

## Likelihood-Based Parameter Estimation for High-Dimensional Phylogenetic Comparative Models: Overcoming the Limitations of “Distance-Based” Methods

ERIC W. GOOLSBY\*

Arnold Arboretum, Harvard University, Boston, MA 02131, USA and Interdisciplinary Toxicology Program, Department of Plant Biology, University of Georgia, Athens, GA 30602, USA.

\*Correspondence to be sent to: Arnold Arboretum, Harvard University, 1300 Centre Street, Boston, MA 02131, USA;  
E-mail: eric.goolsby.evolution@gmail.com.

Received 2 November 2015; reviews returned 21 April 2016; accepted 23 May 2016  
Associate Editor: Luke Harmon

**Abstract.**—Recently, a suite of distance-based multivariate phylogenetic comparative methods has been proposed for studying the evolution of high-dimensional traits, such as morphometric coordinates, gene expression data, and function-valued traits. These methods allow for the statistical comparison of evolutionary rates, assessment of phylogenetic signal, and tests of correlated high-dimensional trait evolution. Simulations reveal that distance-based comparative methods exhibit low statistical power and high Type I error under various evolutionary scenarios. Distance-based methods are also limited to relatively simple model specification (e.g., Brownian motion evolution) due to the lack of a likelihood function for parameter estimation. Here I propose an alternative method for studying high-dimensional trait evolution which overcomes some of the statistical limitations associated with distance-based methods. This framework, based on parametric bootstrapping and maximum pseudolikelihood parameter estimation, opens up the ability to estimate alternative evolutionary models, combine multiple evolutionary hypotheses, and potentially allow missing data and within-species variation. Simulations reveal that pairwise composite likelihood methods demonstrate appropriate Type I error and high statistical power, thus providing a robust framework for studying high-dimensional trait evolution. These methods are implemented in the R package *phylocurve*. [Covariance; distance; evolutionary rate; function-valued trait; high-dimensional; morphometric; multivariate; pairwise composite likelihood; phylogenetic comparative method; phylogenetic generalized least squares; phylogenetic signal.]

Phylogenetic comparative methods provide a framework for testing hypotheses in comparative biology while accounting for statistical nonindependence of hierarchically related species. In recent years, multivariate traits in the context of comparative data have been of increasing interest, leading to the development of multivariate extensions of phylogenetic comparative methods (Felsenstein 1981; Hansen and Martins 1996; Revell and Collar 2009; Bartoszek et al. 2012; Adams 2013). For certain kinds of multivariate traits, such as coordinates of morphometric traits (Adams 2014a–c; Adams and Felice 2014; Adams and Collyer 2015; Denton and Adams 2015), gene expression data (Dunn et al. 2013), and function-valued traits (Goolsby 2015), the number of trait dimensions is often very large relative to the number of species in a study. Such traits, known as *high-dimensional traits*, pose multiple computational and statistical challenges. Namely, as the number of parameters to be estimated increases for a given sample size, statistical power decreases substantially (Adams 2014b). Additionally, when the number of trait dimensions equals or exceeds the number of species in a study, maximum likelihood trait covariance matrices are non-invertible and thus cannot be used for calculations central to most phylogenetic comparative methods (Adams 2014b). Finally, even assuming computational feasibility and adequate statistical power, it may be difficult to draw unified conclusions from individual (and potentially conflicting) trait dimensions. For example, Adams (2014a) considered 11 anatomical landmarks describing head shape evolution in *Plethodon* salamanders, and each landmark is composed of two dimensions ( $x$  and  $y$

coordinates), resulting in a 22-dimensional multivariate trait. Any individual  $x$  or  $y$  coordinate for a given landmark conveys relatively little information about the overall evolution of head shape. Accordingly, it may be helpful to approach comparative analyses of such traits with single generalized multivariate metrics rather than multidimension-specific metrics.

Adams (2014a–c) proposed a suite of multivariate phylogenetic comparative methods for studying high-dimensional traits while maintaining statistical power and providing generalized test statistics for multivariate traits as a whole. The methods, which include multivariate extensions of Blomberg’s  $K$  (Adams 2014a), phylogenetic generalized least squares (PGLS) (Adams 2014b; Adams and Collyer 2015), comparisons of evolutionary rates (Adams 2014c; Denton and Adams 2015), and phylogenetic partial least squares (Adams and Felice 2014), avoid the problem of dealing with non-invertible covariance matrices and can even handle traits in which the number of dimensions far exceeds the number of species in the study. This is accomplished by phylogenetic transformation of the data and subsequent distance-based (Q-mode), rather than conventional covariance-based (R-mode), analyses. When applied to a single univariate trait, distance-based methods and conventional phylogenetic comparative methods provide identical results (Adams 2014a–c). When applied to higher dimensional traits, such as landmark coordinates of morphometric shape data (e.g., leaf shape coordinates obtained from Procrustes analysis (Chitwood et al. 2014) or function-valued traits (e.g., species reaction norms of a phenotypically plastic trait (Goolsby 2015)), distance-based comparative methods fit

a single consensus metric that attempts to capture the variation of the entire high-dimensional trait as a whole.

#### *Limitations of Distance-Based Comparative Methods*

Despite potential advantages, the distance-based comparative framework involves considerable shortcomings. For one, distance-based methods require eigen decomposition and inversion of the phylogenetic covariance matrix, which is extremely time-consuming and inefficient for large phylogenies. However, distance-based (Q-mode) methods can be equivalently expressed in a fully covariance-based (R-mode) framework (for an in-depth comparison of distance-based and covariance-based formulas, see Appendix 1), thus allowing for fast linear-time calculations and thereby reducing computational time by up to several orders of magnitude (Appendix 2; Figs. A1–2) (Felsenstein 1973; FitzJohn 2012; Freckleton 2012; Ho and Ané 2014).

Beyond computational issues, distance-based methods lack a likelihood function. Accordingly, distance-based model parameters can only be computed using closed-form solutions. For example, the distance-based evolutionary rate for high-dimensional traits (Adams 2014c) is simply the arithmetic mean of the univariate Brownian motion rate parameters, which can be computed by averaging the mean squared phylogenetically independent contrasts for each trait separately (Felsenstein 1985; see Appendix 1). However, parameters for many types of models (e.g., non-Brownian evolution, missing data, within-species variation) lack closed-form solutions and are thus excluded from the distance-based framework. Furthermore, model selection is limited to relatively simple hypothesis tests, lacking any clear way to compare complex combinations of model specifications (e.g., simultaneous modeling of fixed effects, phenotypic integration, rate heterogeneity, etc.). Most importantly, distance-based methods suffer from high Type I error and low statistical power under a variety of evolutionary scenarios, as demonstrated in the remainder of this article.

In this article, I propose a new method for studying high-dimensional trait evolution which overcomes some of the limitations associated with distance-based methods. In the “Methods” section, I describe the statistical properties of this method, which is based on parametric bootstrapping (Boettiger et al. 2012), and how it can be used to perform phylogenetic comparative tests that are analogous to existing distance-based methods, including statistically comparing evolutionary rates, testing for evolutionary correlations, and evaluating phylogenetic signal. I also introduce methods for estimating alternative (non-Brownian) evolutionary models, combining multiple evolutionary hypotheses, and studying higher dimensional traits (>1000). In the “Results” section, the statistical performance (Type I error and statistical power) of these methods is compared

to their distance-based counterparts via simulation, and the performance of new methods for estimating alternative evolutionary models and higher dimensional traits is also assessed.

## METHODS

Because there is no likelihood function for distance-based methods, a covariance-based framework is vital to expanding the capabilities of high-dimensional comparative methods. For multivariate phylogenetic comparative models, the log-likelihood function is defined as

$$\log L(\boldsymbol{\theta}|\mathbf{Y}) = -\frac{1}{2}((\mathbf{Y}-E(\mathbf{Y}))^t(\mathbf{W})^{-1}(\mathbf{Y}-E(\mathbf{Y}))) + \log|\mathbf{W}| + NM\log(2\pi) \quad (1)$$

and the restricted log-likelihood is defined as

$$\log L(\boldsymbol{\theta}|\mathbf{Y}) = -\frac{1}{2}(((\mathbf{Y}-E(\mathbf{Y}))^t(\mathbf{W})^{-1}(\mathbf{Y}-E(\mathbf{Y}))) + \log|\mathbf{W}| + \log|\mathbf{X}^t(\mathbf{W})^{-1}\mathbf{X}| + (NM-M)\log(2\pi)) \quad (2)$$

where  $N$  is the number of species,  $M$  is the number of traits (dimensions),  $\mathbf{Y}$  contains species values for each trait stacked into a single  $NM \times 1$  column vector,  $\mathbf{X}$  is an  $NM \times M$  matrix consisting of ones and zeros describing which rows and columns correspond to elements of  $\mathbf{Y}$  ( $X_{ij}=0$  when  $i \neq j$  and  $X_{ij}=1$  when  $i=j$ ), two vertical lines denotes the determinant, and  $\mathbf{W}$  is the species-trait covariance matrix describing the evolutionary model (e.g., for a simple Brownian motion model,  $\mathbf{W}=\mathbf{R} \otimes \mathbf{C}$ , which is the Kronecker product of the  $M \times M$  evolutionary rate matrix ( $\mathbf{R}$ ) and the  $N \times N$  phylogenetic species covariance matrix ( $\mathbf{C}$ )). Unfortunately, the likelihood function become unstable as the number of traits approaches the number of species, and is undefined when  $M \geq N$ . Additionally, the number of parameters to estimate for an  $M \times M$  symmetric matrix is  $(M^2+M)/2$ , which may fail to converge on the maximum likelihood parameters even with moderate trait dimensionality.

#### *Pairwise Composite Likelihood for High-Dimensional Comparative Models*

A potential solution to this dilemma is the substitution of a pseudolikelihood metric into existing likelihood-based estimation methods. It has been shown that the product of the likelihoods (i.e., the sum of the log-likelihoods) for all possible pairwise combinations of variables, termed *pairwise composite likelihood*, shares many desirable properties with the full likelihood function. In particular, maximum pairwise composite likelihood estimates are consistent, unbiased, and asymptotically normal (Cox and Reid 2004; Varin and Vidoni 2005; Fieuwis and Verbeke 2006). It is straightforward then to reduce high-dimensional

problems into multiple small maximum likelihood estimation problems. Specifically, for  $M$  traits,  $(M^2 - M)/2$  pairwise models must be estimated. Although these are a large number of models, each individual estimation problem contains an extremely small number of parameters that can be estimated for each pairwise trait combination using efficient linear-time computations (Felsenstein 1973; Freckleton 2012; Ho and Ané 2014). Additionally, the log-likelihood function for each pairwise trait combination is computationally stable because  $N \gg 2$ . For higher dimensional traits (e.g.,  $M > 1000$ ) in which the total number of pairwise likelihood combinations may be computationally prohibitive, pairwise composite log-likelihood can be approximated using Monte Carlo sampling.

#### *A General Framework for Hypothesis Testing and Model Selection using Pairwise Composite Likelihoods*

Here, I propose a parametric bootstrapping approach adapted from the simulation-based methods developed by Boettiger et al. (2012). A simulation-based approach is essential for the proposed framework, as pairwise likelihoods represent overlapping information (and are, therefore, nonindependent from one another), so pairwise composite likelihoods cannot be used for conventional model selection criteria such as Akaike information criterion (AIC), Bayesian information criterion (BIC), and likelihood ratio tests (Varin and Vidoni 2005). Similarly, standard errors of parameter estimates based on Fisher information matrices (which are also nonindependent among pairwise trait combinations) are uninterpretable (Fieuw and Verbeke 2006). To statistically compare two models, the pairwise composite log-likelihood is calculated for the null model (e.g., simple Brownian motion) and for the alternative model (e.g., Brownian motion with different rates for two groups of species, as in Adams 2014c). Next, data are simulated from the parameters of the null model 1000 (or more) times, and pairwise composite log-likelihoods are estimated for both the null and alternative models (refit to the simulated data). The likelihood ratio test statistic  $\delta = -2(\log L_{\text{null}} - \log L_{\text{alt}})$  is computed for the observed data ( $\delta_{\text{obs}}$ ) and for the data simulated under the null hypothesis ( $\delta_{\text{sim.null}}$ ). For a nominal significance level of  $P \leq 0.05$ , the critical value for the test statistic ( $\delta_*$ ) is set so that 95% of  $\delta_{\text{sim.null}}$  values fall under  $\delta_*$ . The proportion of times  $\delta_{\text{obs}} \geq \delta_{\text{sim.null}}$  provides an approximation of the  $P$ -value for comparing the null model to the alternative model. A similar procedure can be applied to assess the statistical power of a model comparison by simulating under the alternative model many times (e.g., 1000 or more) and calculating  $\delta_{\text{sim.alt}}$ . The proportion of  $\delta_{\text{sim.alt}}$  values greater than or equal to  $\delta_*$  provides an approximation of the statistical power of the test (Boettiger et al. 2012).

