mean structures may ask somewhat different evolutionary questions. For example, the β parameters from Grafen’s and Cheverud et al. ’s methods are not exactly the same, and the two methods are estimating different sorts of evolutionary relationships among characters. Since the problem of statistical nonindependence for comparative data can be solved by incorporating phylogenetic information in either or both ways, equation (1) can be used to describe a wide variety of evolutionary questions, simply by specifying different forms of B, C, and the error term. We recommend starting with a simple mean structure that allows us to estimate the evolutionary relationship of most interest and developing an error structure that corresponds directly to it (as shown later).

Rates of Phenotypic Evolution

A second important issue in phylogenetic comparative studies considers the relationship between time and phenotypic evolution (fig. 3). How quickly do traits evolve? Has one type of trait or group of organisms evolved more or less quickly than another? Has a trait changed more or less quickly than expected, given the phylogeny and mode of evolutionary change? These questions have been asked, for example, by researchers interested in whether behavioral characters are suitable for phylogeny reconstruction or in determining whether the evolution of
Phylogenetic Dictance

A scatterplot depicting a comparative study in which the question involves the relationship between phenotypic evolution and time (as in eq. [2]). The Y-axis measures the phenetic distance (in this case, the variance between fruit sizes) between pairs of phylogenetically related species, and the X-axis measures the phylogenetic distance between those two species (in this case, time). If estimates of phenetic and phylogenetic divergence are available for several taxa, we might use this information to estimate the rate of phenotypic evolution (i.e., the slope of the line) or to determine which model of phenotypic evolution is most appropriate for a set of data. The line depicts the best fit of the model if fruit size is assumed to evolve under a Brownian motion model of evolution (e.g., by random genetic drift alone), as calculated using the method described elsewhere (Martins 1994; a specific case of eq. [2]). Both the scatterplot and regression statistics show that the model fits rather poorly and that special attention should be paid to the divergence between ancestors e and f, which are responsible for the one major outlier. Data and phylogeny are like those depicted in figure 1.
behavior drives the evolution of morphology, or vice versa (e.g., Wilson et al. 1975; Wyles et al. 1983; Wilson 1985; de Queiroz and Wimberger 1993; Gittleman et al. 1996; Irwin 1996; Wimberger and de Queiroz 1996). In these cases, we can use

\[ D_{ij} = P_{ij} \delta + \epsilon_{ij}, \tag{2a} \]

where \( D_{ij} \) is the phenotypic divergence between two taxa (e.g., \( |Z_i - Z_j| \)), where \( Z_i \) is the phenotype of taxon \( i \), \( P_{ij} \) is the expected amount of evolution due to phylogenetic distance (e.g., time) between the two taxa, \( \delta \) is a regression coefficient describing the relationship between phenotypic divergence and time, and \( \epsilon \) is an error term. For example, the relationship between phenotypic divergence \( (D_{ij} = (Z_i - Z_j)^2/2) \) and phylogenetic distance has been considered (Martins 1994) using the simple model expected for phenotypic evolution under random genetic drift (\( \text{fig. 3; } P_{ij} = \text{time} \)) and the more complex, nonlinear model expected if phenotypes are subject to stabilizing selection and also random genetic drift (\( P_{ij} = 1 - \exp[-2at_{ij}], \) where \( \alpha \) is the strength of the restraining force). Several other authors (e.g., Haldane 1949; Turelli et al. 1988; Lynch 1990; Bulmer 1991; Spicer 1993; Uyenoyama 1995) have also explored models that might be considered specific cases of equation (2a). In particular, note that specific versions of equation (2a) can be used to estimate branch lengths on a phylogeny or rates of evolution along particular branches of a phylogeny (Bulmer 1991; Martins 1994; Uyenoyama 1995).

In matrix notation,

\[ D = P\delta + \epsilon, \tag{2b} \]

where \( D \) is a vector of the phenotypic divergences between all pairs of taxa, \( P \) is a matrix of expected similarities due to phylogenetic distance between all pairs of taxa, \( \delta \) is the average rate of phenotypic evolution throughout the clade, and \( \epsilon \) is a vector of error terms.

Again, the phylogeny can be incorporated into either the mean structure (\( P \)) and/or the error structure (\( \epsilon \)), and exact definition of these terms will determine the assumptions of the method. How the correct error term can be calculated when \( D \) is a variance between two species phenotypes and the elements of \( P \) are in units of time has been illustrated elsewhere (Martins 1994). Again, we recommend starting with definitions of phenotypic and phylogenetic divergence that most closely describe the biological question of interest and using the approach described here to determine the error structure most appropriate for those definitions.

