



Methods for the Analysis of Comparative Data in Evolutionary Biology

Michael Lynch

Evolution, Vol. 45, No. 5. (Aug., 1991), pp. 1065-1080.

Stable URL:

<http://links.jstor.org/sici?sici=0014-3820%28199108%2945%3A5%3C1065%3AMFTAOC%3E2.0.CO%3B2-U>

Evolution is currently published by Society for the Study of Evolution.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ssevol.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

EVOLUTION

INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

PUBLISHED BY

THE SOCIETY FOR THE STUDY OF EVOLUTION

Vol. 45

August, 1991

No. 5

Evolution, 45(5), 1991, pp. 1065–1080

METHODS FOR THE ANALYSIS OF COMPARATIVE DATA IN EVOLUTIONARY BIOLOGY

MICHAEL LYNCH

Department of Biology, University of Oregon, Eugene, OR 97403 USA

Abstract.—Inferences regarding phylogenetic patterns and constraints on the evolution of characters often can be derived only from comparisons of extant species. If the phylogeny of these species is known, then the mean phenotypes of taxa can be partitioned into heritable phylogenetic effects and nonheritable residual components. Methods are presented for the estimation of phylogeny-wide means of characters, the variance-covariance structure of the components of taxon-specific means, and the mean phenotypes of ancestral taxa. These methods, which are largely an extension of maximum-likelihood techniques used in quantitative genetics, make an efficient use of the data, are unbiased by phylogenetically uninformative contributions to mean phenotypes, and take into account fully the nonindependence of data resulting from evolutionary relationships. Statistical tests are introduced for evaluating the significance of phylogenetic heritability and of correlations between traits, and expressions are given for the standard errors of ancestral mean phenotype estimates. It is argued that the covariance structure of phylogenetic effects provides a description of a macroevolutionary pattern, whereas that for the residual effects, when corrected for sampling error, is more closely related to a microevolutionary pattern.

Key words.—Comparative analysis, evolutionary constraints, phylogenetic analysis.

Received December 4, 1989. Accepted January 25, 1991.

Due to the inadequacies of the fossil record, inferences about patterns of evolution are often derived from the study of extant species. Behavioral, physiological, life-historical, and soft-tissue attributes, which do not fossilize, can be studied only in this way. Methods for measuring the forces of selection operating within populations (Lande and Arnold, 1983) and for evaluating the genetic properties (components of genetic variance and covariance) that modify the response to selection (Falconer, 1981) are well established. Although their implementation usually requires large sample sizes and the fulfillment of several assumptions (Mitchell-Olds and Rutledge, 1986; Mitchell-Olds and Shaw, 1987), with suitable study organisms and adequate resources, these difficulties can be overcome. Thus, an understanding of the microevolutionary process is within reach. However, it remains to be

seen how useful quantitative genetic analyses within populations will be in interpreting the differences that have evolved among species.

Macroevolutionary patterns (here interpreted broadly to include any patterns that transcend species boundaries) are usually identified by the comparative study of species mean phenotypes. An especially popular approach is to infer constraints on the evolutionary process from the study of bivariate distributions of characters from different species. Such a treatment can result in biased estimates of interspecific patterns when the phylogenetic relationships of taxa (nonindependence of data) are ignored (Felsenstein, 1985). This “degrees of freedom” issue recently has been the focus of considerable attention (Ridley, 1983; Stearns, 1983; Cheverud et al., 1985; Felsenstein, 1985; Harvey and Clutton-Brock, 1985;

Huey, 1987; Bell, 1989; Pagel and Harvey, 1988; Martins and Garland, 1991). A successful resolution of the problem is necessary to insure that emergent patterns from comparative analyses reflect true constraints on the evolutionary process. Such constraints may arise from factors internal or external to the taxa of interest, i.e., from pleiotropy, or from correlated selection and/or random genetic drift.

Although the newly proposed methodologies have several attractive features, they also have some shortcomings. First, several of the methods do not fully account for phylogenetic relationships. Nested analysis of variance (Bell, 1989), for example, ignores the phylogenetic relationships within different taxonomic levels and ignores the possibility that the divergence times between pairs of taxa (species) may differ substantially among higher-order groups (genera). Second, as emphasized by Cheverud et al. (1985), there is some uncertainty as to the biological interpretation of correlations computed by most comparative methods. Generally, the observed mean phenotypes of all taxa are assumed to be fully evolutionarily informative. This ignores the problem of sampling error, the possible confounding of environmental and genetic effects, and the occurrence of taxon-specific evolutionary events that are aberrant with respect to the remainder of the phylogeny. Finally, it is not clear that any of the methods utilize the data in a fully efficient manner.

The method proposed herein attempts to deal simultaneously with all of these issues, while also providing a formal basis for hypothesis testing. At the outset, credit should be given to the applied quantitative geneticists, in particular the late C. R. Henderson, who for rather different reasons developed much of statistical theory to be discussed.

MIXED MODEL METHODOLOGY

We start with the assumption that estimates of phenotypic means are available for characters $c = 1, \dots, k$ in taxa $i = 1, \dots, n$, the phylogeny of which has been established from independent data. Such a data set can be used to address a number of issues. What is the average rate of divergence of mean phenotypes within the phylogeny,

and does this differ among various characters and/or among different phylogenetic lineages? Are there significant correlations between characters? To what extent are such correlations due to phylogeny-wide effects jointly inherited by related species, and to what extent are they caused by taxon-specific properties? What are the most likely mean phenotypes of the unobserved ancestral species?

Each observed mean can be written as a linear function

$$\bar{z}_{ci} = \mu_c + a_{ci} + e_{ci} \quad (1)$$

where μ_c is the grand mean of the c th character over the entire phylogeny, a_{ci} is the heritable additive value (the phylogenetic effect of Cheverud et al., 1985) of the character in the i th taxon (measured as a deviation from the grand mean), and e_{ci} is the residual deviation (the specific effect of Cheverud et al., 1985) from the predicted value $\mu_c + a_{ci}$ caused by nonadditivity of genetic effects, by environmental effects, and by sampling error. All members of the phylogeny share μ_c . The additive values are analogous to breeding values employed in quantitative genetics (Falconer, 1981). Thus, the quantity $\mu_c + a_{ci}$ can be interpreted as the phylogenetically heritable component of the mean phenotype of the i th taxon. Stated in another way, it is the expected mean phenotype of a taxon descendent from species i . Although Cheverud et al. (1985) follow a similar logic in partitioning observed mean phenotypes, their methodology and interpretation of parameters are rather different from that developed below.

The entire set of $k \times n$ equations can be written more compactly in matrix form. For example, if $k = 2$, and \bar{z}_c , \mathbf{a}_c , and \mathbf{e}_c are $n \times 1$ column vectors for the c th character,

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{pmatrix} \quad (2)$$

where $\mathbf{1}$ and $\mathbf{0}$ in the first matrix on the right denote $n \times 1$ column vectors of ones and zeros respectively. Generalizing to an arbitrary number of characters while maintaining the structure of Equation (2),

$$\bar{\mathbf{z}} = \mathbf{X}\boldsymbol{\mu} + \mathbf{a} + \mathbf{e} \quad (3)$$

where \mathbf{X} is an $nk \times k$ incidence matrix.

The variance-covariance structure of the nk additive effects is a function of a matrix of phylogenetic relationships \mathbf{G} . The diagonal elements of this $n \times n$ matrix are all equal to one, while the off-diagonal elements are in the range $0 \leq G_{ij} < 1$, a zero implying that species i and j share no evolutionary history within the phylogeny under consideration (complete independence). For a bivariate analysis, the variance-covariance matrix of additive effects is

$$\text{Var} \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{G}\sigma_a^2(1) & \mathbf{G}\sigma_a(1, 2) \\ \mathbf{G}\sigma_a(2, 1) & \mathbf{G}\sigma_a^2(2) \end{pmatrix} \quad (4)$$

where $\sigma_a^2(1)$ and $\sigma_a^2(2)$ are the variances and $\sigma_a(1, 2) = \sigma_a(2, 1)$ the covariances of additive values among independent branch tips in the phylogeny. Note that the right side of Equation (4) is actually a $2n \times 2n$ matrix ($kn \times kn$ in the general case). Letting \mathbf{A} be the matrix of additive variance and covariance terms, the above expression can be generalized to an arbitrary number of characters,

$$\text{Var}(\mathbf{a}) = \mathbf{D} = \mathbf{A} \otimes \mathbf{G} \quad (5)$$

where \otimes represents the Kronecker product function. This expression simply states that for pairs of taxa i and j with relatedness G_{ij} , the expected covariance of additive values of the trait c is $G_{ij}\sigma_a^2(c)$ whereas the covariance between traits c and d is $G_{ij}\sigma_a(c, d)$. An analogous situation arises in quantitative genetic analysis where $G_{ij} = 2\theta_{ij}$, with θ_{ij} being the coefficient of coancestry. A method for computing \mathbf{G} is presented below.