This procedure is extremely flexible and maintains appropriate Type I error and high statistical power for the comparative methods described here. In particular,

because the likelihood ratio statistic can be calculated for any model, complex combinations of multiple evolutionary hypotheses (e.g., fixed effects, multiple evolutionary rates, phylogenetic signal, etc.) may be incorporated simultaneously, whereas hypothesis tests for various distance-based models (which rely on model-specific metrics, such as ratios of  $F$ -statistics or evolutionary rates) cannot be combined.

The approach described above (as well as the methods described in the following sections) is implemented in the R package *phylocurve*, in which null and alternative models can be fit separately using the *evo.model* function and subsequently compared via parametric bootstrapping (Boettiger et al. 2012) using the *compare.models* function (R Core Team 2016).

#### *Statistically Comparing Evolutionary Rates*

Adams (2014c) proposed a distance-based method for comparing evolutionary rates for high-dimensional traits among *a priori* designated groups of species, hereafter referred to as *regimes*. Comparisons of evolutionary rates may be performed to test a variety of questions involving the interplay between phenotypic diversification and key factors such as morphology, ecological niche, speciation/extinction events, and physiology (O'Meara et al. 2006; Thomas et al. 2006; Adams 2014c). For example, to test whether a high-dimensional character (such as leaf shape morphometrics) evolves faster in one group of species versus another (e.g., annuals vs. perennials), one may test the null hypothesis of equal evolutionary rates for all species in the phylogeny versus the alternative hypothesis of statistically different evolutionary rates between the two regimes. As described later in this article, the distance-based approach to comparing evolutionary rates suffers several statistical drawbacks. To overcome these drawbacks, I propose an alternative method for statistically comparing evolutionary rates among regimes using the parametric bootstrapping method described above.

To simulate the null hypothesis for comparing evolutionary rates among regimes, simulate trait evolution (e.g., using the *sim.char* function in *geiger* (Pennell et al. 2014)) under the restricted maximum likelihood evolutionary rate matrix (equation (A9); for the maximum likelihood rate matrix,  $N-1$  is simply replaced with  $N$ ). To simulate the alternative hypothesis (distinct evolutionary rates among regimes), the following procedure is used:

1. Estimate the evolutionary rate matrix for each regime ( $\mathbf{R}_j$ ) using either the censored method described in O'Meara et al. (2006), the noncensored approach (O'Meara et al. 2006), or by subsetting transformed residuals (Adams 2014c).
2. Determine the proportion of each tree edge to be assigned to each respective regime rate matrix, for instance, by assigning entire clades values of either

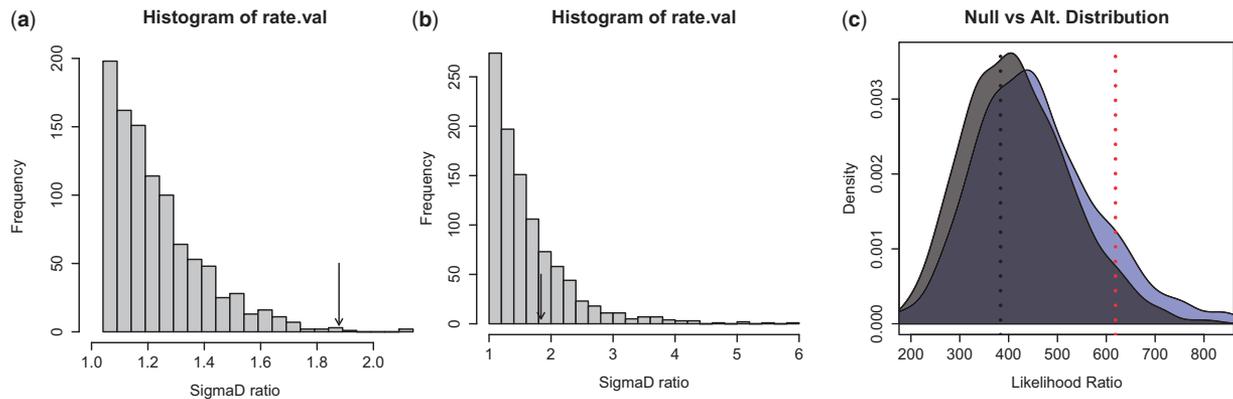


FIGURE 1. Comparison of rate differences among two regimes from nine *Plethodon* salamanders for 11 cranial landmark coordinates (Adams 2014c) using a) distance-based hypothesis testing while failing to account for trait covariance, as in Adams (2014c) (performed using the *phylocurve* function *fast.geomorph.compare.evol.rates* with *force.diag=FALSE*), b) distance-based hypothesis testing while accounting for trait covariance (performed using the *phylocurve* function *fast.geomorph.compare.evol.rates* with *force.diag=FALSE*), and c) pairwise composite likelihood-based hypothesis testing and statistical power assessment (performed using the *phylocurve* function *compare.models* on null and alternative models both fit using the *evo.model* function), where the left distribution is the null distribution (no difference in rate ratio) and the right distribution is the alternative distribution (significant difference in rate ratio); the left vertical dotted line is the observed test statistic and the right vertical line is the critical value for the test statistic (indicating that the observed test statistic falls within the null distribution). While failing to account for trait covariance, the rate ratio between the two hypothesized regimes (1.84) was found to be significantly different from 1.0 ( $P=0.005$ ), identical to the result reported in Adams (2014c). While accounting for trait covariance, the rate ratio (1.84) was found to be nonsignificant ( $P=0.261$ ). Pairwise composite likelihood-based hypothesis testing was also nonsignificant ( $P=0.64$ ) and revealed very little statistical power (0.097, the proportion of simulated alternative models greater than the critical value) to detect a significant rate difference for the given the phylogeny and data set.

zero or one (for known discrete regime shifts), or by reconstructing the probabilities of ancestral regime states (Yang et al. 1995; Pupko et al. 2000; Paradis et al. 2004; Revell 2012).

3. Simulate phenotypic evolution under the evolutionary rate matrix for each regime by scalar multiplication of  $\mathbf{R}_j$  (the evolutionary rate matrix for regime  $j$ ) by the branch lengths of the phylogeny and by the regime-specific proportions determined in (2).
4. Add the resulting simulated phenotypic values together. Note that this procedure differs substantially from the null hypothesis described in Adams (2014c), which assumes a diagonal evolutionary rate matrix and results in high Type I error for correlated traits (see below for more detail). In contrast, the procedure described here results in appropriate Type I error and statistical power to compare evolutionary rates among regimes.

To assess statistical significance and power, compare the pairwise composite likelihood test statistic  $\delta_{\text{obs}}$  to the null and alternative distributions. Although the raw values of evolutionary rates themselves may be of limited usefulness, a statistically significant rate comparison suggests a deviation from simple Brownian motion evolution, whereas a nonsignificant result may be the result of either a lack of a true difference in rates or an unidentifiable signal. The degree of overlap in the null and alternative distributions and the relative positions of  $\delta_{\text{obs}}$  and  $\delta_*$  can help distinguish whether the test

was adequately powered to detect shifts in evolutionary rates, or if the null and alternative hypotheses were statistically indistinguishable (see Fig. 1c)

Denton and Adams (2015) also described a method for statistically comparing evolutionary rates among two high-dimensional traits, which may be used to test a variety of questions regarding phenotypic evolution and modularity. For example, one may test the hypothesis that leaf shape evolution is slower among leaves produced during the plant seedling stage than for leaves produced during plant maturity. To simulate the null hypothesis under this scenario, construct a modified evolutionary rate matrix in which the diagonal of  $\mathbf{R}$  is constrained to equal  $\sigma_{\text{mult}}^2$  (the mean of the diagonal of the evolutionary rate matrix  $\mathbf{R}$ ). For the alternative hypothesis, the diagonal of  $\mathbf{R}$  is divided into trait groups subsets, in which subset  $k$  is set to equal  $\sigma_{\text{trait.sub}_k}^2$  (the mean of the diagonal of the subset of  $\mathbf{R}$  corresponding to traits represented in trait group  $k$ ).  $\delta_{\text{obs}}$  is then statistically compared to the null and alternative distributions. In many cases, the resulting matrix is not positive semidefinite. Following Denton and Adams (2015), the nearest positive definite matrix to the constrained rate matrix is found using the *nearPD* function in the *Matrix* package for trait simulations (Bates and Maechler 2015). As with comparisons of evolutionary rates among regimes, the raw values of evolutionary rates for individual traits may not be particularly meaningful, whereas the relative difference in rates and the statistical significance of the test may corroborate or contradict existing explanations for observed phenotypic patterns. However, it is important

to note that the reliability of this test depends strongly on the appropriateness of the null hypothesis of equal rates. For example, comparisons of traits measured in different units (e.g., mm and cm) or comparisons of non-analogous traits (e.g., morphometrics and function-valued traits) are inappropriate, as there would be no reasonable expectation of equally scaled rates of evolution among traits.

#### Testing for Evolutionary Correlations

Adams (2014b) described a distance-based method for testing the significance of evolutionary correlations between fixed effects (e.g., univariate environmental variables) and high-dimensional traits in a multivariate extension of evolutionary regression called *D*-PGLS. As with univariate evolutionary regression (i.e., PGLS or phylogenetically independent contrasts regressed through the origin), *D*-PGLS is intended to test for the significance of correlations between one or more predictor variables and a response variable (here a high-dimensional trait). However, as with other distance-based methods described in this article, distance-based *D*-PGLS performs poorly under certain scenarios, and an alternative method using parametric bootstrapping is described here. To test the significance of fixed effects (as in *D*-PGLS (Adams 2014b)), simulate both the null and alternative hypotheses under  $\mathbf{R}$ , and add  $\mathbf{X}\beta$  (equation (A8)) to the simulated  $\mathbf{Y}$  under the alternative hypothesis. It should be noted that restricted likelihood (or restricted pairwise composite likelihood) cannot be used for model comparisons in which fixed effects differ between the null and alternative hypotheses. Instead, comparisons of  $\delta_{\text{obs}}$  with the null and alternative distributions must be made using maximum pairwise composite likelihood estimates, and a significant result suggests an evolutionarily conserved relationship between one or more predictor variables and the multivariate response trait.

Adams and Felice (2014) described a method for assessing correlated evolution among two high-dimensional traits (termed *phylogenetic partial least squares*). This is conceptually similar to phylogenetic regression, except that both variables are multivariate. For example, one could test whether the evolution of leaf shape is evolutionarily correlated with the evolution of species reaction norms. However, phylogenetic partial least squares suffers severely elevated Type I error when within-trait covariance is low (see below), so an analogous framework which avoids these issues is described here. To assess correlations between two sets of multivariate traits in the pairwise composite likelihood framework, estimate the evolutionary rate matrix  $\mathbf{R}$  (equation (A9)) and partition  $\mathbf{R}$  into four blocks ( $\mathbf{R}_{11}$ ,  $\mathbf{R}_{21}$ ,  $\mathbf{R}_{12}$ ,  $\mathbf{R}_{22}$ —see equation (A10)) (Adams and Felice 2014; see also Revell and Harrison 2008). Rather than performing singular value decomposition on the evolutionary rate matrix as in Adams and Felice (2014), simulate the null hypothesis by setting all elements

of blocks  $\mathbf{R}_{12}$  and  $\mathbf{R}_{21}$  to zero. For the alternative hypothesis, simulate data under the unconstrained rate matrix ( $\mathbf{R}$ ), and compare  $\delta_{\text{obs}}$  to the null and alternative distributions to assess statistical significance. A significant result indicates evolutionarily correlated changes between the two multivariate traits, known as phenotypic integration. In a similar manner, the presence of significant evolutionary covariation among traits can be tested by setting non-diagonal elements of the rate matrix to zero for null hypothesis simulations.

#### Phylogenetic Signal

Adams (2014a) introduced a distance-based extension of Blomberg's *K* (Blomberg et al. 2003) for testing phylogenetic signal for high-dimensional traits. The parametric bootstrapping framework described here can also be used to test for phylogenetic signal in high-dimensional traits. Rather than phylogenetic permutation (as in Blomberg et al. 2003 and Adams 2014a), I propose the use of phylogenetic simulation for both the null and alternative hypotheses. Under Blomberg's *K*, the null hypothesis is an absence of phylogenetic signal, so data are simulated under  $\mathbf{R}$  on a star phylogeny; for the alternative hypothesis (Brownian motion), data are simulated under  $\mathbf{R}$  on the original phylogeny. For this procedure, *K* is used as the summary statistic (rather than the likelihood ratio  $\delta$ ). The null distribution of *K* is used to calculate the critical value ( $K_*$ ), and the proportion of  $K_{\text{obs}} \geq K_{\text{sim.null}}$  is the *P*-value for the test of phylogenetic signal. The mean of  $K_{\text{sim.alt}}$  provides the expectation of *K* under Brownian motion (which should be approximately 1.0 if the model is correctly specified), and the proportion of  $K_{\text{sim.alt}} \geq K_*$  provides an estimate of the statistical power to detect significant phylogenetic signal. To simplify calculations under complex evolutionary models, the expectation of the ratio of raw to phylogenetic mean squared error (the denominator of  $K_{\text{mult}}$ ) can be approximated by simulation under the alternative hypothesis (see below). This implementation of  $K_{\text{mult}}$  is implemented in the *phylocurve* function *K.mult*.