**Phylogenetic Effects**

Another common question concerns the degree of phylogenetic "effect," "constraint," or "inertia." How much are traits affected by the phylogenetic history of the species that exhibit them? Is a trait evolving freely, responding quickly to the action of direct selection, or is it constrained by the states of other traits in the same organisms? Phylogenetic effect might be described by a statistical model in several ways. For example, a phylogenetic effect might be a measure of the
amount of phenotypic variation among taxa that is explained by the underlying phylogenetic relationships. In this case, a model similar to equation (2) might be appropriate. Specifically, we might develop a mean structure considering the relationship between phenotypic divergence \((D)\) and phylogenetic divergence \((P)\), determine the appropriate error structure for our model, and use a measure of the fit of that model (e.g., the coefficient of determination, \(r^2\)) as an estimate of the degree of phylogenetic effect. A specific example of this approach is described elsewhere (Martins 1994).

Alternatively, phylogenetic effect might refer to a measure of the relationship between characters measured in extant taxa and characters as they existed in the clade’s hypothetical ancestors. In this case, the mean structure in equation (1) might be preferred, with \(B\) defined as the character measured in extant taxa and \(C\) defined as trait values estimated for the hypothetical ancestors of the group, possibly weighted by their phylogenetic distance from the extant taxa. Again, we could use measures of the fit of that model (e.g., the coefficient of determination, \(r^2\)) as an estimate of the degree of phylogenetic effect. Cheverud et al.’s (1985) and Lynch’s (1991) estimators for phylogenetic inertia are special cases of this approach. Both partition variation in a character by examining how well a taxon’s trait value is predicted by measures of other taxa and then define phylogenetic inertia as this measure of prediction ability. Similar estimators could be obtained for any statistical model developed to infer the relationship between a character and its ancestral states using equation (1). Note again, however, that an error structure that matches the mean structure of the model should also be carefully determined (see below).

Reconstruction of Ancestral States

What is the primitive state of a character? What is the sequence and pattern of evolutionary changes that the character has undergone? Did evolutionary changes in one character precede those in a second character? Many comparative studies are concerned with the pattern of evolutionary changes that a trait has undergone. To observe the pattern of phenotypic evolution, we can estimate the ancestral states of the trait at various points along the phylogeny and use these to plot out the sequence of evolutionary changes along the phylogeny. This is certainly the most important set of questions in phylogenetic systematics and is often the basis for testing hypotheses about phenotypic evolution (e.g., Ridley 1983; Brooks and McLennan 1991). In our regression framework, we can predict an ancestral state as a linear combination of the phenotypes of extant taxa using

\[
A_i = \sum w_{ij} Y_j + \epsilon, 
\]

where the summation is over a set of taxa \(j\), \(w_{ij}\) is the expected similarity between the ancestor \(i\) and taxon \(j\) due to the phylogeny, and \(\epsilon\) is an error term. This model predicts each ancestral state as the weighted average of the measures of other taxa on the phylogeny.

In matrix form,

\[
A = WY + \epsilon, 
\]
Illustration of how ancestral states of fruit size might be estimated using the data and phylogeny in figure 1, using equation (3) and assuming that both data and tree are known without error. Under a Brownian motion model of phenotypic evolution, the variance matrix \( \text{var}[Y] \) is identical to that in figure 2, and the grand mean of the clade \( \mu \) was estimated to be 2.49 (using eq. [11]). Ancestral estimates were obtained using \( \hat{A} = \text{var}[A, Y] \text{var}[Y]^{-1}Y \), where \( \text{var}[A, Y] = \{9,9,4,4,2,0,0,0,0\}, \{4,4,4,4,2,0,0,0,0\}, \{4,4,6,6,2,0,0,0,0\}, \{2,2,2,2,0,0,0,0\}, \{0,0,0,0,0,8,8,4,4\}, \{0,0,0,0,0,4,4,4,4\}, \{0,0,0,0,0,4,4,9,9\} \). Standard errors for estimates of ancestral states are given in parentheses.

where \( A \) is a vector of ancestral states, \( Y \) is a vector of measured comparative data, and \( W \) is a matrix describing the expected similarities between the ancestors and the measured taxa due to shared phylogenetic histories.

Equations (3a) and (3b) are quite general and can be applied to characters having evolved under almost any microevolutionary process and along any phylogeny, simply by applying a \( W \) matrix that describes the way in which the characters evolved. Definition of the \( W \) matrix will also influence the choice of an appropriate error structure. Because this way of estimating ancestral states of a phenotype is essentially novel, we provide further details regarding the estimation procedure below and a worked example in figure 4.

A General Model

The above models (eqq. [1]–[3]) are of the same type and can be described in matrix form as

\[
Y = X\beta + \epsilon,
\]

where \( Y \) is a vector of character states or functions of character states for either extant or ancestral taxa; \( X \) is a vector or matrix of character states, environments, phylogenetic distances, or some function of these; \( \beta \) is a vector of regression
coefficients to be estimated; and \( \varepsilon \) is a vector of error terms with a particular, assumed structure. Equation (3) is a predictive form of the same model in which \( X \) and \( \beta \) are known, and we are trying to predict \( Y \).

