doi:10.1111/evo.13492

Phylogenetic ANOVA: Group-clade aggregation, biological challenges, and a refined permutation procedure

Dean C. Adams[1,2,3] and Michael L. Collyer[4]

[1]*Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa*

[2]*Department of Statistics, Iowa State University, Ames, Iowa*

[3]*E-mail: dcadams@iastate.edu*

[4]*Department of Science, Chatham University, Pittsburgh, Pennsylvania*

Received January 11, 2018
Accepted April 12, 2018

Phylogenetic regression is frequently used in macroevolutionary studies, and its statistical properties have been thoroughly investigated. By contrast, phylogenetic ANOVA has received relatively less attention, and the conditions leading to incorrect statistical and biological inferences when comparing multivariate phenotypes among groups remain underexplored. Here, we propose a refined method of randomizing residuals in a permutation procedure (RRPP) for evaluating phenotypic differences among groups while conditioning the data on the phylogeny. We show that RRPP displays appropriate statistical properties for both phylogenetic ANOVA and regression models, and for univariate and multivariate datasets. For ANOVA, we find that RRPP exhibits higher statistical power than methods utilizing phylogenetic simulation. Additionally, we investigate how group dispersion across the phylogeny affects inferences, and reveal that highly aggregated groups generate strong and significant correlations with the phylogeny, which reduce statistical power and subsequently affect biological interpretations. We discuss the broader implications of this phylogenetic group aggregation, and its relation to challenges encountered with other comparative methods where one or a few transitions in discrete traits are observed on the phylogeny. Finally, we recommend that phylogenetic comparative studies of continuous trait data use RRPP for assessing the significance of indicator variables as sources of trait variation.

KEY WORDS: **M*acroevolution*, *morphological evolution*, *multivariate data*, phylogenetic comparative methods.**

Understanding how traits correlate across species is fundamental to evolutionary biology, and evaluating such patterns requires a phylogenetic perspective. Over the past several decades, the phylogenetic comparative toolkit has grown to include a diverse set of analytical methods to evaluate myriad biological hypotheses in a phylogenetic framework (e.g., Felsenstein 1985; Garland et al. 1992; Blomberg et al. 2003; O'Meara et al. 2006; Revell and Harmon 2008; Beaulieu et al. 2012; Blomberg et al. 2012; Pennell and Harmon 2013). Likewise, recent years have seen the advent of statistical approaches for evaluating patterns of trait evolution in multivariate phenotypes (Revell and Harmon 2008; Klingenberg and Gidaszewski 2010; Bartoszek et al. 2012; Klingenberg and Marugán-Lobón 2013; Adams 2014a,b,c; Adams and Felice 2014; Uyeda et al. 2015; Goolsby 2016; Adams and Collyer 2018). Together, these analytical tools are now in standard use in macroevolutionary studies evaluating trends in phenotypic datasets (e.g., Baker et al. 2015; Friedman et al. 2015; Zelditch et al. 2015; Moen et al. 2016; Reynolds et al. 2016; Sherratt et al. 2016; Arbour and López-Fernández 2017; Serb et al. 2017).

For many macroevolutionary hypotheses, patterns in phenotypes are evaluated with respect to one or more independent (predictor) variables using phylogenetic linear models. In this context, most of the theoretical development has centered on phylogenetic regression (simple linear regression models that account for phylogenetic relatedness of covariates). However, because generalized least squares (GLS) estimation of coefficients

for linear models has become a standard method of phylogenetic regression, it has been tacitly assumed that the mathematical and statistical properties observed for implementations of phylogenetic regression correspond equally to implementations for other types of linear models as well. Inasmuch as this assumption is perhaps safe for coefficient estimation (mathematical properties), assumptions about the dispersion of such statistics (i.e., their statistical properties) require more scrutiny over different scenarios. In this article, we introduce a refined method of randomizing residuals in a permutation procedure (RRPP) for statistically evaluating phylogenetic linear models. We demonstrate that this procedure displays appropriate statistical properties such as type I error, power, and other properties under diverse scenarios, and that these properties hold when examining various types of response variables, including both (analysis of variance [ANOVA]) and regression designs. With respect to group differences ANOVA, we compare the statistical performance of RRPP to alternative methods based on phylogenetic simulation (Garland et al. 1993), and find that RRPP exhibits higher power than this alternative. Additionally, we explore how the dispersion of groups across the phylogeny affects statistical and biological inferences. We discover that aggregation of groups to particular sublineages within a phylogeny reduces statistical power, and subsequently affects biological inferences from these data. The broader implications of this phylogeny-group aggregation on our ability to make accurate macroevolutionary inferences are discussed.

A GENERAL PERMUTATION PROCEDURE FOR PHYLOGENETIC LINEAR MODELS

The goal of phylogenetic linear models is to ascertain whether there is an association between species traits and ecological variables while accounting for phylogenetic nonindependence. In phylogenetic regression, the independent variable, X, is a continuous quantitative variable. However, when X represents a factor containing indicator values (e.g., 0 or 1), phylogenetic linear models are conceptually no different in approach, and in this case assess the association of the grouping variable with the dependent variables (henceforth referred to as phylogenetic ANOVA). Most implementations of phylogenetic linear models are based on GLS, where coefficients (slopes) are estimated between sets of trait values ($\mathbf{Y}$) and independent variables (X) while accounting for the nonindependence among observations as described by the phylogeny. Three common algebraic formulations are encountered in the macroevolutionary literature: phylogenetically independent contrasts (PICs: Felsenstein 1985), phylogenetic generalized least squares (PGLSs: Grafen 1989; Martins and Hansen 1997), and phylogenetic transformation (Garland and Ives 2000; Adams 2014a). When implemented properly, all three formulations yield identical parameter estimates and model coefficients (Garland and Ives 2000; Rohlf 2001; Blomberg et al. 2012). Finally, phylogenetic linear models are typically evaluated statistically using parametric procedures such as likelihood ratio tests or the evaluation of derived parametric summary measures such as F ratios; though both permutation and simulation methods are sometimes used, particularly for multivariate data (see discussion in Adams and Collyer 2018).

One approach to evaluating multivariate phylogenetic regression makes use of phylogenetic transformation and permutation procedures. Here, a phylogenetic transformation matrix (Garland and Ives 2000) is used to condition both the independent and dependent variables on the phylogeny, and parameter estimates are obtained from the transformed data. Statistical evaluation is then accomplished by a permutation procedure, in the simplest case permuting the rows of the dependent variables and repeating the transformation and parameter estimation to generate empirical sampling distributions of test statistics for comparison to the observed test value (Adams 2014a; Adams and Collyer 2015). For more complex multifactor designs, RRPPs are preferable, and their use has been explored for ordinary least squares (OLSs) models (Collyer et al. 2015). For the case of phylogenetic regression, randomizing dependent data was previously shown to have appropriate type I error rates under various scenarios (Adams 2014a). However, using simulations with pure birth trees, Goolsby (2016) documented slightly elevated type I error rates for PGLS models evaluated by randomizing dependent data. Given this discrepancy in results, how one should permute data for PGLS designs is revisited here. Specifically, we present a more generalized solution, based on a simple refinement of RRPP, which not only alleviates the challenges identified by Goolsby (2016), but produces appropriate type I error rates and high power regardless of the types of independent variables analyzed, including both regression and ANOVA designs. The approach is as follows.