#### Estimation of Alternative Evolutionary Models

Parameters for alternative evolutionary models, such as Early-Burst (Harmon et al. 2010) or Ornstein–Uhlenbeck (Hansen 1997), or tree transformations such as Pagel's  $\lambda$  (Pagel 1999) can be fit to high-dimensional comparative models by transforming the branch lengths such that a Brownian motion-like process of trait evolution applies on the transformed tree. Estimation of maximum pairwise composite likelihood tree transformation parameters proceeds as follows: (i) transform the phylogeny according to an initial guess for a tree transformation parameter (e.g.,  $\alpha$  for an Ornstein–Uhlenbeck process); (ii) estimate maximum likelihood parameters for each pairwise combination of traits using closed-form solutions (if available) or

by numerical optimization; (iii) sum the pairwise log-likelihoods for each combination of traits; (iv) repeat steps 1–3 with a new guess for the tree transformation parameter until convergence on the maximum pairwise composite likelihood estimate is achieved. To assess whether the evolutionary model provides a significantly better fit than the null model (e.g., Brownian motion), perform parametric bootstrap simulations and compare  $\delta_{\text{obs}}$  to the null distribution, as described in the previous sections.

#### Combining Multiple Evolutionary Hypotheses

The hypothesis testing framework described above (based on Boettiger et al. 2012) allows for straightforward combinations of multiple evolutionary hypotheses by imposing appropriate alterations of or constraints on the evolutionary rate matrix or by adding predicted values based on fixed effects to simulated data ( $\mathbf{Y}_{\text{sim}} + \mathbf{X}\beta$ ). To incorporate multiple models into tests of phylogenetic signal,  $K_{\text{mult}}$ , the expectation of the ratio of raw mean squared error to phylogenetically corrected mean squared error

$$E(\text{MSE}_0/\text{MSE}) = E\left(\frac{\sum_{i=1}^M ((\mathbf{y}_i - E(\mathbf{y}_i))^t (\mathbf{y}_i - E(\mathbf{y}_i)))}{\sum_{i=1}^M ((\mathbf{y}_i - E(\mathbf{y}_i))^t \mathbf{C}^{-1} (\mathbf{y}_i - E(\mathbf{y}_i)))}\right) \quad (3)$$

is estimated by simulation under the hypothesized evolutionary model. Under simple Brownian motion, this should yield an estimate of approximately  $(\text{tr}(\mathbf{C}) - N(1^t \mathbf{C}^{-1} 1)^{-1}) / (N - 1)$  (Blomberg et al. 2003). The observed ratio  $\text{MSE}_0/\text{MSE}$  is then scaled by  $E(\text{MSE}_0/\text{MSE})$  to calculate  $K_{\text{mult}}$ :

$$K_{\text{mult}} = \frac{\sum_{i=1}^M ((\mathbf{y}_i - E(\mathbf{y}_i))^t (\mathbf{y}_i - E(\mathbf{y}_i)))}{\sum_{i=1}^M ((\mathbf{y}_i - E(\mathbf{y}_i))^t \mathbf{C}^{-1} (\mathbf{y}_i - E(\mathbf{y}_i)))} \cdot E(\text{MSE}_0/\text{MSE}) \quad (4)$$

Next, the hypothesized model is simulated on a star phylogeny to obtain the null distribution of  $K_{\text{mult}}$ , which is used to calculate the critical value ( $K_*$ ). As before, the proportion of  $K_{\text{obs}} \geq K_{\text{sim.null}}$  is the  $P$ -value for testing phylogenetic signal, and the proportion of  $K_{\text{sim.alt}} \geq K_*$  provides an estimate of the statistical power of the test. Under simple Brownian motion, equations (A6) and (A7) should be nearly identical to equation (4), whereas under deviations from simple Brownian motion, equation (4) yields a generalization of  $K_{\text{mult}}$  which can incorporate fixed effects (by setting  $E(\mathbf{Y}) = \mathbf{X}\beta$ ), multiple evolutionary rate regimes, non-Brownian evolutionary models, and other model specifications. These features are implemented in the *phylocurve* function *K.mult*.

#### Incorporation of Missing Data and Within-Species Variation

Several methods have been developed to estimate evolutionary trait covariance in the presence of missing data and within-species variation (Ives et al. 2007; Felsenstein 2008; Bruggeman et al. 2009; Hansen and Bartoszek 2012). These methods can all be estimated in a maximum likelihood (or restricted likelihood) framework and can be readily incorporated into pairwise composite likelihood estimation. However, for such methods, there is no algorithmic solution for estimating evolutionary covariances, so step 2 of the outlined method for parameter estimation (see above) requires numerical optimization of covariance parameters. Problematically, estimated evolutionary rates for data sets with missing data or within-species variation will be different if parameters for pairwise trait combinations are estimated separately. One solution is to maximize the pairwise log-likelihood for all parameters in a single optimization routine, which constrains variance parameters for individual traits to be identical across pairwise combinations. This is a potentially large optimization problem which may have difficulty converging. A potential simplification involves separate maximum likelihood estimation for each pairwise trait combination (as described above), followed by averaging each estimate of the evolutionary rate for a given trait. The resulting estimates are not equivalent to the maximum pairwise composite likelihood estimates but should provide reasonable starting parameters for a single optimization routine. Alternatively, the averaged estimates could be used as is to approximate the maximum pairwise composite likelihood parameters. Optimal strategies for incorporating within-species variation and missing data for high-dimensional models require further investigation that is beyond the scope of this article (but see Denton and Adams 2015, in which a bootstrapping approach is performed on randomly sampled within-species individual observations). Approaches for dealing with within-species variation and missing data are not yet (as of the time of publication) implemented in *phylocurve*.

#### Higher-Dimensional Traits

The pairwise composite log-likelihood approach may become excessively cumbersome for extremely high-dimensional traits. For example, for a 1024-dimensional trait there are 523,776 pairwise trait combinations, a prohibitively large number for parameter estimation, as the pairwise log-likelihood must be calculated several times for numerical optimization. For simulation-based hypothesis testing, this optimization procedure must be repeated a large number of times. For instance, supposing numerical optimization of a tree transformation parameter (e.g., Early-Burst rate) requires 50 log-likelihood calculations, and 1000 null hypothesis simulations are performed, a total of  $50 \times 1000 \times 523,776 = 26,188,800,000$  pairwise

log-likelihood evaluations would be required. For such computationally infeasible problems, I propose a simple Monte Carlo-based approach. Rather than computing every possible pairwise log-likelihood, randomly sample a subset of pairwise combinations, then divide the subsetted pairwise composite log-likelihood by the number of random samples, and then multiply by the total number of possible pairwise combinations. As the number of random samples increases, the estimated composite log-likelihood will approach the true pairwise composite log-likelihood. Next, approximate the composite log-likelihood surface for the tree transformation parameter by estimating the composite log-likelihood for several parameter values spanning the range of feasible parameters, and then regress parameters against the approximated log-likelihood surface (e.g., polynomial or Gaussian process regression). Finally, identify the parameter value corresponding to the maximum predicted value of the regression, and use this value as the parameter estimate. This method can be performed in the *phylocurve* function *evo.model* by setting the argument *max.combn* to a value lower than the number of pairwise trait combinations  $((M^2 - M)/2)$ .

#### Simulation Methods

To compare the Type I error and statistical power of distance-based methods with the pairwise composite log-likelihood approaches described here, simulations were performed under a variety of evolutionary scenarios. For each scenario, 1000 data sets were simulated on 32-species pure-birth phylogenies using the *pbtree* function in *phytools* (Revell 2012), and subsequently using the *sim.traits* and *sim.groups* functions in *phylocurve* (both of which call the *sim.char* function in *geiger* (Pennell et al. 2014)). Hypothesis tests were performed using both distance-based and pairwise composite log-likelihood-based parametric bootstrapping simulation (Boettiger et al. 2012), with 1000 bootstrap simulations per simulated data set.

To evaluate the statistical performance of comparing evolutionary rates among regimes, two regimes were assigned 16 species each. For each simulation, an evolutionary rate matrix was simulated by parameterizing the upper triangle of an  $M \times M$  matrix drawn from the standard normal distribution. The transpose of the matrix was then matrix multiplied by itself, resulting in a positive-definite covariance matrix ( $\mathbf{R}_{\text{sim}}$ ). For one of the simulated regimes, traits were simulated under  $\mathbf{R}_{\text{sim}}$ ; the other regime was simulated under  $\mathbf{R}_{\text{sim}}$  scalar multiplied by either 1.0 (for Type I error), 1.5, 2.0, 3.0, or 4.0. Simulations were conducted for 2, 16, 32, and 64 traits. Similarly, the statistical performance of comparing evolutionary rates among traits was evaluated by generating  $\mathbf{R}_{\text{sim}}$ , subdividing the diagonal of  $\mathbf{R}_{\text{sim}}$  into two groups, and setting one group to the mean of the diagonal of  $\mathbf{R}_{\text{sim}}$  and the other group to the mean of the diagonal of  $\mathbf{R}_{\text{sim}}$  multiplied by either

1.0 (for Type I error), 1.5, 2.0, 3.0, or 4.0. Simulations were conducted for 4, 8, 16, 32, and 64 traits. For simulations involving fixed effects (for comparison with *D*-PGLS), the diagonal of  $\mathbf{R}$  (of dimension  $(M+1) \times (M+1)$ ) was set to 1.0 and off-diagonal elements were set to either 0.0 (for Type I error), 0.1, 0.3, 0.5, 0.7, or 0.9, and data were simulated for 2, 5, 10, 25, and 32 traits. To assess the statistical performance of testing evolutionary covariance between two groups of multivariate traits (for comparison with phylogenetic partial least squares), the diagonal of  $\mathbf{R}$  (of dimension  $M \times M$ ) was set to 1.0 and off-diagonal elements were set to either 0.0 (for Type I error), 0.1, 0.3, 0.5, 0.7, or 0.9. To test the power of multivariate phylogenetic signal ( $K_{\text{mult}}$ ), trait data were simulated under  $\mathbf{R}_{\text{sim}}$  on a transformed phylogeny with Pagel's  $\lambda$  set to either 0.0 (for Type I error), 0.05, 0.1, 0.25, 0.5, 0.75, and 1.0 for 2, 5, 10, 32, and 64 traits. For simulations testing evolutionary covariance, traits were simulated under  $\mathbf{R}$  with the diagonal set to 1.0 and off-diagonal elements set to either 0.0 (for Type I error), 0.1, 0.3, 0.5, 0.7, or 0.9 for 2, 5, 10, 32, and 64 traits.

Additionally, alternative evolutionary models were simulated for 2, 5, 10, 25, 31, and 50 traits, and simulations were performed on phylogenies with the following tree transformations: Early-Burst (rate = 0.0 (for Type I error), -0.25, -0.5, -0.75, -1.0), Ornstein-Uhlenbeck  $\alpha = 0.0$  (for Type I error), 0.25, 0.5, 1.0, 2.0), and Pagel's  $\lambda$  ( $\lambda = 1.0$  (for Type I error), 0.75, 0.5, 0.25, 0.0).

R code for performing the simulations used to generate Figures 2, 3, and 4 is provided in the Dryad Supplement available at <http://dx.doi.org/10.5061/dryadfh0mc>. Several *phylocurve* functions beginning with "fast.geomorph." (*fast.geomorph.compare.evol.rates*, *fast.geomorph.compare.multi.evol.rates*, *fast.geomorph.physignal*, *fast.geomorph.phylo.integration*, and *fast.geomorph.procD.pgls*) were developed to provide fast covariance-based implementations of equivalent *geomorph* functions (*compare.evol.rates*, *compare.multi.evol.rates*, *physignal*, *phylo.integration*, and *procD.pgls*). These functions, which produce identical results to distance-based methods, should not be confused with the pairwise composite likelihood framework described in this article, which is implemented in the *phylocurve* functions *evo.model*, *compare.models*, and *K.mult*.