We can combine these models into more complex forms. For example, suppose we want to infer the relationship between trait \( Y \) as it exists today and trait \( Z \) as it existed 10 million yr ago using equation (1). Since we do not have direct measures of the ancestral states of \( Z \), we may choose to estimate them using equation (3). We can combine these two estimation procedures into a single complex model by setting \( C \) in equation (1) equal to \( A \) in equation (3):

\[
y_i = [\sum w_{ij}X_j] \beta + \varepsilon_i. \tag{4b}
\]

By creatively defining \( X \) and \( Y \), we can answer a broad variety of evolutionary questions, and by carefully varying the specified error structure (see below), we can make informed choices among evolutionary assumptions.

\textit{Link Functions}

Although the linear model in equation (4) can accommodate a wide range of situations, some common cases cannot be forced into this form. When the response variable (\( Y \)) is naturally restricted in some way (e.g., body size is always positive; other traits are either present or absent), we can use nonlinear models of the mean structure as in a generalized linear model (GLM; see McCullagh and Nelder 1989; Dobson 1990). In GLM, we create a monotonic and differentiable "link" function of the mean structure \((g[\mu] = X\beta)\) and incorporate this link function into equation (4a):

\[
Y = g^{-1}(X\beta) + \varepsilon. \tag{4c}
\]

For example, when the response variable (\( Y \)) exists only as 1 (presence) and 0 (absence), we can use a logit link function, \( g(\mu) = \log[\mu/(1 - \mu)] \), which for any value of \( X\beta \) yields a mean (\( \mu \)) between 0 and 1. Instead of assuming that \( X \) has an additive effect on \( Y \), we assume that \( X \) has an additive effect on the log odds of \( Y \) being present. Taking the inverse of the logit link function, we obtain a logistic regression model:

\[
Y = \exp[X\beta]/(1 + \exp[X\beta]) + \varepsilon. \tag{4d}
\]

Similarly, a log link function \((g[\mu] = \log[\mu])\) can be used to avoid the problem of negative body sizes. Link functions used to form log-linear, probit, and other models can be found in statistical literature discussing GLM (e.g., McCullagh and Nelder 1989; Dobson 1990). When the link function is the identity function, equation (4c) reduces to the more familiar equation (4a).

\textbf{DETERMINING THE STRUCTURE OF THE ERROR TERM}

\textit{Sources of Error}

Once the mean structure for the analysis has been chosen, we can consider each variable and determine the sources of error that may affect it. Predictions
made using measurements of extant taxa are usually subject to at least three
types of error: within-species variation or measurement error ($\epsilon_m$), error due
to the stochastic nature of evolutionary change along a phylogeny ($\epsilon_s$), and error
due to imprecise specification of the phylogenetic relationships among taxa ($\epsilon_p$).
For example, if we consider the relationship between two evolving traits ($X$ and $Y$), we expect both traits to be subject to measurement error ($\epsilon_{XM}$ and $\epsilon_{YM}$), error
due to stochastic evolution of each trait ($\epsilon_{SX}$ and $\epsilon_{SY}$), and possibly even some
interaction between traits ($\epsilon_{SXY}$).

Once the sources of error have been defined for each variable, we need to
determine the type of variance structures we expect for comparative data subject
to each type of error and combine those structures into an overall error term.
This overall variance structure is usually described as a symmetric $N \times N$ vari-
ance matrix, $\text{var}[\epsilon]$, where $N$ is the number of error terms, the diagonal elements
are the variances of the error terms, and the off-diagonal elements are the covari-
ances between them. In an ordinary least squares regression, we assume
$\text{var}[\epsilon] = \mathbf{I}\gamma$, where $\mathbf{I}$ is the identity matrix and $\gamma$ is a constant. By varying $\text{var}[\epsilon]$ appropriately, we can usually develop a better set of assumptions to suit a particu-
lar set of comparative data and evolutionary question. In essence, by applying
different $\text{var}[\epsilon]$ matrices, we create a new phylogenetic comparative method for
each situation.

**Within-Species Variation and Measurement Error**

If measures of several individuals within each taxon are available, these can
be used to calculate a standard error about the estimate of the mean trait phen-
type for that taxon. An overall description of this source of error for that trait
across taxa (e.g., $\text{var}[\epsilon_{M}]$) can be formed by creating a matrix with the square
of these standard errors for each taxon as the diagonal elements and with off-
diagonal elements equal to 0. If estimates of the standard error are available for
several traits, these matrices can be added together to obtain the variances for
all the error terms (e.g., $\text{var}[\epsilon_{M}] = \text{var}[\epsilon_{MY}] + \beta^2\text{var}[\epsilon_{MX}] + \ldots$). If estimates
of the standard error are not available for any particular trait, then we may
estimate them from the comparative data themselves (using, e.g., one more nuis-
ance parameter in the procedure as discussed below) or assume that this source
of error is negligible such that $\text{var}[\epsilon_{M}]$ for that trait equals 0 and need not be
included in the model.

