

Analyzing Disparity and Rates of Morphological Evolution with Model-Based Phylogenetic Comparative Methods

THOMAS F. HANSEN^{1,*}, GEIR H. BOLSTAD², AND MASAHITO TSUBOI^{1,3}

¹Department of Biology, CEES & Evogene, University of Oslo, Oslo, Norway; ²Norwegian Institute for Nature Research (NINA), NO-7485 Trondheim, Norway, and ³Department of Biology, Lund University, Lund, Sweden

*Correspondence to be sent to: Department of Biosciences, University of Oslo, P. O. Box 1066, Blindern, 0316 Oslo, Norway;
Email: thomas.hansen@bio.uio.no.

Received 07 July 2021; reviews returned 16 September 2021; accepted 23 September 2021
Associate Editor: Simon Ho

Abstract.—Understanding variation in rates of evolution and morphological disparity is a goal of macroevolutionary research. In a phylogenetic comparative methods framework, we present three explicit models for linking the rate of evolution of a trait to the state of another evolving trait. This allows testing hypotheses about causal influences on rates of phenotypic evolution with phylogenetic comparative data. We develop a statistical framework for fitting the models with generalized least-squares regression and use this to discuss issues and limitations in the study of rates of evolution more generally. We show that the power to detect effects on rates of evolution is low in that even strong causal effects are unlikely to explain more than a few percent of observed variance in disparity. We illustrate the models and issues by testing if rates of beak-shape evolution in birds are influenced by brain size, as may be predicted from a Baldwin effect in which presumptively more behaviorally flexible large-brained species generate more novel selection on themselves leading to higher rates of evolution. From an analysis of morphometric data for 645 species, we find evidence that both macro- and microevolution of the beak are faster in birds with larger brains, but with the caveat that there are no consistent effects of relative brain size. [Baldwin effect; beak shape; behavioral drive; bird; brain size; disparity; phylogenetic comparative method; rate of evolution.]

Rates of evolution are highly variable. While some species and traits remain nearly unchanged for tens of millions of years, others undergo dramatic shifts in mere dozens of generations (Simpson 1944; Eldredge and Gould 1972; Hendry and Kinnison 1999; Schluter 2000; Gould 2002; Hendry et al. 2008; Uyeda et al. 2011; Hunt and Rabosky 2014; Gingerich 2019; Reznick et al. 2019; Voje et al. 2020; Zelditch et al. 2020). Explaining these differences is a key challenge for a theory of macroevolution (e.g., Hansen and Houle 2004; Eldredge et al. 2005; Estes and Arnold 2007; Bell 2012; Hansen 2012; Arnold 2014; Jablonski 2017a,b). Yet, the majority of work on rates of trait evolution is descriptive and lacking in quantitative hypotheses and statistical methods derived from first principles.

While substantial progress may have to wait for a more mature quantitative theory of macroevolutionary change, it is important to move beyond mere description toward testing hypotheses about factors influencing rates of evolution. Model-based phylogenetic comparative methods provide a framework for such tests. The models that are used to translate phylogenetic relationships into the statistical covariances needed for statistical analysis usually depend on parameters related to rates of evolution. In particular, the commonly used Brownian-motion model includes trait-specific rates of “diffusion” that are estimated as a part of the analysis. These can be used to estimate rates of evolution in different traits or directions of morphospace (e.g., Felsenstein 1985; Lynch 1991; Adams 2013). Methods have been developed to estimate the Brownian diffusion rate in different parts of the phylogeny, and this allows testing for clade differences in rates of evolution (Garland 1992; Martins 1994; McPeck 1995; Gittleman

et al. 1996; O’Meara et al. 2006; Thomas et al. 2006; Revell 2008; Revell and Harmon 2008; Eastman et al. 2011; Revell et al. 2011; Beaulieu et al. 2012; Adams 2014; Fuentes et al. 2016; Weir and Lawson 2016; Caetano and Harmon 2019; May and Moore 2020). Such approaches make it possible to obtain and compare estimated rates of evolution from standard comparative data (e.g., Pie and Meyer 2017; Alhajeri and Steppan 2018; Chira et al. 2018; Revell et al. 2018). Estimated rates can then be used as data in further analyses to test hypotheses about causes of rate differences, but this approach entails statistical and interpretational difficulties in dealing with estimation error and linking rates with explanatory variables. What is lacking are methods for combining rates of evolution with explanatory variables that are themselves treated as evolving entities in a common model. This would allow direct estimation of causal links in a statistical framework that accounts for stochasticity in both response and predictor variables.

Here, we explore some stochastic-process models for estimating the relationship between rates of evolution in a quantitative trait, formalized either as the diffusion parameter of a Brownian motion or as the variance of recent deviations from a phylogenetic prediction, and potential explanatory variables modeled as evolving quantitative traits on the phylogeny. We illustrate the models with an analysis of rates of evolution in bird beaks based on data from Cooney et al. (2017) and Chira et al. (2018). We relate these rates to measures of absolute and relative brain sizes, thus revisiting the classical analyses of Wyles et al. (1983) to test whether evolution is speeded up by increased behavioral flexibility due to a Baldwin effect (e.g., Baldwin 1896; Popper 1972; Wilson 1985). We argue that the classical tests of this hypothesis

were inadequate, but that current data based on orders of magnitude more species, accurate phylogenies, and better statistical approaches now make it feasible to test such hypotheses.

THEORY

Models for Testing Macroevolutionary Hypotheses about Rates

In contrast to standard models used in comparative methods for continuous traits, most of our models will not be Gaussian, meaning that the resulting distributions of the traits will not be normal (Gaussian). This complicates statistical analyses as standard statistics based on normal distributions cannot be used. Additionally, the joint probability distributions are hard to derive, and this rules out exact likelihood or Bayesian approaches. It is possible, however, to characterize the distributions by deriving moments. This allows the use of methods-of-moments estimators and generalized least squares, which do not require exact distributional assumptions. Approximate and numerical likelihood approaches are also possible but will not be explored here.

For each model, we will start by deriving moments of the trait distributions after a time interval t . We then use this to motivate a regression-based estimator of the influence of a predictor x on the rate of evolution of a trait y . Because the state of a Brownian motion is uncorrelated with its own linear deviation from the starting point, this will take the form of a regression of the squared trait deviation on the predictor. We then derive the variances and covariances of the relevant functions for phylogenetically related species, and use this to characterize the residual variance matrix for the regression. This forms the basis for generalized least-squares (GLS) estimation. We present three models that are all based on the assumption that the variance of the evolutionary rate is linear in the predictor. This assumption helps generate approximately linear regressions of y^2 on x and will often be a reasonable choice because independent components act additively on the variance scale. Some microevolutionary models also predict approximate additive effects of microevolutionary parameters on variances of change (Bolstad et al. 2014).

Model 1: Predictor Evolves as a Brownian motion

In this model, the response variable, y , follows a Brownian motion with a rate variance that is a linear function of the predictor variable, x , which itself follows a standard Brownian motion with constant variance, σ^2 . The model is

$$\begin{aligned} dy &= \sqrt{a+bx}dW_1, \\ dx &= \sigma dW_2, \end{aligned} \quad (1)$$

where the dW_i are uncorrelated white noise, and we assume that the initial conditions are $y_0 = x_0 = 0$. The influence of x on the rate of evolution in y is characterized by the parameter b , and the goal is to find an estimator of b . This model breaks down when $a+bx$ becomes negative, and should be seen as an approximation that may be valid for the range of x values spanned by the species to be used in the analysis.

As there is no causal influence of y on the evolution of x , the unconditional distribution of x is Gaussian with mean equal to the initial condition, x_0 , and variance equal to $\sigma^2 t$, where t is the time from the root to the tip of the phylogeny. The moments of y and the cross-moments between y and x are derived in Appendix A.1 by stochastic integration. This reveals that y and x are indeed uncorrelated, but the squared deviation of y from its mean is correlated with x as

$$\text{Cov}[y^2, x] = \frac{b\sigma^2 t^2}{2}, \quad (2)$$

which suggests that a regression of y^2 on x can be used to estimate the b -parameter. In Appendix B.1, we show that the best linear predictor of the species vector of y^2 from the species vector of the x -variable is

$$(a - v_y)\mathbf{1} + b\left(\mathbf{I} - \frac{1}{2}(\mathbf{T} \circ \mathbf{T})\mathbf{T}^{-1}\right)\mathbf{x}, \quad (3)$$

where \mathbf{x} is the column vector of species predictor variables, $\mathbf{1}$ is a column vector of ones, \mathbf{I} is the identity matrix, \mathbf{T} is the matrix of shared branch lengths between all the species pairs, \circ denotes Hadamard (i.e., elementwise) multiplication, and v_y is the error variance in the prediction of the root value of the y -variable. This model assumes an ultrametric phylogeny scaled to unit height and that the x and y variables are centered on their predicted values. The residual variance matrix for this model is

$$\begin{aligned} \text{Var}[\mathbf{r}] &= 4av_y\mathbf{T} + 2\left(a^2 + b^2v_x\right)\mathbf{T} \circ \mathbf{T} \\ &+ b^2\sigma^2\left(\mathbf{T} \circ \mathbf{T} \circ \mathbf{T} - \frac{1}{4}(\mathbf{T} \circ \mathbf{T})\mathbf{T}^{-1}(\mathbf{T} \circ \mathbf{T})\right), \end{aligned} \quad (4)$$

where v_x is the variance in the predicted root value used to center the x -variable. Equations 3 and 4 can be used to estimate the parameters a and b in an iterated GLS procedure conditionally on the phylogeny and previous estimates of the σ^2 , v_y , and v_x . The units of the regression parameters are $[a] = [y^2]$ and $[b] = [y^2/x]$.

Model 2: Predictor Evolves as a Geometric Brownian motion

In Model 1, we assumed that the predictor x follows a Brownian motion. This yields simplicity and may be realistic in many cases, but there are also cases in which this assumption is problematic. In particular, many candidate predictor variables will be on ratio scale types. Such variables must be positive and tend to follow multiplicative dynamics, which is incompatible with Brownian motions. The usual solution to this problem is

log transformation, as the logarithm of a multiplicative process may follow additive dynamics compatible with Brownian motions, but this does not solve the problem if the causal influence of the predictor variable on the rate of evolution is linear on the original scale. To deal with such situations, we use the model

$$\begin{aligned} dy &= \sqrt{a+bx}dW_1, \\ dx &= \frac{1}{2}\sigma^2 xdt + \sigma x dW_2, \end{aligned} \quad (5)$$

where x follows geometric Brownian motions, which is equivalent to Brownian motions of the logarithm of x . We again assume that the y -variable is centered on an unbiased predictor of its root value. We also assume that the x -variable is standardized with an unbiased predictor of its logarithmic root value. In this model, y is also uncorrelated with x (and any power of x), but after a time interval t ,

$$\text{Cov}[y^2, x] = \frac{2b}{3\sigma^2} \left(e^{2\sigma^2 t} + 2e^{\frac{1}{2}\sigma^2 t} - 3e^{\sigma^2 t} \right), \quad (6)$$

which shows that the b -parameter can be estimated from a regression of the squared y -variables on the x -variables. In Appendix B.1, we show that the best linear predictor of the vector of squared y -variables from the vector of x -variables is

$$(A - v_y)1 + \frac{2b}{\sigma^2} \left(\mathbf{I} - \frac{2}{3}e^{-\frac{1}{2}\sigma^2} \left[e^{\frac{3}{2}\sigma^2 t_a} - 1 \right] \left[e^{\sigma^2 t_a} - 1 \right]^{-1} \right) \mathbf{x}, \quad (7)$$

where \mathbf{x} is centered on its predicted mean (see Appendix B.1), $[f(t_a)]$ denotes a matrix in which the ij 'th element is $f(t_a)$ with t_a being the shared branch length between species i and j on the phylogeny, and A is an intercept equal to a minus a correction term given in the appendix. We have again assumed an ultrametric phylogeny scaled to unit height ($t = 1$).

The residual variance matrix for this model is

$$\begin{aligned} \text{Var}[\mathbf{r}] &= 4av_y \mathbf{T} + \frac{8bv_y e^{\frac{1}{2}v_{\ln x}}}{\sigma^2} \left[e^{\frac{1}{2}\sigma^2 t_a} - 1 \right] \\ &+ 2a^2 \mathbf{T} \circ \mathbf{T} + \frac{8abe^{\frac{1}{2}v_{\ln x}}}{\sigma^2} \mathbf{T} \circ \left[e^{\frac{1}{2}\sigma^2 t_a} - 1 \right] \\ &+ \frac{2b^2 e^{2v_{\ln x}}}{\sigma^4} \left(\frac{8}{3} \left[e^{2\sigma^2 t_a} - e^{\frac{1}{2}\sigma^2 t_a} \right] - \left[e^{2\sigma^2 t_a} - 1 \right] \right. \\ &\left. - \frac{8}{9} \left[e^{\frac{3}{2}\sigma^2 t_a} - 1 \right] \left[e^{\sigma^2 t_a} - 1 \right]^{-1} \left[e^{\frac{3}{2}\sigma^2 t_a} - 1 \right] \right), \end{aligned} \quad (8)$$

where $v_{\ln x}$ is the error variance in predicting $\ln x_0$, the root value of the x -variable on log scale. Note that σ^2 for this model is the predicted variance of the natural logarithm of x across the tips of an ultrametric phylogeny scaled to unit height.

The regression is not linear in the sense that the covariance of y^2 with higher powers of the predictor are

not zero, but these tend to zero as σ^2 increases. In the simulations, however, we found some bias when $\sigma^2 = 2$, but not for $\sigma^2 \leq 1$. Nevertheless, if σ^2 is below 0.5 or so, it may be better to use the simpler Model 1 since the multiplicative dynamics of a ratio-scale variable with moderate variation can be approximated with additive dynamics and Brownian motions.