The variance-covariance structure of the residual deviations can be expressed in a similar way. Letting $\sigma_e^2(1)$, $\sigma_e^2(2)$, $\sigma_e(1, 2) = \sigma_e(2, 1)$ be the variances and covariances of residual deviations for the bivariate case,

$$\text{Var} \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I}\sigma_e^2(1) & \mathbf{I}\sigma_e(1, 2) \\ \mathbf{I}\sigma_e(2, 1) & \mathbf{I}\sigma_e^2(2) \end{pmatrix} \quad (6)$$

where \mathbf{I} is an $n \times n$ identity matrix. Note that with this structure, it is assumed that the residual deviations for different taxa are distributed independently. This assumption can be relaxed by substituting a multiplier matrix with nonzero off-diagonal elements for \mathbf{I} , but it is not obvious how the elements of such a matrix would be obtained, and for

the time being we will focus on the simplest situation. Letting \mathbf{E} be the variance-covariance matrix of residual deviations, Equation (6) can be generalized to an arbitrary number of characters,

$$\text{Var}(\mathbf{e}) = \mathbf{R} = \mathbf{E} \otimes \mathbf{I}. \quad (7)$$

It is assumed here that the elements of \mathbf{a} are uncorrelated with those of \mathbf{e} . (See Cheverud et al., 1985 for a discussion of this matter).

The least-squares solution for the grand means and additive values is given by the mixed-model equation,

$$\begin{pmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{R}^{-1} \\ \mathbf{R}^{-1}\mathbf{X} & \mathbf{R}^{-1} + \mathbf{D}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{\mathbf{a}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{R}^{-1}\bar{\mathbf{z}} \\ \mathbf{R}^{-1}\bar{\mathbf{z}} \end{pmatrix} \quad (8)$$

(Henderson et al., 1959; Henderson, 1963; Harville, 1977), where T denotes transposition, $^{-1}$ denotes inversion, and $\hat{}$ denotes an estimate. The matrix on the left, which is known as an information matrix, can be understood more easily by expanding the bivariate case,

$$\begin{pmatrix} ne^{11} & ne^{12} & \dots & \mathbf{1}^T e^{11} & \mathbf{1}^T e^{12} \\ ne^{21} & ne^{22} & \dots & \mathbf{1}^T e^{21} & \mathbf{1}^T e^{22} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{1}e^{11} & \mathbf{1}e^{12} & \dots & \mathbf{I}e^{11} + \mathbf{G}^{-1}a^{11} & \mathbf{I}e^{12} + \mathbf{G}^{-1}a^{12} \\ \mathbf{1}e^{21} & \mathbf{1}e^{22} & \dots & \mathbf{I}e^{21} + \mathbf{G}^{-1}a^{21} & \mathbf{I}e^{22} + \mathbf{G}^{-1}a^{22} \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix} \quad (9)$$

where e^{cd} and a^{cd} are the elements of the c th row and d th column of the inverse matrices \mathbf{E}^{-1} and \mathbf{A}^{-1} . In the general case, the upper left block of the information matrix is a $k \times k$ matrix equal to the inverse of \mathbf{E} times the number of taxa. The lower left block is a $k \times k$ matrix of column vectors, each consisting of n repetitions of the respective element of \mathbf{E}^{-1} . The upper right block is the transpose of the lower left block; it consists of expanded rows of \mathbf{E}^{-1} . The lower right block is a $k \times k$ matrix of matrices, each of these consisting of the sum of a diagonal matrix of the respective element of \mathbf{E}^{-1} plus the product of \mathbf{G}^{-1} and the respective element of \mathbf{A}^{-1} .

Expression (9) shows that the solution for the estimates of μ and \mathbf{a} does not actually require the inversion of \mathbf{D} or \mathbf{R} . Only the

smaller matrices **G**, **A**, and **E** need to be inverted. Nevertheless, there is still a practical difficulty with Equation (8). Solution for the means and additive values requires that the variance-covariance structures **A** and **E** be known in advance. This is never the case. The only recourse is to estimate the variances and covariances from the data themselves, but this requires estimates of the fixed effects and additive values. Therefore, we require an iterative algorithm.

The first step is to assign initial values to the elements of the variance-covariance matrices **A** and **E**. A simple starting point is to compute, without regard to the phylogenetic relationships, the variances of the elements of each vector $\bar{z}_c, \hat{\sigma}^2(c, 0)$. In this process the initial means $\hat{\mu}_c(0)$ are also computed. In the absence of any prior information, each of these unweighted estimates of the phenotypic variances and covariances is then partitioned arbitrarily into contributions to **A** and **E** (for example, by multiplying by a random number in the range of 0 to 1).

Next, the preliminary estimates of the additive values and residual deviations are computed. Returning to Equation (8), it can be seen that

$$\mathbf{R}^{-1}\mathbf{X}\hat{\mu} + (\mathbf{R}^{-1} + \mathbf{D}^{-1})\hat{\mathbf{a}} = \mathbf{R}^{-1}\bar{\mathbf{z}}. \quad (10)$$

Rearranging, the estimates of the additive effects are found to be

$$\hat{\mathbf{a}}(t) = (\mathbf{R}^{-1} + \mathbf{D}^{-1})^{-1}\mathbf{R}^{-1}[\bar{\mathbf{z}} - \mathbf{X}\hat{\mu}(t)], \quad (11)$$

where t denotes the iteration, and **D** and **R** are evaluated with the current elements of **A** and **E**. By definition, the residual deviations are estimated by

$$\hat{\mathbf{e}}(t) = \bar{\mathbf{z}} - \mathbf{X}\hat{\mu}(t) - \hat{\mathbf{a}}(t). \quad (12)$$

In all subsequent iterations, the variances ($c = d$) and covariances ($c \neq d$) of residual deviations are estimated by

$$\hat{\sigma}_e(c, d, t + 1) = \frac{\hat{\mathbf{e}}_c^T(t)\hat{\mathbf{e}}_d(t) + \text{tr}[\mathbf{C}^{cd}(t)]}{n}. \quad (13)$$

The first term in this expression is the mean of the cross-products (squares in the case of variance estimates) of the estimated residual effects in all species. The second term corrects for the fact that the first term pro-

vides a downwardly biased estimate of the true variance because the additive values have been estimated from the data. The trace (tr) of a matrix is the sum of the diagonal elements. The matrix $\mathbf{C}^{cd}(t)$ is the portion of the inverse of the information matrix $(\mathbf{R}^{-1} + \mathbf{D}^{-1})^{-1}$ that occupies the rows and columns that pertain to characters c and d . It contains the sampling variances and covariances of the relevant additive effects estimates.

The variances and covariances of additive values are estimated by

$$\hat{\sigma}_a(c, d, t + 1) = \frac{\hat{\mathbf{a}}_c^T(t)\mathbf{G}^{-1}\mathbf{a}_d(t) + \text{tr}[\mathbf{G}^{-1}\mathbf{C}^{cd}(t)]}{n}. \quad (14)$$

The first term in this expression is the mean of the cross-products (squares) of the additive values of characters c and d in all possible pairs of species (including self), each weighted by the appropriate element of the inverse of the relationship matrix. Note that if none of the individuals in the phylogeny share any common ancestry subsequent to the root of the tree, $\mathbf{G} = \mathbf{I}$, and the first term is simply the average cross-product (square) of the estimated additive values within taxa. The inverse of the relationship matrix accounts for the bias in variance component estimation resulting from nonindependence of data. The second term in Equation (14) accounts for the additional bias caused by the estimation of the grand means and additive effects.