**Error Due to the Evolutionary Process**

The variance matrix ($\text{var}[\epsilon_{S}]$) describes the expected similarity between taxon
phenotypes due to their shared evolutionary histories. We (Hansen and Martins
1996) developed a procedure for computing this matrix for a given phylogeny and
model of evolutionary change and showed that the covariance between two taxon
phenotypes ($Y_i$ and $Y_j$) equals the covariance between the expected phenotype
of each taxon given the state of their most recent common ancestor ($Y_a$):

$$\text{cov}[Y_i, Y_j] = \text{cov}[E[Y_i|Y_a], E[Y_j|Y_a]]. \quad (5)$$
Given a particular model of evolutionary change for each character, we can calculate the elements of \( \text{var}[e_3] \) using equation (5) as illustrated elsewhere (Hansen and Martins 1996) for several different models of evolution, including directional and stabilizing selection, random genetic drift, and environmental fluctuations.

Unfortunately, in many regression situations, the variance matrices may be quite complicated and intractable. One practical alternative is to employ a simple form of the variance matrix that captures the essential properties of the mode of evolutionary change. Although in doing so much of the explicit evolutionary interpretation is lost, the results provide a useful heuristic and still allow flexibility in the choice of assumptions. For example, the Brownian motion model used in Felsenstein (1985) results in a variance matrix with the following elements:

\[
\text{var}[e_s]_{ij} = \gamma t_{ij},
\]

where \( t_{ij} \) is the distance on the phylogeny between the root and the most recent common ancestor of taxa \( i \) and \( j \) and \( \gamma \) is a constant. This variance matrix is also compatible with other models used to describe phenotypic evolution involving mutation and random genetic drift, fluctuating directional selection, and even certain types of punctuated or burstlike change (Hansen and Martins 1996). The \( \gamma \) parameter is related to the magnitude (variance) of change, and its exact interpretation depends on the model's details. For all models, use of equation (6a) to describe the evolution of a character implies that the magnitude of future evolutionary changes does not depend on the current state of the phenotype.

If phenotypic evolution is thought to have been constrained such that taxa with extreme phenotypic values are more likely to evolve toward less extreme values, a variance matrix with the following elements may be a better choice:

\[
\text{var}[e_s]_{ij} = \gamma \exp[-\alpha t_{ij}],
\]

where \( t_{ij} \) is the phylogenetic distance between taxa \( i \) and \( j \). Equation (6b) may result from evolutionary models that include stabilizing selection or other constraining forces (e.g., Lande 1976; Felsenstein 1988; Hansen 1996; and others discussed in Hansen and Martins 1996). In general, \( \alpha \) can be interpreted as the magnitude of the restraining force or pull toward a central state. The \( \gamma \) parameter can be loosely interpreted as the interspecific variance that would be attained at equilibrium due to the balance between the restraining forces and perturbing forces such as genetic drift and changes in the environmental, selective, or genetic background. Use of an exponential variance structure to describe a component of the error variance implies that the magnitude of future evolutionary changes in the relevant characters depends on the distance of the current states from the combined optimum. Other transformations of phylogenetic distance (e.g., power transformations) have also been suggested (e.g., Grafen 1989; Gittleman and Kot 1990) but have not yet been developed in a microevolutionary framework.

Most microevolutionary models resulting in the variance structures in equations (6a) and (6b) assume that the error variance could be modeled independently of the mean. In many situations, this assumption is not acceptable, either because
of known functional or scaling relationships or because of an observed empirical relationship. For example, the error variance of dichotomous or binary variables (e.g., in a logistic regression) cannot be modeled as above. A binomial relationship between the mean (as expressed in the logit link in eq. [4d]) and the variance must be incorporated. In this case, we might prefer

$$\text{var}[\epsilon_{ij}] = \mu_a(1 - \mu_a) \exp[-\alpha t_{ij}], \quad (6c)$$

where $\mu_a$ is the predicted value of the most recent common ancestor of $i$ and $j$, ideally found by estimating the state of the predictor variables ($X_a$) in the ancestor but possibly estimated as the average of the predictor variables for taxa $i$ and $j$ in an iterative procedure. (This equation can be formally derived from a two-state Markovian model using eq. [5]). As another example, the variance of a size-related character is often expected to scale with the square of the mean. This can be incorporated by multiplying the right side of equation (6a) or (6b) by $\mu^2$. Again, in this case, predicted means of the most recent common ancestors would be most appropriate for the covariance terms.