Model 3: Predictor Affects Rates of Recent Microevolution

In a large data set of morphological traits, Uyeda et al. (2011) observed that recent microevolution was largely decoupled from macroevolution in a so-called “blunderbuss” pattern. The best general model for this pattern was a combination of a white noise with a million-year-scale Gaussian point process. Gaussian point processes converge to Brownian motions on long time scales, and the current phenotype of a species may thus be reasonably described by the sum of a Brownian motion on a macroevolutionary scale and a white noise representing recent evolution on a less than million-year time scale. In contrast to the previous models, which represent rates of cumulative macroevolution over the entire phylogeny, the following model may be used to test if rates of recent microevolution are related to candidate predictor variables. In light of the blunderbuss, we can do this by using the squared deviation of each species from its macroevolutionary prediction as a response variable.

We model the species trait vector as $\mathbf{y} = \mathbf{y}_{\text{macro}} + \mathbf{y}_{\text{micro}}$, where $\mathbf{y}_{\text{macro}}$ is the result of a macroevolutionary process (e.g., Brownian motion) unfolding on the phylogeny and $\mathbf{y}_{\text{micro}}$ is a microevolutionary deviation, independent of the macroevolutionary process, but with a variance that may depend on the predictor variable as

$$\mathbf{y}_{\text{micro}} \sim N(\mathbf{0}, a\mathbf{I} + \text{diag}(b\mathbf{x})), \quad (9)$$

where \mathbf{x} is a vector of species-specific predictor variables. The *diag*-function applied to a vector yields a diagonal matrix with the vector along the diagonal. In this case, only the recent value of the predictor matters and we will assume that it does not interact with the macroevolutionary process.

As the microevolutionary deviation is not directly observable, we model it as $\mathbf{y} - \hat{\mathbf{y}}$, where $\hat{\mathbf{y}}$ is a vector of macroevolutionary predictions for each species based on the other species given the phylogeny and a model of macroevolution. The best linear unbiased predictor for each species based on the other species in the phylogeny is the vector

$$\hat{\mathbf{y}} = \bar{\mathbf{y}} + \left(\mathbf{V} - d(\mathbf{V}^{-1})^{-1} \right) \mathbf{V}^{-1} (\mathbf{y} - \bar{\mathbf{y}}), \quad (10)$$

where the d -function sets to zero all nondiagonal elements of its argument, \mathbf{V} is the variance matrix of the \mathbf{y} vector, and $\bar{\mathbf{y}}$ is the mean vector. Technically, the elements of the mean vector should be calculated separately for each species by not including that species, but in practice with many species it makes little difference to use the

same grand mean for all species. Any fixed effects of the evolutionary model can be included in the mean vector. This equation can be used to make predictions from any evolutionary model by using the phylogenetic variance matrix and mean vector derived from that model. It can be derived from the general equation for best linear unbiased prediction of ancestral states given in [Martins and Hansen \(1997\)](#) to predict the trait value of an hypothetical species in the exact phylogenetic position of the (deleted) species to be predicted. The form of the equation can be verified by elementwise comparison. The deviance from the prediction is then

$$\mathbf{y} - \hat{\mathbf{y}} = d(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1} (\mathbf{y} - \bar{\mathbf{y}}), \quad (11)$$

and the squares of the elements of this vector can be used as a response variable in the regression model. If the evolutionary process is Gaussian, then these deviations are normally distributed with mean zero and variance matrix

$$\text{Var}[\mathbf{y} - \hat{\mathbf{y}}] = \left(d(\mathbf{V}^{-1}) \mathbf{V} d(\mathbf{V}^{-1}) \right)^{-1}. \quad (12)$$

The variance matrix of the \mathbf{y} -vector is

$$\mathbf{V} = \mathbf{V}_{\text{macro}} + \mathbf{V}_{\text{micro}} + \mathbf{V}_{\text{measurement}}, \quad (13)$$

where $\mathbf{V}_{\text{macro}}$ is the phylogenetic variance matrix derived from the assumed macroevolutionary process, $\mathbf{V}_{\text{micro}} (= a\mathbf{I} + \text{diag}(b\mathbf{x}))$ is the variance matrix contributed by recent microevolution (white noise) and $\mathbf{V}_{\text{measurement}}$ contains any measurement variances we wish to add to the model. If we assume that macroevolution is a Brownian-motion process, $\mathbf{V}_{\text{macro}} = \sigma_y^2 \mathbf{T}$, where \mathbf{T} is the matrix of shared branch lengths, but other Gaussian processes, such as an Ornstein-Uhlenbeck process, may also be used.

In this model, the microevolutionary parameters a, b , and macroevolutionary parameters such as σ_y^2 could be estimated with maximum likelihood based on the normal distribution, but in line with the approach of the other models, we focus on a GLS regression of the square of the predicted microevolutionary deviations on the predictor variable as

$$\mathbf{A} + b \left(\left(d(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1} \right) \circ \left(d(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1} \right) \right) (\mathbf{x} - \bar{\mathbf{x}}) \quad (14)$$

as shown in Appendix B.1. The intercept vector is

$$\begin{aligned} \mathbf{A} = & \text{diag} \left[d(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-2} d(\mathbf{V}^{-1})^{-1} \right] a \\ & + \text{diag} \left[d(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1} \mathbf{V}_{\text{macro}} \mathbf{V}^{-1} d(\mathbf{V}^{-1})^{-1} \right] \end{aligned} \quad (15)$$

which is computed by use of the relation $\mathbf{y} - \hat{\mathbf{y}} = \mathbf{y}_{\text{micro}} + (\mathbf{y}_{\text{macro}} - \hat{\mathbf{y}})$, and measurement variance can be added to $\mathbf{V}_{\text{macro}}$. This assumes that micro- and macroevolution are independent, and that the predictor variable only affects the microevolutionary deviation. The latter assumption is not likely to be important in the GLS setting, as the

effects of macroevolution are largely removed by using the deviation from the macroevolutionary prediction as a response variable. The units of these parameters are $[a] = [y^2]$ and $[b] = [y^2/x]$.

Because predictions for the different species are correlated with each other, we need to compute the residual covariances. Assuming that $\mathbf{y} - \hat{\mathbf{y}}$ is normally distributed with zero mean, the vector of their squared elements follows a multivariate χ^2 -distribution with one degree of freedom and covariances

$$\begin{aligned} \text{Cov} \left[(y_i - \hat{y}_i)^2 (y_j - \hat{y}_j)^2 \right] &= 2 \text{Cov} \left[y_i - \hat{y}_i, y_j - \hat{y}_j \right]^2 \\ &= 2 [d(\mathbf{V}^{-1}) \mathbf{V} d(\mathbf{V}^{-1})]_{ij}^2 \end{aligned} \quad (16)$$

Because the predictor variable, x , is regarded as fixed, these covariances are also the residual covariances of the regression model. In matrix notation the residual variance matrix can then be written

$$\text{Var}[\mathbf{r}] = 2 \left(d(\mathbf{V}^{-1}) \mathbf{V} d(\mathbf{V}^{-1}) \right)^{-1} \circ \left(d(\mathbf{V}^{-1}) \mathbf{V} d(\mathbf{V}^{-1}) \right)^{-1}. \quad (17)$$

METHODS

The Statistical Model

To analyze these models we will use the standard framework of phylogenetic generalized least squares (e.g., [Martins and Hansen 1997](#)). The statistical model is

$$\mathbf{Y} = \mathbf{D}\boldsymbol{\beta} + \mathbf{r}, \text{Var}[\mathbf{r}] = \mathbf{V}_r, \quad (18)$$

where \mathbf{Y} is a vector of species observations, which in our case will be the species squared deviation from a grand mean (Models 1 and 2) or prediction (Model 3). The \mathbf{D} is a design matrix with predictor variables, $\boldsymbol{\beta}$ is a vector of parameters to be estimated, and \mathbf{r} a vector of residuals with mean zero and, a not necessarily diagonal, variance matrix, \mathbf{V}_r . If the \mathbf{D} and \mathbf{V}_r are specified, generalized least squares (GLS) estimates can be used to obtain best linear unbiased estimates of the parameter vector $\boldsymbol{\beta}$, as

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{D}^T \mathbf{V}_r^{-1} \mathbf{D} \right)^{-1} \mathbf{D}^T \mathbf{V}_r^{-1} \mathbf{Y}, \quad (19)$$

with estimation variance

$$\text{Var}[\hat{\boldsymbol{\beta}}] = \left(\mathbf{D}^T \mathbf{V}_r^{-1} \mathbf{D} \right)^{-1}, \quad (20)$$

which can be used to form standard errors. Note that these standard errors cannot be translated into confidence intervals when the distribution of the residuals is unknown. In our models, the \mathbf{V}_r matrix will be a function of model parameters such as a, b and σ^2 . The parameter σ^2 in Models 1 and 2 can be obtained by maximum likelihood based on the predictor variable alone (since either the predictor or its logarithm is assumed to be normally distributed). The other parameters can be estimated with an iterated

procedure starting from an initial guess (e.g., based on ordinary least-squares regression). Usually, such schemes converge in a handful of iterations.

The design matrix, \mathbf{D} , will normally contain two columns. For Model 1, we use

$$\mathbf{D} = \left[\mathbf{1}, \left(\mathbf{I} - \frac{1}{2} (\mathbf{T} \circ \mathbf{T}) \mathbf{T}^{-1} \right) \mathbf{x} \right], \quad (21)$$

with \mathbf{x} centered on its predicted root value. This estimates $\hat{\boldsymbol{\beta}} = \{\hat{A}, \hat{b}\}$, and an unbiased estimator of a can be obtained by adding v_y to A . For Model 2, we use

$$\mathbf{D} = \left[\mathbf{1}, \frac{2}{\sigma^2} \left(\mathbf{I} - \frac{2}{3} e^{-\frac{1}{2} \sigma^2} \left[e^{\frac{3}{2} \sigma^2 t_a} - 1 \right] \left[e^{\sigma^2 t_a} - 1 \right]^{-1} \right) \mathbf{x} \right], \quad (22)$$

with \mathbf{x} standardized with its predicted root value and centered on its predicted mean value. Here, we estimate a by adding v_y plus a bias-correction term given in appendix B to the intercept. For Model 3, we use

$$\mathbf{D} = [\text{diag}(\mathbf{d}(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-2} \mathbf{d}(\mathbf{V}^{-1})^{-1}), (\mathbf{d}(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1}) \circ (\mathbf{d}(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1}) \mathbf{x}], \quad (23)$$

with \mathbf{V} as in (13), and with the term $\text{diag}(\mathbf{d}(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1} \mathbf{V}_{\text{macro}} \mathbf{V}^{-1} \mathbf{d}(\mathbf{V}^{-1})^{-1})$ subtracted from the response vector (i.e., a is estimated according to equation 15). It is also possible to add columns to these design matrices to account for various fixed effects that may influence the squared deviations.

Most of our models require that variables are centered on or standardized with predicted values at the root of the phylogeny. Under the assumption of Brownian motion, the expression

$$m = \frac{\mathbf{1}^T \mathbf{T}^{-1} \mathbf{y}}{\mathbf{1}^T \mathbf{T}^{-1} \mathbf{1}}, \quad (24)$$

will be an unbiased predictor of the root value of y . Its variance is $v_y = a/n_e$, where a is the variance parameter of the Brownian motion, and $n_e = \mathbf{1}^T \mathbf{T}^{-1} \mathbf{1}$ is an effective sample size. The same equation can be used for the root value of x or $\ln x$ as needed. In principle better predictors could be found by replacing \mathbf{T} with the estimated variance matrices of the variables, but this is unlikely to make a difference in practice.

The residual variance matrix, \mathbf{V}_r , will contain a biological component with residual variances and covariances as described in equation 4 for Model 1, equation 8 for Model 2, and equation 17 for Model 3. When available, measurement or observation variances for response and predictor variables may also be included in the residual variance matrix and used to correct for attenuation as described in Appendix B.1. Observation variances in comparative studies are often available from standard errors or confidence intervals of the means (or other statistics) used from the individual species. Note that the observation variance of the response variable in this case would be the predicted variance of the square of the mean. If we assume that the

observation error is normally distributed, observation variances of squared means can be obtained as $2se^4$, where se is the standard error of the mean. Hansen and Bartoszek (2012), Garamszegi (2014), and Grabowski et al. (2016) provide further discussion of observation error in comparative studies.

The three models are set up to explain the variance of a variance, and we expect a worse fit as compared to standard linear models of trait means. Indeed, even an R^2 of a few percent could be regarded as a strong signal in these models. To illustrate, let us assume that, as in all our models, the y variable conditional on x is normally distributed with a variance that is linear in x , as in $\text{Var}[y|x] = a + bx$. Then, assuming that y is mean centered, the proportion of variance in the square of y explained by x is

$$\begin{aligned} R^2 &= \frac{\text{Var}[E[y^2|x]]}{\text{Var}[y^2]} = \frac{\text{Var}[E[y^2|x]]}{E[\text{Var}[y^2|x]] + \text{Var}[E[y^2|x]]} \\ &= \frac{b^2 \text{Var}[x]}{2a^2 + 4ab\bar{x} + 2b^2\bar{x}^2 + 3b^2 \text{Var}[x]} < \frac{1}{3}. \end{aligned}$$

Hence, with these assumptions there is a theoretical ceiling for R^2 at less than 1/3. In practice, much less variance will be explained. For example, in a rather ideal situation with two equal-sized groups of species with one group having twice the variance of the other, we get an expected $R^2 = 1/21 \approx 5\%$ for the simple ANOVA set up (i.e., $a = 1, b = 1$, and $\text{Var}[x] = 1/4$ with x as a zero-one indicator variable); even with infinitely larger variance in one group ($b = \infty$) the variance explained is limited to $R^2 = 1/5$.