The next estimates of the means are obtained with

$$\hat{\mu}_c(t + 1) = \frac{1^T[\bar{\mathbf{z}}_c - \hat{\mathbf{a}}_c(t)]}{n}, \quad (15)$$

which is also obtained directly from Equation (8). Equations (11) to (14) can then be solved anew. The entire process is iterated until the elements of $\hat{\mu}$, $\hat{\mathbf{a}}$, $\hat{\mathbf{A}}$, and $\hat{\mathbf{E}}$ have reached a satisfactory degree of stability. This iterative approach to obtaining parameter estimates is an EM (expectation-maximization) algorithm (Dempster et al., 1977; Thompson and Shaw, 1990). It is known that this procedure leads to the joint maximum in the likelihood surface for the means, additive values, and variance-covariance components when convergence oc-

curs. The additive values defined by the convergent solution to Equation (11) are analogous to the best linear unbiased predictors (BLUP) of breeding values commonly employed in animal husbandry (Henderson, 1984a, 1984b).

The EM algorithm usually leads to convergence, but problems do sometimes arise. With small numbers of taxa, convergence may fail to occur, and the possibility of multiple peaks in the likelihood surface cannot be ruled out. The latter problem can be evaluated by simply varying the starting conditions.

In summary, starting with a phylogenetic tree for n species, the mixed model in conjunction with the EM algorithm yields a substantial amount of useful information: k estimates of the phylogeny-wide character means, kn estimates of the additive values and residual deviations, and the $k \times k$ variance-covariance matrices for additive values and residual deviations. The elements of these matrices provide the basis for computing regression coefficients, as outlined below.

HYPOTHESIS TESTING

Relative to issues regarding parameter estimation, little attention has been given to hypothesis testing in the mixed-model literature. In animal science, where most applications of the mixed model have been made, the emphasis has usually been upon the acquisition of the best possible estimates of individual breeding values, with little regard to evaluating hypotheses concerning the distribution of data. The situation is quite different in evolutionary biology where the primary concern is usually with patterns in comparative analyses. The following test procedures rely on the assumption that the data have been scaled appropriately to yield an asymptotically multivariate normal distribution.

In the preceding section, the relationship between the proposed methodology and procedures in applied quantitative genetics was emphasized. One interesting analogy remains. If we take $\hat{\sigma}_T^2 = \hat{\sigma}_a^2 + \hat{\sigma}_e^2$ to be the estimated total variance among independent phenotypic means in a phylogeny, then the ratio

$$H_p^2 = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_T^2} \quad (16)$$

provides an estimate of phylogenetic heritability. H_p^2 estimates the degree to which related taxa provide phenotypic information about each other. Since $G_{ij}H_p^2$ is the expected correlation between mean phenotypes of taxa separated by t_{hi} time units (where h is the ancestor of i and j), differences in H_p^2 among characters within a phylogeny may provide useful information on the degree to which different types of traits are constrained phylogenetically. A low value of H_p^2 implies a high degree of randomness in the expression of a character throughout a phylogeny, i.e., a low degree of resemblance between related species (or low phylogenetic inertia). This is not to say that H_p^2 is a measure of the degree of genetic determinism, because the residual deviations can be caused largely by nonadditive genetic factors.

Assuming the sampling errors of σ_a^2 and σ_e^2 are uncorrelated, an approximate standard error of the phylogenetic heritability is given by

$$SE(H_p^2) = H_p^2 \left(\frac{\hat{\sigma}^2(\hat{\sigma}_a^2)}{\hat{\sigma}_a^4} + \frac{\hat{\sigma}^2(\hat{\sigma}_a^2) + \hat{\sigma}^2(\hat{\sigma}_e^2)}{(\hat{\sigma}_a^2 + \hat{\sigma}_e^2)^2} - \frac{2\hat{\sigma}^2(\hat{\sigma}_a^2)}{\hat{\sigma}_a^2(\hat{\sigma}_a^2 + \hat{\sigma}_e^2)} \right)^{1/2}$$

$\hat{\sigma}^2(\hat{\sigma}_a^2)$ and $\hat{\sigma}^2(\hat{\sigma}_e^2)$, which are the sampling variances of $\hat{\sigma}_a^2$ and $\hat{\sigma}_e^2$, can be estimated by a procedure outlined below. In the case of a univariate analysis, and under the assumption of multivariate normality, Equation (17a) reduces to

$$SE(H_p^2) = \frac{2H_p^2(1 - H_p^2)}{\sqrt{n}} \quad (17b)$$

Equations (17a) and (17b) are "large sample variance" estimators, i.e., their accuracy increases asymptotically with the sample size. Thus, these expressions provide some information about the degree of confidence one can have in a specific estimate of H_p^2 , but it is difficult to be very explicit on that matter without knowledge of the form of the sampling distributions of H_p^2 and $SE(H_p^2)$. However, a likelihood ratio test for

goodness-of-fit (Edwards, 1972), which allows a test of the hypothesis that $H^2_{\hat{p}}$ is significantly greater than zero, can be constructed.

Based on the parameter estimates $\hat{\mu}$, $\hat{\sigma}_a^2$, and $\hat{\sigma}_e^2$ (focusing for the time being on a univariate analysis), the probability density for the entire set of observed taxon-specific means is

$$p = (2\pi)^{-n/2} |\hat{\mathbf{V}}|^{-1/2} \exp \cdot [-1/2(\bar{\mathbf{z}} - \mathbf{X}\hat{\boldsymbol{\mu}})^T \hat{\mathbf{V}}^{-1}(\bar{\mathbf{z}} - \mathbf{X}\hat{\boldsymbol{\mu}})] \quad (18)$$

where $\hat{\mathbf{V}} = \hat{\mathbf{D}} + \hat{\mathbf{R}}$ and $|\cdot|$ denotes a determinant. A more restrictive form of the univariate model can also be evaluated under the assumption that $\sigma_a^2 = 0$. Since this is equivalent to the situation when the data are completely independent, the mean and variance can be estimated by the usual methods for these moments, $\hat{\mu}_0 = \Sigma \bar{z}_i / n$ and $\hat{\sigma}_0^2 = \Sigma (\bar{z}_i - \hat{\mu}_0)^2 / (n - 1)$. The probability density under the null hypothesis of zero phylogenetic heritability is then

$$p_0 = (2\pi\hat{\sigma}_0^2)^{-n/2} \exp\left(-\frac{\Sigma(\bar{z}_i - \hat{\mu}_0)^2}{2\hat{\sigma}_0^2}\right) \\ = (2\pi\hat{\sigma}_0^2)^{-n/2} \exp\left(\frac{1 - n}{2}\right). \quad (19)$$

The likelihood ratio test statistic

$$\Lambda = 2(\ln p - \ln p_0) \quad (20)$$

is asymptotically χ^2 distributed with one degree of freedom. Thus, $\Lambda > 3.84$ or 6.64 implies that the null hypothesis of no phylogenetic heritability can be rejected at the 5% or 1% level respectively.

The above procedure can be generalized to test for the significance of any variance or covariance component (or set of such components) in the model. The likelihood ratio test statistic is always equal to twice the difference between the log likelihoods under the more parameter rich and less parameter rich models, and this statistic is asymptotically χ^2 distributed with degrees of freedom equal to the difference in the number of parameters estimated under the two models (Edwards, 1972; Kendall and Stuart, 1979).

Consider, for example, the hypothesis that the observed mean phenotypes of two characters are uncorrelated at either the addi-

tive-value or residual-deviation levels. For a bivariate analysis, the probability density under the complete model is identical to Equation (18) with n substituted for the exponent $n/2$. Under the null hypothesis, $\sigma_a(1, 2) = \sigma_a(2, 1)$ and $\sigma_e(1, 2) = \sigma_e(2, 1)$ are set equal to zero, such that \mathbf{D} and \mathbf{R} are block-diagonal matrices. Letting $\hat{\mu}_0$ be the vector of means and $\hat{\mathbf{V}}_0$ the total variance-covariance matrix estimated under the more restrictive model,

$$p_0 = (2\pi)^{-n} |\hat{\mathbf{V}}_0|^{-1/2} \exp \cdot [-1/2(\bar{\mathbf{z}} - \mathbf{X}\hat{\boldsymbol{\mu}}_0)^T \hat{\mathbf{V}}_0^{-1}(\bar{\mathbf{z}} - \mathbf{X}\hat{\boldsymbol{\mu}}_0)]. \quad (21)$$

The likelihood ratio test statistic is again given by Equation (20), but in this case it is evaluated against a χ^2 distribution with two degrees of freedom (the critical points for significance at the 5% and 1% levels being 5.99 and 9.21).