Finally, we note that the phylogenetic distance ($t_a$ or $t_b$) is open to interpretation. For example, if one assumes that evolutionary change is mostly concentrated around speciation events, it may be more appropriate to use the number of speciation events separating the species as a measure of phylogenetic distance in the above equations. Of course, use of this interpretation requires knowing the actual (or at least relative) number of speciation events occurring along every branch of a phylogeny, including those leading to extinct taxa. In most cases, these values will be impossible or at least prohibitively difficult to obtain directly, and, to our knowledge, the theory required to estimate this information has not yet been developed.

**Error Due to Unknown or Uncertain Phylogenetic Information**

Whether the phylogenies are obtained from the literature or inferred particularly for the comparative study, they will almost always be subject to some uncertainty, and special care should be taken to understand the phylogeny and the possible ways in which it might be wrong. Consider creating several extreme phylogenies (e.g., Martins 1994) or using computer programs to create random samples from the set of all possible phylogenies (e.g., Losos 1994; Martins 1996). Branch lengths will need to be known or assumed in most applications of phylogenetic comparative statistics. Thus, it is also important to determine whether branch lengths are available for the various phylogenies, and the units in which those branch lengths have been measured.

The variance due to uncertain phylogeny ($\text{var}[\epsilon_p]$) is a measure of the effect of incorrect phylogenetic information, including both the topology and branch lengths (which are required to estimate $\text{var}[\epsilon_s]$ under most models of microevolutionary change described earlier). In the unlikely situation that the phylogeny is known, error due to an uncertain phylogeny can be safely ignored. If the phylogeny has been estimated from morphological or molecular information, estimation error can and should be included in the analysis (e.g., see Felsenstein 1985 for
suggestions). If there are a few possible phylogenies or if a random sample from the set of all possible phylogenies can be generated, the statistical model chosen above can be fit to the data using each of the possible phylogenies separately and the variance of the resulting parameter estimates used as an estimate of \( \text{var}[\epsilon_p] \) (Martins 1996).

**Combining Variance Structures**

If measurement error is independent of the error due to stochastic evolution along a phylogeny, the two variances describing such error can be added together to form a compound error term \( \text{var}[\epsilon_{M+S}] = \text{var}[\epsilon_M] + \text{var}[\epsilon_S] \). Since \( \text{var}[\epsilon_S] \) depends on a known phylogeny, \( \text{var}[\epsilon_p] \) cannot usually be added to the other two. It has been illustrated (Martins 1996) how \( \text{var}[\epsilon_p] \) can be combined with the other components of the error variance to obtain an overall confidence interval for a regression parameter that takes phylogenetic uncertainty into account. This is a nested procedure in which the regression model must be fit several times with \( \epsilon_p \) set to 0 and the results combined to incorporate \( \text{var}[\epsilon_p] \) into the final confidence intervals.

In many cases, it will be difficult to incorporate all of the error terms into the analysis, either because the necessary information is not available or because the variances of the error terms are forbiddingly complex. In these situations, we can still conduct the comparative analysis, simply by assuming that some of the error terms are negligible. Such assumptions should be kept firmly in mind when interpreting the results.

**ESTIMATING PARAMETERS**

As usual, the first step in estimating parameters or conducting hypothesis tests in any statistical analysis is to conduct exploratory analyses. Look at the data, plot them out in various ways, and try to determine whether any interesting patterns emerge. Conduct preliminary analyses using standard regression techniques (nonphylogenetic). Add taxonomic or phylogenetic labels to the points, and look for interesting groupings of data with evolutionary history, or any taxa that are particularly influential. Note that any hypothesis tests conducted during exploratory analyses should be taken into account in determining probabilities in later hypothesis tests.

There are many ways in which the parameters of equation (4) can be estimated statistically (McCullagh and Nelder 1989; Dobson 1990). We focus on generalized least squares (GLS) estimation, because GLS requires only knowledge of the variance of the error terms and may be more robust toward misspecification of the model. The choice of estimation method, however, should depend on the available information and the preferences of the individual researcher. Note that GLS estimates are identical to maximum-likelihood estimates calculated assuming normally distributed error terms and that a GLS estimate is the best quadratic approximation to a maximum-likelihood estimate when only the variance of the distribution is known.