First, the GLS model for phylogenetic ANOVA is found as:

$$\mathbf{Y} = \mathbf{X}\hat{\mathbf{B}} + \mathbf{E}, \tag{1}$$

where $\mathbf{Y}$ is an $n \times p$ matrix of trait values for the n species for p traits, $\mathbf{X}$ is an $n \times k$ design matrix containing optionally an $n \times 1$ vector of ones for an intercept plus one or more independent variables (X), $\hat{\mathbf{B}}$ is a $k \times p$ matrix of model coefficients, and $\mathbf{E}$ is an $n \times p$ matrix of residuals. The residuals of the model ($\mathbf{E}$) are not independent, but have an expected covariance of $\sigma^2\mathbf{C}$ where $\mathbf{C}$ is an $n \times n$ phylogenetic covariance matrix and σ^2 is the Brownian motion rate parameter (Rohlf 2001; Blomberg et al. 2012). One may reformulate the problem above via phylogenetic transformation (Garland and Ives 2000), based on the well-known OLS transformation of GLS models (Judge et al. 1985; Johnston and DiNardo 1997; Rencher 2000; Kariya and Kurata 2004). Here, a phylogenetic transformation matrix is

obtained as: $\mathbf{P} = (\mathbf{U}\mathbf{W}^{-1/2}\mathbf{U}^t)^{-1}$, where **U** and **W** are the matrices of eigenvectors and eigenvalues of **C**, respectively (Garland and Ives 2000; also Adams 2014a; Adams and Collyer 2015). Transforming the model design and data by projection yields: $\mathbf{X}_{\text{phy}} = \mathbf{PX}$, $\mathbf{Y}_{\text{phy}} = \mathbf{PY}$, and $\mathbf{E}_{\text{phy}} = \mathbf{PE}$. The GLS problem may then be reformulated as:

$$\mathbf{PY} = \mathbf{PX}\hat{\mathbf{B}} + \mathbf{PE}, \quad (2)$$

which is equivalently written as: $\mathbf{Y}_{\text{phy}} = \mathbf{X}_{\text{phy}}\hat{\mathbf{B}} + \mathbf{E}_{\text{phy}}$. Coefficients of the model may be found via the OLS solution as: $\hat{\mathbf{B}} = (\mathbf{X}^t_{\text{phy}}\mathbf{X}_{\text{phy}})^{-1}\mathbf{X}^t_{\text{phy}}\mathbf{Y}_{\text{phy}}$, which yields identical estimates to those found from GLS methods. Unlike the residuals from equation (1), which are correlated, it has been shown that phylogenetic transformation yields residuals with uncorrelated error, as one expects with OLS models: $E\{\mathbf{E}_{\text{phy}}\boldsymbol{E}^T_{\text{phy}}\} = E\{\mathbf{PE}(\mathbf{PE})^T\} = \mathbf{P}\sigma^2\mathbf{C}\,\mathbf{P}^T = \sigma^2\,\mathbf{I}$ (Garland and Ives 2000; see also Johnston and DiNardo 1997). In other words, $\mathbf{E}_{\text{phy}}$ are independent of the phylogeny as expressed in **C**. As a consequence, because $\mathbf{E}_{\text{phy}}$ are independent, they are appropriate as the exchangeable units under the null hypothesis in a permutation procedure to assess model significance (*sensu* Collyer et al. 2015). This assignment of exchangeable units differs from the one offered by Adams and Collyer (2015), who argued the untransformed residuals were appropriate exchangeable units, as in the OLS case. However, unlike OLS models, with GLS models an unequivocal definition of null hypothesis exchangeability is not possible (Commenges 2003), and the most appropriate assignment of exchangeable units involves a heuristic investigation (see the Supporting Information for more detail). In the case that p is small relative to n, parametric evaluation of models is possible. However, RRPP permits evaluation of models for multivariate data (including high-dimensional data, where p is large relative to n), but also achieves consistent results compared to parametric evaluation for low-dimensional cases (as we demonstrate in the Supporting Information). Thus, RRPP (using phylogenetically transformed residuals as exchangeable units) provides a more universal paradigm of model evaluation.

To implement RRPP, both full and reduced design matrices ($\mathbf{X}_{\text{f}}$ and $\mathbf{X}_{\text{r}}$) are defined for the statistical model of interest, such that $\mathbf{X}_{\text{r}}$ contains one fewer term as compared to $\mathbf{X}_{\text{f}}$ (for details see Collyer et al. 2015). Next, the design matrices and the matrix of dependent variables (**Y**) are transformed by **P**, conditioning them on the phylogeny. Coefficients ($\hat{\mathbf{B}}$) are then calculated for the reduced model as above and predicted values ($\mathbf{X}_{\text{phy}}\hat{\mathbf{B}}$) and residuals ($\mathbf{E}_{\text{phy}}$) are obtained. These residuals are then permuted, added to the predicted values, and parameters are obtained from the full model using these data. Repeating this procedure many times yields empirical sampling distributions for test statistics (e.g., F) from which statistical significance may be evaluated (see Collyer et al. 2015). It should be emphasized that the main difference between the RRPP approach described here and previous permutation methods is that RRPP permutes residuals from data conditioned on the phylogeny, and it is those residuals that have been rendered independent relative to the phylogeny (see additional discussion in Supporting Information). By inference, any GLS model where data are conditioned on the phylogeny yields residuals that are appropriate for downstream significance testing via ANOVA with RRPP. Therefore, the RRPP approach described here is appropriate for evaluating all manner of phylogenetic linear models, including phylogenetic regression, phylogenetic ANOVA, and phylogenetic factorial models, as it accounts for the correlation between phylogenetic covariance and model design.

COMPLICATING ISSUES: GROUP DISPERSION ON THE PHYLOGENY

When considering phylogenetic ANOVA, one issue that has received relatively little attention is the extent to which phylogenetic covariance matrices and model design matrices (**X**) are correlated. For regression, several authors have examined the relationship between a continuous independent variable and the phylogeny (e.g., Freckleton et al. 2011; Mazel et al. 2016). However, to our knowledge, an equivalent examination of the effect of grouping variables has not been explored. To illustrate the issue, consider the hypothetical scenarios in Figure 1. Here, we have the phylogeny for a group of taxa, along with several ecological grouping variables to which the species belong. The ecological variables include large versus small species (Fig. 1A), the presence or absence of a predator species within the prey species' range (Fig. 1B), species that have invaded a new geographic region (Fig. 1C), and distinct microhabitat use (Fig. 1D). The evolutionary biologist wishes to determine whether multivariate phenotypes differ between species as a function of each ecological group designation while accounting for shared evolutionary history. However, while each hypothesis is biologically reasonable to consider, as are hypotheses that combine these ecological variables, a potential complication is that the manner in which the groups are distributed on the phylogeny differs for each of the ecological variables. How this might affect statistical and biological inferences is currently unknown.

