doi:10.1111/j.1558-5646.2008.00412.x

A COMPARATIVE METHOD FOR STUDYING ADAPTATION TO A RANDOMLY EVOLVING ENVIRONMENT

Thomas F. Hansen,[1,2,3] **Jason Pienaar,**[2,4] **and Steven Hecht Orzack**[5]

[1] *Center for Evolutionary and Ecological Synthesis, Department of Biology, University of Oslo, PB 1066, Blindern, 0316 Oslo, Norway*

[2] *Department of Biological Sciences, Florida State University, Tallahassee, Florida 32306*

[3] *E-mail: thomas.hansen@bio.uio.no*

[5] *Fresh Pond Research Institute, Cambridge, Massachusetts 02140*

Received September 26, 2007
Accepted April 9, 2008

Most phylogenetic comparative methods used for testing adaptive hypotheses make evolutionary assumptions that are not compatible with evolution toward an optimal state. As a consequence they do not correct for maladaptation. The "evolutionary regression" that is returned is more shallow than the optimal relationship between the trait and environment. We show how both evolutionary and optimal regressions, as well as phylogenetic inertia, can be estimated jointly by a comparative method built around an Ornstein–Uhlenbeck model of adaptive evolution. The method considers a single trait adapting to an optimum that is influenced by one or more continuous, randomly changing predictor variables.

KEY WORDS: **Adaptation, maladaptation, optimality, Ornstein–Uhlenbeck process, phylogenetic comparative method, phylogenetic effect, phylogenetic inertia.**

Optimality models and other adaptive hypotheses are often tested by comparing their predictions to the trait values of species in different environments (e.g., Ridley 1983; Harvey and Pagel 1991). Most comparative methods are, however, based on models that are inconsistent with evolution toward an optimum. For example, the method of independent contrasts makes the assumption that traits evolve according to a Brownian-motion process (Felsenstein 1985), but if evolution is governed by this process, the expected trait value of a descendant species must equal the trait value of its ancestor, and there can be no systematic evolution toward an optimal state. If the ancestral species' trait value does not match the optimum, then the trait value of the descendant species is not expected to match it any better. Hansen and Orzack (2005) called this the problem of inherited maladaptation. A multivariate Brownian-motion process can be used to represent correlated evolutionary changes in two or more traits, but correlated evolution is not equivalent to adaptive evolution. For example, suppose we predict that the optimal relation between two traits, x and y, is $y = x$. This simple prediction is incompatible with evolution as a Brownian-motion process. Even if there is a positive correlation between changes in y and x, any deviation from the 1:1 line will be inherited by the descendant species and there will be no systematic tendency to evolve toward the predicted relationship.

This lack of attention to the fundamental nature of adaptive evolution has influenced the application of phylogenetic comparative methods. Such methods are often used erroneously to "correct" for phylogeny when they should only correct for the residual effects of phylogeny that remain after adaptation has been accounted for. A phylogenetic signal in the data can arise both

---

[4] Current address: Department of Zoology, University of Hawaii at Manoa, Honolulu, Hawaii 96822

from the influence of ancestral character states due to a lag or inertia in adaptation to the current niche, and from the effects of adaptation to niche variables that are themselves phylogenetically structured. One should correct only for the former. Correcting for effects stemming from adaptation to phylogenetically structured variables can amount to correcting for the very phenomenon that one wants to study, and may lead to erroneous inferences about both adaptation and inertia.

As reviewed by Blomberg and Garland (2002), terms such as phylogenetic effect and phylogenetic inertia have been given many and often vague meanings. In this article, we will use the term inertia to denote a resistance to or slowness in the adaptation to a specific optimum. Such inertia will generate patterns of correlated maladaptation in related species, which need to be corrected for in the study of adaptation. The causes of inertia could include any constraint on the process of adaptation toward the specified optimum. In contrast, we will use the terms phylogenetic effect (or signal) to denote any statistical influence of phylogeny on the trait. This distinction between inertia and effect is essential to understand the arguments in this article.

We have previously argued that inertia and adaptation must be analyzed in a joint framework in which both are controlled for the effects of the other (Orzack and Sober 2001; Sober and Orzack 2003; Hansen and Orzack 2005). Only in this way can we correctly test hypotheses about adaptation to an optimal state. Furthermore, many comparative studies compound the problem by only testing for presence or absence of phylogenetic effects. It is clear that different traits can display very different levels of phylogenetic effects and presumably very different levels of phylogenetic inertia (Freckleton et al. 2002; Blomberg et al. 2003; Ashton 2004). It is therefore essential to base analyses of adaptation on parametric estimates of inertia.

Few phylogenetic comparative methods deal successfully with these problems (Orzack and Sober 2001; Hansen and Orzack 2005). For continuous response traits, the "adaptation-inertia" method of Hansen (1997) is the only approach that both allows for joint estimation of inertia and adaptation, and is based on an evolutionary model that avoids the problem of inherited maladaptation. This method only works, however, for fixed predictor variables, or "niches," that can be mapped onto a phylogeny (Butler and King 2004). Here we extend the method to include continuous predictor variables that evolve randomly over time.

Model

BACKGROUND: A MODEL OF ADAPTATION TO A PRIMARY NICHE OPTIMUM

It has long been realized that proper statistical analysis of comparative data must take account of the phylogenetic relationships between species (Felsenstein 1985; Harvey and Pagel 1991; Martins and Hansen 1996; Martins 2000; Rohlf 2001). To do so requires phylogenetic information, and also a model of the evolutionary process that unfolds on the phylogeny (Felsenstein 1988; Hansen and Martins 1996; Martins et al. 2002). The choice of process model is crucial, and it is essential that this choice is compatible with the evolutionary hypotheses under test.

Following Hansen (1997), we assume that the response trait, y, evolves toward a "primary" niche optimum, θ, which is predicted by the adaptive hypothesis under test. The simplest stochastic model that allows evolution toward a specific state is the so-called Ornstein–Uhlenbeck process, which can be represented by the stochastic differential equation

$$dy = -\alpha(y - \theta)\,dt + \sigma_y\,dW_y,$$

where dy is the change in y over a time step dt, α is a parameter measuring the rate of adaptation toward the optimum, dW_y is a white-noise process (i.e., independent, normally distributed random changes with mean zero and unit variance), and σ_y is the standard deviation of the random changes. Thus, one component of evolution is a linearly increasing deterministic pull of the trait toward the primary optimum, and the other is a nondirected stochastic change.

We interpret the Ornstein–Uhlenbeck model not as describing evolution toward a fixed fitness optimum, but as describing the movements of the fitness optima themselves, as influenced by a number of unmeasured variables. Adaptation to a particular niche is defined as the approach to the primary optimum defined by this niche. Hansen (1997) defined the primary optimum for a niche as the average fitness optimum that would be reached by a large number of independent species evolving for a long time in this niche. The term "primary" derives from Simpson's (1944) notion of primary and secondary adaptation where primary adaptations might be inhibited by variation in secondary adaptations that exhibit phylogenetic variation. To this extent, the stochastic part of the model represents changes in the fitness optimum generated by changes in unmeasured selective factors and other third variables. Note that the Ornstein–Uhlenbeck model can also be derived from Lande's (1976) model of the dynamics of stabilizing selection and genetic drift, but on the evolutionary time scales relevant to most comparative data, Lande's model predicts essentially instantaneous movements around the optimum, and is therefore incompatible with the presence of phylogenetic inertia (Hansen and Martins 1996; Hansen 1997; but see Estes and Arnold 2007). This shift in interpretation of the Ornstein–Uhlenbeck model from describing dynamics on the adaptive landscape to describing dynamics of the adaptive landscape is a key step toward linking microevolutionary processes to macroevolutionary pattern (see Arnold et al. 2001).

For a fixed primary optimum, we can solve the above stochastic differential equation (see Appendix). The expected trait value of a species that has evolved for a period of time, t, toward the primary optimum, θ, is

$$E[y(t)] = \theta(1 - e^{-\alpha t}) + y_a e^{-\alpha t},$$

where y_a is the ancestral value of the trait. The expected trait value of a species is a weighted average of the influences of the primary optimum and of the ancestral trait value. The weighting factor, $e^{-\alpha t}$, depends on how much time, t, the species has spent evolving in the niche, and on its rate of adaptation, α. The rate of decay of the influence of the ancestral trait value can also be measured as a "phylogenetic half-life," defined as the time it takes for the expected trait value to move half the distance from the ancestral state to the primary optimum (Hansen 1997). In this model the phylogenetic half-life is

$$t_{1/2} = \frac{\ln 2}{\alpha}.$$

If the half-life is short relative to phylogeny, it means that adaptation to the primary optimum is rapid in expectation, and if the half-life is long, it means that ancestral influence lingers, and we expect species to be poorly adapted to the primary niche. A half-life of infinity corresponds to evolution governed by a Brownian-motion process, and there is no tendency to move toward the optimum. The rate of adaptation and the half-life are also related to the phylogenetic correlations between species, as discussed below.