## RESULTS

### Statistically Comparing Evolutionary Rates

For comparisons of evolutionary rates among regimes (Fig. 2b,d) and among groups of traits (Fig. 3a–b) using both distance-based and pairwise composite likelihood-based hypothesis testing, all simulated scenarios displayed appropriate Type I error (approximately 0.05). As expected, as trait dimensionality was increased, statistical power also increased (as in Adams (2014c) and Denton and Adams (2015)), although distance-based approaches exhibited somewhat higher statistical power for rate ratios of 1.5 and 2.0 (both methods

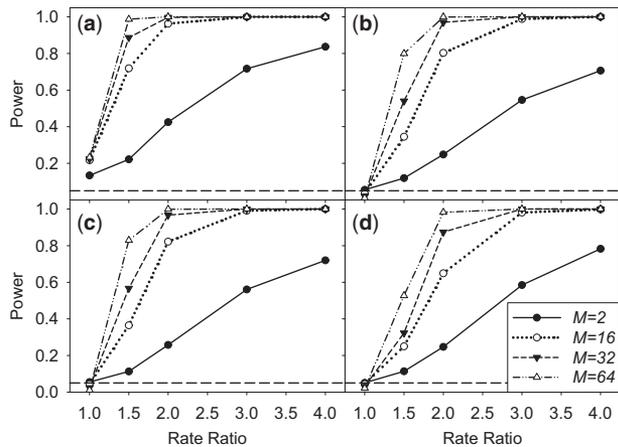


FIGURE 2. Statistical performance of evolutionary rate comparisons among regimes using a) distance-based hypothesis tests assuming a diagonal evolutionary rate matrix as was assumed in Adams (2014c) and implemented in the geomorph function `compare.evol.rates` in versions 1.1.5–2.1.5 (performed using the `fast.geomorph.compare.evol.rates` function in phylocurve with `force.diag=TRUE`), b) distance-based hypothesis tests assuming a fully parameterized evolutionary rate matrix is implemented in geomorph versions 2.1.6 and higher (performed using the `fast.geomorph.compare.evol.rates` function in phylocurve with `force.diag=FALSE`), c) distance-based hypothesis tests using a fully parameterized evolutionary rate matrix with regime-specific rates using the censored method of O’Meara et al. (2006) (performed using the `fast.geomorph.compare.evol.rates` function in phylocurve with `censored=TRUE`), and d) pairwise composite likelihood methods performed using the `compare.models` function in phylocurve. Failing to account for trait covariance (a) results in unacceptably high Type I error. Simulations were performed on 1000 simulated 32-species pure-birth phylogenies with 2, 16, 32, and 64 traits simulated under two different regimes of equal size with a rate ratio of 1.0 (for Type I error), 1.5, 2.0, 3.0, and 4.0.

had statistical power of approximately 1.0 at higher rate ratios). Regime-specific evolutionary rates were also compared using the censored method (Fig. 2c; O’Meara et al. 2006), which displayed appropriate Type I error and statistical power approximately identical to that of the distance-based procedure.

It should be noted that a modification to the null hypothesis described in Adams (2014c) was implemented for comparing rates among regimes. Adams (2014c) simulated the null distribution of rate ratios by setting the null hypothesis to be a diagonal evolutionary rate matrix, with each trait independently evolving under a Brownian motion rate of  $\sigma_{\text{mult}}^2$ . Because this approach fails to account for trait covariance, this procedure results in unacceptably high Type I error rates (Fig. 2a), as trait covariance artificially inflates the observed differences among regime-specific rates relative to independently simulated traits. This Type I error-prone procedure was implemented in the supplemental code provided in Adams (2014c) as well as in the `compare.evol.rates` function in *geomorph* up to version 2.1.5 (Adams and Otárola-Castillo 2013). As of version 2.1.6 of *geomorph*, the null hypothesis has been updated to account for trait covariance (Denton and Adams 2015). Likewise, earlier versions of *phylocurve*

(1.0.0–1.3.0), which implemented fast covariance-based analogues of distance-based approaches, also used the original null hypothesis specified in Adams (2014c). *phylocurve* versions 2.0.0 and higher implement the modified null hypothesis incorporating trait covariance (however, the `fast.geomorph.compare.evol.rates` can be used to perform hypothesis tests under the old null hypothesis, as described in Adams (2014c), by setting the `force.diag` option to `TRUE`).

As an example of the potential consequences of failing to account for trait covariance, data from Adams (2014c) consisting of 11 cranial landmark coordinates from 9 *Plethodon* salamanders were analyzed using the *phylocurve* function `fast.geomorph.compare.evol.rates` while assuming zero-trait covariance (`force.diag` set to `TRUE`). When failing to account for trait covariance, the rate ratio between the two hypothesized regimes (1.84) was found to be significantly different from 1.0 (Fig. 1a;  $P=0.005$ ), identical to the result reported in Adams (2014c). Next, the *phylocurve* function `compare.models` was used to test for significant trait covariation (where the null model was fit using *evo.model* with `diag.phylocov` set to `TRUE` and the alternative model was fit using *evo.model* with `diag.phylocov` set to `FALSE`). The null hypothesis of zero trait covariation was unequivocally rejected ( $P < 0.001$ ). Next, `fast.geomorph.compare.evol.rates` was used while accounting for trait covariance (`force.diag` set to `FALSE`), and the rate ratio (1.84) was found to be nonsignificant (Fig. 1a;  $P=0.261$ ), in conflict with Adams (2014c). To simultaneously assess the statistical significance and statistical power of the test using pairwise composite likelihood methods, the *phylocurve* function `compare.models` was used to compare the null hypothesis of a single rate regime and the alternative hypothesis of two regimes. This test, analogous to the `fast.geomorph.compare.evol.rates` test accounting for trait covariance, was also nonsignificant ( $P=0.64$ ) and revealed essentially no statistical power (0.097) to detect a significant rate difference given the phylogeny and data set (Fig. 1c).

#### Comparing Evolutionary Rates under Violations of Distance-Based Assumptions

Distance-based comparisons of evolutionary rates among regimes operate under the assumption that shifts in evolutionary rates are proportional among individual traits—that is, evolution is assumed to speed up or slow down equally for all traits by the same factor (e.g., half, double, quadruple, etc.). To assess the consequences of deviations from this model, statistical power was assessed under a scenario in which the evolutionary rates among two clades differed substantially but in conflicting ways. Specifically, a 100-dimensional trait was simulated on a 32-species phylogeny divided into two equally sized regimes. Traits 1–50 were simulated with an evolutionary rate of 1.0 for the first regime (species 1–16) and an evolutionary rate of 4.0 for the second regime (species 17–32). Traits 51–100 were

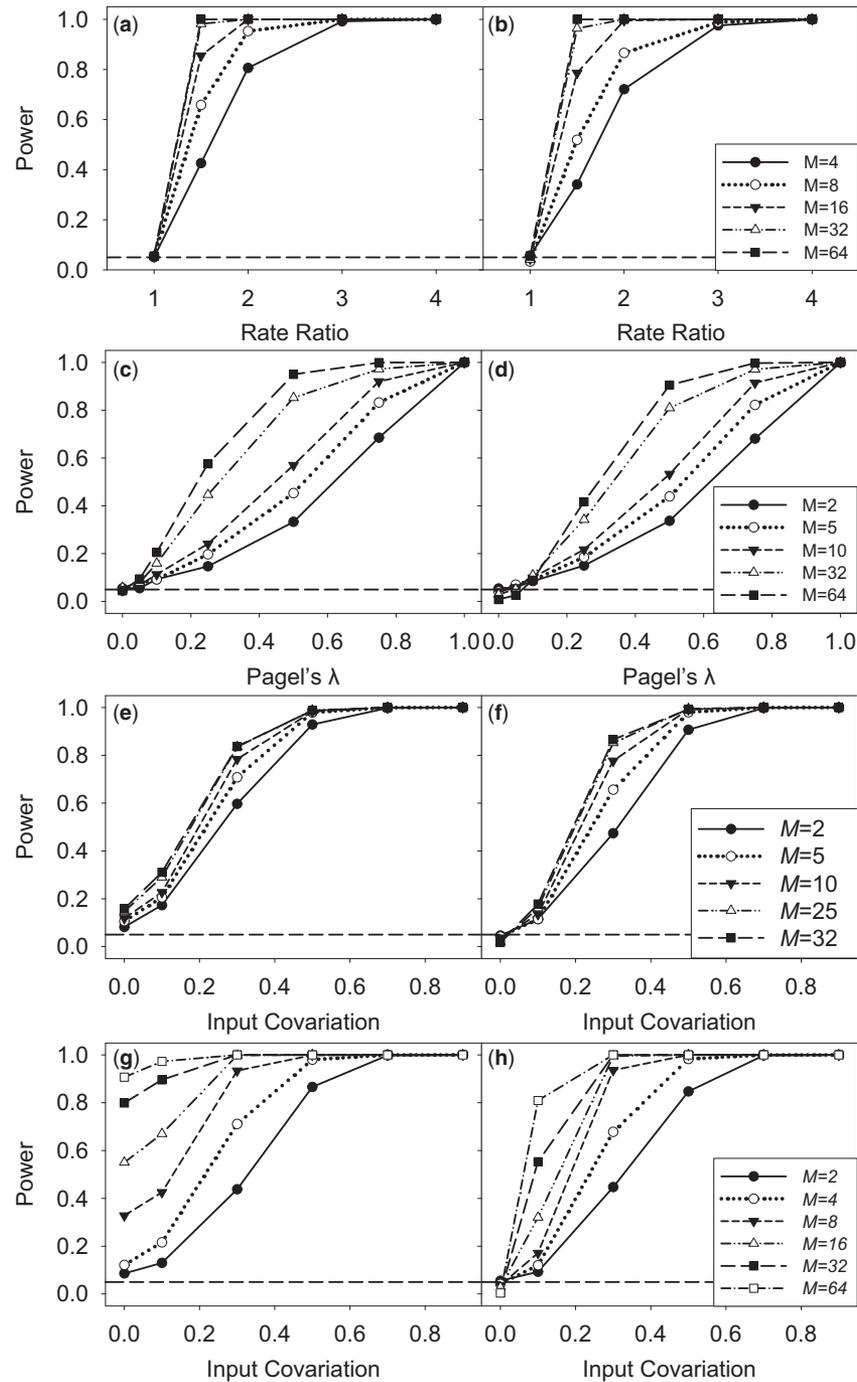


FIGURE 3. Comparisons of the statistical performance of distance-based methods (left) and pairwise composite likelihood methods (right) for comparisons of evolutionary rates among trait groups a–b) using the *geomorph* function *compare.multi.evol.rates* and the *phylocurve* function *compare.models*), tests of significant  $K_{\text{mult}}$ , c–d) using the *geomorph* function *physignal* and the *phylocurve* function *K.mult*), tests for significant covariation with fixed effects, e–f) using the *geomorph* function *procD.pgls* and the *phylocurve* function *compare.models*, and tests for significant covariation among two multivariate traits, g–h) using the *geomorph* function *phylo.integration* and the *phylocurve* function *compare.models*). Comparisons of evolutionary rates (a–b) among trait groups were simulated with rate ratios 1.0 (for Type I error), 1.5, 2.0, 3.0, and 4.0 for simulated data with 4, 8, 16, 32, and 64 traits (where subset trait groups were of size 2, 4, 8, 16, and 32, respectively). Both methods exhibited similar statistical performance. Tests of significant phylogenetic signal (c–d) were performed using Pagel's  $\lambda$  tree transformations of 1.0 (for Type I error), 0.75, 0.5, 0.25, and 0.0 on simulated data with 2, 5, 10, 32, and 64 traits. Statistical performance was similar for both methods, although distance-based tests of phylogenetic signal (c) had higher statistical power at higher simulated levels of Pagel's  $\lambda$ . Multivariate phylogenetic regression (e–f) was simulated with an input covariation among simulated data and fixed effects of 0.0 (for Type I error), 0.1, 0.3, 0.5, 0.7, and 0.9 for data of trait dimension 10, 24, 50, 100, and 250. Distance-based methods (e) exhibited elevated Type I error. Tests for significant covariation among multivariate traits (g–h) were simulated using an input covariation of 0.0 (for Type I error), 0.1, 0.3, 0.5, 0.7, and 0.9 for traits of dimension 2, 4, 8, 16, 32, and 64. Distance-based phylogenetic partial least squares (g) exhibited extremely high Type I error. All simulations were performed on 1000 simulated 32-species pure-birth phylogenies.

simulated with an evolutionary rate of 4.0 for species 1–16 and an evolutionary rate of 1.0 for species 17–32. Thus, the average simulated evolutionary rate for both regimes was 2.5, but substantial rate differences existed between regimes for each trait (where half the traits had a rate ratio of 4.0 and the other half had a rate ratio of 0.25). Simulations were repeated 1000 times, and the proportion of significant results (corresponding to the statistical power of the test) was compared for distance-based and pairwise composite likelihood-based hypothesis testing. As demonstrated in the previous section, Type I error is appropriate for both methods (assuming distance-based methods account for trait covariation). However, the statistical power for distance-based methods was approximately 0.05, suggesting that distance-based rate comparisons are poorly equipped to detect rate differences for the biologically realistic scenario of heterogeneous rate ratios. In contrast, the statistical power for pairwise composite likelihood-based regime rates comparisons under the described scenario was exactly 1.0.

#### Phylogenetic Signal

For testing the significance of multivariate phylogenetic signal ( $K_{\text{mult}}$ ), distance-based testing (Adams 2014a) follows the permutation procedure described in Blomberg et al. (2003). For the approach described here, the null hypothesis is simulated on a star phylogeny (see above). Results for both approaches yielded appropriate statistical power (Fig. 3c–d), although the simulation-based approach exhibited decreasing Type I error as trait dimensionality increased (0.008 for simulations with 64 traits), whereas the distance-based maintained a Type I error of approximately 0.05. Consistent with Type I error findings, statistical power was very similar between the two methods but with slightly higher statistical power for distance-based phylogenetic permutation under high trait dimensionality (although both methods exhibited higher power as the number of traits increased).