Theoretical development of PGLS until now has generally assumed independence between phylogenetic covariance matrices (**C**) and model design matrices (**X**), especially as simulation studies tend to use phylogenetic regression with random normal variables simulated from Brownian motion models of evolution, given a phylogeny, which naturally simulates independence between **C** and a continuous variable, X (e.g., Revell 2010; Goolsby 2016; Adams and Collyer 2018; Uyeda et al. 2018). However, visual inspection of the hypothetical examples in Figure 1

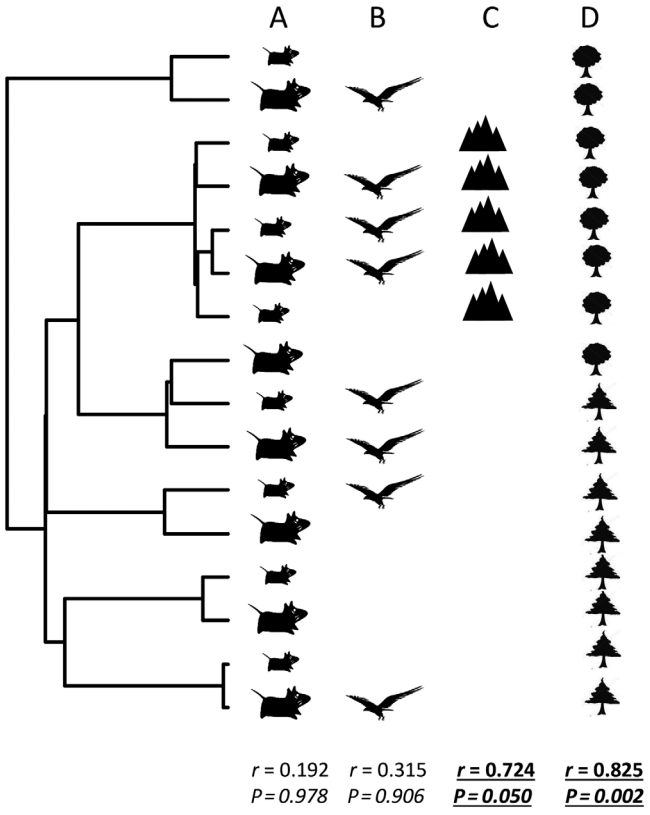

Figure 1. Phylogeny for 16 species in a hypothetical lineage. Several ecological grouping variables are displayed (A) size groups, (B) predator presence/absence, (C) invasion of a new geographic region, and (D) ecoregions. For each ecological scenario, the PLS correlation between the design matrix of that scenario and the phylogenetic covariance matrix is displayed. When all variables are included in the design matrix, the correlation is: $r_{PLS} = 0.816$; $P = 0.034$.

demonstrate that **C** and **X** (where **X** is a matrix that includes both a vector of ones for an intercept and one or more *X* variables) will be inherently correlated. To diagnose the problem statistically, one may assess the multivariate correlation between **C** and **X**. We propose the use of two-block partial least squares to accomplish this task. Partial least squares is a statistical method that identifies the maximal covariation between two sets of variables based on singular value decomposition of the covariance matrix across sets of variables (Jöreskog and Wold 1982; Rohlf and Corti 2000; Abdi and Williams 2013). Applying the approach to the phylogeny and the ecological variables in Figure 1, we observe a significant correlation between the phylogeny and ecological groups in variables C and D, but not for variables A and B. (Correlation is also strong when all ecological variables are combined, suggesting that any variable correlation produces an inherently correlated design matrix; see Fig. 1.) Visual inspection of Figure 1 reveals why: variables C and D display a high degree of aggregation of the ecological groups to particular sublineages of the phylogeny that is not present for groups in variables A and B. This observation reveals a potential concern when performing phylogenetic ANOVA; that the distribution of groups across the phylogeny can result in high or low correlations between the independent (grouping) variable and the phylogeny. Thus, even *before* we consider the relationship between multivariate phenotypes and ecological groups via phylogenetic ANOVA, we should first investigate whether groups are conflated with phylogenetic sublineages. The question then becomes, if such phylogenetic aggregation of groups exists, can this affect our statistical and biological inferences?

Methods and Results

PHYLOGENETIC ANOVA: SIMULATIONS EVALUATING STATISTICAL PERFORMANCE

To evaluate the analytical performance of the proposed RRPP procedure, we conducted a series of computer simulations. First, we examined the type I error and power of the approach for both univariate and multivariate data. For each simulation, we generated 1000 pure-birth trees containing 64 taxa, and on each we generated two distinct grouping variables: one where groups were assigned randomly to species (**C** and **X** not inherently correlated, chance notwithstanding) and a second where groups were assigned in an aggregated manner (the first 32 species were assigned to group A and the second 32 species to group B, such that some degree of inherent correlation between **C** and **X** was imposed). Group designation was represented as a vector of indicator values: $X = 0,1$. Next, we selected the number of dimensions for the dependent (**Y**) variable ($p = 1$, 5, and 10), and selected β_{input} to represent the known group differences ($\beta_{input} = 0.0$, 0.1, 0.25, 0.5, 0.75, 1.0). We then simulated 1000 datasets (1 per phylogeny) of the dependent (**Y**) variable under a Brownian motion model of evolution, with known differences in groups incorporated. This was accomplished by simulating random noise under a Brownian model (Y_{BM}) using a $p \times p$ identity matrix to represent the input covariance matrix. To incorporate known differences among groups, predicted values were obtained by multiplying the indicator variable *X* by the input regression coefficient, and adding this to the BM-simulated data: $Y = \beta X + Y_{BM}$. The resulting data therefore contained trait variation as expected under Brownian motion evolution on the phylogeny, plus a known degree of group difference (when $\beta_{input} > 0.0$).

For each simulated dataset, phylogenetic patterns were evaluated using the refined RRPP procedure proposed here. The proportion of simulated datasets whose significance level was less than the nominal $\alpha = 0.05$ was treated as an estimate of type I error ($\beta_{input} = 0.0$) or power ($\beta_{input} > 0.0$) of the test. Type I error and power were determined for RRPP under both the random and

aggregated group scenarios. In parallel, we conducted phylogenetic ANOVA on the same simulated datasets using a previously proposed approach based on phylogenetic simulation (Garland et al. 1993; generalized to multivariate in *geiger*: Pennell et al. 2014). This approach does not condition the data on the phylogeny during the ANOVA computations, but rather compares results from an OLS ANOVA to those obtained from datasets generated on the phylogeny under Brownian motion. Type I error rates and power were obtained for this method for comparison. Note that we did not include comparisons with parametric evaluation of PGLS for ANOVA. Prior work has demonstrated that parametric evaluation of phylogenetic regression displays decreasing power as the number of dependent trait dimensions (p) increases, and statistical evaluation is not possible when $p \geq (n - k)$, where n and k are the number of observations and model parameters, respectively (Adams 2014a). We confirmed through simulation that this debilitating property likewise holds for parametric evaluations of multivariate phylogenetic ANOVA (results not shown), so this alternative was not included in the simulations.