Step-by-Step Summary of Approach.

1. Do exploratory analysis of both the response (y) and predictor (x) variable to assess their distributional properties and approximate modes of evolution. Phylogenetic signal can be assessed by estimating a phylogenetic heritability (Lynch 1991; Housworth et al. 2004) or a phylogenetic half-life (Hansen 1997). This will tell if the variables (or their logarithms) can be approximated by a Brownian motion and help decide which of the three models is most appropriate.

2. Prepare the data by choosing appropriate scales and units for the analysis to be used. This may involve log transformation and centering on the mean or predicted root value. Be careful about variance standardization as this may obscure the signal of interest.

3. Formulate the hypothesis to test, and pay attention to the form of the relationship between predictor and rates that is predicted by the hypothesis.

4. Fit the predictor variable or its logarithm to a Brownian motion to estimate the parameter σ^2 . In cases of low but nonzero phylogenetic signal, it is possible to fit a mixed model and use the rate parameter (σ^2) from the Brownian-motion component.

5. Choose the model to use. This should be based both on statistical considerations (distribution of variables, model adequacy, e.g., Pennell et al. 2015) and biological considerations (likely or hypothesized modes of evolution and relations between variables). Model 2 should only be used with predictor variables on ratio or log-ratio scale types (see Houle et al. 2011), and log-transformed predictors would not be suitable for this model. Note that Model 3 is testing a different hypothesis than Models 1 and 2. In cases of low phylogenetic signal, Model 3 would be fine to use and will then approximate an ordinary least-squares regression, but Models 1 and 2 would be problematic to interpret as they assume the relationship between the variables is built gradually over the history of the species.

6. Fit the desired models as described above.

7. Calculate the parameters a and b from the regression parameters, and report results with units and estimated uncertainty. Interpret effect sizes. The R^2 of the model should be considered, but be aware that we can only reasonably expect to explain a few percent of the variance. Resampling procedures can be used to further explore model adequacy and uncertainty of the estimates.

We have implemented statistical machinery for most of the operations in Step 4–7 in the function `rate_gls` in the R-package `evolvability` (Bolstad et al. 2014). We have also included functions for simulation of the three evolutionary models (function: `simulate_rate`), parametric bootstrapping (function: `rate_gls_boot`), and a vignette explaining the use of the functions (Analyzing rates of evolution). The package is available at CRAN with the latest version at "github.com/GHBolstad/evolvability."

APPLICATION

Do Beak Shapes Evolve Faster in Large-Brained Birds?

Wyles et al. (1983) argued that animals with larger brains have faster rates of morphological evolution than comparable animals with smaller brains. They attributed this to behavioral drive, a process by which learning allows individuals to explore new ways of life and thereby setting up novel selection pressures on their anatomy and physiology that lead to more extensive divergence on macroevolutionary time scales. This idea is part of a broader conglomerate of hypotheses and theories about how plasticity, learning, preferences, intentions, culture, and niche construction can influence evolution and adaptation by altering selection pressures or allowing organisms to persist in novel environments long enough to adapt to them (e.g., Baldwin 1896; Popper 1972; Wake et al. 1983; Odling-Smee et al. 2003; Frank 2011; Zuk et al. 2014). The effects of such mechanisms on rates of evolution are far from obvious, and it has also been argued that behavioral flexibility may buffer against changes and help explain morphological stasis (e.g., Bogert 1949; Wake et al. 1983; Hansen

and Houle 2004). Paenke et al. (2007) show that both scenarios are theoretically plausible and provide general conditions for plasticity to accelerate or decelerate the response to selection.

The study of Wyles et al. (1983) was based on comparing morphological disparity between a few large clades with different degrees of encephalization such as primates against other mammals, song birds against other birds, and mammals against reptiles. This procedure yields only a handful of comparisons and is vulnerable to spurious third-variable effects and artifacts of measurement stemming from the many biological differences that must exist at such broad phylogenetic scales. It can only be regarded as crudely suggestive.

Application: Methods

To illustrate our models, we capitalize on recent large-scale phylogenies and open-source data bases for bird morphology to test the hypotheses of Wyles et al. (1983) and Wilson (1985) more rigorously. We use 3D morphometric beak-shape data taken from Cooney et al. (2017) and brain- and body-size data from Fristoe et al. (2017) and Tsuboi et al. (2018) for 651 bird species to test if rates of beak-shape evolution are influenced by brain size. We excluded the paleognaths, comprising four species of ratites and one species of tinamou, from the analysis, as this group is both a sister group to the rest and biologically different from most other birds. We also removed the emperor penguin (*Aptenodytes forsteri*), as this species was a strong outlier in terms of both brain and body size. This left us with 645 species that we used for the analyses presented here. Results were qualitatively similar when including all species.

The 3D-landmark data (4 landmarks and 75 semilandmarks) of beak shape were standardized by generalized Procrustes analysis before principal-component analysis in Cooney et al. (2017). We studied the first eight principal components from this analysis, which collectively explain more than 99% of the variance. The principal components are in units of centroid size (i.e., mean linear distance of landmarks to centroid). The phylogeny was taken from Jetz et al. (2014), and we used a maximum credibility tree from a sample of thousand phylogenies based on family-level topologies from Ericson et al. (2006) using TreeAnnotator version 2.4.5.

We tested for the effects of both absolute and relative brain mass. To obtain relative brain mass we performed an allometric analysis of brain mass on body mass (Fig. 1) with the R-package `slouch` (Hansen et al. 2008; version 2.0.0. documented in Kopperud et al. 2020). Following Grabowski et al. (2016), we fitted a direct-effect model and accounted for estimation error in species means based on reported standard errors of the means of logged brain and body masses. The estimated allometric exponent corrected for attenuation due to observation error was $3/5 (= 0.60 \pm 0.01)$, and log body size explained

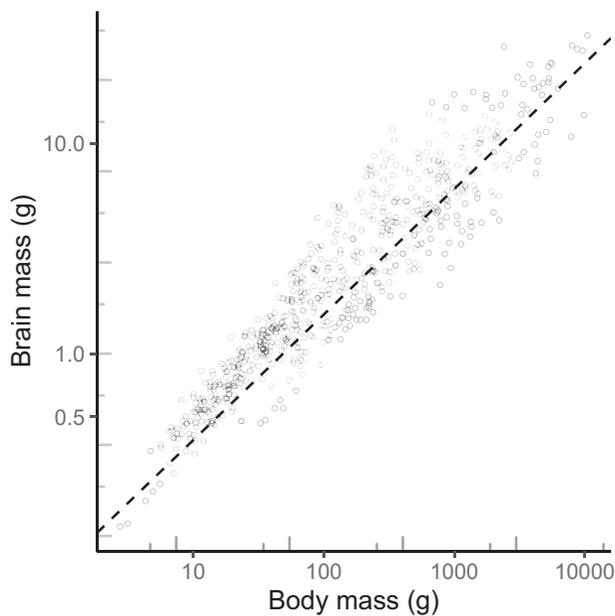


FIGURE 1. Brain–body allometry for the 645 species of birds used in this study. The data points are species means taken from Tsuboi et al. (2018). The regression line $-2.21(\pm 0.12) + 0.598(\pm 0.008) \log(\text{body mass})$, $R^2 = 88.4\%$ was estimated from a “direct-effect” Ornstein–Uhlenbeck model and corrected for estimation error in species means (including attenuation) as described in Grabowski et al. (2016). The phylogenetic half-life and stationary variance of the Ornstein–Uhlenbeck model were $t_{1/2} = 5h$ with 2-unit support interval: $(1.35 - \infty)$ and $\sigma_{ou}^2 = 0.43$ with 2-unit support interval: $(0.14 - 0.49)$. The tree length was $h = 104$ myr. The long half-life means that the residuals evolve similarly to a Brownian motion.

88% of the variance in log brain size. Relative brain mass is based on the residuals from this allometric model.

The basic statistics of the traits are in Table 1. The brain- and body-size data show a strong phylogenetic signal and good fit to Brownian motion on log scale. For the beak-shape data, the first principal component, which is an axis of stoutness to pointedness of the beak, likewise show strong phylogenetic signal and good fit to Brownian motion. The other seven “shape” traits show a moderately strong phylogenetic signal, leaving room for Ornstein–Uhlenbeck-type dynamics or a component of recent evolution. The two-unit support intervals for the phylogenetic halflife exclude both infinity (Brownian motion) and zero (white noise) for all these.

We tested the hypotheses that either absolute or relative brain size influence the rate of evolution of beak shape, and whether this happens on average throughout the phylogeny (Models 1 and 2) or in terms of recent “within-species” microevolution (Model 3). The models were fitted with the functions in the evolvability package as described above. Estimates are shown $\pm se$ if not otherwise indicated. Reported standard errors are based on equation 20. With some exceptions, these were similar to bootstrap estimates (see Supplementary material). The predictor variables in the plots are phylogenetically weighted values that account for correlation with related species (i.e., values correspond to the entries in the design matrices in equations 21–23).

Application: Results

Starting with absolute brain mass, we note that this is a ratio-scale variable that we expect to evolve on a multiplicative scale. This makes the geometric Brownian-motion Model 2 is the most natural choice, and the mean-scaled variance for brain mass of $1.19 (= \sigma^2 t) > 1$ means that Model 1 may be less good on an untransformed scale. The results from this model mix small to moderate positive effects of absolute brain size on the rates of evolution in most shape variables (PCs 2, 3, 7, 8) with a large positive effect on PC 6, and essentially no effects on PCs 1, 4, and 5 (Fig. 2). The phylogenetic R^2 of the positive effects are small in absolute terms, 5% for PC 6 and 0.6–2% for the others, but still give room for substantial influence on rates. The predicted rate variance, in units of squared centroid size (cs) per tree length ($h = 104$ my), for PC 6 is $-7.5 \times 10^{-4} + 12.4 \times 10^{-4}x$, where x is mean-standardized brain mass, which means that a doubling of brain mass from the phylogenetic mean ($x = 1$) would increase the rate variance from $4.9 \times 10^{-4} cs^2/h$ to $17.3 \times 10^{-4} cs^2/h$. This means that expected variance (disparity) in this shape variable of a clade with twice the average brain mass would be expected to be 3.5 times as large as the average. In other words, doubling the brain mass of a lineage would increase the expected change in PC 6 over an arbitrary time period, and thus the classic rate of evolution (i.e., the *darwin*), with a factor of $\sqrt{3.5}$, or 88%. Similar calculations for PCs 2, 3, 7, and 8 yield factors of, respectively, 27%, 25%, 65%, and 60%.

An alternative is to use Model 1 with log brain mass as predictor variable. This is equivalent in terms of its assumptions about the evolution of brain mass, but differs in that the assumed causal link to the rate variance is now linear on a proportional scale (i.e., linear on log brain mass rather than linear on brain mass). The results of this analysis are congruent with the fitting of Model 2 in that the strongest effect is for PC 6, but both this and the other effects are weaker (Supplementary Fig. S1). For PC 6 the predicted rate variance is $9.5 + 10.5 \log$ brain mass, which means that doubling brain mass from the clade mean ($x = 0$) by setting log brain mass to $x = \ln 2$ would increase the rate variance from $9.5 \times 10^{-4} cs^2/h$ to $17 \times 10^{-4} cs^2/h$, or the expected rate of evolution with a factor of 34%. The other “significant” effects, on PC 7 and 8, correspond to increases of 32% and 25%, respectively.

For relative brain mass, which is on an interval scale type, the Brownian-motions-based Model 1 is the natural choice. Fitting this model to the beak-shape variables shows little effect (Fig. 3). The estimated slopes are all positive, but small, with phylogenetic R^2 below 1% in all cases except for PC 7. Taken in isolation the results on PC 6, 7, and 8 are consistent with some effect, but we interpret the results collectively as evidence against any substantial effect of relative brain size on rates of beak-shape evolution.

Fitting Model 3 tests a different hypothesis, namely that the rate of recent (within-species) evolution of beak

TABLE 1. Basic statistics for the among-species data: the phylogenetic heritability, PH^2 , is the fraction of among-species variance attributable to Brownian motion (Lynch 1991). The $\sigma_{\text{Brownian}}^2$ is the rate variance of the Brownian motion in units per tree length ($t = 104$ my), and $\sigma_{\text{residual}}^2$ is the nonphylogenetic residual variance. The phylogenetic half-life, $t_{1/2}$, is the time in units of tree length to lose half the expected ancestral influence in an Ornstein–Uhlenbeck model of evolution (Hansen 1997). Body and brain size are in units of ln (gram), and principal components of beak shape in units of $100 \times$ centroid size (cs). The variances are in units of the squares of the trait units. The phylogenetic heritability and the variance components were estimated with the function *Almer* in the *evolvar* package, and the half-life was estimated with *slouch*.

Trait	PH^2	$t_{1/2}$	$\sigma_{\text{Brownian}}^2$	$\sigma_{\text{residual}}^2$
Log body mass	0.97	4.5 ^a	2.54	0.082
Log brain mass	0.99	$\sim \infty^a$	1.12	0.012
Relative brain size	0.91	2.2 ^a	0.091	0.0095
Beak shape pc1	0.95	8.2	289	15
Beak shape pc2	0.76	0.37	156	48
Beak shape pc3	0.88	0.57	43	6.0
Beak shape pc4	0.88	0.75	20	2.9
Beak shape pc5	0.88	0.55	12	1.7
Beak shape pc6	0.75	0.34	5.0	1.7
Beak shape pc7	0.83	0.54	2.7	0.56
Beak shape pc8	0.69	0.27	1.6	0.75

^aCorrected for estimation error in species means.

shape is influenced by brain size. We used a Brownian-motion-based mixed model to predict the states of the shape variables at the tips of the phylogeny and used the squared deviance of the observed states from these predicted states as response variables in the regression. This model makes no assumptions about the evolution or distribution of the predictor variables, which can then be transformed and fitted on any scale that is deemed biologically reasonable. As with Model 1 there are only tiny, inconsistent effects of relative brain size (Supplementary Fig. S2), but increasing absolute brain size elevates the rates of all PCs except the first and marginally the third with R^2 ranging from 0.5% to 4% (Fig. 4). The lack of an effect on PC 1 is perhaps not surprising given that the strong phylogenetic signal of this variable leaves a lesser fraction of “recent” variation to be explained. The estimates from the other shapes indicate that the predicted residual variance at the tip would increase from 18% to 57%, corresponding to 9% to 25% increases in conventional rates of evolution.