The basic formulas for estimating \( \beta \) and its standard error in most of the above
models are the usual ones used in GLS regression:

\[ \mathbf{b} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \]

and

\[ \text{var}[\mathbf{b}] \approx (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}, \]

where \( \mathbf{b} \) is an estimate of \( \beta \), \( \mathbf{V} \) is the matrix \( \text{var}[\mathbf{e}] \) as defined earlier, a prime denotes transposition, and the latter equation is only asymptotically true. When there is a link function, an iterative procedure can be used instead of the GLS equations above (McCullagh and Nelder 1989). We begin by obtaining starting values of the \( \mathbf{b} \)'s, using the raw data and simple, nonphylogenetic regression procedures. These starting values (\( \mathbf{b}_0 \)) can then be used to calculate an initial value of the mean, \( \mathbf{M}_0 = g^{-1}(\mathbf{X}\mathbf{b}_0) \) and the matrix of derivatives of that inverse link function (\( \mathbf{D}_0 \)) with respect to \( \mathbf{b} \) and evaluated at \( \mathbf{b}_0 \). A second estimate (\( \mathbf{b}_1 \)) is then obtained using

\[ \mathbf{b}_1 = \mathbf{b}_0 + (\mathbf{D}_0'\mathbf{V}_0^{-1}\mathbf{D}_0)^{-1}\mathbf{D}_0'\mathbf{V}_0^{-1}(\mathbf{Y} - \mathbf{M}_0) \]

and

\[ \text{var}[\mathbf{b}_1] \approx (\mathbf{D}_0'\mathbf{V}_0^{-1}\mathbf{D}_0)^{-1}. \]

The overall fit of the model can be monitored using the residual sum of squares (RSS):

\[ \text{RSS} = (\mathbf{Y} - \mathbf{M})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{M}), \]

where \( \mathbf{M} \) is the predicted value of \( \mathbf{Y} \) (i.e., an estimate of \( \mu \)).

If this procedure is repeated several times, replacing \( \mathbf{b}_0 \) by \( \mathbf{b}_1 \) with each iteration, the RSS, \( \text{var}[\mathbf{b}] \), and \( \mathbf{b} \) should all converge. As usual, the variances of \( \mathbf{b} \) can be used to generate confidence intervals about estimates of \( \beta \). Coefficients of determination may also be useful in model fitting, and choices between full and reduced models can be made by conducting log-likelihood ratio tests. These procedures can be conducted using any advanced statistical program (e.g., SAS). For some specific models, GLS analyses of comparative data are implemented in COMPARE (Martins 1995a).

The above results assume that the \( \mathbf{V} \) matrix is known. In most cases, however, the elements of \( \text{var}[\mathbf{e}] \) will be functions of one or more parameters that are unknown (e.g., \( \gamma \) in eq. [6]) or even the parameters we are trying to estimate (e.g., \( \beta \)). Equations (7) and (8) provide the best linear unbiased estimates of \( \beta \), and they will be unbiased regardless of the accuracy of the \( \mathbf{V} \) matrix. Estimates of the variance of \( \mathbf{b} \), however, may be very biased and inaccurate if the \( \mathbf{V} \) matrix is specified incorrectly. When \( \mathbf{V} \) is a function of the parameters being estimated, a second iterative procedure can be developed, again using the RSS to monitor convergence (McCullagh and Nelder 1989). Other unknown or “nuisance” parameters in the \( \mathbf{V} \) matrix can be estimated using maximum-likelihood procedures or by minimizing the RSS. (For examples of estimation of parameters in the variance matrix in a comparative method context, see Grafen 1989; Gittleman and Kot 1990; Martins 1994; Hansen 1996).
The above procedure is useful for estimating the parameters of the models in equations (1) and (2). In inferring ancestral states (eq. [3]), however, the aim is to make predictions using the chosen model. The best linear unbiased predictor can be obtained by estimating the ancestral state as a deviation from the grand mean of the clade \((M_G)\) and using weights:

\[
W = \text{var}[A, Y]\text{var}[Y]^{-1},
\]

where the elements of \(\text{var}[A, Y]\) describe the expected similarity (covariance) between the phenotype of the ancestor \((A)\) and that of each extant taxon \((Y)\), and \(\text{var}[Y]^{-1}\) is the inverse of a variance matrix of all pairs of extant taxa. The grand mean can be obtained using generalized least squares (GLS) regression procedures:

\[
M_G = (J'\text{var}[Y]^{-1}J)^{-1}J'\text{var}[Y]^{-1}Y,
\]

where \(J\) is a column of 1's (such that \(J'\text{var}[Y]^{-1}J\) equals the sum of the elements of \(\text{var}[Y]^{-1}\); fig. 4). The variance of ancestral estimates derived using equations (3) and (7) is

\[
\text{var}[\hat{A}] = \text{var}[A] - \text{var}[A, Y]^{'}\text{var}[Y]^{-1}\text{var}[Y, A],
\]

where \(\text{var}[\hat{A}]\) is a variance matrix describing the expected similarity (covariance) among ancestral states given the phylogeny and model of evolutionary change. All of the above variance matrices can be obtained by specifying a model of the evolutionary process and combining sources of error as described above in determining \(\text{var}[\epsilon]\).