RESULTS

Results of our simulations are found in Figure 2. The results reveal several important patterns. First, both phylogenetic ANOVA evaluated by simulation (Garland et al. 1993) and the revised RRPP procedure display appropriate type I error rates at the nominal $\alpha = 0.05$ for both random and aggregated groups, and for both univariate and multivariate datasets. Second, the statistical power of both procedures increases as the known differences among groups increases (Fig. 2). However, evaluating such patterns using RRPP displays higher statistical power as compared to using phylogenetic simulation (Fig. 2), revealing that RRPP is more capable of detecting group differences in a phylogenetic context when they are present. Additionally, it should be noted that because approaches based on phylogenetic simulation use OLS methods (Garland et al. 1993; Pennell et al. 2014), parameter estimates are incorrect, as they are not conditioned on the phylogeny. Additionally, because statistical evaluation is based on parametric test statistics such as Wilks' λ, the phylogenetic simulation approach cannot be used for highly multivariate data, as such test measures cannot be calculated when $p \geq (n - k)$. Therefore, when evaluating phenotypic differences among groups in a phylogenetic context, the new RRPP approach is preferred.

Inspection of the results illustrated in Figure 2 reveals another important pattern; namely that statistical power is compromised when groups are aggregated on the phylogeny. Comparing power curves obtained between randomly distributed and aggregated groups, it is discovered that power decreases when groups are more highly aggregated on the phylogeny. This reveals an important point: that the manner in which independent grouping variables associate with the phylogeny can affect downstream inferences on how dependent ($\mathbf{Y}$) and independent (X) variables associate with one another (see also Fig. 1). This point will be revisited in more detail in the Discussion below.

Finally, we performed further simulations to evaluate additional statistical properties of the RRPP procedure proposed here. The results of these simulations are found in the Supporting Information. Summarizing those results: for phylogenetic regression, the RRPP method yields identical model coefficients and statistics (β and F) as those obtained from other GLS implementations, and produces significance estimates that are highly correlated with earlier permutation procedures (Supporting Information). Additionally, the RRPP method displays low bias in parameter estimation, and generates empirical sampling F distributions that match theoretical expectation (Supporting Information). And following the theoretical requirements outlined in Adams and Collyer (2018) for multivariate phylogenetic comparative methods, the approach is rotation-invariant, insensitive to differing levels of covariation in the dependent ($\mathbf{Y}$) variables, and is robust to increasing trait dimensionality (Supporting Information). Taken together, these results allow one to argue through consilience that the RRPP method described here provides a useful refinement of existing phylogenetic comparative methods that is appropriate for evaluating patterns in phenotypes as described by regression, ANOVA, and other linear models. The new method is implemented in *geomorph* 3.0.6 (Adams et al. 2018).

A Biological Example

To illustrate the utility of the RRPP approach described above, we examined patterns of phenotypic variation (body proportions) in salamanders of the genus *Plethodon.* This lineage of terrestrial salamanders is widely distributed in North American forests, and extensive ecological research has shown that interspecific competition is widespread (Jaeger 1970; Hairston 1980; Anthony et al. 1997). Further, competitive interactions affect patterns of community composition (Adams 2007), and result in phenotypic differences between populations and species (e.g., Adams and Rohlf 2000; Adams 2010). In eastern North America, most species belong to one of two functional guilds, which also correspond to differences in overall body size between taxa (Highton 1995). Thus, an interesting ecological hypothesis to consider is whether multivariate phenotypes differ between functional guilds, and whether the allometric relationship for multivariate phenotypes relative to body size is similar across guilds.

For the present example, we selected 37 species representing members of two functional guilds, which are commonly referred to as the *P. cinereus* and *P. glutinosus* species complexes (e.g., Highton 1995). A recent phylogeny for the group is found in

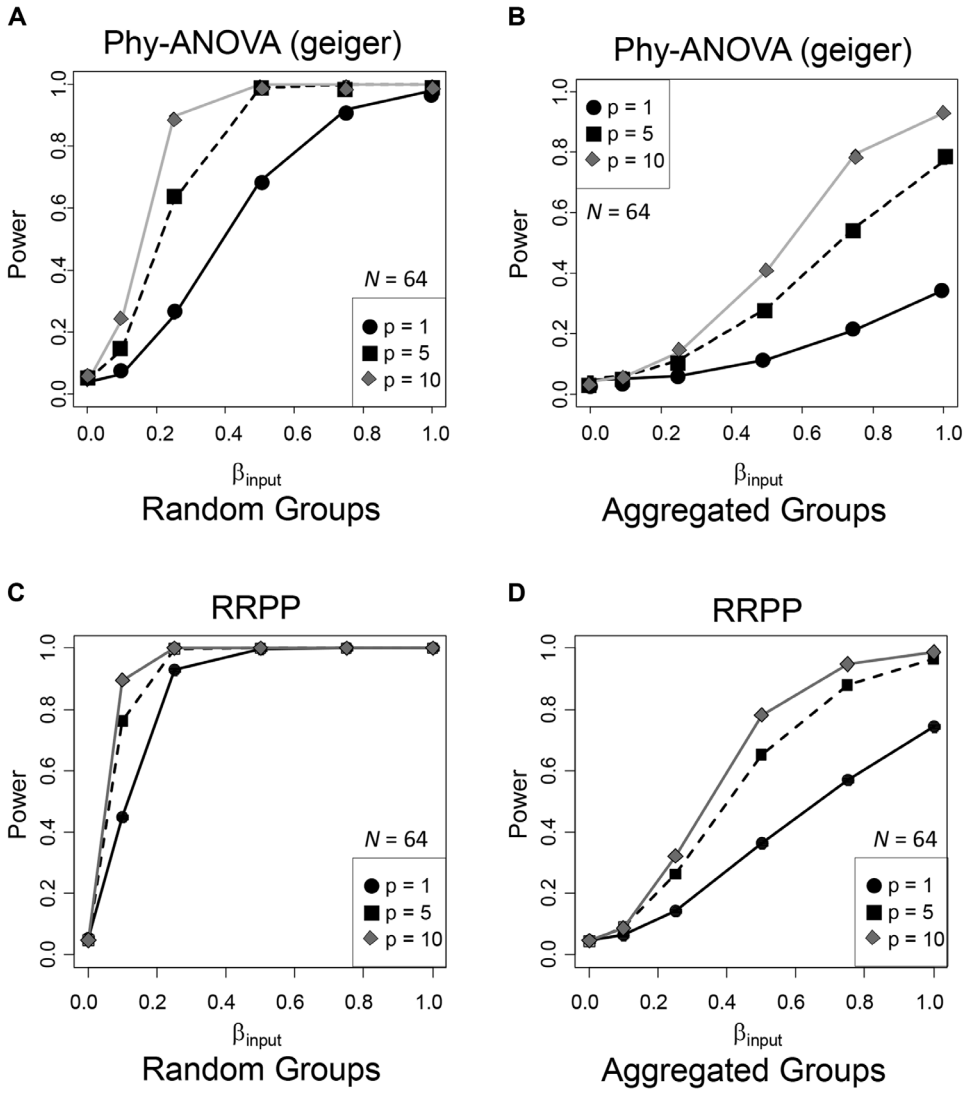

Figure 2. Results from statistical simulations evaluating methods for two implementations of phylogenetic ANOVA. Type I error and power curves for phylogenetic ANOVA using OLS plus simulation (Garland et al. 1993) when groups are (A) randomly distributed, or (B) aggregated across the phylogeny. Type I error and power curves for RRPP methods when groups are (C) randomly distributed, or (D) aggregated across the phylogeny.