The general lack of effect of relative brain size may suggest that the effects of absolute brain size may be caused by body size. For this reason, we also tested body mass as an explanatory variable. With Models 1 and 2, we found generally weaker effects of body mass than we found with brain mass although 3.4% of the variance in PC 6 was explained with the geometric Brownian-motion Model 2 (results not shown). This is still weaker than the effect of brain mass, but it is possible that body size may explain some of the large effect of brain mass on PC 6 in particular. For the recent-evolution model, however, the effects of body mass were similar to those of brain mass (Supplementary Fig. S3).

DISCUSSION

The variation in tempo and mode of evolution is such that even the basic concept of a rate of evolution defined as character change per time, as in the classic *darwin* and *haldane* measures, is problematic. Beyond a

few generations, evolutionary change rarely resembles a deterministic trend, but instead becomes stochastic with huge variations in both rate and direction (e.g., Gingerich 1983, 2019; Bookstein 1987; Hunt 2012; Voje 2016). These variations are such that one cannot compare rates of change across different time intervals; in fact, below species-level time scales the amount of change is often independent of time span (Gingerich 1983, 1993, 2001, 2009; Roopnarine 2003; Uyeda et al. 2011; Voje et al. 2018) rendering conventional rates meaningless. A relationship between change and time span appears at longer, among-species, time scales, but change remains stochastic and often with scaling-relationships resembling Brownian motion (Uyeda et al. 2011). This means that rates of macroevolution cannot be measured on the scale of the trait. Instead, it is necessary to measure rates as “intensities,” variance of change per time. Over the last decades, it has become common to use the diffusion parameter of a fitted Brownian motion as a measure of rate of evolution on macroevolutionary time scales (e.g., Hunt and Carrano 2010, and references in the introduction).

Estimating rates as variances of change has a downside as compared to the estimation of a trend in that it is less statistically efficient and demands larger samples to reach comparable levels of accuracy (e.g., Slater and Pennell 2014). Our theoretical analysis illustrates this point. Variables affecting rates of evolution are not expected to correlate with the state of the evolving traits in the absence of directional trends. Correlations only appear with disparity on the scale of the squared traits and statistical analysis of causal effects on the tempo of evolution must be conducted on this scale. We have shown that even strong effects on the rates of change only generate weak correlations with the causal variable implying that we do not expect to explain more than a few percent of the among-species variance in squared deviations. This means that large samples of species are necessary to detect an effect. Fortunately, databases of morphological traits and phylogenetic relationships are

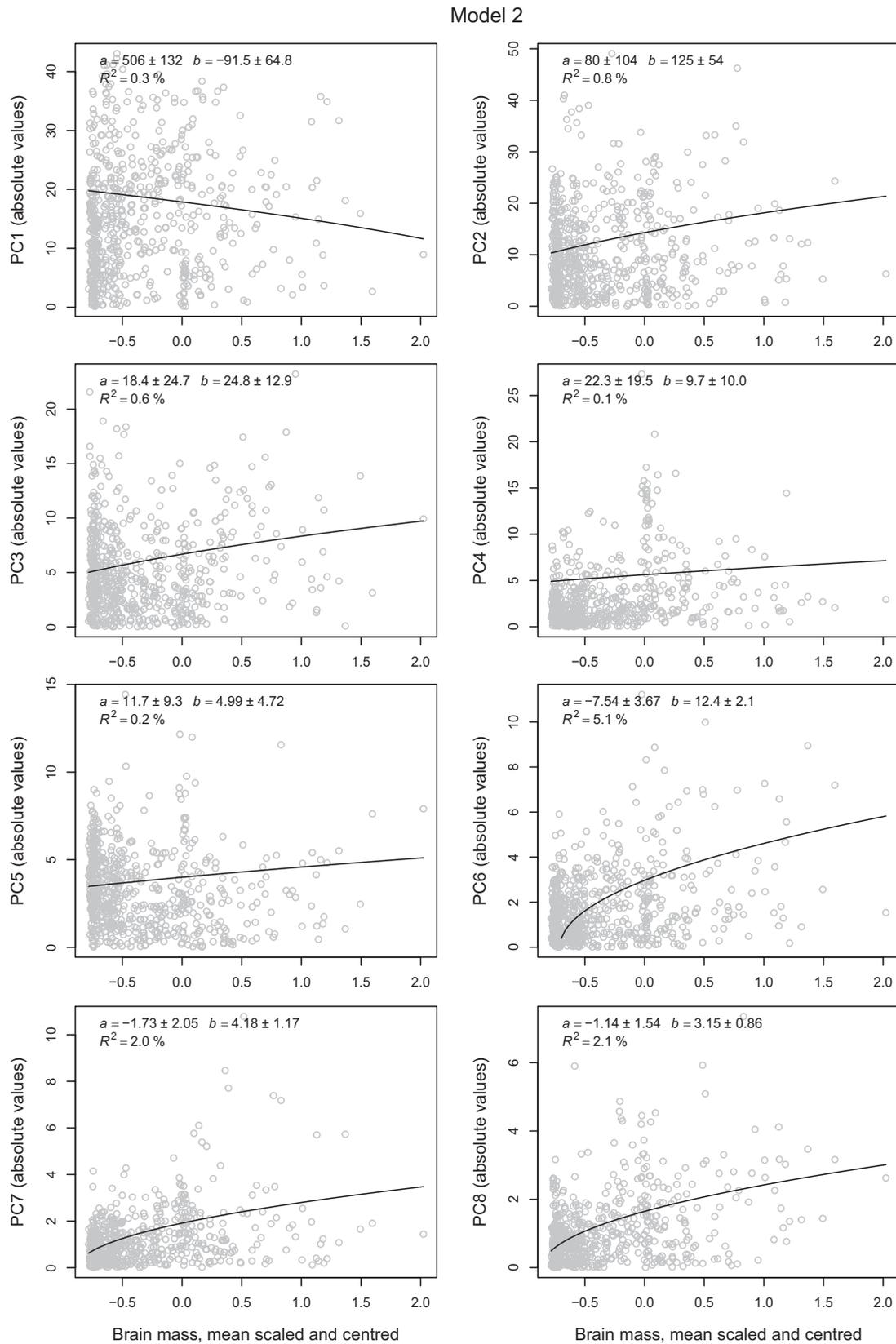


FIGURE 2. The effect of absolute brain size on beak-shape evolution. The plot shows the magnitude of the deviance from clade mean of beak-shape variables in units of centroid size ($\times 100$) plotted against (phylogenetically weighted) log brain mass for 645 bird species. The fitted line and parameter estimates (\pm SE) are based on Model 2 in the main text assuming that brain mass evolves like a geometric Brownian motion. The estimated slopes (b) of PCs 2, 6, 7, and 8 are statistically significant in the sense that they are more than two standard errors above zero.

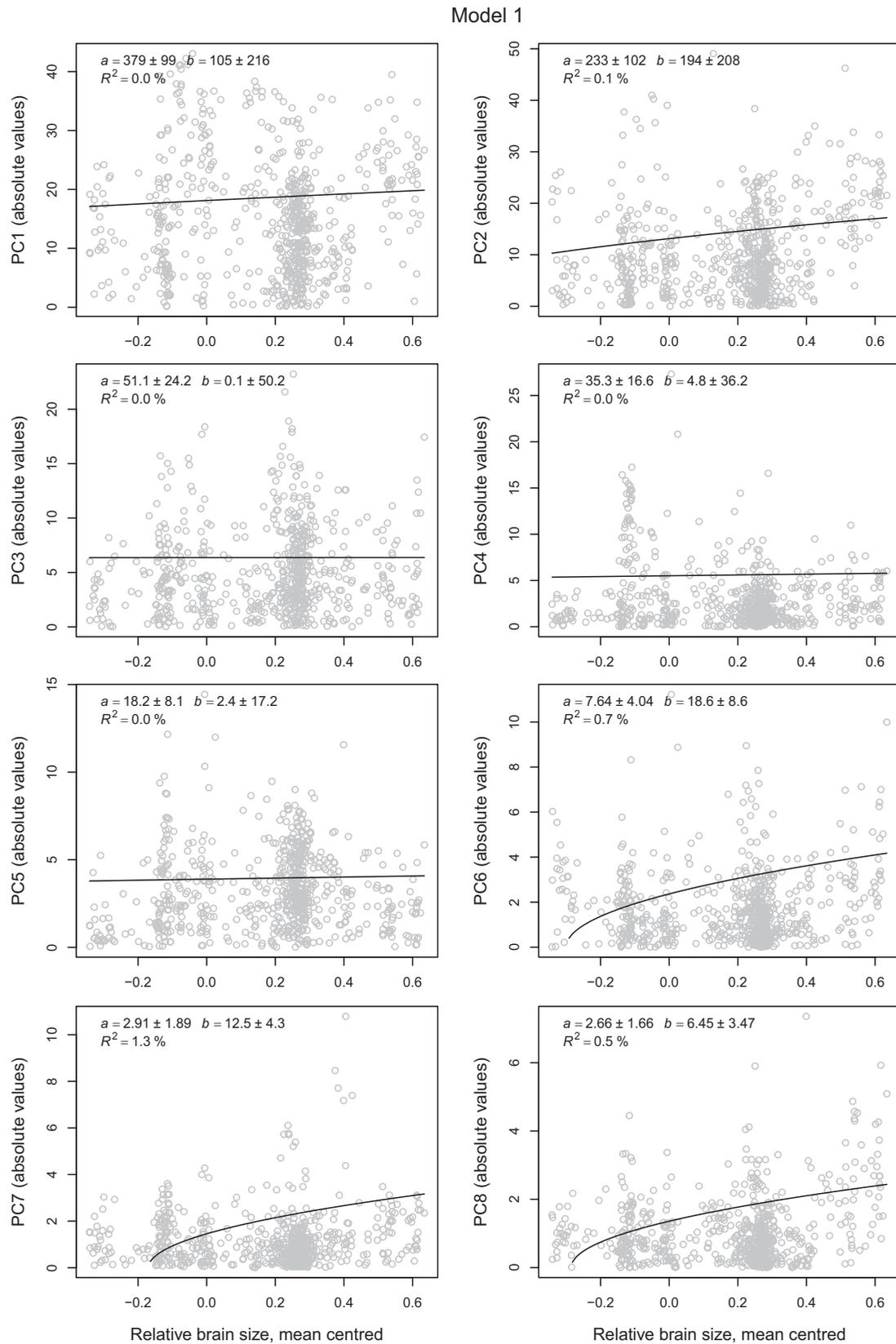


FIGURE 3. The effect of relative brain size on beak-shape evolution. The plot shows the magnitude of the deviance from clade mean of beak-shape variables in units of centroid size ($\times 100$) plotted against relative brain size for 645 bird species. The fitted line and parameter estimates are based on Model 1 in the main text assuming that relative brain size evolves like a Brownian motion. The estimated slopes (b) of PCs 6 and 7 are statistically significant in the sense that they are more than two standard errors above zero.