Given fixed niches mapped onto a phylogeny, one can use a general linear model or likelihood framework to obtain estimates of the primary optima in these niches, as well as estimates of phylogenetic inertia (Hansen 1997; Butler and King 2004).

ADAPTATION TO A RANDOMLY CHANGING OPTIMUM

The statistical framework developed by Hansen (1997) and Butler and King (2004) has the assumption that the niches and their primary optima are fixed on the phylogeny. Here, we will assume instead that the optimum at any point on the phylogeny is a function of a randomly changing predictor variable, x, for which we only know the value at the tips of the phylogeny. We will assume that the predictor variable evolves as if by a Brownian-motion process. This is a reasonable choice for a variable influenced by a large number of unknown stochastic factors, which due to the central limit theorem will combine to generate normally distributed random changes at any given time.

We now have coupled stochastic differential equations:

$$\begin{aligned} dy &= -\alpha(y - \theta(x))\,dt + \sigma_y\,dWy, \\ dx &= \sigma_x\,dW_x, \end{aligned}$$

where the primary optimum is a function of the predictor variable, x. Note that the predictor only indirectly influences the trait through its influence on the fitness optimum for the trait. In the Appendix we obtain the moments of this model. If we assume that all species are extant and that the relationship of the primary optimum to the predictor variable is a simple linear regression, $\theta = b_0 + b_1 x$, then the predicted "evolutionary regression" of y on x is

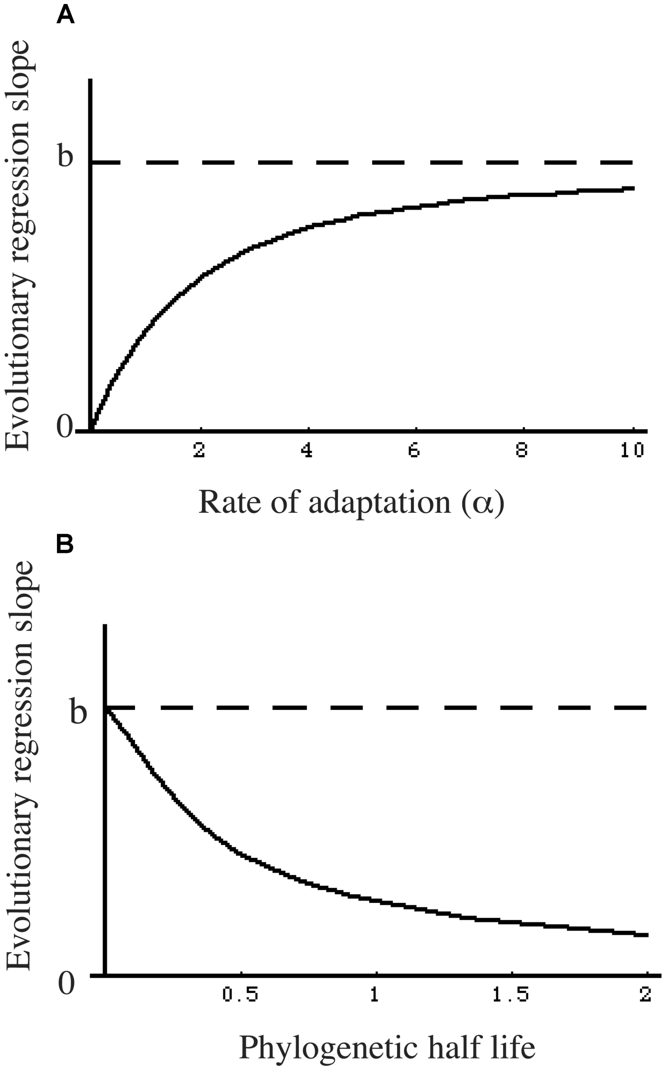

Figure 1. The evolutionary regression slope as a function of inertia: (A) the expected evolutionary regression as a function of the rate of adaptation (α). (B) The expected evolutionary regression as a function of the phylogenetic half-life ($t_{1/2}$). The total phylogeny length is scaled to one, and *b* denotes the value of the optimal regression slope.

$$E[y|x] = k + \left(1 - \frac{(1 - e^{-\alpha t})}{\alpha t}\right) b_1 x,$$

where k is a constant (see the Appendix). Note that the evolutionary regression slope is a product of a phylogenetic correction factor, $\rho(\alpha t) = (1 - (1 - e^{-\alpha t})/\alpha t)$, due to inertia, and of the underlying "optimal regression" slope, b_1, which describes the relation between the optimum and the predictor variable. Figure 1 illustrates the effects of phylogenetic inertia on the evolutionary regression. Note also that the intercept k is not identical to optimal intercept b_0 and that we cannot recover an independent estimate of b_0 because k is influenced by the ancestral values of y and x.

INTERPRETING THE OPTIMAL AND THE EVOLUTIONARY REGRESSIONS

The optimal regression describes the relationship between the trait and the predictor that we would observe if the trait was able to evolve fast enough to perfectly track the changes of the predictor. It is an estimate of the optimal relationship free of constraints. In contrast, the evolutionary regression is the standard (generalized) least squares regression of trait on the predictor, and is influenced by both adaptation and inertia. The evolutionary and the optimal regressions are not alternative models, but estimates of different aspects of the same model. A prediction from an optimality model should ideally be tested against the optimal regression, and not against the evolutionary regression. The difference between the two regressions tells us about the influence of evolutionary lag on the relationship between the predictor and the response trait.

PHYLOGENETIC COVARIANCES

To estimate the regression parameters, we need the variances and covariances of the residuals from the model. In the Appendix we derive the following formula for the residual covariance between two species i and j:

$$\begin{aligned}\mathrm{Cov}[r_i, r_j] = {} & \left(\frac{\sigma_\theta^2 + \sigma_y^2}{2\alpha}\right)\left(1 - e^{-2\alpha t_a}\right)e^{-\alpha t_{ij}} \\ & + \sigma_\theta^2 t_a \left(\left(\frac{1 - e^{-\alpha t}}{\alpha t}\right)^2 \right. \\ & \left. -2\left(\frac{1 - e^{-\alpha t}}{\alpha t}\,\frac{1 - e^{-\alpha t_a}}{\alpha t_a}\right)e^{-\alpha t_{ij}/2}\right),\end{aligned}$$

where t is the elapsed time from the base of the phylogeny to the present, t_a is the time from the base to the most recent common ancestor of the two species, and t_{ij} is the time separating the two species. The parameter $\sigma_\theta^2 = b_1^2\,\sigma_x^2$ is the instantaneous variance of the optimum. We assume that all species have the same value of t, as would be the case when all species are extant. The residual variances are found by setting $i = j$, such that $t_{ij} = 0$ and $t_a = t$.

EXTENSION TO MULTIPLE REGRESSION

We can also assume that the primary optimum is a linear combination of predictor variables, x_i, each of which evolves as if by a Brownian motion, such that $\theta = b_0 + \sum_i b_i x_i$. To recover the b_i through a regression of y on the x's, we use the following regression:

$$E[y|x_1, x_2, \ldots] = k + \rho(\alpha t)\sum_i b_i x_i,$$

in which each optimal regression coefficient is scaled by the phylogenetic correction factor, $\rho(\alpha t)$. The residual covariances for this model are given in the Appendix.

The Method

GENERAL SET UP

Estimates of the vector of regression parameters can be derived from the following model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{r}, \quad \mathbf{r} \sim N(\mathbf{0}, \mathbf{V}),$$

where $\mathbf{y}$ is an n-dimensional vector of species means, $\mathbf{X}$ is an $n \times q$ design matrix, $\mathbf{b}$ is a vector of q regression parameters, $\mathbf{r}$ is a n-vector of residual deviances from the model, assumed to follow a normal distribution with zero mean vector, and $n \times n$ variance matrix, $\mathbf{V}$, containing the residual variances and covariances given above. The entries in the design matrix define the regression model. For a regression with one predictor variable, the design matrix has two columns: each entry in the first column is 1, whereas entries in the second column are $\rho(\alpha t)x_i$, where x_i is the observed value of the predictor variable for species i. This design matrix is identical to the design matrix of a standard multiple regression except that each predictor variable is multiplied by $\rho(\alpha t)$. The residual variance matrix can also contain the measurement variances of the response and predictor variables as described in Martins and Hansen (1997).