Figure 3A, which represents a time-calibrated maximum clade credibility tree obtained from a sampling of the posterior distribution of trees from a Bayesian analysis of a multigene dataset (Bonett and Blair 2017). Note that each of the functional guilds is reciprocally monophyletic, implying that species in these two groups are perfectly aggregated into sublineages. A partial least squares analysis confirmed this visual finding ($r_{PLS} = 0.99$; $P < 0.0001$), demonstrating a strong correlation between ecological groups and the phylogeny. Thus, this example represents the "worst case" scenario with respect to group aggregation on the phylogeny.

Phenotypic means for each species were obtained from previous sources (Adams and Church 2011; Blankers et al. 2012; Adams 2013) and included body size (snout-vent length), head length, snout-eye distance, body width, forelimb length, and hindlimb length (Fig. 3B). The latter five measurements were then divided by body size and were treated as a multivariate dataset representing the relative body proportions of each species. We performed analyses in three principal ways to illustrate the advantages of the proposed RRPP procedure for empirical investigations (using ANOVA via RRPP) with: (1) OLS estimation of coefficients, ignoring phylogeny; (2) PGLS estimation of coefficients, using untransformed residuals in RRPP; and (3) PGLS estimation of coefficients, using transformed residuals in RRPP (as described above); henceforth OLS, PGLS-u, and PGLS-t, respectively, for simplicity and to distinguish between untransformed (-u) and transformed (-t) residuals used in RRPP. In each of these analytical methods, we calculated single-factor, type I sums of

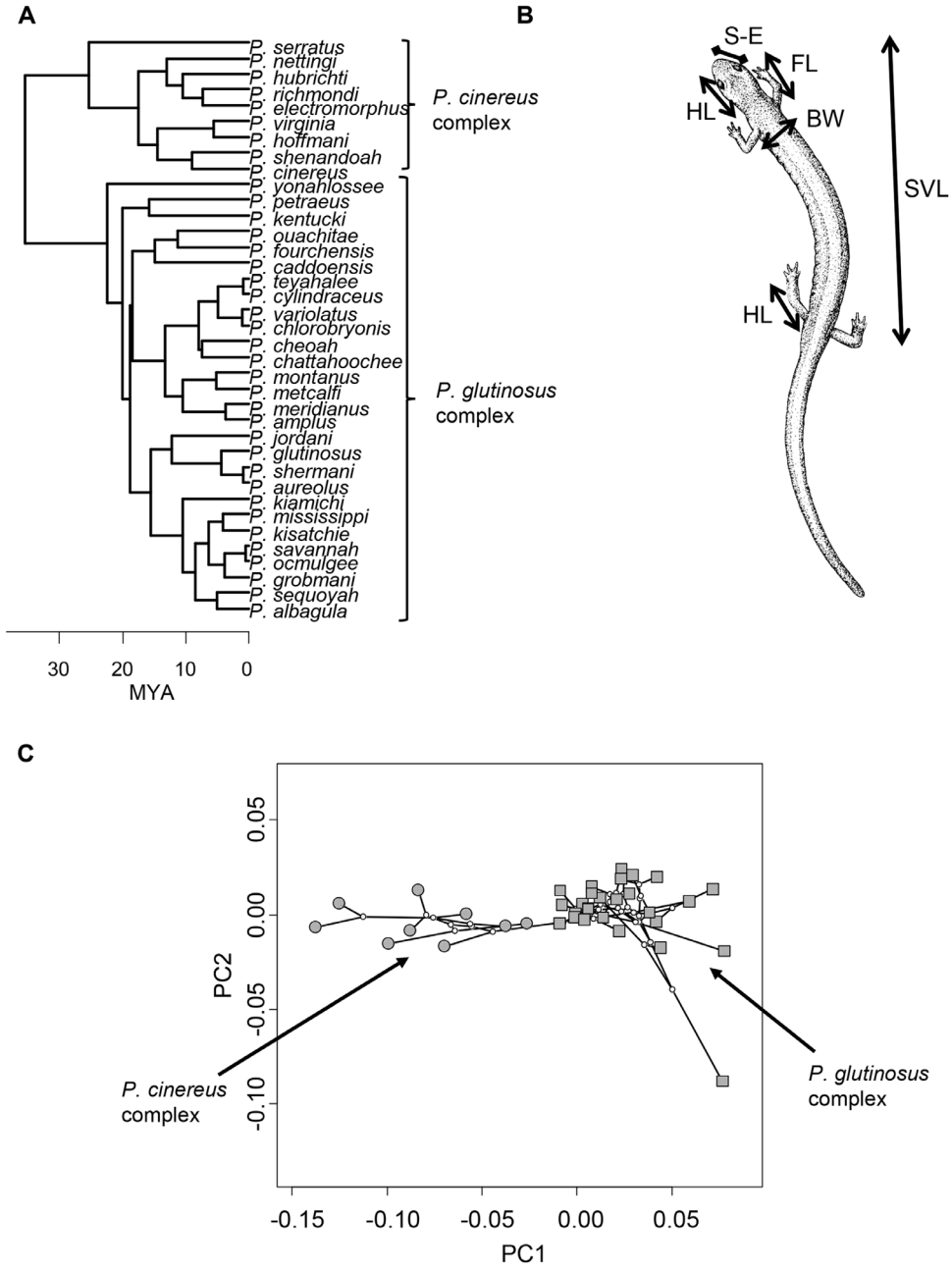

Figure 3. Results from empirical example. (A) A time-calibrated molecular phylogeny for the group, (B) the linear measurements used in this study (image redrawn from Adams (2013)), (C) phylomorphospace representing the variation in body proportions for the species. The first two dimensions explain 82% of the variation.

squares and cross-products (*SSCP*, with *SS* equal to the trace of the *SSCP* matrix) for the main effects, *Size* and *Guild*, as solitary model effects. The RRPP procedure in the case of single model effect for both the OLS and the PGLS methods is functionally equivalent to randomizing the row vectors of the observed data matrix $\mathbf{Y}$ (Collyer et al. 2015) and for PGLS-u, specifically, is the same as the original D-PGLS method introduced by Adams (2014a). PGLS-t is a functional analog of randomizing the row vectors of $\mathbf{Y}_{phy}$ in this case, as randomizing the transformed residuals, $\mathbf{E}_{phy}$, preserves the global exchangeability (constant mean and variance through permutations; sensu Commenges 2003) of the transformed data (see Supporting Information for details). Additionally, these models allowed parametric multivariate analysis of variance (MANOVA) with exact *F* values, which we performed (generalizing from Roy's maximum root) to compare nonparametric and parametric *P* values. This additional step helped to emphasize divergence from theoretical expectation for any of the three methods.