In a similar manner, one can test for a significant correlation of additive (or residual) effects alone by performing an analysis with only $\sigma_a(1, 2) = \sigma_a(2, 1)$, or only $\sigma_e(1, 2) = \sigma_e(2, 1)$, constrained to be zero. Equations (21) and (20) are again employed, but Λ is treated as χ^2 distributed with one degree of freedom. It should be noted that significant covariance can exist at the subsidiary levels but not at the phenotypic level if the additive- and residual-effects covariances are of opposite sign.

A more quantitative statement about the dependence between two characters can be achieved by use of the regression of character 2 on character 1, $\hat{b} = \hat{\sigma}(1, 2)/\hat{\sigma}^2(1)$, which can be estimated at both the levels of additive and residual effects. The standard error of \hat{b} can be estimated by using the first-order Taylor approximation for the large-sample variance of a ratio,

$$SE(\hat{b}) \\ = \hat{b} \left(\frac{\hat{\sigma}^2[\hat{\sigma}(1, 2)]}{\hat{\sigma}^2(1, 2)} - \frac{2\hat{\sigma}[\hat{\sigma}^2(1), \hat{\sigma}(1, 2)]}{\hat{\sigma}(1) \cdot \hat{\sigma}(1, 2)} \right. \\ \left. + \frac{\hat{\sigma}^2[\hat{\sigma}^2(1)]}{\hat{\sigma}^4(1)} \right)^{1/2}, \quad (22)$$

where $\hat{\sigma}^2[\hat{\sigma}(1, 2)]$ is the sampling variance of the covariance of characters 1 and 2, $\hat{\sigma}[\hat{\sigma}^2(1), \hat{\sigma}(1, 2)]$ is the sampling covariance of the variance of character 1 and the co-

variance of the two traits, and $\hat{\sigma}^2[\hat{\sigma}^2(1)]$ is the sampling variance of the variance of character 1.

The standard errors for the variance and covariance components can be estimated by solving for the elements of the matrix of second-order partial derivatives of the log-likelihood with respect to the parameter estimates (Searle, 1970). The sampling variances and covariances of the parameter estimates are approximated by the negative of the elements of the inverse of this matrix. For the bivariate case, the sampling variance-covariance matrix for the variance of character 1, the covariance of 1 and 2, and the variance of 2 can be shown as in Equation (23) below where $\delta = \sigma^2(1)\sigma^2(2) - \sigma^2(1, 2)$. This expression can be applied separately to the elements of A and E, substituting observed for expected values. The diagonal elements (top left to lower right) of the solution to (23) are the estimates of $\hat{\sigma}^2[\hat{\sigma}^2(1)]$, $\hat{\sigma}^2[\hat{\sigma}(1, 2)]$, and $\hat{\sigma}^2[\hat{\sigma}^2(2)]$. The off-diagonal elements of the solution are the sampling covariances between these three variance-covariance estimates; they have nonzero expectations unless the two characters are uncorrelated.

The asymptotic sampling variances and covariances of the estimated means and additive values are obtained by substituting fitted for expected values in the information matrix given in (9) and inverting. The square roots of the diagonal elements of this matrix provide estimates of the standard errors of the parameter estimates—the first k for the means, the next n for the additive values of the first character, etc.

Finally, we come to the problem of predicting the phenotypes of ancestral species. A nodal taxon in a phylogenetic tree represents an evolutionary bifurcation. Thus, the additive value of such a taxon can be estimated by a weighted average of the additive values of its two descendent lineages. Under the assumption that the expected variance of additive values increases linearly with time, the weights are taken to be

the inverses of the times (t) between the node and the descendent taxa (Felsenstein, 1985). The mean phenotypes of all nodal taxa leading to two branch tips are computed first. For example, if h is the immediate ancestor of branch-tip species i and j , the predicted additive value of character c is

$$\hat{a}_{ch} = \frac{w_{hi}\hat{a}_{ci} + w_{hj}\hat{a}_{cj}}{w_{hi} + w_{hj}} \quad (24)$$

where $w_{hi} = 1/t_{hi}$. The predicted mean phenotype is $\hat{z}_{ch} = \hat{\mu}_c + \hat{a}_{ch}$. Once the most distal nodal taxa have been characterized, the more remote ancestral nodes can be dealt with progressively, until the root of the tree has been predicted. Equation (24) is utilized throughout this process. However, if i represents a node rather than a branch-tip taxon, the weight is not $1/t_{hi}$ but $[t_{hi} + (w_{i1} + w_{i2})^{-1}]^{-1}$ where 1 and 2 represent the taxa from which the prediction for i was derived (Felsenstein, 1985). The decreased weights for nodal taxa account for the sampling error of their estimated additive values. If Equation (24) is to yield unbiased estimates of ancestral additive values, "parallel" evolution in daughter taxa should occur no more often than expected by chance. This is difficult to verify if data are only available for extant species.

Note that in following this procedure, the residual deviations in the daughter lineages do not contribute to the predicted mean phenotypes of their ancestors. This is a departure from the approaches of Kluge and Farris (1969), Felsenstein (1985), and Huey and Bennett (1987). It is, however, consistent with the treatment of additive values as the phylogenetically inherited portions of the taxon-specific mean phenotypes.

The standard error of a predicted nodal phenotype can be estimated by use of the formula for the variance of a sum,

$$\hat{\sigma}^2(\hat{z}_{ch}) = \hat{\sigma}^2(\hat{\mu}_c) + \hat{\sigma}^2(\hat{a}_{ch}) + 2\hat{\sigma}(\hat{\mu}_c, \hat{a}_{ch}) \quad (25)$$

where as noted above $\hat{\sigma}^2(\hat{\mu}_c)$ is obtained from the inverse of the information matrix,

$$\frac{\delta^2}{n} \begin{pmatrix} \sigma^4(2)/2 & -\sigma(1, 2)\sigma^2(2) & \sigma^2(1, 2)/2 \\ -\sigma(1, 2)\sigma^2(2) & \sigma^2(1)\sigma^2(2) + \sigma^2(1, 2) & -\sigma^2(1)\sigma(1, 2) \\ \sigma^2(1, 2)/2 & -\sigma^2(1)\sigma(1, 2) & \sigma^4(1)/2 \end{pmatrix}^{-1} \quad (23)$$

$$\hat{\sigma}^2(\hat{a}_{ch}) = \frac{w_{hi}^2 \hat{\sigma}^2(\hat{a}_{ci}) + w_{hj}^2 \hat{\sigma}^2(\hat{a}_{cj}) + 2w_{hi}w_{hj} \hat{\sigma}(\hat{a}_{ci}, \hat{a}_{cj})}{(w_{hi} + w_{hj})^2}, \tag{26}$$

and

$$\hat{\sigma}(\hat{\mu}_c, \hat{a}_{ch}) = \frac{w_{hi} \hat{\sigma}(\hat{\mu}_c, \hat{a}_{ci}) + w_{hj} \hat{\sigma}(\hat{\mu}_c, \hat{a}_{cj})}{w_{hi} + w_{hj}}. \tag{27}$$

The sampling variances and covariances on the right sides of Equations (26) and (27) are taken directly from the inverse of the information matrix in the case of branch-tip species. For nodal species, they must be progressively computed using expressions for the variance and covariance of linear functions, as demonstrated in the example provided below.

A WORKED EXAMPLE

To clarify some of the practical aspects of the application of the mixed model, a small data set will now be analyzed. The data to be examined, mean adult weight (kg) and maximum longevity (years) in five species of primates (Fig. 1), were obtained from Appendices 2 and 3 in Eisenberg (1981), and these were natural-log transformed pri-

or to analysis. The phylogenetic tree (Fig. 1) is based on evidence from the fossil record (Fig. 2 in Gingerich, 1984). Data are available for many more species than presented here, so no special significance should be attached to this analysis, which is provided solely for heuristic purposes.