#### Testing for Evolutionary Correlations: Comparisons with D-PGLS

The statistical performance of testing for fixed effects (described above) in which the null hypothesis (without fixed effects) is compared to the alternative hypothesis (with fixed effects) was compared to the performance of D-PGLS. As with phylogenetic signal, D-PGLS relies on a phylogenetic permutation testing procedure. The simulation-based pairwise composite likelihood method described here exhibited similar statistical power to D-PGLS (Fig. 3e–f). However, D-PGLS exhibited high Type I error, which increased as trait dimensionality was increased (with Type I errors as high as 0.158). In contrast, the simulation-based pairwise composite likelihood method exhibited appropriate Type I error

which decreased with increasing trait dimensionality (0.018 for simulations with 32 traits). The reason for the high Type I error for D-PGLS in these simulations is unknown, as D-PGLS exhibited appropriate Type I error in Adams (2014b). One possible explanation is the type of phylogenies generated for simulations: here, pure-birth phylogenies were generated using the *pbtree* function in *phytools*, but simulations using phylogenies with random splits (using the *rtree* function in *ape*) and branch lengths computed using Grafen's method (1989) (implemented in the *compute.brlen* function in *ape*) appear to yield appropriate Type I error for D-PGLS, suggesting a possible sensitivity of D-PGLS to tree topology and branch lengths.

#### Testing for Evolutionary Correlations: Comparisons with Phylogenetic Partial Least Squares

Simulations were performed to assess the statistical performance of pairwise composite likelihood-based tests for evolutionary covariance among multiple traits. This test is analogous to phylogenetic partial least squares, as the null hypothesis for both methods is zero evolutionary covariance among two groups of traits ( $\mathbf{R}_{12} = \mathbf{R}_{21} = 0$ ) and the alternative hypothesis is  $\mathbf{R}_{12} \neq \mathbf{R}_{21} \neq 0$ . As with tests for phylogenetic signal and fixed effects, phylogenetic partial least squares relies on phylogenetic permutation, whereas pairwise composite likelihood-based tests use Monte Carlo simulations. However, the results of Adams and Felice (2014) could not be replicated for the simulations performed here, as phylogenetic partial least squares suffered from extremely high Type I error that was exacerbated by increasing trait dimensionality (for simulations with 64 traits, Type I error exceeded 0.90) (Fig. 3g–h). In contrast, the Type I error and statistical power for pairwise composite likelihood-based tests was appropriate under all simulated scenarios, and statistical power increased with increases in trait dimensionality (which reflected the statistical performance originally reported for phylogenetic partial least squares by Adams and Felice (2014)). In light of these findings, simulations were performed under a variety of scenarios in order to identify conditions under which distance-based phylogenetic least squares display appropriate Type I error, and it was determined that the statistical performance of phylogenetic partial least squares depends at least partially on the input evolutionary rate matrix ( $\mathbf{R}$ ) for simulations under the null hypothesis. Whereas the simulations for Type I error used to generate Figure 3g–h assumed independently evolving traits (diagonal  $\mathbf{R}$ ), a fully parameterized  $\mathbf{R}$  with high covariance within blocks  $\mathbf{R}_{11}$  and  $\mathbf{R}_{22}$  (and  $\mathbf{R}_{12}$  and  $\mathbf{R}_{21}$  subsequently set to zero) yielded appropriate Type I error and high statistical power for both distance-based and pairwise composite likelihood-based methods, suggesting that the robustness of distance-based phylogenetic partial least squares is highly sensitive to low within-group covariation.

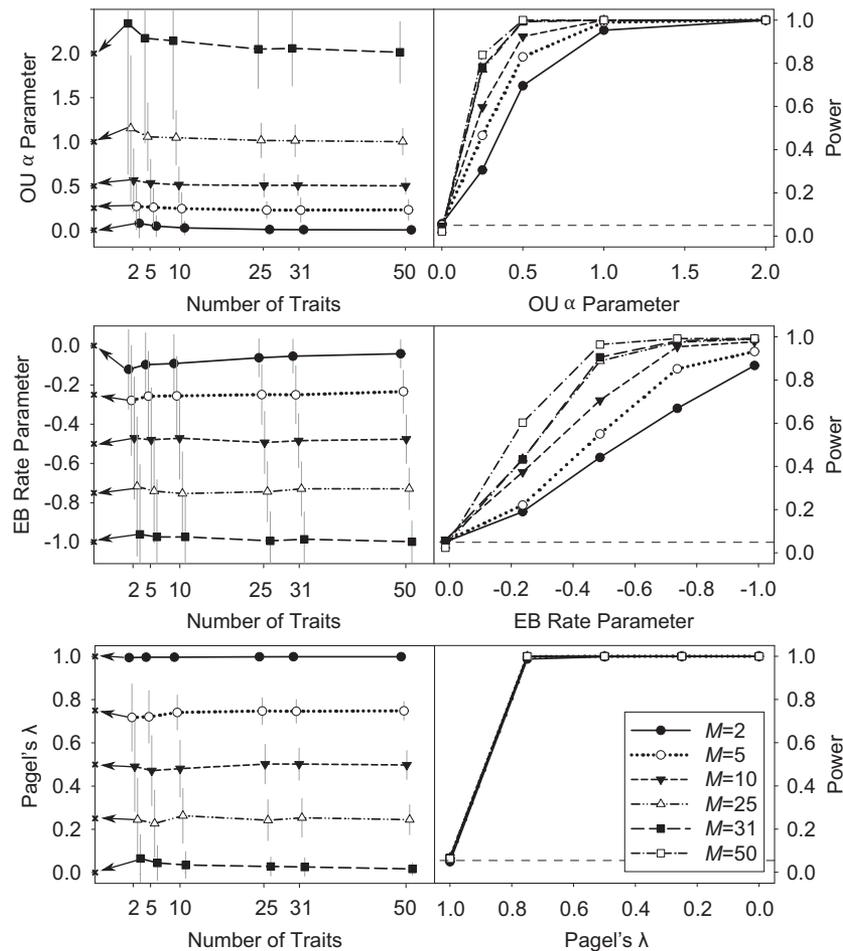


FIGURE 4. Results of maximum pairwise composite likelihood tree transformation parameters for the Ornstein-Uhlenbeck (a-b), Early-Burst (c-d), and Pagel's  $\lambda$  (e-f) models. Simulations were performed on 1000 simulated pure-birth phylogenies with data of trait dimension 2, 5, 10, 25, 31, and 50. Simulated tree transformation parameters were as follows: Ornstein-Uhlenbeck  $\alpha=0.0$  (for Type I error), 0.25, 0.5, 1.0, and 2.0; Early-Burst rate=0.0 (for Type I error), -0.25, -0.5, -0.75, and -1.0; Pagel's  $\lambda=1.0$  (for Type I error), 0.75, 0.5, 0.25, and 0.0. Left panels show mean parameter estimates of 1000 simulations (gray bars represent standard deviation) corresponding to the tree transformation parameter indicated by arrows. Right panels indicate Type I error and statistical power under the range of simulated conditions.

#### Testing for Evolutionary Covariation

Simulations were performed to assess the statistical performance of testing for the presence of evolutionary covariance in a multivariate trait. For this test, the null hypothesis is simulated by constraining the off-diagonal elements of the evolutionary rate matrix to zero, whereas the alternative hypothesis uses an unconstrained evolutionary rate matrix. As expected, this test exhibited statistical performance similar to that of tests for evolutionary covariance between trait groups (Fig. 3h), displaying appropriate Type I error rates and increasing statistical power as trait dimensionality increased.

#### Estimation of Alternative Evolutionary Models

Type I error and statistical power of maximum pairwise composite likelihood estimation was evaluated

for three common tree transformations: Ornstein-Uhlenbeck, Early-Burst, and Pagel's  $\lambda$  (Fig. 4). Bias and error of parameter estimates was also assessed (Fig. 4). Estimates of Page's  $\lambda$ , the Ornstein-Uhlenbeck parameter  $\alpha$ , and the Early-Burst rate parameter were generally unbiased, with increasing accuracy and precision as the number of trait dimensions was increased, although estimates for the Ornstein-Uhlenbeck parameter  $\alpha$  were extremely variable when simulated under larger values  $\alpha > 1$ ). As expected, model tests exhibited appropriate Type I error and increasing statistical power with increases in trait dimensionality.

#### Higher Dimensional Traits

A random 32-species phylogeny was generated and multivariate data with 2, 32, 64, and 1024 traits with random covariance were simulated under an Early-Burst model with a rate parameter of -0.5. The

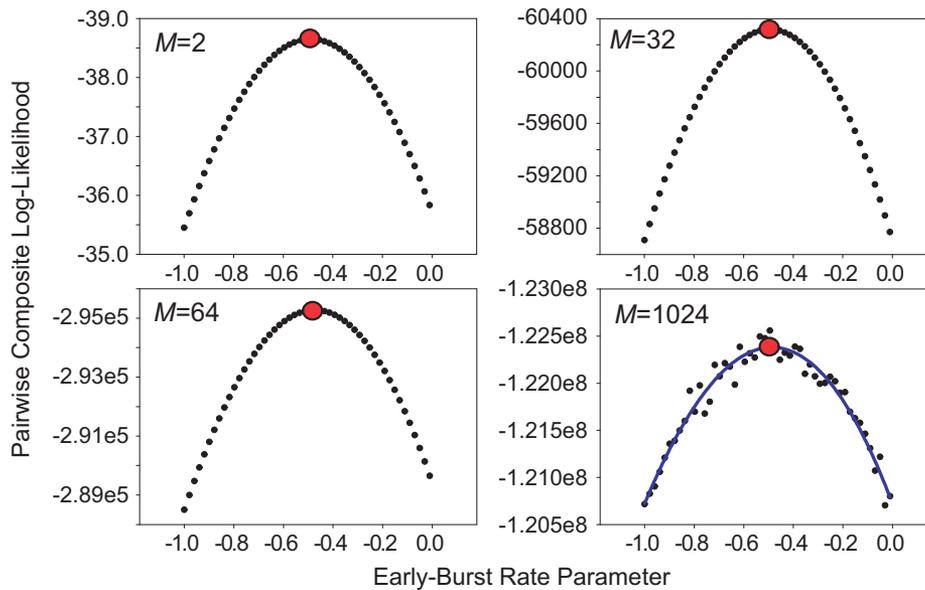


FIGURE 5. The pairwise composite log-likelihood surface for the Early-Burst rate transformation for trait data of dimension 2, 32, 64, and 1024, which were simulated on a 32-species phylogeny under an Early-Burst model of phenotypic evolution with a rate parameter of  $-0.5$ . For the 1024-trait data set, a Monte Carlo approach was performed using 10,000 random pairwise combinations to approximate the pairwise composite log-likelihood surface, and polynomial regression was performed to obtain a smoothed estimate of the likelihood surface. Regardless of the number of traits, the simulated Early-Burst rate of  $-0.5$  was strongly identifiable and was recovered via maximum pairwise composite likelihood estimation.

pairwise composite log-likelihood surface was assessed for 50 values (spanning  $-1$  to  $0$ ) of the Early-Burst rate parameter for the simulated 2, 32, and 64-dimensional traits. For the 1024-dimensional trait, the Monte Carlo-based approach (described above) was performed using 10,000 random pairwise combinations to approximate the composite log-likelihood surface. Polynomial regression was then performed to obtain a smoothed estimate of the pairwise composite log-likelihood surface. The rate value corresponding to the maximum predicted value of the resulting polynomial regression was  $-0.49$ , consistent with simulated conditions and the rates recovered for simulated data sets of dimension 2, 32, and 64 (rate =  $-0.5$ ) (Fig. 5).

#### CONCLUSION

Distance-based multivariate comparative methods offer a framework for testing evolutionary hypotheses for high-dimensional traits while avoiding several problems associated with high dimensionality (e.g., singular covariance matrices and low statistical power). However, extremely large data sets may be computationally infeasible using a distance-based approach. An equivalent covariance-based approach is described in Appendices 1–2 which allows fast linear-time implementation (implemented in the R package *phylocurve*), thus avoiding most of the computational challenges associated with distance-based methods.

A novel approach based on pairwise composite likelihood is also proposed. These methods (also

implemented in *phylocurve*) allow for greater flexibility than distance-based methods, such as the ability to incorporate multiple evolutionary hypotheses and alternative evolutionary models. An approach for incorporating within-species variation and missing data is also discussed. Simulations revealed high statistical power and appropriate Type I error for these methods, whereas some distance-based approaches ( $D$ -PGLS and phylogenetic partial least squares) exhibited elevated Type I error under various scenarios, and regime-specific rate comparisons exhibited extremely low statistical power when rate ratios differed among traits.