Given a design matrix and a residual variance matrix, generalized least squares (GLS) can be used to provide estimates of the regression parameters. The GLS estimates of regression parameters are, however, conditional on knowing the various parameters that appear in the design and variance matrices (i.e., α, σ_y^2, σ_x^2). This can be solved by estimated GLS (Martins and Hansen 1997; Butler et al. 2000), where nuisance parameters and regression parameters are estimated conditionally on each other in an iterative manner.

THE ALGORITHM

Assuming the predictor variables evolve independently, their variances are estimated as

$$\hat{\sigma}_x^2 = \frac{(\mathbf{x} - E[\mathbf{x}])^T\mathbf{T}^{-1}(\mathbf{x} - E[\mathbf{x}])}{n - 1},$$

where $\mathbf{x}$ is the vector of species values for the predictor variable in question, and $\mathbf{T}$ is the matrix of shared branch lengths. The estimate of the mean of the x variables is $E[x] = (\mathbf{1}^{\mathrm{T}}\mathbf{T}^{-1}\mathbf{1})^{-1}\mathbf{1}^{\mathrm{T}}\mathbf{x}$, where $\mathbf{1}$ is a column vector of ones. To estimate α, σ_y^2, and the regression parameters, we generate a grid based on the transformed parameters, $t_{1/2} = \ln(2)/\alpha$, and $v_y = \sigma_y^2/2\alpha$. The parameter v_y is the equilibrium variance of the Ornstein–Uhlenbeck process. It describes the among-species variance we expect for a set of species that have evolved for a long time in a constant niche. For each set of values we first estimate the regression parameters with GLS. Because the regression slope also enters into the variance matrix, this is done iteratively. We start with an ordinary least-squares estimate of the regression parameters (i.e., $\hat{\mathbf{b}}_0 = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$). We use this initial estimate to obtain a GLS estimate of the regression parameters as

$$\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y},$$

where b, $t_{1/2}$, v_y, and $\sigma^2{}_x$ enter into $\mathbf{V}$, and $t_{1/2}$ enters into $\mathbf{X}$. This iterative procedure is continued until $\hat{\mathbf{b}}$ no longer changes. At this point we evaluate the support (= log likelihood) function

$$S(\alpha, v_y) = -\frac{n}{2}\mathrm{Ln}[2\pi] - \frac{1}{2}\mathrm{Ln}[\mathrm{Det}[\mathbf{V}]] - \frac{1}{2}(\mathbf{y} - \mathbf{Xb})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Xb}).$$

When all values of $t_{1/2}$ and v_y have been evaluated, we compare their support. The maximum-likelihood values of $t_{1/2}$ and v_y, and their associated $\hat{\mathbf{b}}$ are returned.

The final step is to obtain measures of uncertainty for the parameter estimates. Estimation variances and covariances for $\hat{\mathbf{b}}$ can be computed as $\mathrm{Var}[\hat{\mathbf{b}}] = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$, and be used to form standard errors and confidence intervals. The uncertainty of $\hat{\alpha}$ and $\hat{v}_y$ can be obtained from the likelihood function. We recommend reporting support sets; that is all the values of $\hat{t}_{1/2}$ and $\hat{v}_y$ that are within two support units of the best estimate (Edwards 1992).

MODEL EVALUATION, COMPARISON, AND DIAGNOSTICS

We second the recommendation of Butler and King (2004) to use information criteria for model evaluation. For comparative studies, we recommend the following small-sample version of AIC (Burnham and Anderson 2004):

$$\mathrm{AIC}_c = -2S[\hat{\mathbf{p}}] + \frac{2mn}{n - m - 1},$$

where $S[\hat{\mathbf{p}}]$ is the log likelihood of the parameter vector $\hat{\mathbf{p}}$ containing all parameters in the model, m is the number of parameters, and n is the number of taxa. A lower value of AIC_c signifies better fit. In the present model, the number of parameters, m, is two (α and v_y) plus the number of parameters in the regression model (1 + number of predictor variables). We do not count σ_x^2 and the mean of the predictor variables, which are estimated independently of the response data.

The user needs to assess whether the data and research questions are compatible with our model of trait evolution. The model will not be appropriate for all traits, especially because of our assumption of one-way causation, as reflected in our use of a regression model, as opposed to a correlational model. We do not expect the method to be very sensitive to violations of the Brownian-motion assumption for the predictor variables, but note that adaptation cannot take place if the predictor variables are changing much faster than the rate of adaptation. It is advisable to compute the $t_{1/2}$ for each predictor variable (derived from a model only including an intercept), and check that this is approximately equal to or larger than the $t_{1/2}$ for the adaptive process with respect to the predictor variable. It is also advisable to visually inspect the normality of the residuals from the fitted model.

Another important assumption of our approach is that the evolutionary process is homogeneous across the phylogeny. For a large phylogeny, it is possible to check for homogeneity by fitting the model separately to different parts of the phylogeny. O'Meara et al. (2006) show how hypotheses of rate variation across the phylogeny can be evaluated with likelihood. They focus on Brownian-motion models, but the approach could be extended to an Ornstein–Uhlenbeck model.

Parameters such as $t_{1/2}$ are estimated with relatively low accuracy, and may not be significantly different from zero even when the best estimate indicates that there is substantial phylogenetic inertia. Accordingly, inferences about inertia should be based on an estimate and the support interval, and not on statistical significance. For small datasets, we recommend doing the analysis on a set of fixed $t_{1/2}$ values with reasonable biological meaning.

COMBINING FIXED AND RANDOM VARIABLES IN THE MODEL

It is possible to include fixed effects in the model in the manner described by Hansen (1997) and Butler and King (2004). This involves mapping the state of the fixed effects onto the phylogeny, and then computing the influence of each state on each species. This influence will depend on the degree of inertia, with older states being weighted less when there is less inertia. The equations to include in the design matrix are given in Hansen (1997) and Butler and King (2004). Methods for random categorical predictors remain to be developed.

UNEQUAL BRANCH LENGTHS

When species have unequal total branch lengths, as is likely when the data include extinct species, each species has its own phylogenetic correction factor and intercept based on its own individual elapsed time back to the base of the phylogeny. The residual variance matrix also becomes more complicated. The equations to use for nonultrametric phylogenies are given in the Appendix. When the unequal total branch lengths are due to estimation error, it may be more logical to transform the phylogeny to ultrametric form, or at least use the same $\rho(\alpha t)$ for all species. It may, however, still be advisable to use the best estimate of the branch lengths when computing the residual covariances, as these may more accurately estimate the distances between the species.

Implementation and Example

SOFTWARE

To implement the method we have developed an extension to Butler and King's (2004) OUCH program in R. This program, called SLOUCH (Stochastic Linear Ornstein-Uhlenbeck models for Comparative Hypotheses), is available, along with a user manual, from http://folk.uio.no/thomasha/Programs/, or http://www.freshpond.org/software/SLOUCH. The input format of SLOUCH is similar to OUCH. SLOUCH requires a phylogeny with branch lengths and vectors with values for dependent and independent variables. The user must specify a set of candidate values for some of the parameters (α or $t_{1/2}$, and v_y), while the program calculates the best estimates of the other parameters (regression coefficients and σ_x^2).

EXAMPLE: AN ANALYSIS OF RENSCH'S RULE IN PRIMATES

Rensch's rule states that sexual size dimorphism increases with body size when the female is the smaller sex (Rensch 1959, p. 157–159; Abouheif and Fairbairn 1997). In primates, Rensch's rule has been the subject of numerous comparative studies with varying results (see for example Cheverud et al. 1985; Plavcan and van Schaik 1992; Smith and Cheverud 2002; Thoren et al. 2006). To illustrate our method, we analyze the relationship between sexual size dimorphism, measured as the log of the ratio of male to female body weight, and log female body weight in primates. We use data and a molecular phylogeny for 105 species reported in Smith and Cheverud (2002). As we wish to merely illustrate the method, we will not control for mating system, correct for measurement error and spurious correlation, or fully assess heterogeneity in different parts of the phylogeny. These and other extensions are left for a future contribution.

The phylogeny of the 105 primate species is shown in Figure 2. In Figure 3 we show the support surface for the phylogenetic half-life, $t_{1/2}$, and the equilibrium variance parameter, v_y. The best estimate of the half-life is $t_{1/2} = 8.17$ million years (myr), and the two-unit support region is from $t_{1/2} = 5.10$ myr to $t_{1/2} = 14.74$ myr. Thus, according to the best estimate it would take 8.17 myr for the average species to evolve half the distance toward a new primary optimum. This indicates fairly strong inertia in the evolution of body-size dimorphism, but given the length of the phylogeny (63 myr), the phylogenetic correction factor corresponding to the best estimate of the half-life is still as high as $\rho = 0.81$. Thus, the evolutionary regression slope is 81% of the underlying optimal slope (Fig. 4). The best estimate of the optimal regression slope is 0.081 ± 0.020, which means that a doubling of body size would give about 6% increase in the optimal sexual size dimorphism. Although highly significant, this explains only 13% of the among-species variation. Thus, our analysis supports size-dependent sexual size dimorphism in primates, but not as a strong pattern.