For each analytical method, we also calculated type II *SS* for main effects and their interaction; that is, $\mathbf{Y} \sim Size + Guild + Size \times Guild$. Type II *SS* are important in this case because each main effect *SS* is conditioned on the *SS* of the alternative main effect. Thus, we were able to determine if the variable, *Guild*, which confounds phylogeny and ecology, was meaningful in spite of phylogeny and species' size. In all cases, the coefficient of determination (R^2), *F*-value, effect size (*Z*), and *P*-value were estimated. The *Z* score for effect size was estimated as the SD of the observed *F*-value in the distribution of RRPP-generated values (after log-transformation for normalization). The *P*-value is the probability of finding a larger RRPP-generated *F*-value than the observed, by chance. In all cases, an acceptable type I error rate of $\alpha = 0.05$ was used as a significance level. All analyses were performed in R 3.4.1 (R Development Core Team 2017) using the RRPP approach described above as implemented in *geomorph* 3.0.6 (Adams et al. 2018: data and R scripts available in DRYAD https://doi.org/doi:10.5061/dryad.2s8d0f9).

RESULTS

We consider our results, summarized in Table 1, on multiple levels. First, the *Guild* effect was understandably large in size for OLS evaluation and comparatively small in PGLS-u and PGLS-t evaluations. OLS estimation of coefficients and subsequent *SS* calculations failed to account for the most prominent source of variation, evolutionary history. For type I *SS* effects, both OLS and PGLS-t produced *P* values similar to parametric *P* values, at least in terms of evaluating the significance of model effects, but in the case of PGLS-u, the *Guild* effect was large and significant ($Z = 3.34$, $P = 0.001$) despite a small *F*-value calculation (1.70), suggesting a type I error, consistent with our simulation experiments. Comparatively, the type I *SS* ANOVA statistics were quite similar for *Size* between PGLS-u and PGLS-t, emphasizing the mitigating nature of a continuous quantitative covariate. Together, the type I *SS* ANOVA statistics emphasize that not transforming residuals can exacerbate inferential error for ecophylogenetically correlated variables.

The potential for inferential error was further increased when considering type II *SS* analyses, but these analyses were also more enlightening. In the case of OLS analysis, the reduction of R^2 for the *Guild* effect when conditioning it on the *Size* effect (from $R^2 = 0.62$ to $R^2 = 0.45$) revealed the multicollinearity between *Size* and *Guild*, which is not surprising for *Plethodon*, as species in these two ecological guilds also differ in overall body size. However, when accounting for phylogeny and conditioning the *Guild* effect on the *Size* effect, R^2 increased (from $R^2 = 0.04$ to $R^2 = 0.09$). This result suggests that after accounting for phylogeny and the covariation between morphology and species' size, the *Guild* effect is not merely redundant with phylogeny, as the 9% explained variation was significant in either case. However,

Table 1. ANOVA results with varied methods describing variation in body proportions as a function of overall size, functional groups, and their interaction.

	OLS					PGLS-u					PGLS-t				
	Type I *SS*		Type II (conditional) *SS*			Type I *SS*		Type II (conditional) *SS*			Type I *SS*		Type II (conditional) *SS*		
Statistic	Size	Guild	Size	Guild	Size: Guild	Size	Guild	Size	Guild	Size: Guild	Size	Guild	Size	Guild	Size: Guild
R^2	0.21885	*0.62333*	0.0430	*0.4474*	0.0078	0.1731	*0.0464*	0.2170	*0.0904*	0.0011	**0.1731**	***0.0464***	**0.2170**	***0.0904***	**0.0011**
F	9.8058	57.919	4.3494	45.3008	0.7871	7.3251	1.7049	9.7371	4.0557	0.0483	**7.3251**	**1.7049**	**9.7371**	**4.0557**	**0.0483**
Z	2.2239	3.5976	2.7529	3.7276	1.2270	1.6553	3.3420	2.8266	3.5756	−0.4385	**2.1588**	**1.0069**	**2.4020**	**1.9435**	**−2.0972**
P	0.005	0.001	0.001	0.001	0.104	0.024	0.001	0.001	0.001	0.679	**0.009**	**0.173**	**0.004**	**0.023**	**0.972**
P_{param}	<0.001	<0.001				0.023	0.483				0.023	0.483			

Three general methods used were as follows: OLS, PGLS-u, and PGLS-t (see text for definitions), and within type I and type II, *SS* were calculated for main effects, and type II *SS* for their interaction. Effect sizes (*Z*) and *P* values were obtained from RRPP, and parametric *P* values were also estimated for type I *SS* cases, for comparison. The preferred method, PGLS-t, is represented in bold face and the coefficient of determination for the Guild effect is italicized for comparative emphasis.

only in the case of PGLS-t do the ANOVA statistics seem reasonable. The $R^2 = 0.09$ result corresponded to a modest F-value of 4.06. For PGLS-t, this translated to an effect size of 1.94 SDs greater than expected by chance and a P-value of 0.023 (which is also rather consistent with an area under a standard normal curve of 0.026 for $z = 1.94$). By contrast, the effect size of 3.34 SDs greater than expectation and a P-value of 0.001 from PGLS-u are quite implausible for $R^2 = 0.09$ and $F = 4.06$, again reiterating the folly of not transforming residuals prior to RRPP. These discrepancies were mitigated for the *Size* effect, but even in this case, the PGLS-t method resulted in much more intuitive correspondence among ANOVA statistics. Finally, although there were no significant *Size:Guild* interactions observed, limiting comparative interpretations of methods, the effect size was much smaller with PGLS-t, emphasizing a perhaps more conservative method (and less prone to type I error).

Overall, the PGLS-t result implied that a common allometric pattern was present across both guilds of species. Interpreting differences in body proportions between guilds through inspection of principal component loadings revealed that species in the *Plethodon glutinosus* functional group displayed relatively larger body proportions, especially in their limbs. This pattern is visually confirmed in the phylomorphospace plot (Fig. 3C), where there was a clear separation of the two functional groups in morphospace. Additionally, the significant evolutionary allometry was easily visualized in this plot, as the small species of the *P. cinereus* functional group were located toward the negative side of PC1, whereas the larger species of the *P. glutinosus* group were found toward the positive side of PC1.