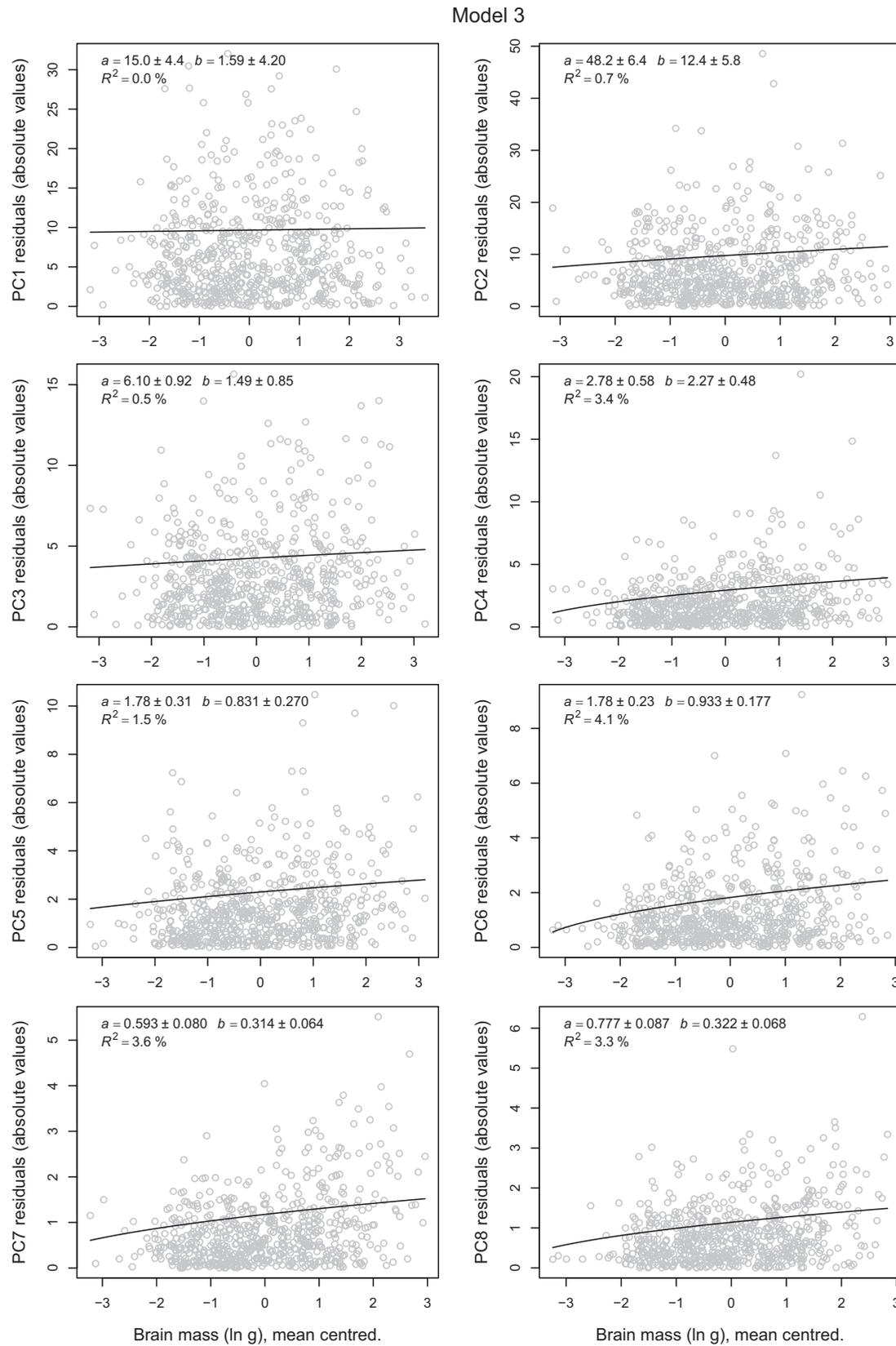


FIGURE 4. The effect of log brain size on recent beak-shape evolution. The plot shows the magnitude of the deviance of beak-shape variables from the macroevolutionary prediction in units of centroid size ($\times 100$) in Model 3 plotted against (predicted) log brain mass for 645 bird species. The fitted line and parameter estimates are based on Model 3 in the main text, and the intercept used for plotting is the average of the prediction in equation 15. The estimated slopes (b) of PCs 2 and 4 to 8 are statistically significant in the sense that they are more than two standard errors above zero.

accumulating and it is becoming increasingly possible to do comparative analyses with the hundreds of species necessary to test hypotheses about causal effects on rates of evolution.

Despite the mathematical complexity, the underlying idea of our method is simple: The effect of a variable on rates of evolution can be detected by a regression of squared trait deviance on the variable. The mathematics yields two refinements to a naive regression of this type. First, it shows how we can translate the coefficients of the raw regression to effects on underlying rates of evolution (the a and b parameters), and second, it shows how to calculate residual covariances so as to use statistically more efficient GLS estimation that accounts for phylogenetic relatedness. As with most comparative methods, both interpretation and the optimal statistical model depend on assumptions about the underlying evolutionary process. Our three models differ in these assumptions and apply under different circumstances.

Do Large-Brained Birds Evolve Faster?

We have illustrated our methods with a data set of 645 species of birds from [Cooney et al. \(2017\)](#), which we used to test the hypothesis that brain size influences rates of evolution in beak shape. This hypothesis, derived from [Wyles et al. \(1983\)](#), may be motivated as the outcome of a Baldwin effect in which presumptively more behaviorally flexible larger-brained species are more able to discover new sources of food or novel uses of their beaks, which may introduce more frequent novel patterns of selection acting on their beaks and thus more frequent evolutionary changes in morphology. This is consistent with the recent finding of increased speciation rates in large-brained birds ([Sayol et al. 2019](#)), and with findings of increased body-size disparity in bird families with large brains ([Sol and Price 2008](#)).

Our results support this idea in finding indications that absolute brain size is related to the rate of evolution of some aspects of beak shape, but with the caveat that relative brain size had little effect. Our models never explained more than a few percent of the variance in beak-shape disparity, but in the cases explaining more than a percent or so of the variance, the best estimates of effects usually indicated that the variance in shape change and thus the disparity would increase substantially. For example, assuming that brain mass evolved as a geometric Brownian motion and had a linear effect on the rate variance of the shape variables leads to the prediction that a clade of birds with twice the brain mass of another would have between 1.5 and 3.5 times the disparity (among-species variance) in many of the beak-shape variables. The recent-evolution model (Model 3) also showed rather consistent positive effects, which indicates that the rapid stationary fluctuations that characterize evolution on short below-species-level time scales are more pronounced in larger-brained birds.

The absence of consistent effects of relative brain size is surprising, but consistent with the idea that it is absolute

more than relative brain size that matters for intelligence (e.g., [MacLean et al. 2014](#); and discussed in [Striedter 2005](#)). There is a serious caveat, however, in that brain size is related to many other ecological and behavioral factors in birds (e.g., [Burish et al. 2004](#); [Shultz and Dunbar 2010](#); [Vincze 2016](#)), and we cannot exclude that some factor other than intelligence mediates the causal effect on the tempo of evolution. We tested the effect of body size on rates and found that these could not fully explain the effect of absolute brain size in Model 2, but that they mirrored the effects of absolute brain size when using the recent-evolution model. It is therefore possible that the elevated microevolutionary rates of large-brained species are mediated through some other correlate of larger bodies.

Relation to Other Methods

As mentioned in the introduction, there are several methods for studying rates of evolution in phylogenies or fossil time series based on stochastic models. The approach most similar to ours is the likelihood-based method pioneered by [O'Meara et al. \(2006\)](#) and [Thomas et al. \(2006\)](#). Here, Brownian motions with different rate parameters are fitted to different sections (regimes) of a phylogeny. This is similar to our Model 1, but with the predictor variable as a categorical fixed effect mapped onto the phylogeny. This approach is superior to ours when the predictor categories can be mapped reliably onto the phylogeny, because the assumption of a fixed (nonevolving) predictor allows a Gaussian process and hence the application of likelihood methods based on the normal distribution. Our models should be considered an alternative to this only when the predictor is continuous or uncertain such that mapping is unreliable. In this situation a full stochastic model for the joint evolution of both response and predictor variables is preferable. In fact, [Revell \(2013\)](#) has shown that rate estimates from different regimes are biased to be more similar when mapping of regimes is uncertain. Also be aware that the low power we have illustrated is not a particular flaw of our models but apply similarly to all statistical estimators of variance of change.

Another approach to the study of rate involves the estimation of ancestral states on a phylogeny and then computation of rates over ancestor–descendant pairs or contrasts between such estimates. This is often done without consideration of statistical uncertainty and is problematic because estimated ancestral states and derived contrasts are generally uncertain, heteroscedastic, and correlated ([Martins and Hansen 1997](#)). As discussed in [Hansen \(2014\)](#), this approach is at best a roundabout way of estimating rate parameters with many pitfalls and no advantages to a direct approach.

A related approach is to correlate disparity at higher taxonomic levels with trait means at those levels, as was done by [Wyles et al. \(1983\)](#) and [Sol and Price \(2008\)](#). See also [Weir and Lawson \(2016\)](#) for a pairwise-contrasts

method to study the effects of predictors on evolutionary rates with an application to song complexity in birds.

Our three basic models are far from exhaustive, and there are many conceivable modifications and extensions. Some flexibility can be achieved by scale transformations of the variables, such as use of log scale or not. A more challenging extension would be to multivariate rates. In our example, we decomposed a multivariate character into principal components and analyzed these separately with our univariate models. It is well known, however, that principal-component analysis of comparative data can be misleading because it mixes phylogenetic structure with trait dependence (Uyeda et al. 2015). The fundamental problem with multivariate extensions, however, is that the number of rate parameters will increase, and given the low power of the univariate approaches, any method not involving some form of dimensionality reduction is unlikely to be useful. Extension to more than one predictor variable in the fashion of a multiple regression is also possible but increases mathematical complexity and would likely need to assume independent evolution of the predictors to keep a manageable number of parameters.

The Ornstein–Uhlenbeck process is a common extension to Brownian motion in the stochastic modeling of evolution. Parameters in the Ornstein–Uhlenbeck process are sometimes used to describe rates of evolution (e.g., Martins 1994) or rates of adaptation (Hansen 1997, 2012). These parameters have a less direct interpretation as rates, however, and are perhaps better considered as collectively describing dynamically changing tempo and mode of evolution in relation to selective optima. Nevertheless, one could consider models with either the “rate of adaptation” or the stochastic diffusion of an Ornstein–Uhlenbeck model being influenced by an evolving trait. The moments of such a process would be calculable, but more complicated than in the models we have presented. The regressions would also be nonlinear. Likelihood methods for estimating the influence of mapped “fixed effects” on parameters in Ornstein–Uhlenbeck models have been presented by Beaulieu et al. (2012). Another, simpler, extension would be to allow the predictor variable to follow an Ornstein–Uhlenbeck process in place of the Brownian motion or geometric Brownian motion. Finally, May and Moore (2020) have recently proposed a model similar to ours but with a discrete predictor variable evolving according to a Markov chain. This complements the models in the present article.

The recent-evolution model provides a different perspective on rates of evolution than in most approaches derived from comparative methods or fossil time-series analysis. Instead of studying rates of evolution as change per time over phylogenetic branches or time intervals, this approach focuses on recent evolution after removing the effects of past change. In fact, the different models we have presented resemble two positions from the early days of the phylogenetic comparative-methods literature (reviewed

in Martins and Hansen 1996). One position, exemplified by the autoregressive approach of Cheverud et al. (1985), sought to remove the phylogenetic component and study adaptation in the remainder, while another position, exemplified by Lynch’s (1991) mixed model, sought to remove the species-specific component and study adaptation in the phylogenetic component. While the latter position, being implicit in Felsenstein’s (1985) independent-contrasts and related process-based phylogenetic comparative methods, won out for conventional analysis of trait evolution, our analysis illustrates how this can be seen as a distinction between studying evolution at two different time scales, a macroevolutionary, above species-level time scale, and a microevolutionary, below species-level time scale. As argued by Uyeda et al. (2011) and Hansen (2012), evolutionary rates behave differently at micro- and macroevolutionary time scales and likely reflect different evolutionary mechanisms. This justifies the development of distinct models to study rates at different time scales.

The recent-evolution model also resembles comparative methods designed to make inferences about changes at single branches; typically aiming to test if particular changes are unusual as compared to the overall rate of evolution on the phylogeny (e.g., McPeck 1995; Revell 2008; Nunn and Zhu 2014). In fact, our equation 11 could be used to pick out the deviance for any species or subset of species, and test these for conformity to an expected Gaussian distribution with the predicted variance matrix in equation 12. See also Uyeda et al. (2018) for more extensive discussion of the role of singular events in comparative studies.

Our motivation for developing these models was to explore the potential for testing hypotheses about factors influencing rates of evolution or the disparity of a clade. For example, as reviewed in Bolstad et al. (2014), there have been many claims that rates of divergence and among-species variance are related to within-species variance, or to more specific measures of species evolvability. Such claims are usually based on comparing clade variances to estimates of evolvability from one or a few species across different traits, and thus do not fit the framework we have set up. It is, however, desirable to develop direct comparative tests of how measures of evolvability and other microevolutionary parameters influence rates of divergence, and the models presented here illustrate both how this can be done and the limitations in doing so.

SUPPLEMENTARY MATERIAL

Data and supplementary analyses are available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.gb5mkkwq9>. Software is available at CRAN: <https://CRAN.R-project.org/package=evolvability>.

ACKNOWLEDGMENTS

We thank Julien Clavel, Simon Ho, Matt Pennell, and an anonymous reviewer for thorough and helpful comments on the manuscript.

FUNDING

This work was partially funded by a Marie Skłodowska-Curie Fellowship (747338) and Vetenskapsrådet (the Swedish Research Council) International Postdoc Grant (2016-06635) to M.T., and a research grant from Norwegian Research Council (275862) to G.H.B.

A.1 CALCULATION OF MOMENTS

In this appendix, we compute the necessary moments for the different models by integration. The general strategy is to use the multivariate Itô formula (e.g., Gard 1988, p. 59) to derive stochastic differential equations for powers of the variables x and y and then integrate over x and y to form differential equations for moments and cross moments. We then solve these differential equations by standard integration over time.

The Model 1 was

$$\begin{aligned} dy &= \sqrt{a+bx}dW_1, \\ dx &= \sigma dW_2, \end{aligned}$$

with initial conditions y_0 and x_0 . Since x follows standard univariate Brownian motions, its moments at time t are

$$E[(x-x_0)^k] = \begin{cases} \sigma^k t^{k/2} (k-1)!! & \text{if } k \text{ is even} \\ 0 & \text{if } k \text{ is odd} \end{cases},$$

where the !! signifies semifactorial. We illustrate the computation of the other moments by an example, $E[y^2x]$. The Itô formula yields

$$dy^2x = (a+bx)xdt + 2yx\sqrt{a+bx}dW_1 + \sigma y^2dW_2.$$

Taking the expectation, and using the fact that the dW_i have mean zero and are independent of previous history, yields

$$\begin{aligned} dE[y^2x] &= E[(a+bx)x]dt = (aE[x] + bE[x^2])dt \\ &= (a+bx_0)x_0 + b\sigma^2tdt. \end{aligned}$$

Integrating yields

$$E[y^2x] = y_0^2x_0 + (a+bx_0)x_0t + \frac{b\sigma^2t^2}{2},$$

and since $E[y^2] = y_0^2 + (a+bx_0)t$, we get

$$\text{Cov}[y^2, x] = \frac{b\sigma^2t^2}{2}.$$

The differential equation for a general moment obtained from the Itô formula is

$$\begin{aligned} \frac{dE[y^m x^k]}{dt} &= \frac{k(k-1)\sigma^2}{2} E[y^m x^{k-2}] \\ &\quad + \frac{m(m-1)}{2} E[(a+bx)y^{m-2}x^k], \end{aligned}$$

which can be solved iteratively in m and k . In the following we specifically need the fourth moment of y , which is

$$E[y^4] = y_0^4 + 6(a+bx_0)y_0^2t + 3(a+bx_0)^2t^2 + b^2\sigma^2t^3,$$

which yields

$$\text{Var}[y^2] = 4(a+bx_0)y_0^2t + 2(a+bx_0)^2t^2 + b^2\sigma^2t^3.$$

Note that this shows that the distribution of y is leptokurtic and not Gaussian.