First, consider the results that would be obtained if an analysis were performed on the basis of the observed mean phenotypes without regard to the underlying phylogenetic relationships. The mean adult weight (character 1) and mean longevity (character 2) for the five members of the phylogeny are 2.13 ± 0.87 and 3.33 ± 0.36 on the logarithmic scale. The estimated variances of the two traits are 4.77 and 0.81 respectively, while their covariance is 1.63. On this basis, the allometric regression of lon-

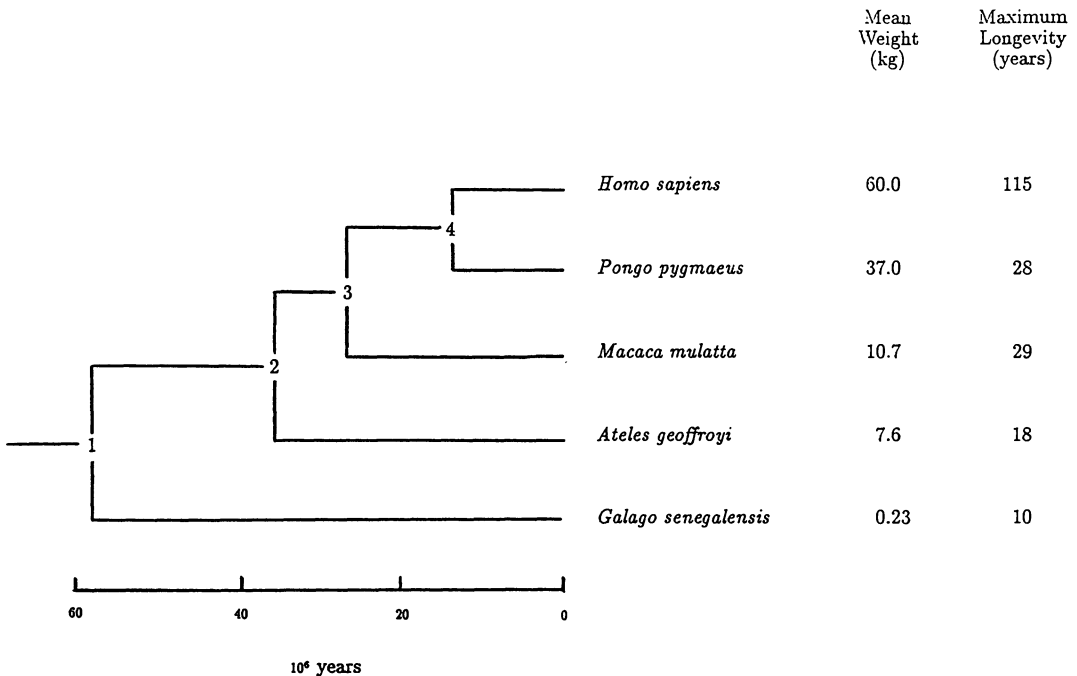


FIG. 1. An evolutionary tree for five species of primates based on data from Gingerich (1984). The associated body weights and longevity are taken from Eisenberg (1981).

evity on body weight is 0.34 ± 0.10 (Fig. 2). However, ignoring the nonindependence of the data, there would be three degrees of freedom, and the correlation ($r = 0.83$) falls just short of 0.88, which is the requirement for significance at the 5% level.

As noted above, the mixed model accounts for the nonindependence of data through the use of the relationship matrix, \mathbf{G} . There are numerous ways to assign values to the elements of \mathbf{G} . Cheverud et al. (1985), for example, proposed the use of a phylogenetic connectivity matrix that employs rather arbitrary weights to contrasts at different taxonomic levels (species, genera, families, etc.). In the following analysis, each element of \mathbf{G} is taken to be the fraction of evolutionary time from the base of the phylogeny that is shared by species i and j ,

$$G_{ij} = \frac{T - t_{hi}}{T}$$

where T is the age of the taxon at the base of the tree and $t_{hi}(=t_{hj})$ is the time separating i and j from their most recent common ancestor h . For example, since the root of the tree in Figure 1 is placed at 58 MYA and a split between *Homo* and *Pongo* is denoted at 13 MYA, the element G_{HP} is $(58 - 13)/58 = 0.78$. *Galago* is an out-group with respect to the remaining four species, so its relationship to each of the latter (internal to the phylogeny) is equal to zero. The complete relationship matrix and its inverse are shown below. Under the assumption that the expected variance of additive values among independent taxa increases linearly with time, which is the case under a commonly employed neutral model (Lande, 1979; Lynch, 1991) and under certain forms

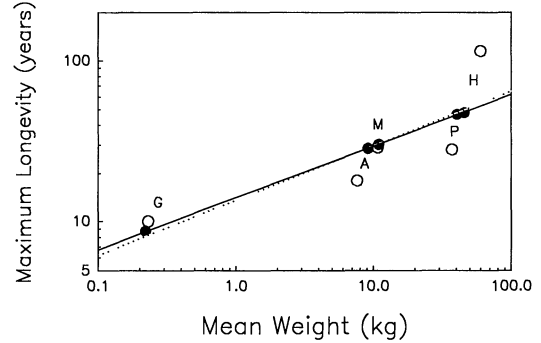


FIG. 2. The regression of longevity on body size in five primates. Open points are the observed mean phenotypes, while the solid points are the estimated mean phenotypes. The dashed line is the regression for the unweighted analysis. The solid line takes into account the phylogenetic relationships among species and is uninfluenced by the nonadditive contributions to the observed means. The letters refer to the genera given in Figure 1.

of fluctuating selection (Felsenstein, 1988a), each element of \mathbf{G} is equivalent to the expected correlation between additive values of pairs of taxa separated $(T - t_{hi})$ time units from the base of the phylogeny. In principle, this method for computing the elements of \mathbf{G} can be performed in the absence of an absolute time scale for the phylogeny, provided the units of branch length are proportional to time.

To start the analysis, random numbers were drawn from a uniform distribution with a range of 0 to +1 and multiplied by the unweighted variances of the observed means of the respective characters (given above) to obtain the diagonal elements of the initial additive-value variance-covariance matrix.

$$\mathbf{G} = \begin{pmatrix} G_{HH} & G_{HP} & G_{HM} & G_{HA} & G_{HG} \\ G_{PH} & G_{PP} & G_{PM} & G_{PA} & G_{PG} \\ G_{MH} & G_{MP} & G_{MM} & G_{MA} & G_{MG} \\ G_{AH} & G_{AP} & G_{AM} & G_{AA} & G_{AG} \\ G_{GH} & G_{GP} & G_{GM} & G_{GA} & G_{GG} \end{pmatrix} = \begin{pmatrix} 1.00 & 0.78 & 0.53 & 0.38 & 0.00 \\ 0.78 & 1.00 & 0.53 & 0.38 & 0.00 \\ 0.53 & 0.53 & 1.00 & 0.38 & 0.00 \\ 0.38 & 0.38 & 0.38 & 1.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{pmatrix},$$

$$\mathbf{G}^{-1} = \begin{pmatrix} 2.67 & -1.79 & -0.40 & -0.18 & 0.00 \\ -1.79 & 2.67 & -0.40 & -0.18 & 0.00 \\ -0.40 & -0.40 & 1.54 & -0.28 & 0.00 \\ -0.18 & -0.18 & -0.28 & 1.24 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{pmatrix}.$$

The initial residual-deviation variances were taken to be the remaining portion of the unweighted variances. The starting values for both types of covariance were set equal to zero. Thus, for $t = 1$, $\hat{\sigma}_a^2(1, 1) = 3.390$, $\hat{\sigma}_a(1, 2, 1) = 0.000$, $\hat{\sigma}_a^2(2, 1) = 0.693$, $\hat{\sigma}_e^2(1, 1) = 1.380$, $\hat{\sigma}_e(1, 2, 1) = 0.000$, and $\hat{\sigma}_e^2(2, 1) = 0.120$. These starting values, which are the elements of the matrices $\hat{\mathbf{A}}(1)$ and $\hat{\mathbf{E}}(1)$, lead to the preliminary inverses

$$\hat{\mathbf{A}}^{-1}(1) = \begin{pmatrix} 0.295 & 0.000 \\ 0.000 & 1.443 \end{pmatrix},$$

$$\hat{\mathbf{E}}^{-1}(1) = \begin{pmatrix} 0.725 & 0.000 \\ 0.000 & 8.333 \end{pmatrix}.$$

The inverse for the initial variance-covariance matrix of additive values is then (to two decimal places) $\hat{\mathbf{D}}^{-1}(1) = \hat{\mathbf{A}}^{-1}(1) \otimes \hat{\mathbf{G}}^{-1} =$

$$\begin{pmatrix} 0.80 & -0.54 & -0.12 & -0.05 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ -0.54 & 0.80 & -0.12 & -0.05 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ -0.12 & -0.12 & 0.45 & -0.08 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ -0.05 & -0.05 & -0.08 & 0.37 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.30 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 3.91 & -2.65 & -0.57 & -0.26 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -2.65 & 3.91 & -0.57 & -0.26 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -0.57 & -0.57 & 2.20 & -0.40 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -0.26 & -0.26 & -0.40 & 1.80 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.44 \end{pmatrix},$$

and for residual effects is $\hat{\mathbf{R}}^{-1}(1) = \hat{\mathbf{E}}^{-1}(1) \otimes \mathbf{I} =$

$$\begin{pmatrix} 0.72 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.72 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.72 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.72 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.72 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 8.33 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 8.33 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 8.33 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 8.33 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 8.33 \end{pmatrix}.$$