To illustrate a more dramatic influence of phylogenetic inertia we consider only the 47 Catarrhine species. For this group the

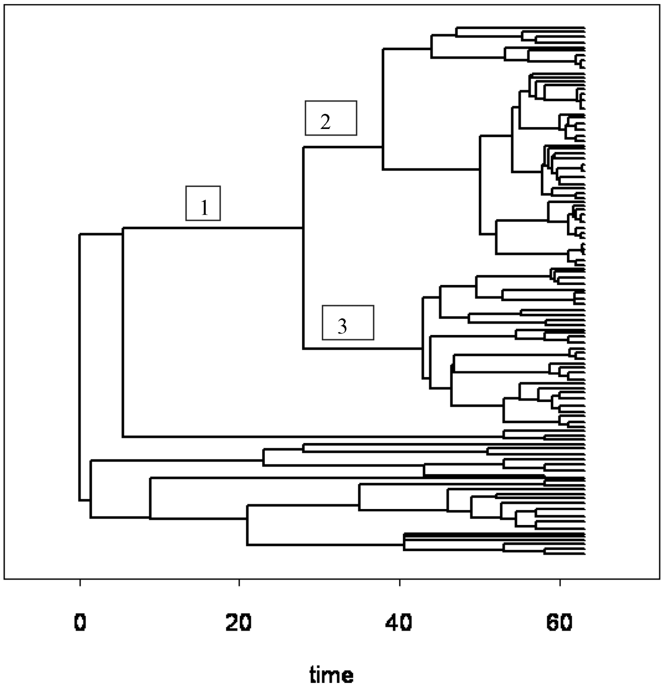

Figure 2. A phylogeny of 105 species of primates from Smith and Cheverud (2002). The branch leading to the Haplorhini is marked as 1, the branch leading to Catarrhini is marked as 2, and the branch leading to the Platyrrhini is marked as 3. The time scale is in millions of years.

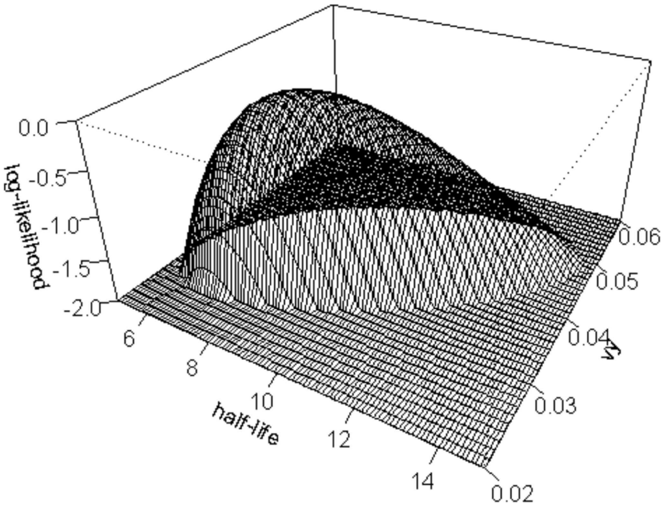

Figure 3. Support surface for the phylogenetic half-life ($t_{1/2}$) and the stationary variance (v_y) for a regression of ln(male body weight/female body weight) on ln(female body weight) for all primates. The maximum-likelihood estimate (the peak of the surface) is $t_{1/2} = 8.17$ myr and $v_y = 0.035$. The elevated (tear-drop shaped) area shows all points that are within two support units of the best estimate. The evidence against a model without phylogenetic inertia ($t_{1/2} = 0$) is 30.47 support units, and the evidence against a Brownian motion (implying no adaptation) is 13.58 support units. These results are conditional on the best estimate of the instantaneous variance of the independent variable ($\sigma_x^2 = 0.396$).

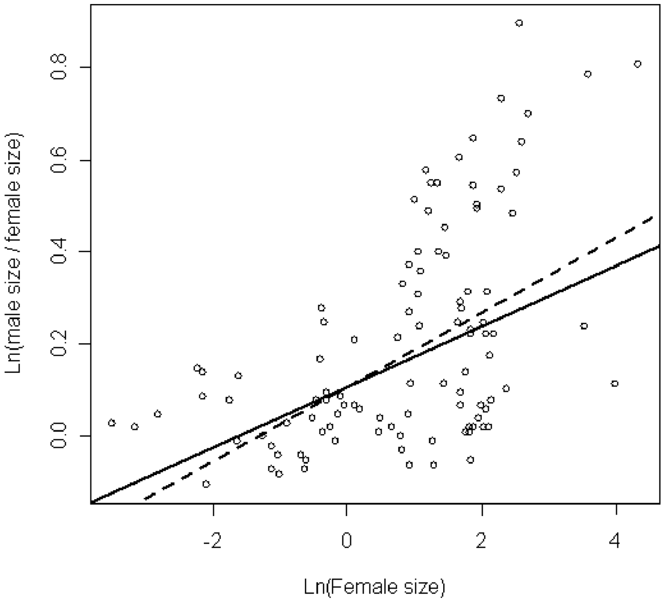

Figure 4. Evolutionary (—) and optimal (- - - -) regressions of sexual dimorphism (ln[male weight/female weight]) on female body size (ln[female weight]) for the primate data from Smith and Cheverud (2002). The optimal regression is $\theta = 0.108 + 0.081x$, and the evolutionary regression is $y = 0.108 + 0.066x$. The variance explained is $R^2 = 13.0\%$.

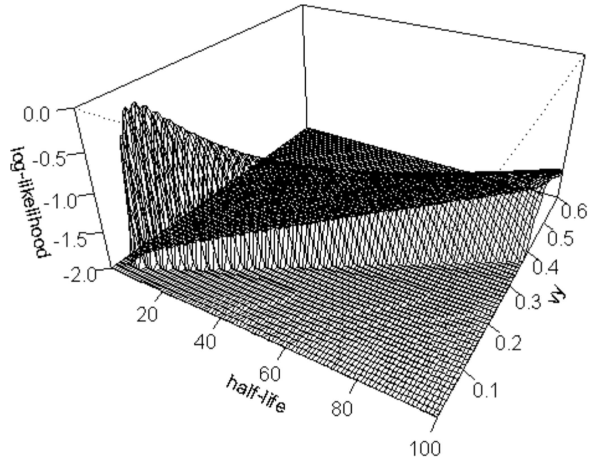

Figure 5. Support surface for the phylogenetic half-life ($t_{1/2}$) and the stationary variance (v_y) for a regression of ln[male weight/female weight] on ln[female weight] for the Catarrhines. The maximum-likelihood estimate is $t_{1/2} = 10.30$ myr and $v_y = 0.068$. The elevated area shows all points that are within two support units of the best estimate. The evidence against a model of no phylogenetic inertia ($t_{1/2} = 0$) is 17.12 support units, and the evidence against Brownian motion (implying no adaptation) is 1.54 support units (the heavy tail extends out to $t_{1/2} = \infty$). These results are conditional on the best estimate of the instantaneous variance of the independent variable ($\sigma_x^2 = 0.043$).

best estimate of $t_{1/2}$ is 10.30 myr, but the two-unit support interval now extends from 4.12 myr to infinity (Fig. 5). Although, we cannot decisively conclude that inertia is stronger for this group than for the entire set of primates, we get a much larger adaptive lag between the evolutionary and the optimal regression, because the Catarrhines have only had 21 myr to diversify (Fig. 6). The best estimate of the optimal regression slope is 0.253 ± 0.106, which indicates that there is selection for a stronger size dependence of sexual size dimorphism in the Catarrhines than in the rest of the primates, but the evolutionary regression is only about 50% of this value. In fact, Rensch's rule predicts a stronger positive relationship within the Catarrhines, because the entire set of primates contains many species in which females are larger than males (see Fig. 4); their inclusion violates the condition of the rule.

PERFORMANCE OF THE METHOD

We tested our estimation algorithm by applying it to datasets created by simulating the evolutionary model with known parameters. The ability to create such simulated datasets is a feature of SLOUCH, as detailed in the SLOUCH manual.