To obtain the necessary cross-species covariances needed for the GLS regression, we use an approach from Hansen and Martins (1996) based on conditioning on the values of the most recent common ancestor of the two species in question, and assuming that they evolve independently after they split apart. This requires the following conditional expectations:

$$\begin{aligned} E[y_i^2 | y_a, x_a] &= y_a^2 + (a+bx_a)t_{ai}, \\ E[x_i | y_a, x_a] &= x_a, \end{aligned}$$

where y_a and x_a are the values of an ancestor of i and t_{ai} is the time that separates the ancestor a and the species i . For an ultrametric tree, $t_{ai} = t_{aj} = t - t_a$ with t_a the time from root to the most recent common ancestor of i and j . From this, we get

$$\begin{aligned} \text{Cov}[y_i^2, y_j^2] &= \text{Cov}[E[y_i^2 | y_a, x_a], E[y_j^2 | y_a, x_a]] \\ &= \text{Var}[y_a^2 + (a+bx_a)(t-t_a)] \\ &= \text{Var}[y_a^2] + b^2(t-t_a)^2 \text{Var}[x_a] \\ &\quad + 2b(t-t_a) \text{Cov}[y_a^2, x_a] \\ &= 4(a+bx_0)y_0^2t_a + 2(a+bx_0)^2t_a^2 \\ &\quad + b^2\sigma^2t_a^3 + b^2(t-t_a)^2\sigma^2t_a + b^2\sigma^2(t-t_a)t_a^2 \\ &= 4(a+bx_0)y_0^2t_a + 2(a+bx_0)^2t_a^2 \\ &\quad + b^2\sigma^2t_a(t^2 - t_a t + t_a^2), \end{aligned}$$

$$\begin{aligned} \text{Cov}[y_i^2, x_j] &= \text{Cov}[E[y_i^2 | y_a, x_a], E[x_j | y_a, x_a]] \\ &= \text{Cov}[y_a^2 + (a+bx_a)(t-t_a), x_a] \\ &= \frac{1}{2}b\sigma^2t_a^2 + b(t-t_a)\sigma^2t_a = b\sigma^2\left(t - \frac{t_a}{2}\right), \end{aligned}$$

$$Cov[x_i, x_j] = Cov[E[x_i|x_a], E[x_j|x_a]] = Var[x_a] = \sigma^2 t_a.$$

These moments are all we need to construct the necessary variance and covariance matrices for GLS regression. This is detailed in Appendix B.1.

The Model 2 was

$$dy = \sqrt{a + bx} dW_1, \\ dx = \frac{1}{2} \sigma^2 x dt + \sigma x dW_2,$$

which implies that x follows a log-normal distribution with moments

$$E[x^k] = x_0^k e^{\frac{1}{2} k^2 \sigma^2 t}.$$

Using the Itô formula and taking moments yield, the following differential equation for an arbitrary moment of the process:

$$\frac{dE[y^m x^k]}{dt} = \frac{k^2 \sigma^2}{2} E[y^m x^k] + \frac{m(m-1)}{2} E[(a + bx) y^{m-2} x^k],$$

which can be solved iteratively in m . Most pertinently

$$E[yx^k] = y_0 x_0^k e^{\frac{1}{2} k^2 \sigma^2 t},$$

$$E[y^2 x^k] = y_0^2 x_0^k e^{\frac{1}{2} k^2 \sigma^2 t} + a x_0^k t e^{\frac{1}{2} k^2 \sigma^2 t} + \frac{2bx_0^{k+1}}{(1+2k)\sigma^2} \left(e^{\frac{1}{2}(k+1)^2 \sigma^2 t} - e^{\frac{1}{2} k^2 \sigma^2 t} \right),$$

$$E[y^3 x^k] = y_0^3 x_0^k e^{\frac{1}{2} k^2 \sigma^2 t} + 3ay_0 x_0^k t e^{\frac{1}{2} k^2 \sigma^2 t} + \frac{6by_0 x_0^{k+1}}{(1+2k)\sigma^2} \left(e^{\frac{1}{2}(k+1)^2 \sigma^2 t} - e^{\frac{1}{2} k^2 \sigma^2 t} \right),$$

$$E[y^4 x^k] = y_0^4 x_0^k e^{\frac{1}{2} k^2 \sigma^2 t} + 3ax_0^k (2y_0^2 + at) t e^{\frac{1}{2} k^2 \sigma^2 t} + \frac{12bx_0^{k+1} (at + y_0^2)}{(2k+1)\sigma^2} \left(e^{\frac{1}{2}(k+1)^2 \sigma^2 t} - e^{\frac{1}{2} k^2 \sigma^2 t} \right) + \frac{6b^2 x_0^{k+2}}{(2k+3)(2k+1)(k+1)\sigma^4} \left((2k+3)e^{\frac{1}{2} k^2 \sigma^2 t} + (2k+1)e^{\frac{1}{2}(k^2+4k+4)\sigma^2 t} - 4(k+1)e^{\frac{1}{2}(k+1)^2 \sigma^2 t} \right),$$

which shows that $Cov[y, x^k] = 0$ and $Cov[y^3, x^k] = 0$ for all k . The even powers of y covary with x , however, and this justifies the regression of the square of y on x . Using these moments we can compute the cross-species covariances with the same method as for Model 1. To do this, we

specifically need

$$E[y_i | y_a, x_a] = y_a, \\ E[y_i^2 | y_a, x_a] = y_a^2 + at_{ai} + \frac{2b}{\sigma^2} \left(e^{\frac{\sigma^2}{2} t_{ai}} - 1 \right) x_a$$

This yields

$$Cov[y_i^2, y_j^2] = Cov[E[y_i^2 | y_a, x_a], E[y_j^2 | y_a, x_a]] = Var \left[y_a^2 + a(t-t_a) + \frac{2b}{\sigma^2} \left(e^{\frac{\sigma^2}{2} (t-t_a)} - 1 \right) x_a \right] = Var[y_a^2] + \frac{4b^2}{\sigma^4} \left(e^{\frac{\sigma^2}{2} (t-t_a)} - 1 \right)^2 Var[x_a] + \frac{4b}{\sigma^2} \left(e^{\frac{\sigma^2}{2} (t-t_a)} - 1 \right) Cov[y_a^2, x_a],$$

$$Cov[y_i^2, x_j] = Cov[E[y_i^2 | y_a, x_a], E[x_j | y_a, x_a]] = Cov \left[y_a^2 + at_{ai} + \frac{2b}{\sigma^2} \left(e^{\frac{\sigma^2}{2} t_{ai}} - 1 \right) x_a, e^{\frac{1}{2} \sigma^2 t_{aj}} x_a \right] = e^{\frac{1}{2} \sigma^2 (t-t_a)} Cov[y_a^2, x_a] + \frac{2b}{\sigma^2} \left(e^{\sigma^2 (t-t_a)} - e^{\frac{1}{2} \sigma^2 (t-t_a)} \right) Var[x_a],$$

$$Cov[x_i, x_j] = Cov[E[x_i | x_a], E[x_j | x_a]] = e^{\sigma^2 (t-t_a)} Var[x_a].$$

The specific moments needed to solve this are

$$Var[y^2] = 4ay_0^2 t + 2a^2 t^2 + \frac{8bx_0 (at + y_0^2)}{\sigma^2} \left(e^{\frac{1}{2} \sigma^2 t} - 1 \right) + \frac{2b^2 x_0^2}{\sigma^4} \left(e^{\sigma^2 t} - 1 \right)^2,$$

$$Cov[y^2, x] = \frac{2bx_0^2 e^{\frac{1}{2} \sigma^2 t}}{3\sigma^2} \left(2 + e^{\frac{3}{2} \sigma^2 t} - 3e^{\frac{1}{2} \sigma^2 t} \right),$$

$$Var[x] = x_0^2 e^{\sigma^2 t} \left(e^{\sigma^2 t} - 1 \right),$$

which yields

$$Cov[y_i^2, y_j^2] = 4ay_0^2 t_a + 2a^2 t_a^2 + \frac{8bx_0}{\sigma^2} \left(y_0^2 + at_a \right) \left(e^{\frac{1}{2} \sigma^2 t_a} - 1 \right) + \frac{2b^2 x_0^2}{\sigma^4} \left(\left(e^{\sigma^2 t} - 1 \right)^2 - \left(e^{\sigma^2 t} - e^{\sigma^2 t_a} \right)^2 \right) + \frac{8}{3} \left(1 - e^{\frac{3}{2} \sigma^2 t_a} \right) \left(e^{\frac{1}{2} \sigma^2 t} - e^{\frac{1}{2} \sigma^2 t_a} \right),$$

$$Cov[y_i^2, x_j] = \frac{2bx_0^2}{\sigma^2} e^{\sigma^2 t} \left(e^{\sigma^2 t_a} - 1 - \frac{2}{3} \left(e^{\frac{1}{2} \sigma^2 (3t_a - t)} - e^{-\frac{1}{2} \sigma^2 t} \right) \right),$$

$$\text{Cov}[x_i, x_j] = x_0^2 e^{\sigma^2 t} (e^{\sigma^2 t_a} - 1),$$

which is what is needed to set up the GLS regression in Appendix B.1.

B.1 CONSTRUCTING THE REGRESSION MATRICES

In this appendix, we derive the variance and covariance matrices for the GLS regression of squared y -values on the predictor x -variables. Let \mathbf{y}^2 denote the column vector of species squared y -values. Proposition 2.22 in Eaton (1983) then gives

$$\hat{\mathbf{y}}^2 = E[\mathbf{y}^2] + \text{Cov}[\mathbf{y}^2, \mathbf{x}] \text{Var}[\mathbf{x}]^{-1} (\mathbf{x} - E[\mathbf{x}]),$$

as the best linear predictor of the \mathbf{y}^2 vector from the \mathbf{x} vector in the least-squares sense. This has residual variance matrix

$$\begin{aligned} \text{Var}[\mathbf{r}] &= \text{Var}[\mathbf{y}^2 - \hat{\mathbf{y}}^2] \\ &= \text{Var}[\mathbf{y}^2] - \text{Cov}[\mathbf{y}^2, \mathbf{x}] \text{Var}[\mathbf{x}]^{-1} \text{Cov}[\mathbf{y}^2, \mathbf{x}]^T, \end{aligned}$$

where $\text{Cov}[\mathbf{y}^2, \mathbf{x}]$ is the matrix with elements $\text{Cov}[y_i^2, x_j]$, $\text{Var}[\mathbf{x}]$ is the matrix with elements $\text{Cov}[x_i, x_j]$, and $\text{Var}[\mathbf{y}^2]$ is the matrix with elements $\text{Cov}[y_i^2, y_j^2]$.

For Model 1, the moments derived in Appendix A.1 then gives:

$$\begin{aligned} \hat{\mathbf{y}}^2 &= (a + bx_0 + y_0^2) \mathbf{1} + b \left(\mathbf{I} - \frac{1}{2} (\mathbf{T} \circ \mathbf{T}) \mathbf{T}^{-1} \right) \mathbf{x}, \\ \text{Var}[\mathbf{r}] &= 4(a + bx_0) y_0^2 \mathbf{T} + 2(a + bx_0)^2 \mathbf{T} \circ \mathbf{T} \\ &\quad + b^2 \sigma^2 \left(\mathbf{T} \circ \mathbf{T} \circ \mathbf{T} - \frac{1}{4} (\mathbf{T} \circ \mathbf{T}) \mathbf{T}^{-1} (\mathbf{T} \circ \mathbf{T}) \right), \end{aligned}$$

where $\mathbf{T} = [t_a]$ is the matrix of shared branch lengths in units of total tree length, \mathbf{I} is the identity matrix, $\mathbf{1}$ is a column vector of ones, and y_0 and x_0 are the initial values of y and x at the root of the tree.

Centering the variables around estimates of the root values would in principle justify setting $y_0 = x_0 = 0$, but the estimation error in the root values can be substantial if there are only a few branches near the root of the phylogeny. To account for this we assume that the estimates of the root values are unbiased and uncorrelated with variance v_y for y_0 and v_x for x_0 . Furthermore, we have to account for the fact that the error in estimating y_0 is correlated with y^2 . To do this, assume the y -variable has evolved as a Brownian motion, so that the y -vector is distributed as $\mathbf{y} \sim N(y_0 \mathbf{1}, a\mathbf{T})$. Let

$$m = \frac{\mathbf{1}^T \mathbf{T}^{-1} \mathbf{y}}{\mathbf{1}^T \mathbf{T}^{-1} \mathbf{1}}$$

be our (unbiased) estimator of y_0 . Then for species i ,

$$\begin{aligned} E[(y_i - y_0 - (m - y_0))^2] &= E[(y_i - y_0)^2] + \text{Var}[m] \\ &\quad - 2\text{Cov}[y_i, m] = E[(y_i - y_0)^2] - v_y, \end{aligned}$$

which follows because $\text{Cov}[y_i, m] = \text{Var}[y_i] / \mathbf{1}^T \mathbf{T}^{-1} \mathbf{1} = v_y$. Hence, to get an unbiased estimate of a from the intercept in the regression we subtract v_y . To correct the residual variance we take the expectation over y_0 and x_0 . This yields

$$\begin{aligned} \hat{\mathbf{y}}^2 &= (a - v_y) \mathbf{1} + b \left(\mathbf{I} - \frac{1}{2} (\mathbf{T} \circ \mathbf{T}) \mathbf{T}^{-1} \right) \mathbf{x}, \\ \text{Var}[\mathbf{r}] &= 4av_y \mathbf{T} + 2(a^2 + b^2 v_x) \mathbf{T} \circ \mathbf{T} \\ &\quad + b^2 \sigma^2 \left(\mathbf{T} \circ \mathbf{T} \circ \mathbf{T} - \frac{1}{4} (\mathbf{T} \circ \mathbf{T}) \mathbf{T}^{-1} (\mathbf{T} \circ \mathbf{T}) \right), \end{aligned}$$

and thus equations 3 and 4 in the main text.