The preliminary estimates for the additive values and residual effects are obtained by use of Equations (11) and (12), substituting $\hat{\mathbf{R}}^{-1}(1)$ and $\hat{\mathbf{D}}^{-1}(1)$ from above, \bar{z} from the values in Figure 1 (log transformed), $\hat{\mu}^T(1) = [2.13, 3.33]$ from the initial estimates of the means, and

$$\mathbf{X}^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

To obtain estimates of the additive values, the information matrix must first be inverted

$$[\hat{\mathbf{R}}^{-1}(1) + \hat{\mathbf{D}}^{-1}(1)]^{-1} = \begin{pmatrix} 0.78 & 0.28 & 0.11 & 0.06 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.28 & 0.77 & 0.11 & 0.06 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.11 & 0.11 & 0.88 & 0.08 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.06 & 0.06 & 0.08 & 0.93 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.98 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.09 & 0.02 & 0.01 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.02 & 0.09 & 0.01 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.01 & 0.01 & 0.10 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.10 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.10 \end{pmatrix}.$$

The initial estimates are then found to be

$$\hat{\mathbf{a}}_1(1) = \begin{pmatrix} 1.41 \\ 1.24 \\ 0.42 \\ 0.10 \\ -2.56 \end{pmatrix} \quad \hat{\mathbf{a}}_2(1) = \begin{pmatrix} 1.01 \\ 0.22 \\ 0.09 \\ -0.32 \\ -0.87 \end{pmatrix}$$

$$\hat{\mathbf{e}}_1(1) = \begin{pmatrix} 0.55 \\ 0.24 \\ -0.18 \\ -0.20 \\ -1.04 \end{pmatrix} \quad \hat{\mathbf{e}}_2(1) = \begin{pmatrix} 0.41 \\ -0.21 \\ -0.05 \\ -0.11 \\ -0.15 \end{pmatrix}$$

The next round of estimates for the variance-covariance matrices of additive values and residual effects are obtained by substituting the preceding vectors into Equations (13) and (14). The additional terms involving traces require the use of $[\hat{\mathbf{R}}^{-1}(1) + \hat{\mathbf{D}}^{-1}(1)]^{-1}$. For example, in estimating the residual-effects variance for the first character,

$$\text{tr}[\mathbf{C}^{11}(1)] = 0.767 + 0.767 + 0.878 \\ + 0.928 + 0.981 = 4.321.$$

The elements of $\hat{\mathbf{A}}(t)$ and $\hat{\mathbf{E}}(t)$ for $t = 2$ and later iterations are given in Table 1. All of the parameter estimates are stable to two decimal places after a few hundred iterations, and the same results were obtained

regardless of the starting conditions. Note also the progressive increase in the log likelihood en route to the asymptotic solution.

The final estimates for the means and additive values are

$$\hat{\boldsymbol{\mu}}^T = (1.19 \pm 1.07 \quad 3.03 \pm 0.41),$$

$$\hat{\mathbf{a}}_1^T(1) = (2.55 \pm 1.06 \quad 2.63 \pm 1.06 \\ 1.20 \pm 1.06 \quad 1.02 \pm 1.06 \\ -2.72 \pm 1.06),$$

$$\hat{\mathbf{a}}_2^T(1) = (0.81 \pm 0.34 \quad 0.84 \pm 0.34 \\ 0.38 \pm 0.34 \quad 0.32 \pm 0.34 \\ -0.87 \pm 0.34),$$

where the standard errors are the square roots of the respective diagonal elements of the inverse of the full information matrix,

$$\begin{pmatrix} 1.13 & 0.38 & -1.12 & -1.12 & -1.12 & -1.12 & -1.12 & -0.36 & -0.36 & -0.36 & -0.36 & -0.36 \\ 0.38 & 0.17 & -0.36 & -0.36 & -0.36 & -0.36 & -0.36 & -0.11 & -0.11 & -0.11 & -0.11 & -0.11 \\ -1.12 & -0.36 & 1.12 & 1.12 & 1.12 & 1.12 & 1.12 & 0.36 & 0.36 & 0.36 & 0.36 & 0.36 \\ -1.12 & -0.36 & 1.12 & 1.12 & 1.12 & 1.12 & 1.12 & 0.36 & 0.36 & 0.36 & 0.36 & 0.36 \\ -1.12 & -0.36 & 1.12 & 1.12 & 1.12 & 1.12 & 1.12 & 0.36 & 0.36 & 0.36 & 0.36 & 0.36 \\ -1.12 & -0.36 & 1.12 & 1.12 & 1.12 & 1.12 & 1.12 & 0.36 & 0.36 & 0.36 & 0.36 & 0.36 \\ -0.36 & -0.11 & 0.36 & 0.36 & 0.36 & 0.36 & 0.36 & 0.11 & 0.11 & 0.11 & 0.11 & 0.11 \\ -0.36 & -0.11 & 0.36 & 0.36 & 0.36 & 0.36 & 0.36 & 0.11 & 0.11 & 0.11 & 0.11 & 0.11 \\ -0.36 & -0.11 & 0.36 & 0.36 & 0.36 & 0.36 & 0.36 & 0.11 & 0.11 & 0.11 & 0.11 & 0.11 \\ -0.36 & -0.11 & 0.36 & 0.36 & 0.36 & 0.36 & 0.36 & 0.11 & 0.11 & 0.11 & 0.11 & 0.11 \\ -0.36 & -0.11 & 0.36 & 0.36 & 0.36 & 0.36 & 0.36 & 0.11 & 0.11 & 0.11 & 0.11 & 0.11 \end{pmatrix}$$

This matrix is obtained by inverting (9) after substituting the elements of the inverses of the final variance-covariance matrices from Table 1.

The mixed-model estimates of μ_1 and μ_2 are rather different from the standard means (2.13 and 3.33) due to the unequal weighting of the five taxa. In addition, the standard errors of $\hat{\mu}_1$ and $\hat{\mu}_2$ are increased above the unweighted estimates due to the loss of “degrees of freedom” in the weighted analysis. Since the additive values are estimated as deviations from the estimated means, their standard errors are of the same order of magnitude as those of the latter. However, due to the negative sampling covariance between the means and additive effects (first two rows and columns in the inverse of the information matrix), the standard errors of the total evolutionary values ($\hat{\mu}_c + \hat{a}_{ci}$) are substantially smaller than that of $\hat{\mu}_c$ and \hat{a}_{ci} . For example, from the elements of the inverse of the information matrix, the stan-

dard error of the total evolutionary value of adult weight in *Homo sapiens* is

$$\begin{aligned} SE(\hat{\mu}_1 + \hat{a}_{1H}) &= [\hat{\sigma}^2(\hat{\mu}_1) + 2\hat{\sigma}(\hat{\mu}_1, \hat{a}_{1H}) \\ &\quad + \hat{\sigma}^2(\hat{a}_{1H})]^{1/2} \\ &= [1.132 - (2 \times 1.124) \\ &\quad + 1.124]^{1/2} \\ &= 0.089. \end{aligned}$$

From the elements of the variance-covariance matrices (Table 1) and Equation (17a), the phylogenetic heritabilities and their standard errors are found to be $H^2_{\bar{p}} = 0.99 \pm 0.01$ for ln(adult weight) and $H^2_{\bar{p}} = 0.54 \pm 0.15$ for ln(longevity). A univariate analysis gave essentially the same result for adult weight, but converged to $H^2_{\bar{p}} = 0.00$ for longevity. The likelihood ratio test statistics for both characters, computed from the univariate analyses, are 2.57 and 0.12. Neither of these is statistically significant, but this is not surprising for such a small data set.

TABLE 1. The elements of the variance-covariance matrices and the log-likelihood as a function of iteration number.