For the full primate phylogeny and for the Catarrhine subclade we generated 100 datasets, and for each we estimated the regression slopes and the phylogenetic half-life. These simulations were based on the best estimates from the original data above. For the full phylogeny the estimates of $t_{1/2}$ appeared almost unbiased and only 11 estimates were (marginally) outside the two-unit support set reported above (Fig. 7). Estimates of the optimal regression slope had a moderate downward bias (although still much less biased than the evolutionary regression), but it is encouraging that all but three estimates were positive despite the small true value.

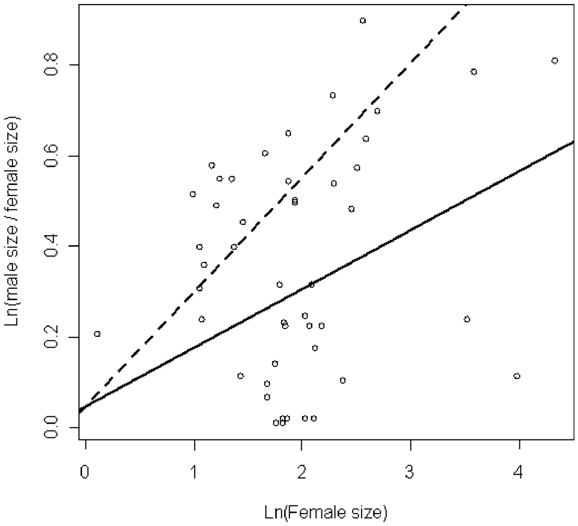

Figure 6. Evolutionary (—) and optimal (- - - -) regressions of sexual dimorphism on female body size for the Catarrhines. The optimal regression is $\theta = 0.046 + 0.253x$, and the evolutionary regression is $y = 0.048 + 0.130x$. The variance explained is $R^2 = 10.8\%$.

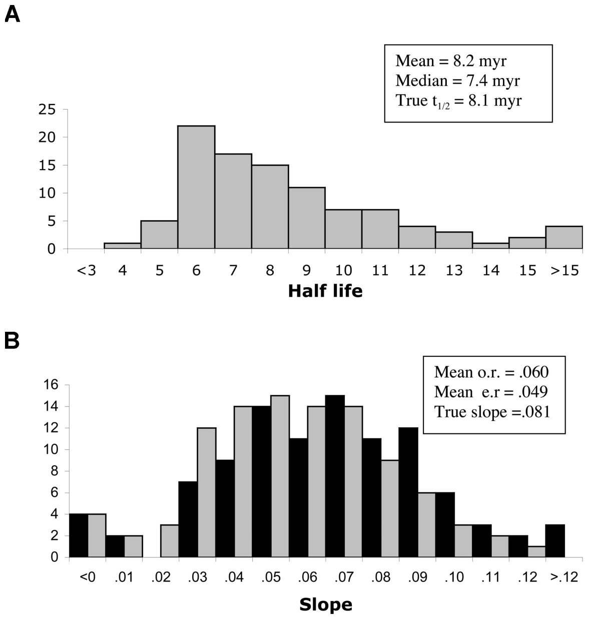

Figure 7. Distribution of estimates from simulations on the primate phylogeny: Simulations based on the best-estimated parameters. (A) The phylogenetic half-life. (B) In black, the optimal regression (o.r.) slope. In gray, the evolutionary regression (e.r.) slope.

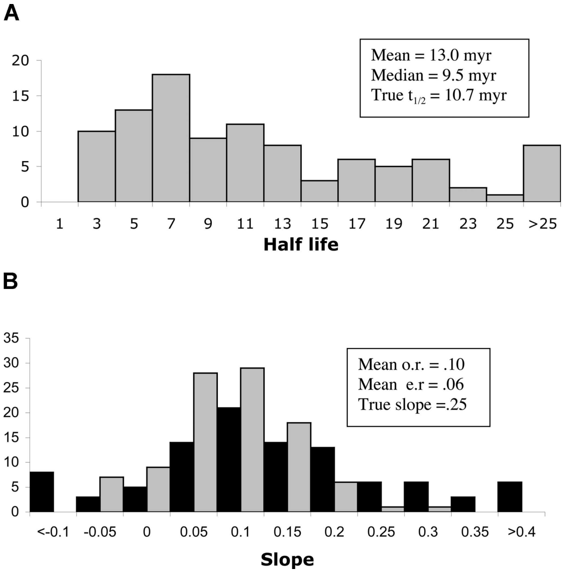

Figure 8. Distribution of estimates from simulations on the Catarrhine phylogeny: Simulations based on the best-estimated parameters. (A) The phylogenetic half-life. (B) In black, the optimal regression (o.r.) slope. In gray, the evolutionary regression (e.r.) slope.

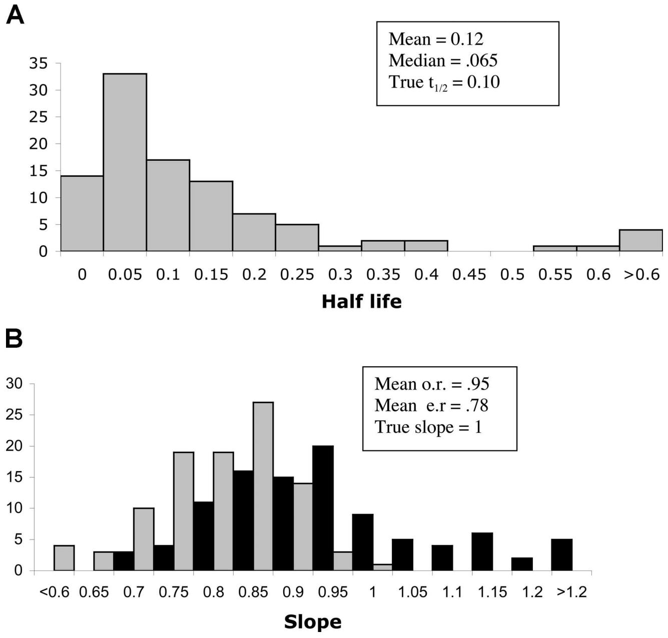

Figure 9. Distribution of estimates from simulations with weak inertia on a symmetric 64-species phylogeny: Parameter values for simulations are $\sigma_x^2 = 1$, $v_y = 0.1$, $t_{1/2} = 0.1$, $b_0 = 0$, $b_1 = 1$, ancestral values of *x* and *y* are zero, and total length of phylogeny is one. (A) The phylogenetic half-life. (B) In black, the optimal regression (o.r.) slope. In gray, the evolutionary regression (e.r.) slope.

The results for the Catarrhine subclade were similar, but less accurate (Fig. 8). Estimates of the half-life had slight positive bias with 10 estimates outside the support set, but estimates were always larger than zero, and the largest value was 77 myr, even though the support set included infinity. Estimates of the optimal regression were downwardly biased, and in this case 16 estimates were negative. Estimates of the other parameters, such as the regression intercept and the v_y, were more accurate (not shown).

We also created simulated data for a symmetric 64-species phylogeny with equal branch lengths. In this case estimates of the optimal regression slope were nearly unbiased for $t_{1/2} = 0.1$ (Fig. 9), and $t_{1/2} = 0.5$ (Fig. 10). This suggests that the bias we observed on the primate phylogenies may be due to their asymmetries. In particular, both the full primate phylogeny and Catarrhine subclade have a few relatively long branches from the base that could make the simulated data sensitive to random events along these branches. The regression within the clades stemming from these long basal branches will be shallower than the predicted regression across the whole dataset. If the phylogeny is asymmetric, the estimate of the regression slope may be dominated by the shallow regression through the larger subclade.

The simulations on the symmetric phylogeny showed good performance of the estimators of half-life and regression slopes with weak phylogenetic inertia ($t_{1/2} = 0.1$, Fig. 9), but performance was less good with strong phylogenetic inertia ($t_{1/2} = 0.5$, Fig. 10). In the latter case, there was an upward bias in the estimates of the half-life, and the estimates of the optimal regression slope were imprecise. Qualitative conclusions would still be reliable, as the method invariably reported strong phylogenetic inertia and a positive regression slope.