For Model 2, the moments derived in appendix A give the following matrix form of the regression

$$\begin{aligned} \hat{\mathbf{y}}^2 &= \left(a + \frac{2b}{\sigma^2} (e^{\frac{1}{2}\sigma^2} - 1) x_0 + y_0^2 \right) \mathbf{1} \\ &\quad + \frac{2b}{\sigma^2} \left(\mathbf{I} - \frac{2}{3} e^{-\frac{1}{2}\sigma^2} [e^{\frac{3}{2}\sigma^2 t_a} - 1] [e^{\sigma^2 t_a} - 1]^{-1} \right) \\ &\quad (\mathbf{x} - E[\mathbf{x}]), \end{aligned}$$

$$\begin{aligned} \text{Var}[\mathbf{r}] &= 4ay_0^2 \mathbf{T} + \frac{8by_0^2 x_0}{\sigma^2} [e^{\frac{1}{2}\sigma^2 t_a} - 1] + 2a^2 \mathbf{T} \circ \mathbf{T} \\ &\quad + \frac{8abx_0}{\sigma^2} \mathbf{T} \circ [e^{\frac{1}{2}\sigma^2 t_a} - 1] \\ &\quad + \frac{2b^2 x_0^2}{\sigma^4} \left(\frac{8}{3} [e^{2\sigma^2 t_a} - e^{\frac{1}{2}\sigma^2 t_a}] - [e^{2\sigma^2 t_a} - 1] \right) \\ &\quad - \frac{8}{9} [e^{\frac{3}{2}\sigma^2 t_a} - 1] [e^{\sigma^2 t_a} - 1]^{-1} [e^{\frac{3}{2}\sigma^2 t_a} - 1], \end{aligned}$$

where the notation $[f(t_a)]$ means a matrix in which the ij 'th element is $f(t_a)$ with t_a being the shared branch length between species i and j . In this notation $\mathbf{T} = [t_a]$. As for Model 1, we center the y -variable on its estimated root value, but for the x -variable we standardize by dividing with the predicted root value. To model the error in this root value, we assume that $\ln(x_0)$ is normally distributed with zero mean and variance $v_{\ln x}$. Then

$$\begin{aligned} E[x_0] &= e^{\frac{1}{2} v_{\ln x}}, \\ E[x_0^2] &= e^{2v_{\ln x}}, \\ E[y_0^2 x_0] &= v_y e^{\frac{1}{2} v_{\ln x}}. \end{aligned}$$

Using these expectations, we get

$$\begin{aligned} \hat{\mathbf{y}}^2 &= \left(a + \frac{2be^{\frac{1}{2}v_{\ln x}}}{\sigma^2} (e^{\frac{1}{2}\sigma^2} - 1) - v_y \right) \mathbf{1} \\ &\quad + \frac{2b}{\sigma^2} \left(\mathbf{I} - \frac{2}{3} e^{-\frac{1}{2}\sigma^2} [e^{\frac{3}{2}\sigma^2 t_a} - 1] [e^{\sigma^2 t_a} - 1]^{-1} \right) \\ &\quad \left(\mathbf{x} - e^{\frac{\sigma^2}{2} - \frac{v_{\ln x}}{2}} \mathbf{1} \right), \end{aligned}$$

$$\begin{aligned} \text{Var}[\mathbf{r}] = & 4av_y\mathbf{T} + \frac{8bv_y e^{\frac{1}{2}v_{\ln x}}}{\sigma^2} \left[e^{\frac{1}{2}\sigma^2 t_a} - 1 \right] + 2a^2\mathbf{T} \circ \mathbf{T} \\ & + \frac{8abe^{\frac{1}{2}v_{\ln x}}}{\sigma^2} \mathbf{T} \circ \left[e^{\frac{1}{2}\sigma^2 t_a} - 1 \right] \\ & + \frac{2b^2 e^{2v_{\ln x}}}{\sigma^4} \left(\frac{8}{3} \left[e^{2\sigma^2 t_a} - e^{\frac{1}{2}\sigma^2 t_a} \right] - \left[e^{2\sigma^2 t_a} - 1 \right] \right. \\ & \left. - \frac{8}{9} \left[e^{\frac{3}{2}\sigma^2 t_a} - 1 \right] \left[e^{\sigma^2 t_a} - 1 \right]^{-1} \left[e^{\frac{3}{2}\sigma^2 t_a} - 1 \right] \right), \end{aligned}$$

assuming that y is centered and x standardized on their predicted root values. This yields equations 7 and 8 in the main text. Note that the second term in the intercept of regression depends on the regression slope. This caused numerical problems in our estimation procedure, which we solved by treating the whole intercept (except v_y) as a constant, A , and the subtracting the second term at the end to estimate a .

For Model 3, let $(\mathbf{y} - \hat{\mathbf{y}}) \circ (\mathbf{y} - \hat{\mathbf{y}})$ be the column vector of squared deviations from the prediction as given in equation 11. The linear regression of the squared deviations vector on the vector \mathbf{x} is

$$\begin{aligned} (\mathbf{y} - \hat{\mathbf{y}}) \circ (\mathbf{y} - \hat{\mathbf{y}}) = & E[(\mathbf{y} - \hat{\mathbf{y}}) \circ (\mathbf{y} - \hat{\mathbf{y}})] \\ & + \text{Cov}[(\mathbf{y} - \hat{\mathbf{y}}) \circ (\mathbf{y} - \hat{\mathbf{y}}), \mathbf{x}] \text{Var}[\mathbf{x}]^{-1} (\mathbf{x} - \bar{\mathbf{x}}) + \mathbf{r}, \end{aligned}$$

where \mathbf{r} is a vector of residuals. To compute this we use

$$\begin{aligned} E[(\mathbf{y} - \hat{\mathbf{y}}) \circ (\mathbf{y} - \hat{\mathbf{y}})] & = \text{diag}[\text{Var}[(\mathbf{y} - \hat{\mathbf{y}})]] \\ & = \text{diag}\left[\text{Var}\left[d(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1} (\mathbf{y} - \hat{\mathbf{y}})\right]\right] \\ & = \text{diag}\left[d(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1} \text{Var}[(\mathbf{y} - \hat{\mathbf{y}})] \mathbf{V}^{-1} d(\mathbf{V}^{-1})^{-1}\right] \\ & = \text{diag}\left[d(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1} (\mathbf{V}_{\text{macro}} + \mathbf{V}_{\text{micro}}) \mathbf{V}^{-1} d(\mathbf{V}^{-1})^{-1}\right] \\ & = \text{diag}\left[d(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1} \mathbf{V}_{\text{macro}} \mathbf{V}^{-1} d(\mathbf{V}^{-1})^{-1}\right] \\ & + \text{diag}\left[d(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1} (a\mathbf{I} \right. \\ & \left. + E[\text{diag}(\mathbf{x} - \bar{\mathbf{x}})]) \mathbf{V}^{-1} d(\mathbf{V}^{-1})^{-1}\right] \end{aligned}$$

and

$$\begin{aligned} \text{Cov}[(\mathbf{y} - \hat{\mathbf{y}}) \circ (\mathbf{y} - \hat{\mathbf{y}}), \mathbf{x}] & = \text{Cov}\left[d(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1} (\mathbf{y} - \hat{\mathbf{y}}) \circ d(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1} (\mathbf{y} - \hat{\mathbf{y}}), \mathbf{x}\right] \\ & = b \left(d(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1} \circ d(\mathbf{V}^{-1})^{-1} \mathbf{V}^{-1} \right) \text{Var}[\mathbf{x}], \end{aligned}$$

which can be verified by elementwise comparison using the assumption that $\text{Cov}[y_r y_k, x_i] = \text{Cov}[E[y_r y_k | \mathbf{x}], x_i] =$

$b\text{Cov}[x_i, x_j]$ if $r = k = j$, and zero if $r \neq k$. Note that we have treated predictor as fixed, and \mathbf{V} as a constant in the macroevolutionary prediction. This yields equations 14 and 15 in the main text.

Observation, that is measurement, error in the species means can also be added to the residual variance matrices. Let \mathbf{V}_{my} and \mathbf{V}_{mx} be the measurement variance matrices for the response (i.e., the y-squared vector) and predictor (i.e. the x-vector). These matrices will typically be diagonal, but in principle measurement covariances can also be included. The terms added to the residual variance will be then be

$$\mathbf{V}_{\text{my}} + b^2 \left(\mathbf{I} - \frac{1}{2} (\mathbf{T} \circ \mathbf{T}) \mathbf{T}^{-1} \right) \mathbf{V}_{\text{mx}} \left(\mathbf{I} - \frac{1}{2} (\mathbf{T} \circ \mathbf{T}) \mathbf{T}^{-1} \right)^T$$

for Model 1 and

$$\mathbf{V}_{\text{my}} + \frac{4b^2}{\sigma^4} \left(\mathbf{I} - \frac{2}{3} e^{-\frac{1}{2}\sigma^2} \left[e^{\frac{3}{2}\sigma^2 t_a} - 1 \right] \left[e^{\sigma^2 t_a} - 1 \right]^{-1} \right)$$

$$\mathbf{V}_{\text{mx}} \left(\mathbf{I} - \frac{2}{3} e^{-\frac{1}{2}\sigma^2} \left[e^{\frac{3}{2}\sigma^2 t_a} - 1 \right] \left[e^{\sigma^2 t_a} - 1 \right]^{-1} \right)^T$$

for Model 2. For Model 3, measurement variance can be added as described in the main text.

Observation error in the predictor variable may also cause attenuation, a downward bias in the estimated regression slope. Attenuation is not corrected by including observation error in the residual variance. Expressions for attenuation with correlated data are given in Hansen and Bartoszek (2012). Applying this to our situation yields the attenuation factor

$$K = 1 - \frac{(\mathbf{x} - \mathbf{1})^T \left(\text{Cov}[\mathbf{y}^2, \mathbf{x}] \text{Var}[\mathbf{x}]^{-1} \right)^T \text{Cov}[\mathbf{y}^2, \mathbf{x}] \text{Var}[\mathbf{x}]^{-1} \text{Var}[\mathbf{r}]^{-1} \text{Cov}[\mathbf{y}^2, \mathbf{x}] \text{Var}[\mathbf{x}]^{-1}}{\mathbf{V}_{\text{mx}} \text{Var}[\mathbf{x}]^{-1} (\mathbf{x} - \mathbf{1})^T \left(\text{Cov}[\mathbf{y}^2, \mathbf{x}] \text{Var}[\mathbf{x}]^{-1} \right)^T \text{Var}[\mathbf{r}]^{-1} \text{Cov}[\mathbf{y}^2, \mathbf{x}] \text{Var}[\mathbf{x}]^{-1} (\mathbf{x} - \mathbf{1})}$$

Dividing the estimated slope with K would give an approximately unbiased estimate of the true slope.

REFERENCES

- Adams D.C. 2013. Comparing evolutionary rates for different phenotypic traits on a phylogeny using likelihood. *Syst. Biol.* 62:181–192.
- Adams D.C. 2014. Quantifying and comparing phylogenetic evolutionary rates for shape and other high-dimensional phenotypic data. *Syst. Biol.* 63:166–177.
- Alhajeri B.H., Stepan S.J. 2018. Disparity and evolutionary rate do not explain diversity patterns in muroid rodents (*Rodentia: Muroidea*). *Evol. Biol.* 45:324–344.
- Arnold S.J. 2014. Phenotypic evolution: the ongoing synthesis. *Am. Nat.* 183:729–746.
- Baldwin J.M. 1896. A new factor in evolution. *Am. Nat.* 30:441–451, 536–553.
- Beaulieu J.M., Jhwueng D.-C., Boettiger C., O'Meara B.C. 2012. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution* 66:2369–2383.