<i>t</i>	$\hat{\sigma}^2_{d(1)}$	$\hat{\sigma}_{d(1,2)}$	$\hat{\sigma}^2_{d(2)}$	$\hat{\sigma}^2_{e(1)}$	$\hat{\sigma}_{e(1,2)}$	$\hat{\sigma}^2_{e(2)}$	ln <i>p</i>
2	3.060	0.771	0.764	1.168	0.072	0.144	-16.839
4	3.227	1.218	0.724	0.748	0.122	0.183	-14.605
6	3.192	1.208	0.604	0.555	0.141	0.228	-13.804
11	3.012	1.084	0.437	0.366	0.169	0.285	-12.990
21	3.071	1.016	0.359	0.224	0.155	0.287	-12.580
51	3.043	0.980	0.323	0.113	0.126	0.276	-12.352
101	3.031	0.970	0.314	0.075	0.115	0.272	-12.283
201	3.025	0.966	0.310	0.057	0.109	0.269	-12.251
501	3.021	0.963	0.308	0.047	0.106	0.268	-12.233
1,001	3.020	0.963	0.307	0.044	0.105	0.268	-12.227
5,001	3.019	0.962	0.307	0.041	0.104	0.267	-12.222
25,001	3.019	0.962	0.307	0.040	0.104	0.267	-12.221

The slope of the additive-value regression of longevity on adult weight is estimated to be $\hat{b} = 0.962/3.019 = 0.32$, which is very similar to the value obtained by unweighted least-squares analysis. The residual-deviation regression is substantially higher, 2.57. The correlations at both levels converged to 1.00. Since this singularity results in $\delta = 0$, it precludes the estimation of standard errors of the regression coefficients with Equation (22). The likelihood ratio test statistics are 2.65 for the additive-value and 0.20 for the residual-deviation regressions, neither of which is significant.

At this point, it is worth emphasizing that the perfect correlation between the additive values of both traits in this example is a consequence of the small number of taxa (relative to numbers of parameters estimated). It is not a general feature of mixed-model analysis. Nor is such a high concordance between weighted and unweighted regression coefficients generally expected.

To illustrate the prediction of mean phenotypes of nodal taxa, we will consider adult weight in taxa 4 and 3. The predicted additive value for nodal species 4 is

$$\hat{a}_{14} = \frac{w_{4H}\hat{a}_{1H} + w_{4P}\hat{a}_{1P}}{w_{4H} + w_{4P}} = \frac{\frac{2.55}{13} + \frac{2.63}{13}}{\frac{1}{13} + \frac{1}{13}} = 2.59$$

where 13 (MYA) is the evolutionary time from taxon 4 to both *Homo* (*H*) and *Pongo* (*P*). The predicted mean phenotype is therefore $\hat{z}_{14} = \hat{\mu}_1 + \hat{a}_{14} = 1.19 + 2.59 = 3.78$. The sampling variance of \hat{z}_{14} is

$$\hat{\sigma}^2(\hat{z}_{14}) = \hat{\sigma}^2(\hat{\mu}_1) + \hat{\sigma}^2(\hat{a}_{14}) + 2\hat{\sigma}(\hat{\mu}_1, \hat{a}_{14})$$

where $\hat{\sigma}^2(\hat{\mu}_1) = 1.132$ from the inverse of the information matrix,

$$\hat{\sigma}^2(\hat{a}_{14}) = \frac{w_{4H}^2\hat{\sigma}^2(\hat{a}_{1H}) + w_{4P}^2\hat{\sigma}^2(\hat{a}_{1P}) + 2w_{4H}w_{4P}\hat{\sigma}(\hat{a}_{1H}, \hat{a}_{1P})}{(w_{4H} + w_{4P})^2} = \frac{(0.077^2 \cdot 1.124) + (0.077^2 \cdot 1.124) + (2 \cdot 0.077 \cdot 0.077 \cdot 1.124)}{(0.077 + 0.077)^2} = 1.124$$

also substituting from the inverse of the information matrix, and

$$\hat{\sigma}(\hat{\mu}_1, \hat{a}_{14}) = \frac{w_{4H}\hat{\sigma}(\hat{\mu}_1, \hat{a}_{1H}) + w_{4P}\hat{\sigma}(\hat{\mu}_1, \hat{a}_{1P})}{w_{4H} + w_{4P}} = \frac{(0.077 \cdot -1.124) + (0.077 \cdot -1.124)}{0.077 + 0.077} = -1.124.$$

Summing up terms, $\hat{\sigma}^2(\hat{z}_{14}) = 0.008$, which yields a standard error of \hat{z}_{14} equal to $\sqrt{0.008} = 0.09$.

The prediction of the mean phenotype of taxon 3 is a little more involved. First, we

need the weights $w_{3M} = 1/27 = 0.037$ and $w_{34} = [14 + (0.077 + 0.077)^{-1}]^{-1} = 0.049$. Using these and the value of \hat{a}_{14} derived above, we obtain

$$\hat{a}_{13} = \frac{w_{34}\hat{a}_{14} + w_{3M}\hat{a}_{1M}}{w_{34} + w_{3M}} = \frac{(0.049 \cdot 2.59) + (0.037 \cdot 1.20)}{0.049 + 0.037} = 1.99$$

which leads to the predicted mean phenotype $\hat{z}_{13} = \hat{\mu}_1 + \hat{a}_{13} = 3.18$. The sampling variance of \hat{z}_{13} is

$$\hat{\sigma}^2(\hat{z}_{13}) = \hat{\sigma}^2(\hat{\mu}_1) + \hat{\sigma}^2(\hat{a}_{13}) + 2\hat{\sigma}(\hat{\mu}_1, \hat{a}_{13})$$

where

$$\hat{\sigma}^2(\hat{a}_{13}) = \frac{w_{3M}^2\hat{\sigma}^2(\hat{a}_{1M}) + w_{34}^2\hat{\sigma}^2(\hat{a}_{14}) + 2w_{3M}w_{34}\hat{\sigma}(\hat{a}_{1M}, \hat{a}_{14})}{(w_{3M} + w_{34})^2}$$

with

$$\hat{\sigma}(\hat{a}_{1M}, \hat{a}_{14}) = \frac{w_{4H}\hat{\sigma}(\hat{a}_{1M}, \hat{a}_{1H}) + w_{4P}\hat{\sigma}(\hat{a}_{1M}, \hat{a}_{1P})}{w_{4H} + w_{4P}},$$

and

$$\hat{\sigma}(\hat{\mu}_1, \hat{a}_{13}) = \frac{w_{3M}\hat{\sigma}(\hat{\mu}_1, \hat{a}_{1M}) + w_{34}\hat{\sigma}(\hat{\mu}_1, \hat{a}_{14})}{w_{3M} + w_{34}}.$$

Summing up terms, $\hat{\sigma}^2(\hat{z}_{13}) = 0.008$, yielding a standard error of \hat{z}_{13} equal to 0.09.

DISCUSSION

The methods introduced in this paper take the approach that each member of a phylogeny provides some information about all the others. Common to all members of the phylogeny are the phylogenetic means. The deviations of individual taxa from these means are partitioned into a phylogenetically heritable portion, responsible for the resemblance between related taxa within the phylogeny, and a residual deviation that is phylogenetically uninformative. This distinction between additive evolutionary values and residual deviations is analogous to the distinction made between genotypic values and environmental deviations in quantitative genetic analysis. Contrary to previous approaches to predicting ancestral phenotypes and to estimating the variances and covariances of characters among taxa, the residual deviations are removed in the course of the analysis. This minimizes the bias in the prediction of ancestral phenotypes that can be caused by branch-tip phenotypes that are greatly modified by environmental effects, by measurement error, or by evolutionary processes that are exaggerated with respect to the remainder of the phylogeny. It also helps insure that the estimated variances and covariances of characters are those that are relevant to macroevolutionary patterns. However, a bias can occur in this sort of analysis if closely related species occupy similar environments that modify phenotypic expression in a parallel fashion (as in quantitative genetics, when sibs are exposed to a common familial environment).

Another advantage to the proposed technique is the utilization of the relationship matrix, which allows an efficient and simultaneous use of all of the data to obtain

the set of parameter estimates that is at or close to the maximum in the likelihood surface. This avoids the necessity of computing and standardizing the independent contrasts suggested by Felsenstein (1985), and is an advance over the methods employed in Lynch (1988, 1991) and Bell (1989) to estimate the divergence rates of characters within phylogenies. In a multivariate analysis, the use of the relationship matrix also takes advantage of the fact that correlated characters provide information about each other. Provided their phylogenetic affinity is known, fossil taxa can readily be incorporated into the relationship matrix and included in the analysis.