If the above variance matrices are computed using the assumption that evolution occurs as if by Brownian motion, the ancestral states predicted using the above algorithm are identical to estimates obtained using the sum-of-squared-changes parsimony algorithm as described by Maddison (1991). Thus, squared-changes parsimony also provides the best linear unbiased prediction of ancestral states of a character evolving under Brownian motion. The above equations provide an extension of Maddison’s (1991) approach that allow for variation in the microevolutionary assumptions, simply by varying the form of the variance matrices. They also provide a direct way of computing estimates of ancestral states (as opposed to the recursive algorithm provided in Maddison 1991) and estimates of the standard errors used to generate confidence intervals about estimated ancestral states.

As with any statistical analysis, it is always important to examine residual plots and measures of the fit of the model to determine whether the assumptions of the analysis correspond well to the data. Standardized residuals \((\text{var}[Y]^{-1}[Y - M])\) can be obtained for the above procedure and used in various residual plots, as usual. In comparative analyses, it may also be important to consider various aspects of the model independently. For example, to consider the impact of phylogenetic information, estimates can be obtained on each of several possible phylogenies in a type of sensitivity analysis. Comparisons among the results will show whether and how the final results depend on the available phylogenetic information. Similarly, it may be useful to vary the underlying microevolutionary
model in various ways and to determine whether the results hold true under more than one type of microevolutionary assumption.

DISCUSSION

Recent criticisms of the comparative method have focused on various issues regarding whether comparative studies can and should be used to infer adaptation (e.g., Frumhoff and Reeve 1994; Leroi et al. 1994). Some of the criticisms made by these authors are general problems in scientific inference such as distinguishing causation from correlation and verifying hypotheses. Our opinion is that these problems are not more severe in comparative studies than in most other studies in evolutionary biology (Hansen 1996). Other issues are not critical problems with the comparative approach in general but rather problems with specific comparative methods and interpretations that have been made in the literature from the results of such analyses. Comparative analyses do share a number of potential problems, but these can be solved in various ways by incorporating phylogenetic information into a statistical framework and by specifying the underlying microevolutionary assumptions explicitly. The approach described here suggests how these problems can be addressed by designing the phylogenetic comparative analysis that is most appropriate for a particular set of data and evolutionary question.

The most common problem pointed out in the discussion of the comparative method is that the residuals or error terms (e) of the statistical model (e.g., eqq. [1]–[3]) are not independent of one another as required by most standard hypothesis tests and estimation procedures. This problem has been discussed at length in the evolutionary literature and is addressed by most modern phylogenetic comparative methods (see, e.g., Felsenstein 1985; Harvey and Pagel 1991; Martins and Hansen 1996b). A number of additional statistical problems, although common in the statistical literature, have not been discussed much in the context of comparative studies. For example, in most comparative studies both response (Y) and predictor (X) variables are subject to error. As discussed here, if the X's are trait values measured in extant species, they may be subject to measurement error, error due to stochastic evolution, and error due to inaccurate specification of the phylogeny. All three of these sources of error should be taken into account in a complete phylogenetic analysis. The result is an issue similar to that underlying discussion of alternative or "Model II" regression methods such as major axis or geometric mean regression (e.g., Sokal and Rohlf 1981). In a phylogenetic analysis, careful consideration of each variable in the regression model and specification of the correct, compound error term for each model (as discussed earlier) eliminates the need to conduct geometric mean instead of standard regression and provides an acceptable solution to this potential problem.

There are also a number of statistical problems that can be solved by using the GLM procedures discussed here. First, the relationship between response and predictor variables in comparative studies is not always linear. Although most recent comparative studies have examined linear relationships between characters or between characters and the environments in which they are exhibited,
nonadditive relationships arise frequently in various situations. For example, many biomechanical models predict allometric relationships between characters that are not well described by the linear model in equation (4a). Nonlinear models are also needed to consider categorical ("state") or discreet ("count") characters or characters with a limited range (e.g., body sizes restricted to positive values) in regression analyses without making evolutionary interpretations exceedingly complex. Finally, certain microevolutionary models of phenotypic change may predict nonlinear relationships. For example, as mentioned earlier, in a previous method (Martins 1994) for estimating the rate of change under a model of phenotypic evolution under stabilizing selection and random genetic drift, the relationship between phenotypic divergence and time is nonlinear and requires special treatment. The link functions included in our general approach allow for incorporation of explicitly nonlinear relationships in the statistical analysis of comparative data and should be explored more thoroughly along these lines.