- Bell M.A. 2012. Adaptive landscapes, evolution, and the fossil record. In: Svensson E.I., Calsbeek R., editors. *The adaptive landscape in evolutionary biology*. Oxford: Oxford University Press. p. 243–256.
- Bogert C.M. 1949. Thermoregulation in reptiles, a factor in evolution. *Evolution* 3:195–211.
- Bolstad G.H., Hansen T.F., Pélabon C., Falahati-Anbaran M., Perez-Barrales R., Armbruster W.S. 2014. Genetic constraints predict evolutionary divergence in *Dalechampia* blossoms. *Philos. Trans. R. Soc. B* 369:20130255.
- Bookstein F.L. 1987. Random walk and the existence of evolutionary rates. *Paleobiology* 13:446–464.
- Burish M.J., Kueh H.Y., Wang S.S. 2004. Brain architecture and social complexity in modern and ancient birds. *Brain. Behav. Evol.* 63:107–124.
- Caetano D.S., Harmon L.J. 2019. Estimating correlated rates of trait evolution with uncertainty. *Syst. Biol.* 68:412–429.
- Cheverud J.M., Dow M.M., Leutenegger W. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weights among primates. *Evolution* 39:1335–1351.
- Chira M., Cooney C.R., Bright J.A., Capp E.J.R., Hughes E.C., Moody C.J.A., Nouri L.O., Varley Z.K., Thomas G.H. 2018. Correlates of rate heterogeneity in avian ecomorphological traits. *Ecol. Lett.* 21:1505–1514.
- Cooney C.R., Bright J.A., Capp E.J.R., Chira A.M., Hughes E.C., Moody C.J.A., Nouri L.O., Varley Z.K., Thomas G.H. 2017. Meta-evolutionary dynamics of the adaptive radiation of birds. *Nature* 542:344–347.
- Eastman J.M., Alfaro M.E., Joyce P., Hipp A.L., Harmon L.J. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution* 65:3578–3589.
- Eaton M.L. 1983. *Multivariate statistics: a vector space approach*. New York: Wiley.
- Eldredge N., Gould S.J. 1972. Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf T., editor. *Models in paleobiology*. San Francisco: Freeman.
- Eldredge N., Thompson J.N., Brakefield P.M., Gavrilits S., Jablonski D., Jackson J.B.C., Lenski R.E., Lieberman B.S., McPeck M.A., Miller W.I. 2005. The dynamics of evolutionary stasis. *Paleobiology* 31:133–145.
- Ericson P.G., Anderson C.L., Britton T., Elzanowski A., Johansson U.S., Kallersjö M., Ohlson J.I., Parsons T.J., Zuccon D., Mayr G. 2006. Diversification of neoaves: integration of molecular sequence data and fossils. *Biol. Lett.* 2:543–547.
- Estes S., Arnold S.J. 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.* 125:1–15.
- Frank S.A. 2011. Natural selection. II. Developmental variability and evolutionary rate. *J. Evol. Biol.* 24:2310–2320.
- Fristoe T.S., Iwaniuk A.N., Botero C.A. 2017. Big brains stabilize populations and facilitate colonization of variable habitats in birds. *Nat. Ecol. Evol.* 1:1706–1715.
- Fuentes J.A., Housworth E.A., Weber A., Martins E.P. 2016. Phylogenetic ANCOVA: estimating changes in evolutionary rates as well as relationships between traits. *Am. Nat.* 188:615–627.
- Garamszegi L.Z. 2014. Uncertainties due to within-species variation in comparative studies: measurement errors and statistical weights. In: Garamszegi L.Z., editor. *Modern phylogenetic comparative methods and their application in evolutionary biology*. Berlin: Springer. p. 157–199.
- Gard T.C. 1988. *Introduction to stochastic differential equations*. New York: Marcel Dekker, Inc.
- Garland T. Jr. 1992. Rate tests for phenotypic evolution using phylogenetically independent contrasts. *Am. Nat.* 140:509–519.
- Gingerich P.D. 1983. Rates of evolution: effects of time and temporal scaling. *Science* 222:159–161.
- Gingerich P.D. 1993. Quantification and comparison of evolutionary rates. *Am. J. Sci.* 293A:453–478.
- Gingerich P.D. 2001. Rates of evolution on the time scale of the evolutionary process. *Genetica* 112/113:127–144.
- Gingerich P.D. 2009. Rates of evolution. *Ann. Rev. Ecol. Syst.* 40:657–675.
- Gingerich P.D. 2019. *Rates of evolution: a quantitative synthesis*. Cambridge: Cambridge University Press.
- Gittleman J.L., Anderson C.G., Kot M., Luh H.-K. 1996. Phylogenetic lability and rates of evolution: a comparison of behavioral, morphological and life history traits. In: Martins E.P., editor. *Phylogenies and the comparative method in animal behavior*. Oxford: Oxford University Press. p. 166–205.
- Gould S.J. 2002. *The structure of evolutionary theory*. Cambridge MA: Belknap.
- Grabowski M., Voje K.L., Hansen T.F. 2016. Evolutionary modeling and correcting for observation error support a 3/5 brain-body allometry for primates. *J. Hum. Evol.* 94:106–116.
- Hansen T.F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341–1351.
- Hansen T.F. 2012. Adaptive landscapes and macroevolutionary dynamics. In: Svensson E.I., Calsbeek R., editor. *The adaptive landscape in evolutionary biology*. Oxford: Oxford University Press. p. 205–226.
- Hansen T.F. 2014. Use and misuse of comparative methods in the study of adaptation. In: Garamszegi L.Z., editor. *Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice*. Berlin Heidelberg: Springer. p. 351–379.
- Hansen T.F., Bartoszek K. 2012. Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. *Syst. Biol.* 61:413–425.
- Hansen T.F., Houle D. 2004. Evolvability, stabilizing selection, and the problem of stasis. In: Pigliucci M., Preston K., editors. *Phenotypic integration: studying the ecology and evolution of complex phenotypes*. Oxford: Oxford University Press. p. 130–150.
- Hansen T.F., Martins E.P. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 50:1404–1417.
- Hansen T.F., Pienaar J., Orzack S.H. 2008. A comparative method for studying adaptation to a randomly evolving environment. *Evolution* 62:1965–1977.
- Hendry A.P., Kinnison M.T. 1999. The pace of modern life: measuring rates of contemporary microevolution. *Evolution* 53:1637–1653.
- Hendry A.P., Farrugia T.J., Kinnison M.T. 2008. Human influences on rates of phenotypic change in wild animal populations. *Mol. Ecol.* 17:20–29.
- Houle D., Pélabon C., Wagner G.P., Hansen T.F. 2011. Measurement and meaning in biology. *Quart. Rev. Biol.* 86:3–34.
- Housworth E.A., Martins E.P., Lynch M. 2004. The phylogenetic mixed model. *Am. Nat.* 163:84–96.
- Hunt G. 2012. Measuring rates of phenotypic evolution and the inseparability of tempo and mode. *Paleobiology* 38:351–373.
- Hunt G., Carrano M. 2010. Models and methods for analyzing phenotypic evolution in lineages and clades. In: Alroy J., Hunt G., editors. *Quantitative methods in paleobiology*. Boulder (CO): The Paleontological Society p. 254–269.
- Hunt G., Rabosky D.L. 2014. Phenotypic evolution in fossil species: pattern and process. *Annu. Rev. Earth Planet. Sci.* 42: 421–441.
- Jablonski D. 2017a. Approaches to macroevolution: 1. General concepts and origin of variation. *Evol. Biol.* 44:427–450.
- Jablonski D. 2017b. Approaches to macroevolution: 2. Sorting of variation, some overarching issues, and general conclusions. *Evol. Biol.* 44:451–475.
- Jetz W., Thomas G.H., Joy J.B., Redding D.W., Hartmann K., Mooers A.O. 2014. Global distribution and conservation of evolutionary distinctness in birds. *Curr. Biol.* 24:919–930.
- Kopperud B.T., Pienaar J., Voje K.L., Orzack S.H., Hansen T.F. 2020. SLOUCH: Stochastic Linear Ornstein-Uhlenbeck Comparative Hypotheses. R package version 2.1.4. Available from: <https://CRAN.R-project.org/package=slouch>.
- Lynch M. 1991. *Methods for the analysis of comparative data in evolutionary biology*. *Evolution* 45:1065–1080.
- MacLean E.L. et al. 2014. The evolution of self-control. *Proc. Natl. Acad. Sci. USA* 111:E2140–E2148.
- Martins E.P. 1994. Estimating rates of character change from comparative data. *Am. Nat.* 144:193–209.
- Martins E.P., Hansen T.F. 1996. The statistical analysis of interspecific data: a review and evaluation of phylogenetic comparative methods. In: Martins E.P., editor. *Phylogenies and the comparative*

- method in animal behavior. Oxford: Oxford University Press. p. 22–75.
- Martins E.P., Hansen T.F. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* 149:646–667.
- May M.R., Moore B.R. 2020. A Bayesian approach for inferring the impact of a discrete character on rates of continuous-character evolution in the presence of background-rate variation. *Syst. Biol.* 69:530–544.
- McPeck M. 1995. Morphological evolution mediated by behavior in the damselflies of two communities. *Evolution* 49:749–769.
- Nunn C.L., Zhu L. 2014. Phylogenetic prediction to identify “evolutionary singularities”. In: Garamszegi L.Z., editor. *Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice*. Berlin, Heidelberg: Springer. p. 481–514.
- Odling-Smee F.J., Laland K.N., Feldman M.W. 2003. *Niche construction: the neglected process in evolution*. Monographs in population biology. Princeton: Princeton University Press.
- O’Meara B.C., Ane C., Sanderson M.J., Wainwright P.C. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60:922–933.
- Paenke I., Sendhoff B., Kawecki T.J. 2007. Influence of plasticity and learning on evolution under directional selection. *Am. Nat.* 170:E47–E58.
- Pennell M.W., FitzJohn R.G., Cornwell W.K., Harmon L.J. 2015. Model adequacy and the macroevolution of angiosperm functional traits. *Am. Nat.* 186:E33–E50.
- Pie M.R., Meyer A.L.S. 2017. The evolution of range sizes in mammals and squamates: heritability and differential evolutionary rates for low- and high-latitude limits. *Evol. Biol.* 44:347–355.
- Popper K.R. 1972. Evolution and the tree of knowledge, Chapter 7. In: *Objective knowledge: an evolutionary approach*. Revised edition. Oxford: Oxford University Press.
- Revell L.J. 2008. On the analysis of evolutionary change along single branches in a phylogeny. *Am. Nat.* 172:140–147.
- Revell L.J. 2013. A comment on the use of stochastic character maps to estimate evolutionary rate variation in a continuously valued trait. *Syst. Biol.* 62:339–345.
- Revell L.J., Harmon L.J. 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. *Evol. Ecol. Res.* 10:311–331.
- Revell L.J., Gonzalez-Valenzuela L.E., Alfonso A., Castaellanos-Garcia L.A., Guarnizo C.E., Crawford A.J. 2018. Comparing evolutionary rates between trees, clades and traits. *Methods Ecol. Evol.* 9: 994–1005.
- Revell L.J., Mahler D.L., Peres-Neto P.R., Redelings B.D. 2011. A new phylogenetic method for identifying exceptional phenotypic diversification. *Evolution* 66:135–146.
- Reznick D.N., Losos J., Travis J. 2019. From low to high gear: there has been a paradigm shift in our understanding of evolution. *Ecol. Lett.* 22:233–244.
- Sayol F., Lapiedra O., Ducatez S., Sol D. 2019. Large brains spur species diversification in birds. *Evolution* 73:2085–2093.
- Schluter D. 2000. *The ecology of adaptive radiation*. Oxford: Oxford University Press.
- Shultz S., Dunbar R.I.M. 2010. Social bonds in birds are associated with brain size and contingent on the correlated evolution of life-history and increased parental investment. *Biol. J. Linn. Soc.* 100:111–123.
- Simpson G.G. 1944. *Tempo and mode in evolution*. New York: Columbia University Press.
- Slater G.J., Pennell M.W. 2014. Robust regression and posterior predictive simulation increase power to detect early bursts of trait evolution. *Syst. Biol.* 63: 293–308.
- Sol D., Price T.D. 2008. Brain size and the diversification of body size in birds. *Am. Nat.* 172:170–177.
- Striedter G.F. 2005. *Principles of brain evolution*. Sunderland (MA): Sinauer Associates.
- Thomas G.H., Freckleton R.P., Szekely T. 2006. Comparative analysis of the influence of developmental mode on phenotypic diversification rates in shorebirds. *Proc. R. Soc. B* 273:1619–1624.
- Tsuboi M., van der Bijl W., Kopperud B.T., Erritzoe J., Voje K.L., Kotrschal A., Yopak K.E., Collin S.P., Iwaniuk A.N., Kolm N. 2018. Breakdown of brain-body allometry and the encephalization of birds and mammals. *Nat. Ecol. Evol.* 2:1492–1500.
- Uyeda J.C., Caetano D.S., Pennell M.W. 2015. Comparative analysis of principal components can be misleading. *Syst. Biol.* 64:677–689.
- Uyeda J.C., Hansen T.F., Arnold S.J., Pienaar J. 2011. The million-year wait for macroevolutionary bursts. *Proc. Natl. Acad. Sci. USA* 108:15908.
- Uyeda J.C., Zenil-Ferguson R., Pennell M.W. 2018. Rethinking phylogenetic comparative methods. *Syst. Biol.* 67:1091–1109.
- Vinze O. 2016. Light enough to travel or wise enough to stay? Brain size evolution and migratory behavior in birds. *Evolution* 70:2123–2133.
- Voje K.L. 2016. Tempo does not correlate with mode in the fossil record. *Evolution* 70:2678–2689.
- Voje K.L., Di Martino E., Porto A. 2020. Revisiting a landmark study system: no evidence for a punctuated mode of evolution in *Metrarabdotos*. *Am. Nat.* 195:899–917.
- Voje K.L., Starrfelt J., Liow L.H. 2018. Model adequacy and microevolutionary explanations for stasis in the fossil record. *Am. Nat.* 191:509–523.
- Wake D.B., Roth G., Wake M. 1983. On the problem of stasis in organismal evolution. *J. Theor. Biol.* 101: 211–224.
- Weir J.T., Lawson A. 2016. Evolutionary rates across gradients. *Methods Ecol. Evol.* 6:1278–1286.
- Wilson A.C. 1985. The molecular basis of evolution. *Sci. Am.* 253:164–173.
- Wyles J.S., Kunkel J.G., Wilson A.C. 1983. Birds, behavior, and anatomical evolution. *Proc. Natl. Acad. Sci. USA* 80:4394–4397.
- Zelditch M.L., Li J., Swiderski D.L. 2020. Stasis of functionally versatile specialists. *Evolution* 74:1356–1377.
- Zuk M., Bastiaans E., Langkilde T., Swanger E. 2014. The role of behavior in the establishment of novel traits. *Anim. Behav.* 92:333–344.