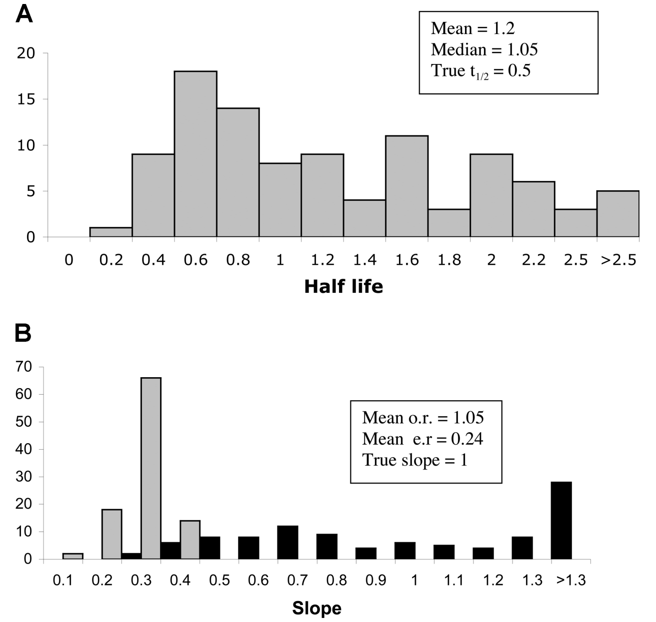

Figure 10. Distribution of estimates from simulations with strong inertia on a symmetric 64-species phylogeny: Parameter values for simulations are $\sigma_x^2 = 1$, $v_y = 0.1$, $t_{1/2} = 0.5$, $b_0 = 0$, $b_1 = 1$, ancestral values of x and y are zero, and total length of phylogeny is one. (A) The phylogenetic half-life. (B) In black, the optimal regression (o.r.) slope. In gray, the evolutionary regression (e.r.) slope.

We conclude that the method performs reasonably well on larger phylogenies, but estimates of phylogenetic inertia are rather inaccurate, and especially so when there is strong inertia. Thus, estimates of $t_{1/2}$ and the optimal regression slopes should be interpreted with caution. We suggest that only qualitative conclusions should be made for phylogenies with less than, say, 20–30 species, depending on the data quality and the shape of the phylogeny. Instead of relying solely on the best estimate of the half-life, we suggest doing the analysis conditionally on a set of biologically reasonable choices of $t_{1/2}$, and discussing the results as if-then propositions; if strong inertia, then conclusion A, but if weak inertia then conclusion B, etc. It is also important to consider phylogeny shape when judging the results. In cases with few widely separated clades, we recommend doing the analysis separately for each clade, because the evolutionary process may differ in the different clades, and because the long branches separating the clades may bias the results if inertia is strong.

Discussion

The standard test of an adaptive hypothesis is to see if the species' trait values fit the prediction from an optimality model. The idea behind the adaptation-inertia model is to develop a comparative method that is consistent with this idea at the same time as it controls for inertia. At the core of the method is a simple regression of a "primary" adaptive optimum, θ, on one or more predictor variables, *x*. If the adaptive hypothesis predicts that the optimum should depend on *x*, this can be tested by estimating the regression of θ on *x*. This regression is not the same as the "evolutionary regression" of species means, *y*, on *x*. The evolutionary regression is expected to be shallower than the "optimal regression" when phylogenetic inertia is present. To this extent, not accounting for inertia could lead one to falsely conclude that the prediction of an optimality hypothesis is not consistent with the data. Although there are many comparative methods that can account for phylogenetic correlations between species, the adaptation-inertia method is the only one that can provide a relatively unbiased estimate of the underlying optimal relationship in the presence of inertia, and estimate how far species lag behind their predicted optima.

The method provides several refinements in the estimation of phylogenetic inertia. One of these is that phylogenetic inertia is estimated jointly with adaptation. Just as an analysis of adaptation must control for inertia, an analysis of inertia must control for the effects of adaptation (Orzack and Sober 2001). If predictor variables also display phylogenetic effects, adaptation automatically generates phylogenetic effects in the response variables. If these effects are estimated without controlling for the adaptive relationship, they may falsely suggest phylogenetic inertia in the response trait. Accordingly, phylogenetic effects estimated independently of the adaptive hypothesis cannot be used to "correct" for phylogeny when testing this hypothesis.

As reviewed in Blomberg and Garland (2002) and Hansen and Orzack (2005), various phylogenetic autoregression methods (e.g., Cheverud et al. 1985; Diniz-Filho et al. 1998), and the mixed-model method of Lynch (1991) can provide controlled estimates of phylogenetic effects, although they are rarely used in this way. In particular, Lynch's phylogenetic heritability is a logical measure of phylogenetic effects. Freckleton et al. (2002) and Housworth et al. (2004) show how it can be estimated when evolution follows a Brownian-motion process. Phylogenetic heritability is a measure of phylogenetic effects without regard to the underlying cause of the effects, which may well be adaptation. In contrast, the phylogenetic half-life is a measure of inertia relative to adaptation toward an optimum, and should be used when adaptation involving stabilizing selection is studied. The phylogenetic half-life may also be used to estimate general phylogenetic effects by including only a single, fixed optimum.

Our adaptation-inertia model assumes that the predictor variables follow a Brownian-motion process. How restrictive is this assumption? We first note that if there are no phylogenetic effects in the predictor variables, then there can be no phylogenetic inertia in the response trait, because any adaptation must happen instantaneously on the relevant time scale. Thus, the adaptation-inertia model is not suitable if the phylogenetic effect in the predictor variable is much weaker than the phylogenetic inertia in the response trait (relative to the predictor variable). We expect, however, that the model is robust toward variation in the strength of phylogenetic effects in the predictors. One could also develop models in which the predictor variables change as if by an Ornstein-Uhlenbeck process. Whether and when a such increase in model complexity will improve inferences about adaptation and inertia remains to be seen. A more important restriction is the assumption of one-way causation from predictor to response. If the response trait also influence the predictor variable, we get a qualitatively different process. Bivariate Ornstein–Uhlenbeck models of this type have been considered in Hansen and Martins (1996) and Martins et al. (2002).

There is a common belief that phylogenetic comparative methods in general, and the method of independent contrasts in particular, are robust against violations of their assumptions, either in terms of error in the phylogeny or in terms of departures from the assumed process model. This is supported by several simulation studies (e.g., Martins and Garland 1991; Martins et al. 2002). It is important to realize that these simulation studies only consider robustness toward error in the phylogenetic covariance structure, and not toward error in the mean-structure of the model. The relative robustness toward error in the variance structure stems from the robustness of GLS, which is unbiased even when the variance structure is misspecified. The precision of estimates is affected, but with the moderate error levels generally considered in simulation studies, the effects are not dramatic (Rohlf 2001, 2006). In the simulation study most relevant to our model, Martins et al. (2002) studied the performance of several comparative methods for data generated by correlated stochastic departures from the fixed optimum of a bivariate Ornstein–Uhlenbeck process. This model describes correlated evolution of two traits around a joint optimum, and not adaptation of one trait to the other. To this extent, it is not surprising that the method of independent contrasts, which is designed for correlated evolution (Felsenstein 1985), showed reasonable performance. We reiterate that inertia not only affects phylogenetic covariances, but also generates a lag in adaptation that can have dramatic consequences for the interpretation of the results and the estimation of optimal states. This lag is not accounted for by the method of independent contrasts.

The time dependency of our phylogenetic correction factor, $\rho(\alpha t)$, sheds light on the classical problem of explaining why allometric slopes are typically shallower at lower taxonomic levels. Martin and Harvey (1985) found that the slope of mammalian brain-body size allometries are much more shallow on species and genus levels than on family and order level. There are two main explanations for this phenomenon. It has been argued to be a statistical artifact, resulting from higher levels of measurement error at lower taxonomic levels (e.g., Pagel and Harvey 1988). In contrast, Riska (1991) argued that real biology is responsible in that biological "error," that is biological deviances from the allometric slope, are likely to swamp the effects of measurement error. In the case of the brain-body size allometry, biological explanations have centered on qualitatively different patterns of evolution at different levels. For example, it has been argued that body-size differences at low taxonomic levels are due to changes in postnatal growth, which do not affect brain growth (Lande 1979; Riska 1989). Although phylogenetic inertia, or evolutionary lag, has also been suggested as an explanation (e.g., Burt 1989), it has not received the attention it deserves. Our model shows that inertia makes the evolutionary regression slope more shallow when less time has been available for adaptation. We note that Burt (1989) and Deaner and Nunn (1999) have proposed methods to test for phylogenetic inertia based upon how regression slopes from pairwise species contrasts change with the phylogenetic distance of those contrasts.

Simulations and the observed support surfaces from our examples show that accurate estimates of inertia require many species. For less than, say, 20–30 species, only a qualitative assessment will be possible. This does not mean that the method should be avoided with smaller phylogenies, only that conclusions about inertia and the optimal regression must be made carefully. Even with smaller sample sizes, estimates of the evolutionary regression should not be worse than those provided by alternative comparative methods. With smaller datasets it may be prudent to put less emphasis on the best estimate of the phylogenetic half-life, and base conclusions on a range of plausible half-lives. Regardless of method one should bear in mind that the evolutionary regression may be a downwardly biased estimate of the adaptive optimum.