It is important to note that despite promising results from statistical simulations, both distance-based methods and the new methods described here make several critical assumptions regarding the evolution of high-dimensional traits. Most notably, all models were simulated such that model parameters applied equally to all traits of interest. However, the consequences of violating these assumptions are unknown. For instance, consider phylogenetic regression on a high-dimensional trait in which a handful of traits are strongly driven by a predictor variable but the remaining traits are independent of the predictor variable. Regardless of the overall  $P$ -value obtained, the biological relevance of the result is unclear: a significant finding suggests correlation where none exists for the majority of traits, whereas a nonsignificant result fails to detect potentially important correlations for the handful of affected traits. Under such circumstances, generalized high-dimensional metrics may not be ideal for testing evolutionary hypotheses, although methods to detect

this type of situation are currently lacking. While high-dimensional comparative methods provide a potentially powerful framework for approaching the study of complex phenotypic evolution, this framework will benefit substantially from the development of model diagnostic techniques, such as high-dimensional multivariate methods for detecting violations of model assumptions and for assessing model adequacy (Pennell et al. 2015).

#### SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.fh0mc>.

#### FUNDING

This work was supported in part by the National Science Foundation [DEB-1501215].

#### ACKNOWLEDGEMENTS

The author wishes to thank Chase Mason and Cécile Ané for comments and discussion, Frank Anderson, Luke Harmon, Julien Clavel, and two anonymous reviewers for extremely helpful feedback, as well as Joe Felsenstein, Robert Freckleton, and Dean Adams for helpful suggestions on earlier versions of this manuscript.

#### APPENDIX 1. A COVARIANCE-BASED (R-MODE) RE-EXPRESSION OF MULTIVARIATE DISTANCE-BASED (Q-MODE) COMPARATIVE METHODS

To address issues of statistical power and non-invertible matrices, and to provide a framework for estimating generalized statistics for high-dimensional multivariate traits, Adams (2014 a–c) proposed several phylogenetic comparative methods based on a distance-based (Q-mode) approach. Consider a phylogeny with  $N$  extant species on which  $M$  traits are observed, given as an  $N \times M$  data matrix ( $\mathbf{Y}$ ). The  $N \times N$  phylogenetic covariance matrix ( $\mathbf{C}$ ) is parameterized by branch lengths and the specified evolutionary model (e.g., Brownian motion). The eigenvectors ( $\mathbf{U}$ ) and diagonal matrix of eigenvalue square roots ( $\mathbf{V}^{1/2}$ ) of  $\mathbf{C}$  are used to construct the phylogenetic transformation matrix ( $\mathbf{T}$ )

$$\mathbf{T} = (\mathbf{U}\mathbf{V}^{1/2}\mathbf{U}^t)^{-1} \quad (\text{A1})$$

which is then matrix multiplied by relevant matrices of interest (e.g.,  $\mathbf{TX}$ ,  $\mathbf{TY}$ ,  $\mathbf{T}(\mathbf{Y}-E(\mathbf{Y}))$ ) to remove phylogenetic covariance from the data (Garland and Ives 2000). Next, the Euclidean distances of phylogenetically transformed data from the origin are calculated ( $\mathbf{PD}$ ), resulting in an  $N$ -length vector which is then used for multivariate comparative calculations.

#### Evolutionary Rates: Distance-Based (Q-mode) Methods

Various methods have been proposed to quantify and compare evolutionary rates for univariate traits, including contrast-based (Garland 1992), generalized least squares-based (Martins and Hansen 1997), and likelihood-based (O'Meara et al. 2006; Thomas et al. 2006), as well as Bayesian methods (Rabosky et al. 2014). For high-dimensional data, Adams (2014c) proposed an estimate called  $\sigma_{\text{mult}}^2$ , which is a single consensus evolutionary rate for the entire multivariate trait.  $\sigma_{\text{mult}}^2$  is estimated by dividing the Euclidean distance of the cross-product of phylogenetically transformed residuals from the origin is by the total number of observations (NM):

$$\sigma_{\text{mult}}^2 = \frac{\mathbf{PD}_{\mathbf{Y}-E(\mathbf{Y}),0}^t \mathbf{PD}_{\mathbf{Y}-E(\mathbf{Y}),0}}{NM} \quad (\text{A2})$$

The estimate of  $\sigma_{\text{mult}}^2$  can be used to statistically compare evolutionary rates among groups of species (Adams 2014c) or among groups of traits (Denton and Adams 2015). To test whether  $\sigma_{\text{mult}}^2$  differs among species groups (as implemented in the *geomorph* function *compare.evol.rates*), subset groups of size  $N_{\text{species.sub}_j}$  are used to estimate  $\sigma_{\text{species.sub}_j}^2$  for each species group by replacing  $\mathbf{PD}_{\mathbf{Y}-E(\mathbf{Y}),0}$  with  $(\mathbf{PD}_{\mathbf{Y}-E(\mathbf{Y}),0})_{\text{species.sub}_j}$  and  $N$  with  $N_{\text{species.sub}_j}$  in equation (2). The observed ratio of regime-specific  $\sigma_{\text{species.sub}_j}^2$  values is compared to the null distribution of  $\sigma_{\text{species.sub}_j}^2$  ratios via phylogenetic simulation (Adams 2014c). Similarly, to test for differences in  $\sigma_{\text{mult}}^2$  among groups of traits (as implemented in the *geomorph* function *compare.multi.evol.rates*),  $\sigma_{\text{trait.sub}_k}^2$  is calculated for each trait group by replacing  $M$  with  $M_{\text{traits.sub}_k}$  and  $\mathbf{PD}_{\mathbf{Y}-E(\mathbf{Y}),0}$  with  $\mathbf{PD}_{\mathbf{Y}_k-E(\mathbf{Y}_k),0}$  in equation (2) (where  $\mathbf{Y}_k$  contains  $M_{\text{trait.sub}_k}$  variables), and the observed ratio of group-specific  $\sigma_{\text{trait.sub}_k}^2$  values is compared to the null distribution of  $\sigma_{\text{trait.sub}_k}^2$  via phylogenetic simulation (Denton and Adams 2015).

#### Evolutionary Rates: Covariance-Based (R-mode) Methods

The calculation of  $\sigma_{\text{mult}}^2$  can be simplified considerably using a covariance-based approach, as  $\sigma_{\text{mult}}^2$  is simply the arithmetic mean of maximum likelihood evolutionary rates for each trait considered individually:

$$\sigma_{\text{mult}}^2 = \frac{\sum_{i=1}^M ((\mathbf{y}_i - E(\mathbf{y}_i))^t \mathbf{C}^{-1} (\mathbf{y}_i - E(\mathbf{y}_i)))}{NM} \quad (\text{A3})$$

Similarly,  $\sigma_{\text{trait.sub}_k}^2$  for a trait group subset  $k$  is simply the arithmetic mean of evolutionary rates for individual traits in  $\mathbf{Y}_k$ .

The calculation of regime-specific  $\sigma_{\text{species.sub}}^2$  can be calculated as the mean maximum likelihood variance of the phylogenetically transformed residuals ( $\mathbf{T}(\mathbf{Y} - E(\mathbf{Y}))_{\text{species.sub}}$ ). Alternatively, the computation of  $\mathbf{T}$  can be avoided using either the “noncensored” or “censored” approach described by O’Meara et al. (2006), the latter of which provides an efficient closed-form estimate of  $\sigma_{\text{species.sub}}^2$ , in which equation (3) is applied to a phylogeny and data set pruned to only contain the species represented in regime  $j$ . The censored estimate is not identical to the distance-based estimate of  $\sigma_{\text{species.sub}}^2$  (the censored approach estimates a separate  $E(\mathbf{Y})$  for each pruned tree). However, the censored approach is a close approximation, retains appropriate Type I error and statistical power, and is far more efficient than the eigendecomposition of  $\mathbf{C}$  and subsequent inversion of an  $N \times N$  matrix (which is required in equation (1) in order to construct the phylogenetic transformation matrix  $\mathbf{T}$ ).

If phylogenetic transformation cannot be avoided, an alternative to equation (1) which avoids matrix inversion can be used to construct the phylogenetic transformation matrix:

$$\mathbf{T} = \mathbf{UV}^* \mathbf{U}^t \quad (\text{A4})$$

where  $\mathbf{V}^*$  is an  $N \times N$  diagonal matrix with  $\sqrt{1/\mathbf{v}}$  along the diagonal, where  $\mathbf{v}$  contains the eigenvalues of  $\mathbf{C}$  (Li 2007). If  $\mathbf{C}$  is singular, infinite or undefined values in  $\mathbf{V}^*$  may be replaced with zero (although the consequences of proceeding with comparative analyses on singular matrices are largely untested).

Functions for comparing evolutionary rates using the described covariance-based approach are implemented in the *phylocurve* functions *fast.geomorph.compare.evol.rates* and *fast.geomorph.compare.multi.evol.rates*. These functions (and all other *phylocurve* functions that begin with “fast.geomorph”) are implemented to demonstrate the equivalence and application of fast linear-time computations for analogous distance-based functions in *geomorph* (see Appendix 2 for a discussion of fast covariance-based approaches).

#### Phylogenetic Signal (Blomberg’s $K$ )

Blomberg’s (2003)  $K$ , which in univariate form is calculated as:

$$K = \frac{(\mathbf{y} - E(\mathbf{y}))^t (\mathbf{y} - E(\mathbf{y}))}{(\mathbf{y} - E(\mathbf{y}))^t \mathbf{C}^{-1} (\mathbf{y} - E(\mathbf{y}))} \bigg/ \frac{\text{tr}(\mathbf{C}) - N(1^t \mathbf{C}^{-1} \mathbf{1})^{-1}}{N - 1} \quad (\text{A5})$$

can be extended to multivariate form  $K_{\text{mult}}$  (Adams 2014a, as implemented in the *geomorph* function

*physignal*) using the distance-based formula

$$K_{\text{mult}} = \frac{\mathbf{D}_{\mathbf{Y}-E(\mathbf{Y}),0}^t \mathbf{D}_{\mathbf{Y}-E(\mathbf{Y}),0}}{\mathbf{PD}_{\mathbf{Y}-E(\mathbf{Y}),0}^t \mathbf{PD}_{\mathbf{Y}-E(\mathbf{Y}),0}} \bigg/ \frac{\text{tr}(\mathbf{C}) - N(1^t \mathbf{C}^{-1} \mathbf{1})^{-1}}{N - 1} \quad (\text{A6})$$

where  $\mathbf{D}_{\mathbf{X},0}$  is the non-phylogenetically transformed Euclidean distance between some matrix and the origin. As with  $\sigma_{\text{mult}}^2$ , the value  $K_{\text{mult}}$  can be calculated using a covariance-based approach (implemented in the *phylocurve* function *fast.geomorph.physignal*) by considering the sums of squared residuals for each individual trait:

$$K_{\text{mult}} = \frac{\sum_{i=1}^M ((\mathbf{y}_i - E(\mathbf{y}_i))^t (\mathbf{y}_i - E(\mathbf{y}_i)))}{\sum_{i=1}^M ((\mathbf{y}_i - E(\mathbf{y}_i))^t \mathbf{C}^{-1} (\mathbf{y}_i - E(\mathbf{y}_i)))} \bigg/ \frac{\text{tr}(\mathbf{C}) - N(1^t \mathbf{C}^{-1} \mathbf{1})^{-1}}{N - 1} \quad (\text{A7})$$

As with univariate Blomberg’s  $K$ , significance of  $K_{\text{mult}}$  is determined by phylogenetic permutation (Blomberg 2003; Adams 2014a).

#### Phylogenetic Generalized Least Squares

Distance-based phylogenetic generalized least squares ( $D$ -PGLS, as implemented in the *geomorph* function *procD.pgls*) regression can be performed by regressing  $\mathbf{EY} \sim \mathbf{EX}$  and  $\mathbf{EY} \sim \mathbf{E1}$  to obtain predicted values  $\hat{\mathbf{Y}}_{\mathbf{X}}$  and  $\hat{\mathbf{Y}}_1$ , which is then used to calculate summary statistics including sums of squares,  $F$ -ratios, and  $R^2$ , and significance is determined by phylogenetic permutation (Adams 2014b). To perform the covariance-based equivalent of  $D$ -PGLS regression (*phylocurve* function *fast.geomorph.procD.pgls*), first, a  $p \times M$  (where  $p$  is the number of regression coefficients to be estimated for each trait dimension) matrix of dimension-specific regression coefficients is calculated as

$$\boldsymbol{\beta} = (\mathbf{X}^t \mathbf{C}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{C}^{-1} \mathbf{Y} \quad (\text{A8})$$

where  $\mathbf{X}$  is the  $M \times p$  model matrix for PGLS regression, typically consisting of a column of ones and one or more columns of univariate predictor variables (Martins and Hansen 1997; Revell 2010). Note that  $\mathbf{Y}$  is an  $N \times M$  matrix in equation (8), unlike univariate PGLS in which  $\mathbf{Y}$  is an  $N \times 1$  vector. Next,  $N \times M$  matrices of predicted species values from the regression  $\hat{\mathbf{Y}}_{\mathbf{X}} = \mathbf{X}\boldsymbol{\beta}$  and null model  $\hat{\mathbf{Y}}_1 = \mathbf{1}E(\mathbf{Y})$  are used to calculate residuals, and sums of squares are obtained from  $SS_{\hat{\mathbf{e}}_{\mathbf{X}}} = \sum_{i=1}^M ((\mathbf{y}_i - \hat{\mathbf{y}}_{i\mathbf{X}})^t \mathbf{C}^{-1} (\mathbf{y}_i - \hat{\mathbf{y}}_{i\mathbf{X}}))$  and  $SS_{\hat{\mathbf{e}}_1} = \sum_{i=1}^M ((\mathbf{y}_i - \hat{\mathbf{y}}_{i1})^t \mathbf{C}^{-1} (\mathbf{y}_i - \hat{\mathbf{y}}_{i1}))$  to calculate mean squared error,  $F$ -ratios, and  $R^2$ . Phylogenetic permutation is then performed to determine significance, and results

are identical to *D*-PGLS regression (Adams 2014b; Adams and Collyer 2015). It should be noted that although phylogenetically independent contrasts can be used to calculate these quantities (Felsenstein 1985), phylogenetic permutation must be performed on raw permuted values (*not* contrasts), and then independent contrasts must be recalculated for each permutation (Adams and Collyer 2015). However, in contrast to the findings of Adams and Collyer (2015), an appropriately and efficiently implemented phylogenetically independent contrasts-based approach is indeed faster than *D*-PGLS (Fig. A2c–d).