Second, the residuals or error terms (e) may not be normally distributed, as assumed by most statistical procedures and by several phylogenetic comparative methods (e.g., Felsenstein 1985). One common way of dealing with this practical problem is to use logarithmic or power transformations of the X and Y variables (e.g., Garland et al. 1992). Unfortunately, such power transformations lead to a modification of the model of phenotypic evolution underlying the characters. For example, the Brownian motion model underlying Felsenstein's (1985) independent contrasts method is a reasonable model for evolution of a continuous trait under random genetic drift or evolution under fluctuating directional selection. Under such a model, we expect a character to be normally distributed after long-term evolution along a phylogeny. If the data show that the character is not normally distributed, then we have evidence that the character has not evolved under Brownian motion and that it violates one of the basic assumptions of Felsenstein's (1985) method. If we simply do a log transformation of the Y variables before proceeding with the analysis, we have created a nonlinear relationship between X and Y, and the error terms should be altered accordingly in making any specific interpretations about the evolutionary relationship between these variables (see Martins 1995b; Martins and Hansen 1996a for more discussion of this point). Again, application of GLM statistics as described in the current study may provide a more reasonable solution to this problem.

Third, the variance of the response variable (Y) may be related to its mean, for both biological and statistical causes. Biologically, the variance often increases with the mean because of scaling effects; elephants have larger variance in size than mice simply because they are larger. The mean and the variance of the response variable may also be related to one another if the predictor variables (X) are stochastically evolving phenotypic traits. For example, as illustrated here for measurement error, if the variance of the error term (\text{var}[e]) includes terms for both predictor and response variables, it will also include regression parameters (\beta). The mean and variance may also be linked for purely statistical reasons. Most nonnormal distributions (e.g., the binomial distribution underlying the logistic regression above) involve linked means and variances. Such problems are usually resolved by conducting logarithmic or other transformations of the re-
sponse variable, which may be reasonable but, as mentioned earlier, require that
the underlying microevolutionary model also be developed and interpreted on the
transformed scale. Alternatively, the mean can be incorporated explicitly into the
variance matrix \( \text{var}[\epsilon] \) of a GLM estimation procedure (eq. \([6c]\)).

In addition to these potential statistical problems, there are also important
conceptual issues that arise from the interpretation of regression parameters and
the choice of predictor variables in comparative analyses. For example, as illus-
trated earlier, the general approach presented in the current article shows how
there are at least two different ways of incorporating phylogenetic information
into comparative analyses. The first is to incorporate the phylogeny into the
structure of the error term, as proposed in Grafen's (1989) regression, which
uses GLS to estimate the relationship among traits. An alternative, illustrated by
Cheverud et al.'s (1985) method, is to incorporate phylogenetic information into
the mean structure of the model. This difference in approach can be critical,
as it may change the interpretation of the evolutionary parameter that is being
estimated.

Another important issue arises in the study of adaptation. Adaptive hypotheses
are often studied by regressing trait values on an environmental factor measured
in the same taxa. Although most modern phylogenetic comparative methods (e.g.,
Felsenstein 1985) may give reasonable estimates of the relationship between phe-
notype and environment, they were not designed to do so, and interpretations of
adaptation from the results of such analyses may be questionable (Lauder et al.
1993; Frumhoff and Reeve 1994; Leroi et al. 1994). A better approach is to de-
velop methods that explicitly incorporate selective forces into the underlying
microevolutionary models and test specifically for adaptive relationships. For
example, a model in which a continuous character \( (Y) \) evolves under stabilizing
selection in a changing environment has been proposed (Hansen 1996). It uses
measures of the species phenotype \( (Y) \) and historical information about the envi-
ronment to estimate the effect of the environment on the optimum phenotype.
The model explicitly describes adaptive evolution along a phylogeny and can be
used to infer various aspects of adaptation from comparative data. Other models
need to be developed for alternative views and definitions of adaptation.

In summary, comparative studies provide a useful tool for studying the evolu-
tion of characters that do not appear in the fossil record and/or that are not
amenable to experimental manipulation. They provide a window onto broad-scale
evolutionary patterns that cannot be seen in any other way. Nevertheless, as
with any statistical procedure, comparative analyses should be conducted cau-
tiously, with explicit descriptions of the assumptions going into any conclusions.
Nonparametric procedures usually require fewer assumptions but also usually
require a focus on hypothesis testing rather than parameter estimation. We be-
lieve parameter estimation is the more important and more interesting question.
Only in parameter estimation can we make powerful use of comparative method
in inferring aspects of the evolutionary process from comparative data. The ap-
proach outlined in this article provides a framework for conducting phylogenetic
comparative analyses and for determining the underlying assumptions of any
proposed comparative method. By examining existing and new methods in this
framework, we can better understand the similarities and differences among them and make informed decisions regarding the choice of statistical method. By considering the approach itself in greater detail, we discover a number of new methods that will allow for the analysis of most sets of comparative data with assumptions appropriate to the characters and organisms of interest.

ACKNOWLEDGMENTS

We would like to thank J. Kingsolver, J. Thorne, T. Titus, B. Walsh, P. Zani, and an anonymous reviewer for helpful discussions and/or comments on the manuscript. This work was supported by a grant from the National Science Foundation to E.P.M. (DEB-9406964).

LITERATURE CITED


*Associate Editor: J. Bruce Walsh*