The OUCH program (Butler and King 2004) has increased use of the adaptation-inertia method (Hansen 1997; Hansen et al. 2000; Butler and King 2004; Valiente-Banuet et al. 2006; Verdu and Gleiser 2006; Gomez and Thery 2007; Hipp 2007; Whittall and Hodges 2007; Labra et al., unpubl. ms.; see also Garland et al. 1993; Butler et al. 2000; Martins et al. 2004; Ord and Martins 2006 for use of Ornstein–Uhlenbeck processes to model residual variance). A simplified version of the method, allowing only a single shift in the optimum, can also be fit with the software package COMPARE (Martins 2004). The inclusion of information-based model selection has made the method more flexible in evaluating hypotheses about phylogenetic niche structure (Butler and King 2004), and in this article we have shown how to analyze random continuous predictor variables in addition to fixed categorical ones. We hope that these developments and the associated software, SLOUCH, will make the approach applicable and available for use in a wider range of circumstances. An application to thermophysiological adaptation in lizards is presented.

ACKNOWLEDGMENTS

JP was supported by NSF DEB-0344417 to David Houle and Thomas Hansen. SO was partially supported by NSF SEI 0513910, NIH R01 DA015789-01A2, and NIA P01-AG0225000-01. We thank M. Butler, A. Labra, and an anonymous reviewer for discussions and comments on the manuscript.

LITERATURE CITED

Abouheif, E., and D. J. Fairbairn. 1997. A comparative analysis of allometry for sexual size dimorphism: assessing Rensch's rule. Am. Nat. 149:540–562.

Arnold, S. J., M. E. Pfrender, and A. G. Jones. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. Genetica 112/113:9–32.

Ashton, K. G. 2004. Comparing phylogenetic signal in intraspecific and interspecific body size data sets. J. Evol. Biol. 17:1157–1161.

Blomberg, S. P., and T. Garland. 2002. Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. J. Evol. Biol. 15:899–910.

Blomberg, S. P., T. Garland, and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717–745.

Burnham, K. P., and D. R. Anderson. 2004. Model selection and inference: a practical information—theoretic approach. Sec. Ed. Springer, New York.

Burt, A. 1989. Comparative methods using phylogenetically independent contrasts. Pp. 33–53 *in* P. H. Harvey and L. Partridge, eds. Oxford surveys in evolutionary biology. Vol. 6. Oxford Univ. Press, Oxford.

Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modelling approach for adaptive evolution. Am. Nat. 164:683–695.

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

Cheverud, J. M., M. M. Dow, and W. Leutenegger. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses—sexual dimorphism in body-weights among primates. Evolution 39:1335–1351.

Deaner, R. O., and C. L. Nunn. 1999. How quickly do brains catch up with bodies? A comparative method for detecting evolutionary lag. Proc. R. Soc. Lond. B 266:687–694.

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

Edwards, A. W. F. 1992. Likelihood: expanded edition. Johns Hopkins, Baltimore.

Estes, S., and S. J. Arnold. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. Am. Nat. 169:227–244.

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 126:1–25.

———. 1988. Phylogenies and quantitative characters. Annu. Rev. Ecol. Syst. 19:445–471.

Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. Am. Nat. 160:712–726.

Gard, T. C. 1988. Introduction to stochastic differential equations. Marcel Dekker, Inc.

Garland, T. Jr., A. W. Dickerman, C. M. Janis, and J. A. Jones. 1993. Phylogenetic analysis of covariance by computer simulation. Syst. Biol. 42:265–292.

Gomez, D., and M. Thery. 2007. Simultaneous crypsis and conspiciousness in color patterns: comparative analysis of a Neotropical rainforest bird community. Am. Nat. 169: S42–S61.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351.

Hansen, T. F., and E. P. Martins. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50:1404–1417.

Hansen, T. F., and S. H. Orzack. 2005. Assessing current adaptation and phylogenetic inertia as explanations of trait evolution: the need for controlled comparisons. Evolution 59:2063–2072.

Hansen, T. F., W. S. Armbruster, and L. Antonsen. 2000. Comparative analysis of character displacement and spatial adaptations as illustrated by the evolution of *Dalechampia* blossoms. Am. Nat. 156: S17–S34.

Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford Univ. Press, New York.

Hipp, A. L. 2007. Nonuniform processes of chromosome evolution in sedges (*Carex*: Cyperaceae). Evolution 61:2175–2194.

Housworth, E. A., E. P. Martins, and M. Lynch. 2004. The phylogenetic mixed model. Am. Nat. 163:84–96.

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

———. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33:402–416.

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

Martin R. D., and P. H. Harvey. 1985. Brain size allometry: ontogeny and phylogeny. Brain size allometry: ontogeny and phylogeny. Pp. 147–173 *in* W. L. Jungers, ed. Size and scaling in primate biology. Plenum, New York.

Martins, E. P. 2000. Adaptation and the comparative method. Trends Ecol. Evol. 15:296–299.

———. 2004. COMPARE, version 4.6b. Computer programs for the statistical analysis of comparative data. Distributed by the author at http://compare.bio.indiana.edu/. Department of Biology, Indiana University, Bloomington, IN.

Martins, E. P., and T. Garland. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. Evolution 45:534–557.

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

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

Martins, E. P., J. A. F. Diniz-Filho, and E. A. Housworth. 2002. Adaptive constraints and the phylogenetic comparative method: a computer simulation test. Evolution 56:1–13.

Martins, E. P., A. Labra, M. Halloy, and J. T. Thompson. 2004. Large-scale patterns of signal evolution: an interspecific study of *Liolaemus* lizard headbob displays. Anim. Behav. 68:453–463.

O'Meara, B. C., C. Ane, M. J. Sanderson, and P. C. Wainwright 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922–933.

Ord, T. J., and E. P. Martins. 2006. Tracing the origins of signal diversity in anole lizards: phylogenetic approaches to inferring the evolution of complex behaviour. Anim. Behav. 71:1411–1429.

Orzack, S. H., and E. Sober. 2001. Adaptation, phylogenetic inertia, and the method of controlled comparisons. Pp. 45–63 *in* S. H. Orzack and E. Sober, eds. Adaptationism and optimality. Cambridge Univ. Press, New York.

Pagel, M. D., and P. H. Harvey. 1988. The taxon-level problem in the evolution of mammalian brain size: facts and artifacts. Am. Nat. 132:344–359.

Plavcan, J. M., and C. P. Van Schaik. 1992. Intrasexual competition and canine dimorphism in Anthropoid Primates. Am. J. Phys. Anthro. 87:461–477.

Rensch, B. 1959. Evolution above the species level. Columbia Univ. Press, New York.

Ridley, M. 1983. The explanation of organic diversity: the comparative method and adaptations for mating. Oxford Univ. Press, New York.

Riska, B. 1989. Composite traits, selection response, and evolution. Evolution 43:1172–1191.

———. 1991. Regression models in evolutionary allometry. Am. Nat. 138:283–299.

Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. Evolution 55:2143–2160.

———. 2006. A comment on phylogenetic correction. Evolution 60:1509–1515.

Simpson, G. G. 1944. Tempo and mode in evolution. Columbia Univ. Press, New York.

Smith R. J., and J. M. Cheverud. 2002. Scaling of sexual dimorphism in body mass: a phylogenetic analysis of Rensch's rule in primates. Int. J. Primatol. 23:1095–1135.

Sober, E., and S. H. Orzack. 2003. Common ancestry and natural selection. Brit. J. Philos. Sci. 54:423–437.

Thoren, S., P. Lindenfors, and P. M. Kappeler. 2006. Phylogenetic analyses of dimorphism in primates: evidence for stronger selection on canine size than on body size. Am. J. Phys. Anthro. 130:50–59.

Valiente-Banuet, A., A. V. Rumbe, M. Verdu, and R. M. Callaway. 2006. Modern quaternary plant lineages promote diversity through facilitation of ancient tertiary lineages. Proc. Natl. Acad. Sci. USA 103:16812–16187.

Verdu, M., and G. Gleiser. 2006. Adaptive evolution of reproductive and vegetative traits driven by breeding system. New Phytol. 169:409–417.

Whittall, J. B., and S. A. Hodges. 2007. Pollinator shifts drive increasingly long nectar spurs in columbine flowers. Nature 447:706–709.