#### Phylogenetic Partial Least Squares

In addition to the methods discussed above, Adams and Felice (2014) proposed a distance-based method for evaluating covariation between two multivariate traits  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  called phylogenetic partial least squares (PLS). First, the evolutionary rate matrix  $\mathbf{R}$  for  $\mathbf{Y}$  is calculated using the generalized least squares restricted maximum likelihood estimator:

$$\mathbf{R} = \frac{(\mathbf{Y} - \mathbf{X}E(\mathbf{Y}))^t \mathbf{C}^{-1} (\mathbf{Y} - \mathbf{X}E(\mathbf{Y}))}{N - 1} \quad (\text{A9})$$

The covariance of  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  is partitioned into four blocks:

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix} \quad (\text{A10})$$

and singular-value decomposition is subsequently performed on  $\mathbf{R}_{12}$ . Next, the values  $\mathbf{Y} - E(\mathbf{Y})$  are projected onto the phylogenetic transformation matrix  $\mathbf{T}$ , and  $\mathbf{T}(\mathbf{Y}_1 - E(\mathbf{Y}_1))$  and  $\mathbf{T}(\mathbf{Y}_2 - E(\mathbf{Y}_2))$  are matrix multiplied by the left ( $\mathbf{U}$ ) and right ( $\mathbf{V}$ ) singular vectors of  $\mathbf{R}_{12}$ , respectively. The first two columns of the resulting scores are regressed to determine the evolutionary correlation between  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$ , and phylogenetic permutation is used to assess significance of the PLS regression (Adams and Felice 2014; *geomorph* function *phylo.integration*). To perform the covariance-based equivalent of phylogenetic PLS (as implemented in the *phylocurve* function *fast.geomorph.phylo.integration*),  $(\mathbf{Y}_1 - E(\mathbf{Y}_1))\mathbf{U}_1$  and  $(\mathbf{Y}_2 - E(\mathbf{Y}_2))\mathbf{V}_1$  are regressed using phylogenetically independent contrasts (Felsenstein 1985) regressed through the origin, or equivalently using PGLS. The resulting regression correlation is equivalent to the PLS correlation obtained from distance-based phylogenetic PLS, and significance is assessed using phylogenetic permutation (Adams and Felice 2014).

#### APPENDIX 2. FAST COMPUTATIONS FOR HIGH-DIMENSIONAL COMPARATIVE MODELS

##### Efficient Calculations for High-Dimensional Data

For repeated calculations on a given phylogeny (e.g., as is associated with bootstrapping procedures), a

substantial portion of the computational burden of phylogenetically independent contrasts (a linear-time algorithm) is associated with redundant operations (e.g., reordering internal edges, conversion of data types, etc.). In some cases, repeating these steps can account for over 90% of computational time. To minimize redundant operations, all preparation steps can be performed in a single function call, and then relevant quantities (e.g., phenotypic data or branch lengths) can be updated as needed prior to calling the independent contrasts algorithm. Additionally, phylogenetically independent contrasts can be performed simultaneously on all traits of interest by performing relevant calculations on the full matrix  $\mathbf{Y}$  rather than on individual vectors. Refer to Appendix 3 for details on implementing these methods.

The quantity  $(\mathbf{Y} - E(\mathbf{Y}))^t \mathbf{C}^{-1} (\mathbf{Y} - E(\mathbf{Y}))$ , which is simply the cross-product of the matrix of independent contrasts of  $\mathbf{Y}$ , can be calculated efficiently in linear time using phylogenetically independent contrasts. The independent contrasts algorithm also automatically calculate  $E(\mathbf{Y})$ , so a separate ancestral reconstruction is unnecessary (see Appendix 3). Contrast variances for each node ( $\mathbf{v}$ ) and the length of the two edges extending from the root of the independent contrast-transformed tree ( $\mathbf{x}$ ) can be used to calculate  $1^t \mathbf{C}^{-1} \mathbf{1} = \prod \mathbf{x} / \sum \mathbf{x}$  and  $\log |\mathbf{C}| = \log(\prod \mathbf{x} / \sum \mathbf{x}) + \sum \log \mathbf{v}$  in linear time (Felsenstein 1985; Freckleton 2012). Using these formulas, any quantity of the form  $\mathbf{L}^t \mathbf{C}^{-1} \mathbf{R}$  (corresponding to left and right matrices of compatible dimensions, such as  $\mathbf{X}$  and  $\mathbf{Y}$ ) can be computed in linear time as  $\mathbf{L}^t \mathbf{C}^{-1} \mathbf{R} = \mathbf{L}_{pic}^t \mathbf{R}_{pic} + E(\mathbf{L})E(\mathbf{R})^t (1^t \mathbf{C}^{-1} \mathbf{1})$  (for a related linear-time algorithm, see Ho and Ané 2014). Quantities required for the multivariate log-likelihood may also be calculated efficiently:  $\log |\mathbf{R} \otimes \mathbf{C}| = N \log |\mathbf{R}| + M \log |\mathbf{C}|$  and  $\log |\mathbf{X}^t (\mathbf{R} \otimes \mathbf{C})^{-1} \mathbf{X}| = M \log |(\mathbf{X}_{pic}^t \mathbf{X}_{pic} + \mathbf{X}_{anc} \mathbf{X}_{anc}^t (1^t \mathbf{C}^{-1} \mathbf{1}))| - p \log |\mathbf{R}|$  (where  $p$  is the number of columns of  $\mathbf{X}$ , and the matrix subscripts *pic* and *anc* denote phylogenetically independent contrasts and the estimated root state for each column of a given matrix, respectively).

A more complex approach is required to accommodate multiple regimes. The following algorithm, modified from the linear-time algorithm described by Ho and Ané (2014), calculates quantities of the form  $\mathbf{Q} = \mathbf{L}^t \mathbf{W}^{-1} \mathbf{R}$  and  $\log |\mathbf{W}|$  in linear time relative to the number of species, where  $\mathbf{W}$  is the  $NM \times NM$  matrix describing species-trait covariance for multiple regimes (under a single regime,  $\mathbf{W} = \mathbf{R} \otimes \mathbf{C}$ ), and  $\mathbf{L}$  and  $\mathbf{R}$  are matrices of compatible dimensions (e.g.,  $\mathbf{X}$  and  $\mathbf{Y}$ ). These quantities can then be used for log-likelihood calculations (refer to page 406 of Ho and Ané (2014) for details of the original approach on which the following is based). The following algorithm traverses the tree postorder from the tips to the root:

1. For a tree with a single tip, let  $\mathbf{Z}^{(e)} = \sum_{a=1}^r (t_e p_{a,e} \mathbf{R}_a)$  be an  $M \times M$  matrix representing the length of edge  $e$  scaled by regime-specific rate matrices and

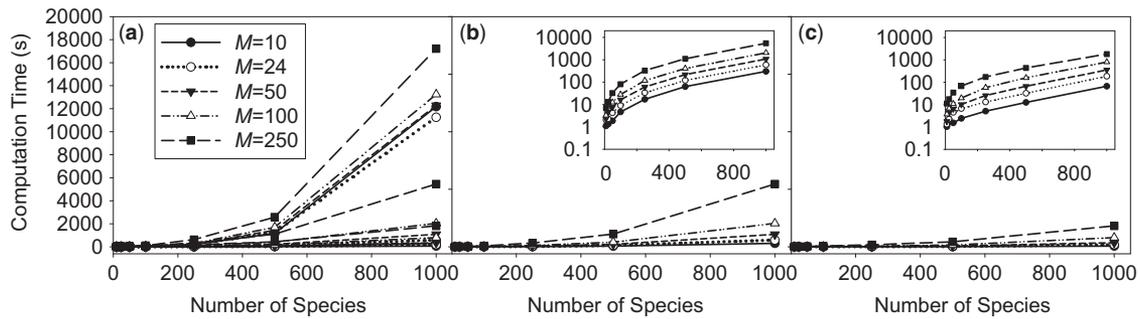


FIGURE A1. Speed comparisons for rate comparisons among species rate regimes using a) the distance-based *geomorph* (version 2.17) function *compare.evol.rates*, b) the covariance-based *phylocurve* function *fast.geomorph.compare.evol.rates* (based on phylogenetic transformation of residuals; *censored=FALSE*), and c) the covariance-based *phylocurve* function *fast.geomorph.compare.evol.rates* using the censored method of O'Meara et al. (2006) (*censored=TRUE*). Speed comparisons were performed on pure-birth phylogenies containing 10, 24, 50, 100, 250, 500, and 1000 species with simulated data containing 10, 24, 50, 100, and 250 traits. Computation times on a log<sub>10</sub>-scaled axis are also plotted for *phylocurve* functions.

their respective proportions assigned to each edge.

Then  $\mathbf{p} = \mathbf{Z}^{(e)-1}$ ,  $\mathbf{U}_l^t = \mathbf{L}^t \mathbf{p}$ ,  $\mathbf{V}_r = (\mathbf{R}^t \mathbf{p})^t$ ,  $\mathbf{Q} = \mathbf{L}^t \mathbf{p} \mathbf{R}$ , and  $\log |\mathbf{W}| = \log |\mathbf{Z}^{(e)}|$ .

- For a tree with two or more tips and root edge  $e$ , again let  $\mathbf{Z}^{(e)} = \sum_{a=1}^r (t_e p_{a,e} \mathbf{R}_a)$ . Then  $\mathbf{p}_A = \sum_s \mathbf{p}_s$ ,  $\log |\mathbf{W}| = \sum_s \log |\mathbf{W}_s| + \log |\mathbf{I} + \mathbf{Z}^{(e)} \mathbf{p}_A|$ ,  $\mathbf{p} = \mathbf{p}_A (\mathbf{I} + \mathbf{Z}^{(e)} \mathbf{p}_A)^{-1}$ ,  $\mathbf{Q} = \left( \sum_s \mathbf{Q}_s \right) - \mathbf{U}_{l,s}^t (\mathbf{I} + \mathbf{Z}^{(e)} \mathbf{p}_A)^{-1} \mathbf{Z}^{(e)} \mathbf{V}_{r,s}$ ,  $\mathbf{U}_l^t = \left( \sum_s \mathbf{U}_{l,s}^t \right) (\mathbf{I} + \mathbf{Z}^{(e)} \mathbf{p}_A)^{-1}$ , and  $\mathbf{V}_r^t = \left( \left( \sum_s \mathbf{V}_{r,s}^t \right) (\mathbf{I} + \mathbf{Z}^{(e)} \mathbf{p}_A)^{-1} \right)^t$ .
- At the root of the full tree, return  $\mathbf{Q}$  and  $\log |\mathbf{W}|$ .

#### Fast Covariance-Based Implementations of Distance-Based Methods

The computational performance of distance-based methods implemented in *geomorph* version 2.1.7 (*compare.evol.rates*, *compare.multi.evol.rates*, *physignal*, *phylo.pls* (now *phylo.integration* as of version 3.0), and *procD.pgls*) were compared to fast covariance-based *phylocurve* functions (*fast.geomorph.compare.evol.rates*, *fast.geomorph.compare.multi.evol.rates*, *fast.geomorph.physignal*, *fast.geomorph.phylo.integration*, and *fast.geomorph.procD.pgls*). Computational times for each method were compared using 10, 24, 50, 100, 250, 500, and 1000 species and 10, 24, 50, 100, and 250 traits (Figs. A1–2). For data sets with a low number of species (<50), the speed of distance-based methods was generally comparable to covariance-based methods. For data sets with higher numbers of species, covariance-based methods were consistently faster (in some cases up to ~1000 times faster). The most time-consuming steps of distance-based methods in *geomorph* include inverting the phylogenetic covariance matrix and performing an eigendecomposition of the phylogenetic covariance matrix, both of which are avoided by covariance-based methods (with the exception of *fast.geomorph.compare.evol.rates*, which

requires an eigendecomposition for computing the phylogenetic transformation matrix, unless the censored approach is used). Covariance-based functions that rely on phylogenetic permutation for hypothesis tests (*fast.geomorph.procD.pgls*, *fast.geomorph.physignal*, and *fast.geomorph.phylo.integration*) operate in linear time relative to the number of species. Although *fast.geomorph.compare.evol.rates* and *fast.geomorph.compare.multi.evol.rates* are substantially faster than *compare.evol.rates* and *compare.multi.evol.rates* (respectively) for data sets with a large numbers of species, all four functions rely on the *sim.char* function in *geiger* (Pennell et al. 2014) for hypothesis testing, which operates with polynomial increases in time as the number of species increases and is thus a nonlinear rate-limiting step for *fast.geomorph.compare.multi*.