Perhaps the most useful aspect of the mixed model in comparative analysis is its utility in hypothesis testing. Log-likelihood procedures provide a sound basis for testing hypotheses regarding components of variance and covariance, regression coefficients, etc. However, as noted in the preceding example, the power of likelihood tests may be rather low with the small number of taxa that are usually employed in comparative analyses. It would be worthwhile to explore this issue in future studies. Small numbers of taxa, as well as the complex variance-covariance structure, normally will also preclude powerful tests of the multivariate normality assumption upon which likelihood tests are based. Nevertheless, investigators would be wise to explore whether various scale transformations for the phenotypic means or for divergence times can significantly improve the likelihood of the model. In Lynch (1991), for example, it is suggested that the interspecific variance in mammalian skeletal traits increases with the square root of time, unlike the linear scaling expected under the neutral model.

A major limitation of all comparative methods is the need for an accurate phylogenetic tree. Methods of molecular anal-

ysis are rapidly improving our ability to address phylogenetic issues. However, when more than a few taxa are involved, as is often the case in comparative studies of quantitative traits, the likelihood of recovering the correct topology is extremely low even with very large sets of molecular data (Nei, 1987; Felsenstein, 1988*b*). Felsenstein (1988*a*, 1988*b*) has discussed the possibility of using resampling procedures to produce multiple trees from the same data set. Following this approach, each tree could be used to construct a relationship matrix that could then be applied to the mixed model. Such a procedure would be computationally intense, but the variances and covariances of parameter estimates derived from the different analyses could be used to account for the uncertainty in the phylogeny. Such sampling variances and covariances would need to be added to those described above. We will never know phylogenies with absolute certainty, but recent work (Martins and Garland, 1991) indicates that the application of even crude phylogenies is substantially superior to the reliance on the independence assumption in comparative analysis.

In the models given above, the sampling errors of the phenotypic means were assumed to be zero. This will often not be true since species-specific means are usually based on a small number of individuals from one or a few populations, and individual variation will generally be present within and between populations. If the sampling variances and covariances of the means are available, they should be subtracted from the respective elements of the final residual effects variance-covariance matrix $\hat{\mathbf{E}}$ to remove the bias in the elements of $\hat{\mathbf{E}}$ caused by measurement error.

Cheverud et al. (1985) have argued that correlations between residual deviations ($\hat{\epsilon}$) provide evidence of coadaptation of characters. This interpretation may be misleading for two reasons. First, such correlations can arise in the absence of selection via random genetic drift of genes with pleiotropic effects (Lande, 1979; Lynch and Hill, 1986). Second, unless sampling error has been accounted for properly, the variances and covariances of residual effects will be biased to an unknown degree by environmental ef-

fects and measurement error. If the latter problem cannot be eliminated, correlations at the level of residual deviations should be interpreted with caution.

For situations in which the contribution from sampling error can be eliminated, $\hat{\mathbf{E}}$ and $\hat{\mathbf{A}}$ may be viewed as estimates of pattern on the microevolutionary (taxon-specific) and macroevolutionary (phylogeny-wide) time scales. A comparison of these two variance-covariance matrices can indicate whether the constraints on multivariate evolution are consistent across evolutionary time scales. For instance, in the preceding example, the regression between additive effects was much less substantial than that between residual effects.

In closing, it should be emphasized that the techniques outlined above provide only a descriptive analysis of phylogenetic patterns. Identification of the genetic and ecological factors responsible for such patterns is a matter of empirical investigation.

ACKNOWLEDGMENTS

This work profited greatly from comments by W. Bradshaw, R. Cantet, J. Cheverud, L. Heisler, P. Jarrell, R. Lande, E. Martins, R. Shaw, K. Spitze, and B. Sullender. Many thanks are also due to R. Shaw and E. Thompson for access to unpublished work, which helped clarify my understanding of mixed-model methodology. This study was supported by PHS grant R01 GM36827-01 and NSF grants BSR 86-00487 and 89-11038.

LITERATURE CITED

- BELL, G. 1989. A comparative method. *Am. Nat.* 133:553-571.
- CHEVERUD, J. M., M. M. DOW, AND W. LEUTENEGGER. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: Sexual dimorphism in body weight among primates. *Evolution* 39:1335-1351.
- DEMPSTER, A. P., N. M. LAIRD, AND D. B. RUBIN. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* 39:1-38.
- EDWARDS, A. W. F. 1972. *Likelihood*. Cambridge Univ. Press, Cambridge, UK.
- EISENBERG, J. F. 1981. *The Mammalian Radiations*. Univ. Chicago Press, IL.
- FALCONER, D. S. 1981. *Introduction to Quantitative Genetics*. 2nd Ed. Longman, N.Y.
- FELSENSTEIN, J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1-15.

- . 1988a. Phylogenies and quantitative characters. *Annu. Rev. Ecol. Syst.* 19:445–472.
- . 1988b. Phylogenies from molecular sequences: Inference and reliability. *Annu. Rev. Genet.* 22: 521–565.
- GINGERICH, P. D. 1984. Primate evolution: Evidence from the fossil record, comparative morphology, and molecular biology. *Yearbook of Phys. Anthropol.* 27:57–72.
- HARVEY, P. H., AND T. H. CLUTTON-BROCK. 1985. Life history variation in primates. *Evolution* 39: 559–581.
- HARVILLE, D. A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.* 72:320–338.
- HENDERSON, C. R. 1963. Selection index and expected genetic advance, pp. 141–163. *In* *Statistical genetics and plant breeding*. Natl. Acad. Sci.—Natl. Res. Council Publ. No. 982.
- . 1984a. Applications of Linear Models in Animal Breeding. Univ. Guelph, Canada.
- . 1984b. ANOVA, MIVQUE, REML, and ML algorithms for estimation of variances and covariances, pp. 257–280. *In* H. A. David and H. T. David (eds.), *Statistics: An Appraisal*. Iowa State Univ. Press, Ames.
- HENDERSON, C. R., O. KEMPTHORNE, S. R. SEARLE, AND C. N. VON KROSIGK. 1959. Estimation of genetic and environmental trends from records subject to culling. *Biometrics* 13:192–218.
- HUEY, R. B. 1987. Phylogeny, history, and the comparative method, pp. 76–98. *In* M. E. Feder, A. F. Bennett, W. W. Burggren, and R. B. Huey (eds.), *New Directions in Ecological Physiology*. Cambridge Univ. Press, UK.
- HUEY, R. B., AND A. F. BENNETT. 1987. Phylogenetic studies of coadaptation: Preferred temperatures versus optimal performance temperatures of lizards. *Evolution* 41:1098–1115.
- KENDALL, M. G., AND A. STUART. 1979. *The Advanced Theory of Statistics*. Vol. 2. Inference and Relationship. Griffin and Co., London, UK.
- KLUGE, A. G., AND J. S. FARRIS. 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* 24:244–256.
- LANDE, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* 33:402–416.
- LANDE, R., AND S. J. ARNOLD. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–1226.
- LYNCH, M. 1988. The rate of polygenic mutation. *Genet. Res.* 51:137–148.
- . 1991. The rate of morphological evolution in mammals from the standpoint of the neutral theory. *Am. Nat.* 136:727–741.
- LYNCH, M., AND W. G. HILL. 1986. Phenotypic evolution by neutral mutation. *Evolution* 40:915–935.
- MARTINS, E. P., AND T. GARLAND, JR. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: A simulation study. *Evolution* 45:534–557.
- MITCHELL-OLDS, T., AND J. J. RUTLEDGE. 1986. Quantitative genetics in natural plant populations: A review of the theory. *Am. Nat.* 127:379–402.
- MITCHELL-OLDS, T., AND R. G. SHAW. 1987. Regression analysis of natural selection: Statistical inference and biological interpretation. *Evolution* 41: 1149–1161.
- NEI, M. 1987. *Molecular Evolutionary Genetics*. Columbia Univ. Press, N.Y.
- PAGEL, M. D., AND P. H. HARVEY. 1988. Recent developments in the analysis of comparative data. *Q. Rev. Biol.* 63:413–440.
- RIDLEY, M. 1983. *The Explanation of Organic Diversity: The Comparative Method and Adaptations for Mating*. Oxford Univ. Press, UK.
- SEARLE, S. R. 1970. Large sample variances of maximum likelihood estimators of variance components using unbalanced data. *Biometrics* 26:505–524.
- STEARNS, S. C. 1983. The influence of size and phylogeny on patterns of covariation among life-history traits in mammals. *Oikos* 41:173–187.
- THOMPSON, E. A., AND R. G. SHAW. 1990. Pedigree analysis for quantitative traits: Variance components without matrix inversion. *Biometrics* 46:399–413.

Corresponding Editor: J. Arnold