Associate Editor: K. Crandall

Appendix

THE OU-BM MODEL

Let $y(t)$ denote the response trait at time t, and let $\theta(t)$ denote the primary optimum of y at time t. We assume that $y(t)$ and $\theta(t)$ obey the stochastic differential equations

$$\begin{aligned} dy/dt &= -\alpha(y-\theta(t)) + \sigma_y\, dW_y/dt, \\ d\theta/dt &= \sigma_\theta\, dW_\theta/dt, \end{aligned}$$

where dW_y and dW_θ are independent white noise processes, α measures the rate of adaptation to the optimum, σ_y measures the size of stochastic perturbations of y, and σ_θ measures the size of stochastic perturbations of θ. Differential equations for the first and second moments of this process can be derived by using Ito's formula to obtain differentials for the power terms and then taking expectations to yield

$$\begin{aligned} dE[y]/dt &= -\alpha(E[y] - E[\theta]), \\ dE[\theta]/dt &= 0, \\ dE[y^2]/dt &= -2\alpha E[y^2] + 2\alpha E[y\theta] + \sigma_y^2, \\ dE[y\theta]/dt &= -\alpha E[y\theta] + \alpha E[\theta^2], \\ dE[\theta^2]/dt &= \sigma_\theta^2, \end{aligned}$$

with initial conditions $y(0) = y_0$ and $\theta(0) = \theta_0$. The solution to this set of ordinary differential equations is

$$\begin{aligned} E[y](t) &= \theta_0(1-e^{-\alpha t}) + y_0 e^{-\alpha t}, \\ E[\theta](t) &= \theta_0, \\ E[y^2](t) &= \sigma_\theta^2 t + ((y_0-\theta_0)^2 - \left(\sigma_\theta^2+\sigma_y^2\right)/2\alpha)e^{-2\alpha t} \\ &\quad +2\left(y_0\theta_0 + \sigma_\theta^2/\alpha - \theta_0^2\right)e^{-\alpha t} + \theta_0^2 + \sigma_y^2/2\alpha - 3\sigma_\theta^2/2\alpha, \\ E[y\theta](t) &= \sigma_\theta^2 t + \left(\theta_0^2 - \sigma_\theta^2/\alpha\right)(1-e^{-\alpha t}) + y_0\theta_0 e^{-\alpha t}, \\ E[\theta^2](t) &= \sigma_\theta^2 t + \theta_0^2. \end{aligned}$$

From these equations we can derive variances and covariances,

$$\begin{aligned} \mathrm{Var}[y](t) &= ((\sigma_\theta^2+\sigma_y^2)/2\alpha)(1-e^{-2\alpha t}) \\ &\quad +\sigma_\theta^2 t(1-2(1-e^{-\alpha t})/\alpha t), \\ \mathrm{Cov}[y,\theta](t) &= \sigma_\theta^2 t(1-(1-e^{-\alpha t})/\alpha t), \\ \mathrm{Var}[\theta](t) &= \sigma_\theta^2 t. \end{aligned}$$

Because the stochastic differential equations are linear, we know that the distribution of the two variables at time t will be a bivariate normal (Gard 1988). Therefore the first and second moments completely specify the distribution. From these moments we can compute the regression of the trait on the optimum as

$$\begin{aligned} E[y|\theta](t) &= \theta_0(1-e^{-\alpha t}) + y_0 e^{-\alpha t} \\ &\quad +(1-(1-e^{-\alpha t})/\alpha t)(\theta-\theta_0). \end{aligned}$$

The regression coefficient is $(1-(1-e^{-\alpha t})/\alpha t)$, and because the true relationship is 1, we see that there is a bias that decreases with time.

COVARIANCE BETWEEN SPECIES

Hansen and Martins (1996) showed that the covariance between two species that diverged at time t_a and evolved independently thereafter can be expressed as

$$\mathrm{Cov}[y_i, y_j] = \mathrm{Cov}[E[y_i|y_a], E[y_j|y_a]],$$

where y_a is the trait value of the most recent common ancestor of species i and j. This assumes that the species evolve independently after they diverged. The conditional expectations are such that

$$\begin{aligned} \mathrm{Cov}[y_i, y_j] &= \mathrm{Var}\left[\theta_a\left(1-e^{-\alpha t_{ij}/2}\right) + y_a e^{-\alpha t_{ij}/2}\right] \\ &= (1-e^{-\alpha t_{ij}/2})^2\mathrm{Var}[\theta_\mathrm{a}] + e^{-\alpha t_{ij}}\mathrm{Var}[y_a] \\ &\quad + 2(e^{-\alpha t_{ij}/2} - e^{-\alpha t_{ij}})\mathrm{Cov}[y_a, \theta_a], \end{aligned}$$

where t_{ij} is the time separating the two species (assuming they are both extant). Using the expressions from above we get

$$\begin{aligned} \mathrm{Cov}[y_i, y_j] &= \left(\left(\sigma_\theta^2+\sigma_y^2\right)/2\alpha\right)(1-e^{-2\alpha t_a})e^{-\alpha t_{ij}} \\ &\quad + \sigma_\theta^2 t_a\left(1-2e^{-\alpha t_{ij}/2}(1-e^{-\alpha t_a})/\alpha t_a\right), \end{aligned}$$

where t_a is time from the base of the phylogeny to the most recent common ancestor of the two species. This covariance has two components, one due to variation in the ancestral state of the trait, and another due to variation in the ancestral position of the optimum. If there are unequal branch lengths from root to the tip species, the covariance is

$$\begin{aligned}\mathrm{Cov}[y_i, y_j] = {} & \left(\left(\sigma_\theta^2 + \sigma_y^2\right)/2\alpha\right)(1 - e^{-2\alpha ta})e^{-\alpha tij} \\ & + \sigma_\theta^2 t_a\left(1 - (e^{-\alpha tia} + e^{-\alpha tja})(1 - e^{-\alpha ta})/\alpha t_a\right),\end{aligned}$$

where t_{ia} is time from species i to the most recent common ancestor of the two.

COVARIANCE BETWEEN RESIDUALS

A GLS analysis requires the use of the covariances between residuals from the regression. That is

$$\mathrm{Cov}[r_i, r_j] = \mathrm{Cov}[y_i - E[y_i \mid \theta_i], y_j - E[y_j \mid \theta_j]]$$

Using the expression for the regression, $E[y_i \mid \theta_i]$, derived above, we compute

$$\begin{aligned}\mathrm{Cov}[r_i, r_j] = {} & \left(\frac{\sigma_\theta^2 + \sigma_y^2}{2\alpha}\right)(1 - e^{-2\alpha t_a})e^{-\alpha t_{ij}} \\ & + \sigma_\theta^2 t_a \left(\frac{1 - e^{-\alpha t_i}}{\alpha t_i}\frac{1 - e^{-\alpha t_j}}{\alpha t_j} - \frac{1 - e^{-\alpha t_a}}{\alpha t_a}\right. \\ & \left.\times\left(e^{-\alpha t_{ia}}\frac{1 - e^{-\alpha t_j}}{\alpha t_j} + e^{-\alpha t_{ja}}\frac{1 - e^{-\alpha t_i}}{\alpha t_i}\right)\right),\end{aligned}$$

where t_i means time from the base of the phylogeny to species i. The special case when all species have the same elapsed time to the base of the phylogeny is given in the main text.

Measurement variance of the response variables can be added to the residual variance. Measurement variance of a predictor variable is multiplied by the square of the regression parameter on that variable and then added to the residual variance.

MODELING THE OPTIMUM

We assume that the optimum is a linear combination of predictor variables, such that $\theta = b_0 + \sum_i b_i x_i$. If each of the x_i evolves as if by an independent Brownian-motion process, then θ evolves as if by a Brownian-motion process with $\sigma_\theta^2 = \sum_i b_i^2 \sigma_{xi}^2$. If the goal is to recover the b_i through a regression of y on the x's, then we use

$$\begin{aligned}E[y|x_1, x_2, \ldots] = {} & e^{-\alpha t} y_a + (1 - e^{-\alpha t})\left(b_0 + \sum\nolimits_i b_i x_{ia}\right) \\ & + \left(1 - \frac{1 - e^{-\alpha t}}{\alpha t}\right)\sum\nolimits_i b_i(x_i - x_{ia}),\end{aligned}$$

which is a standard linear regression except that each regression coefficient is scaled by a function of elapsed time. The parameters y_a and x_{ia} are the ancestral (i.e., root) values of the y and x variables; the terms containing these and the optimal intercept, b_0, are absorbed into the general intercept, k in the main text. The residual variances of the regression are as given above, but substituting $\sum_i b_i^2 \sigma_{xi}^2$ for σ_θ^2.

If the predictor variables are not independent, but evolve as a multivariate Brownian-motion process with variance matrix $\mathbf{M}$, then θ evolves as if by a Brownian-motion process with $\sigma_\theta^2 = \mathbf{b}^T\mathbf{M}\mathbf{1}$, where $\mathbf{b}$ is the vector of regression variables, and $\mathbf{1}$ is a vector of ones. To estimate parameters, we would estimate the variance of the predictor variables (i.e., $\mathbf{M}$), and use this to predict σ_θ^2.