Discussion

In this article, we investigated how the dispersion of groups across the phylogeny affects statistical and biological inferences when evaluating multivariate phenotypes using phylogenetic ANOVA. Our study revealed several key points. First, with respect to evaluating phylogenetic linear models, we introduced a refined RRPP that permutes residuals from data conditioned on the phylogeny, and therefore represents appropriate exchangeable units for GLS models (Supporting Information). We demonstrated that the approach displays appropriate type I error rates and statistical power (Fig. 2) and retains numerous desirable properties when evaluating both regression and ANOVA models (Supporting Information). Goolsby (2016) documented elevated ANOVA type I error rates on PGLS models (for phylogenetic regression) using simulations with pure birth trees, which differed from the results of Adams (2014a) using simulations with random split trees. In the current analyses, we used pure birth trees for both regression and ANOVA models and found type I error rates were appropriate. (We also verified type I error rates were appropriate with random split trees but did not report these results.) Although Adams and Collyer (2018) reasoned that the elevated type I error rates were probably associated with the method of random tree generation, it appears that transforming residuals was an essential missing component of the RRPP procedure that may have contributed to elevated type I error rates in certain scenarios. Our empirical example helped to elucidate the issue, and along with our simulated examples confirmed that previous accounts of elevated type I error rates were a result of the permutation method used.

Under broad conditions, the appropriate type I error rates and power with RRPP performed on transformed residuals suggests that accounting for phylogenetic relatedness in residuals prior to RRPP is a heuristically vetted need. This outcome is not surprising, as Adams and Collyer (2018) also found that phylogenetic transformation of datasets prior to randomization in two-block partial least squares analysis resulted in appropriate type I error rates, contrary to the elevated rates pertaining to postrandomization transformation (as reported by Goolsby 2016). These convergent outcomes demonstrate that "transformation first, randomization second," is the appropriate paradigm for permutation procedures involving phylogenetic transformation. This is not true, however, when using PICs (see Adams and Collyer 2015).

Second, we demonstrated that highly aggregated groups on the phylogeny can generate very strong and significant correlations with the phylogenetic covariance matrix. This represents an unappreciated concern when investigating patterns of group differences phylogenetically. Further, we argue that this concern is biologically relevant, as strong group–phylogeny correlations will likely be encountered in many macroevolutionary studies. In evolutionary studies, the groups that tend to draw our attention are typically not random, but often represent distinct sublineages or monophyletic clades within a broader phylogeny (e.g., a sublineage that has colonized an island or a new habitat relative to a parent lineage). Such instances substantiate the requirement that methods evaluating group differences in a phylogenetic context display the highest power possible. As shown in Figure 2, the RRPP procedure proposed here meets that requirement. Indeed our empirical example provided one such instance where robust statistical evaluation under group aggregation is required to arrive at proper biological inference, as the ecological groups were perfectly aggregated in two sublineages within the broader phylogeny (Fig. 3A). However, whenever species have diversified into new ecological niches, or when a clade colonizes a new habitat, such phylogenetic patterns will likely arise. As such, the statistical challenges associated with group aggregation on the phylogeny cannot be ignored.

One consequence of this realization is that before evaluating multivariate patterns using phylogenetic ANOVA, one must first

determine the extent to which the ecological groups of interest and are confounded with the branching patterns of the phylogeny. As a diagnostic for this problem, we recommend the use of two-block partial least squares analysis. This provides a simple statistical measure (correlation) between the phylogenetic covariance matrix and the ecological grouping variable to evaluate the extent to which the two are correlated in empirical datasets. Identifying such group aggregations on the phylogeny is critical, as correct biological interpretation depends upon the degree to which such aggregation is present. At one extreme, when groups are perfectly aggregated phylogenetically, understanding any phenotypic similarity of species within groups rests largely on inferences based on shared evolutionary history and a common ecological shift on the phylogeny (e.g., Wilke et al. 2010). At the other extreme, when groups are dispersed across the phylogeny, phenotypic similarity of species may identify instances of evolutionary convergence associated with a common ecological niche (e.g., Losos 1992; Harmon et al. 2005; Alvarado-Cárdenas et al. 2013; Serb et al. 2017). Because both extremes are biologically possible, we recommend that partial least squares analysis becomes a standard component of the phylogenetic comparative toolkit for empirical studies where understanding phenotypic differences among groups is a concern.

Third, we posit that any phylogenetic GLS model where data are conditioned on the phylogeny may be statistically evaluated using the RRPP procedure described here. For instance, in the empirical example, we demonstrated its ability to accommodate multiple explanatory factors by evaluating evolutionary hypotheses using both phylogenetic MANOVA and multivariate analysis of covariance (MANCOVA). In the Supporting Information, we demonstrate its efficacy for phylogenetic regression. Because of this analytical flexibility, we recommend that future phylogenetic comparative studies of multivariate trait evolution use this refined RRPP approach.

Fourth, it has not escaped our notice that OLS transformation of GLS models (Judge et al. 1985; Rencher 2000), combined with the refined RRPP approach we propose here, is not restricted to phylogenetic comparative analyses. Rather, it has clear application to a much wider array of statistical problems commonly found in ecology and evolutionary biology where model residuals are not independent, but display covariation that may characterized by an expected covariance matrix (i.e., a GLS model). For instance, analyses within species can make use of the migration matrix (**M**), which describes the lack of independence among populations resulting from gene flow (Felsenstein 2002; Stone 2011). Conditioning the data on **M** via phylogenetic transformation, and using the refined RRPP procedure above facilitates statistical comparisons of phenotypes among groups of populations while accounting for their lack of independence due to migration. In like manner, one may condition the data on a spatial covariance matrix (**S**: Cressie 2015) and use the refined RRPP procedure here to statistically evaluate biological trends across objects whose expected covariation is proportional to their spatial proximity. We contend that OLS transformation, combined with refined RRPP, provides a general statistical solution to many biological problems that can be described using GLS models.

Finally, our investigation reveals a deeper connection between phylogenetic ANOVA and other phylogenetic comparative methods that deserves comment. Recently, it has been shown that trait-dependent diversification methods such as BiSSE (Maddison et al. 2007) can display unacceptably high type I error rates under certain conditions (Rabosky and Goldberg 2015). One reason for this pattern is the fact that in some circumstances, statistical associations are inferred from data where the discrete character of interest displays only one or a few evolutionary changes across the phylogeny (Maddison and FitzJohn 2015). Thus, singular, and often unreplicated events drive the patterns of association, and can lead to support for an incorrect model hypothesis (Uyeda et al. 2018). Considering phylogenetic ANOVA in this light, we note the near perfect correspondence between the problems that challenge BiSSE methods and those that affect inferences based on phylogenetic ANOVA. That is, when groups are aggregated across the phylogeny, there are only one or very few evolutionary changes in group "state," which subsequently leads to challenges in statistical and evolutionary inference regarding ANOVA. This observation should give the evolutionary biologist pause, because in such instances it is difficult to unravel the patterns of trait covariation we wish to investigate from the influence of singular evolutionary events. As such, we echo the recommendations of Uyeda et al. (2018) that a greater emphasis on understanding the origination of singular evolutionary events, and the downstream consequences of those events on patterns of trait evolution, should play a more prominent role in future macroevolutionary studies.