#### APPENDIX 3. EFFICIENT REPEATED CALCULATIONS OF PHYLOGENETICALLY INDEPENDENT CONTRASTS

```
require(phylocurve)
require(ape)
nspecies <- 1000
ntraits <- 500
tree <- rtree(n = nspecies) # Simulate
a random 1000-species phylogeny
Y <- matrix(rnorm(nspecies*ntraits),
ncol=ntraits) # Simulate data
rownames(Y) <- tree$tip.label
# Call prep_multipic (a phylocurve
function)
prep.Y <- prep_multipic(x = Y,
phy = tree)
# Calculate phylogenetically independent
contrasts using pic (ape package)
Y.pics.ape <- apply(X = Y, MARGIN =
2, FUN = pic, phy = tree)
# Calculate PICs using multipic
# returns a list with contrasts,
sum_invV, log_detV, root
# 'contrast' is a matrix of PICs
# 'sum_invV' is equal to
sum(solve(vcv(tree)))
```

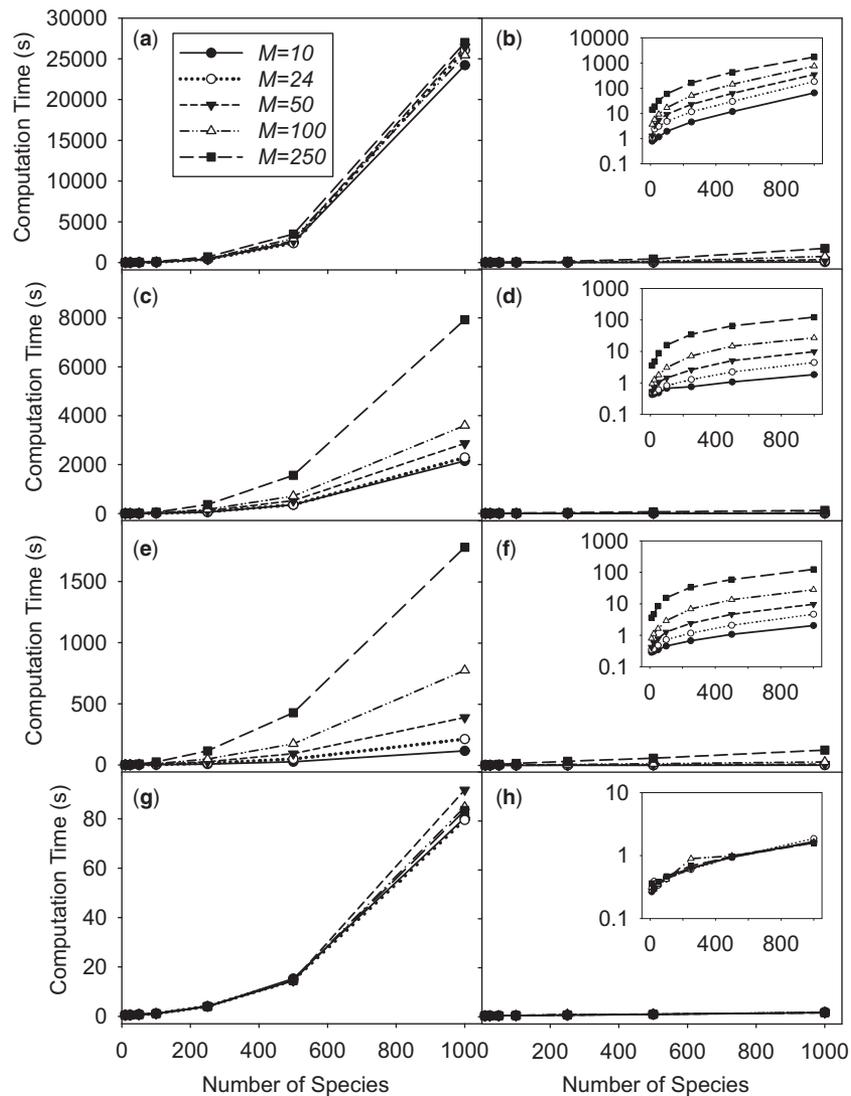


FIGURE A2. Speed comparisons for rate comparisons among trait groups using a) the distance-based *geomorph* (version 2.1.7) function *compare.multi.evol.rates* and b) the covariance-based *phylocurve* function *fast.geomorph.compare.multi.evol.rates*; tests of  $K_{\text{mult}}$  using c) the distance-based *geomorph* (version 2.1.7) function *physignal* and d) the covariance-based *phylocurve* function *fast.geomorph.physignal*; multivariate phylogenetic regression using e) the distance-based *geomorph* (version 2.1.7) function *procD.pgls* and f) the covariance-based *phylocurve* function *fast.geomorph.procD.pgls*; and phylogenetic partial least squares using g) the distance-based *geomorph* (version 2.1.7) function *phylo.pls* (now *phylo.integration* in versions  $\geq 3.0.0$ ) and h) the covariance-based *phylocurve* function *fast.geomorph.phylo.integration*. Speed comparisons were performed on pure-birth phylogenies containing 10, 24, 50, 100, 250, 500, and 1000 species with simulated data containing 10, 24, 50, 100, and 250 traits. Computation times on a  $\log_{10}$ -scaled axis are also plotted for *phylocurve* functions.

```
# 'log_detV' is equal to
log(det(vcv(tree)))
# 'root' is the maximum likelihood
phenotypic value at the root
Y.pics.phylocurve <-
do.call(multipic,prep.Y)
# Verify that results are identical
range(Y.pics.ape -
Y.pics.phylocurve$contrasts)
# Simulate 50 data sets
niter <- 50
randomY <- vector(mode = "list",
length = niter)
for(i in 1:niter)
```

```
{
randomY[[i]] <- matrix
(rnorm(nspecies*ntraits),ncol=ntraits)
rownames(randomY[[i]]) <-
tree$tip.label
}
#####
# Compare time to calculate PICs on
50 simulated data sets using pic
vs multipic
#
##### pic function
system.time(for(i in 1:niter) apply
(X = randomY[[i]],MARGIN = 2,
```

```

FUN = pic, phy = tree))
##### user system elapsed
##### 18.35 0.23 18.61
##### multipic function
system.time(for(i in 1:niter)
{
prep.Y$phe[1:nspecies,] <- randomY[[i]]
# update prep.Y$phe with new data
do.call(multipic,prep.Y)
})
##### user system elapsed
##### 1.38 0.14 1.52
#
#####

```

## 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., Otárola-Castillo E. 2013. geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods Ecol. Evol.* 4:393–399.
- Adams D.C. 2014a. A generalized K statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. *Syst. Biol.* 63:685–697.
- Adams D.C. 2014b. A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. *Evolution* 68:2675–2688.
- Adams D.C. 2014c. Quantifying and comparing phylogenetic evolutionary rates for shape and other high-dimensional phenotypic data. *Syst. Biol.* 63:166–177.
- Adams D.C., Felice R. 2014. Assessing phylogenetic morphological integration and trait covariation in morphometric data using evolutionary covariance matrices. *PLOS One* 9(4):e94335.
- Adams D.C., Collyer M. 2015. Permutation tests for phylogenetic comparative analyses of high-dimensional shape data: what you shuffle matters. *Evolution* 69:823–829.
- Bartoszek K., Pienaar J., Mostad P., Andersson S., Hansen T.F. 2012. A phylogenetic comparative method for studying multivariate adaptation. *J. Theor. Biol.* 314:204–215.
- Bates D., Maechler M. 2015. Matrix: sparse and dense matrix classes and methods. R package, Version 1.2-2. Available from <http://cran.R-project.org> (last accessed April 20, 2016).
- Blomberg S.P., Garland T., Ives A.R. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717–745.
- Boettiger C., Coop G., Ralph P. 2012. Is your phylogeny informative? Measuring the power of comparative methods. *Evolution* 66:2240–2251.
- Bruggeman J., Heringa J., Brandt B.W. 2009. PhyloPars: estimation of missing parameter values using phylogeny. *Nucleic Acids Res.* 37:W179–W184.
- Chitwood D.H., Ranjan A., Martinez C.C., Headland L.R., Thiem T., Kumar R., Covington M.F., Hatcher T., Naylor D.T., Zimmerman S., Downs N., Raymundo N., Buckler E.S., Maloof J.N., Aradhya M., Prins B., Li L., Myles S., Sinha N.R. 2014. A modern ampelography: a genetic basis for leaf shape and venation patterning in grape. *Plant Physiol.* 164:259–272.
- Cox D.R., Reid N. 2004. A note on pseudolikelihood constructed from marginal densities. *Biometrika* 91:729–737.
- Denton J.J., Adams D.C. 2015. A new phylogenetic test for comparing multiple high-dimensional evolutionary rates suggests interplay of evolutionary rates and modularity in lanternfishes (Myctophiformes; Myctophidae). *Evolution* 69:2425–2440.
- Dunn C.W., Luo X., Wu Z. 2013. Phylogenetic analysis of gene expression. *Integr. Comp. Biol.* 53:847–856.
- Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Gen.* 25:471–492.
- Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17:368–376.
- Felsenstein J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15.
- Felsenstein J. 2008. Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *Am. Nat.* 171:713–725.
- Fieuws S., Verbeke G. 2006. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics.* 62:424–431.
- FitzJohn R.G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* 3:1084–1092.
- Freckleton R.P. 2012. Fast likelihood calculations for comparative analyses. *Methods Ecol. Evol.* 3:940–947.
- Garland T.J. 1992. Rate tests for phenotypic evolution using phylogenetically independent contrasts. *Am. Nat.* 171:713–725.
- Garland T., Ives A.R. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *Am. Nat.* 155:346–364.
- Goolsby E.W. 2015. Phylogenetic comparative methods for evaluating the evolutionary history of function-valued traits. *Syst. Biol.* 64:568–578.
- Hansen T.F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341–1351.
- 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., Martins E.P. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 1404–1417.
- Harmon L.J., Losos J.B., Jonathan Davies T., Gillespie R.G., Gittleman J.L., Bryan Jennings W., Kozak K.H., McPeck M.A., Moreno-Roark F., Near T.J., Purvis A., Ricklefs R.E., Schluter D., Schulte J.A. II, Seehausen O., Sidlauskas B.L., Torres-Carvajal O., Weir J.T., Mooers A. 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64:2385–2396.
- Ho L.S.T., Ané C. 2014. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* 63:397–408.
- Ives A.R., Midford P.E., Garland T. Jr. 2007. Within-species variation and measurement error in phylogenetic comparative methods. *Syst. Biol.* 56:252–270.
- Li L. 2007. Sparse sufficient dimension reduction. *Biometrika* 94:603–613.
- 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.
- O'Meara B.C., Ané C., Sanderson M.J., Wainwright P.C. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60:922–933.
- Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–884.
- Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Pennell M.W., Eastman J.M., Slater G.J., Brown J.W., Uyeda J.C., FitzJohn R.G., Alfaro M.E., Harmon L.J. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 15:2216–2218.
- Pennell M.W., FitzJohn W.K., Cornwell W.K., Harmon L.J. 2015. Model adequacy and the macroevolution of angiosperm functional traits. *Am. Nat.* 186:E33–E50.
- Pupko T., Pe'er I., Shamir R., Graur D. 2000. A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Mol. Biol. Evol.* 17:890–896.
- R Core Team. 2016. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available at URL: <https://www.R-project.org/>.
- Rabosky D.L., Donnellan S.C., Grundler M., Lovette I.J. 2014. Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. *Syst. Biol.* 63:610–627.
- Revell L.J., Collar D.C. 2009. Phylogenetic analysis of the evolutionary correlation using likelihood. *Evolution* 63:1090–1100.

- Revell L.J., Harrison A.S. 2008. PCCA: a program for phylogenetic canonical correlation analysis. *Bioinformatics* 24:1018–1020.
- Revell L.J. 2010. Phylogenetic signal and linear regression on species data. *Methods Ecol. Evol.* 1:319–329.
- Revell L.J. 2012. *phytools*: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3: 217–223.
- Thomas G.H., Freckleton R.P., Székely T. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. *Proc. R. Soc. B.* 273:1619–1624.
- Varin C., Vidoni P. 2005. A note on composite likelihood inference and model selection. *Biometrika.* 92:519–528.
- Yang Z., Kumar S., Nei M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141:1641–1650.