AUTHOR CONTRIBUTIONS

Both authors conceived of the project, performed statistical simulations and empirical analyses, and wrote the manuscript.

ACKNOWLEDGMENTS

This work was sponsored in part by the U.S. National Science Foundation grants DEB-1556379 (to DCA) and DEB-1737895 (to MLC). We thank J. Clavel for discussions that inspired us to pursue this line of research, and E. Baken, B. Juarez, A. Kaliontzopoulou, and two anonymous reviewers for comments on the manuscript.

DATA ARCHIVING

Data archived in DRYAD https://doi.org/10.5061/dryad.2s8d0f9.

LITERATURE CITED

Abdi, H., and L. J. Williams. 2013. Partial least squares methods: partial least squares correlation and partial least square regression.

Pp. 549–579 *in* B. Reisfeld and A. Mayeno, eds. Methods in molecular biology: computational toxicology. Springer Verlag, New York.

Adams, D. C. 2007. Organization of *Plethodon salamander* communities: guild-based community assembly. Ecology 88:1292–1299.

———. 2010. Parallel evolution of character displacement driven by competitive selection in terrestrial salamanders. BMC Evol. Biol. 10: 1–10.

———. 2013. Comparing evolutionary rates for different phenotypic traits on a phylogeny using likelihood. Syst. Biol. 62:181–192.

———. 2014a. A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. Evolution 68:2675–2688.

———. 2014b. Quantifying and comparing phylogenetic evolutionary rates for shape and other high-dimensional phenotypic data. Syst. Biol. 63:166–177.

———. 2014c. A generalized K statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. Syst. Biol. 63:685–697.

Adams, D. C., and J. O. Church. 2011. The evolution of large-scale body size clines in Plethodon: evidence of heat-balance or species-specific artifact? Ecography 34:1067–1075.

Adams, D. C., and M. L. Collyer. 2015. Permutation tests for phylogenetic comparative analyses of high-dimensional shape data: what you shuffle matters. Evolution 69:823–829.

———. 2018. Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. Syst. Biol. 67:14–31.

Adams, D. C., and R. Felice. 2014. Assessing phylogenetic morphological integration and trait covariation in morphometric data using evolutionary covariance matrices. PLoS ONE 9:e94335.

Adams, D. C., and F. J. Rohlf. 2000. Ecological character displacement in Plethodon: biomechanical differences found from a geometric morphometric study. Proc. Natl. Acad. Sci. U.S.A. 97:4106–4111.

Adams, D. C., M. L. Collyer, and A. Kaliontzopoulou. 2018. Geomorph: Software for geometric morphometric analysis. R package version 3.0.6. Available at http://CRAN.R-project.org/package=geomorph.

Alvarado-Cárdenas, L. O., E. Martínez-Meyer, T. P. Feria, L. E. Eguiarte, H. M. Hernández, and G. Midgley. 2013. To converge or not to converge in environmental space: testing for similar environments between analogous succulent plants of North America and Africa. Ann. Bot. 111:1125–1138.

Anthony, C. D., J. A. Wicknick, and R. G. Jaeger. 1997. Social interactions in two sympatric salamanders: effectiveness of a highly aggressive strategy. Behaviour 134:71–88.

Arbour, J. H., and H. López-Fernández. 2017. Continental cichlid radiations: functional diversity reveals the role of changing ecological opportunity in the Neotropics. Proc. R. Soc. B 283:20160556.

Baker, J., A. Meade, M. Pagel, and C. Venditti. 2015. Adaptive evolution towards larger size in mammals. Proc. Natl. Acad. Sci. USA 112:5093–5098.

Bartoszek, K., J. Pienaar, P. Mostad, S. Andersson, and T. F. Hansen. 2012. A phylogenetic comparative method for studying multivariate adaptation. J. Theor. Biol. 314:204–215.

Beaulieu, J. M., D.-C. Jhwueng, C. Boettiger, and B. C. O'Meara. 2012. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. Evolution 66:2369–2383.

Blankers, T., D. C. Adams, and J. J. Wiens. 2012. Ecological radiation with limited morphological diversification in salamanders. J. Evol. Biol. 25:634–646.

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

Blomberg, S. P., J. G. Lefevre, J. A. Wells, and M. Waterhouse. 2012. Independent contrasts and PGLS regression estimators are equivalent. Syst. Biol. 61:382–391.

Bonett, R. M., and A. L. Blair. 2017. Evidence for complex life cycle constraints on salamander body form diversification. Proc. Natl. Acad. Sci. USA 114:9936–9941.

Collyer, M. L., D. J. Sekora, and D. C. Adams. 2015. A method for analysis of phenotypic change for phenotypes described by high-dimensional data. Heredity 115:357–365.

Commenges, D. 2003. Transformations which preserve exchangeability and application to permutation tests. Journal of nonparametric statistics. 15:171–185.

Cressie, N. A. C. 2015. Statistics for spatial data: revised edition. Wiley Classics Publisher, New York.

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

———. 2002. Contrasts for a within-species comparative method. Pp. 118–129 *in* M. Slatkin, and M. Veuille, eds. Modern developments in theoretical population genetics. Oxford Univ. Press, Oxford, U.K.

Freckleton, R. P., N. Cooper, and W. Jetz. 2011. Comparative methods as a statistical fix: the dangers of ignoring an evolutionary model. Am. Nat. 178:E10–E17.

Friedman, S. T., S. A. Price, A. S. Hoey, and P. C. Wainwright. 2015. Ecomorphological convergence in plantktivorous surgeonfishes. J. Evol. Biol. 29:965–978.

Garland, T. Jr., and A. R. Ives. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. Am. Nat. 155:346–364.

Garland, T. Jr., P. H. Harvey, and A. R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. Syst. Biol. 41:18–32.

Garland, T. Jr., A. W. Dickerman, C. M. Janis, and J. A. Jones. 1993. Phylogenetic analysis of covariance by computer simulation. Syst. Biol. 43:265–292.

Goolsby, E. W. 2016. Likelihood-based parameter estimation for high-dimensional phylogenetic comparative models: overcoming the limitations of "distance-based" methods. Syst. Biol. 65:852–870.

Grafen, A. 1989. The phylogenetic regression. Phil. Trans. R. Soc. Lond. B 326:119–157.

Hairston, N. G. 1980. Evolution under interspecific competition: field experiments on terrestrial salamanders. Evolution 34:409–420.

Harmon, L. J., J. J. Kolbe, J. M. Cheverud, and J. B. Losos. 2005. Convergence and the multidimensional niche. Evolution 59:409–421.

Highton, R. 1995. Speciation in eastern North American salamanders of the genus *Plethodon*. Ann. Rev. Ecol. Syst. 26:579–600.

Jaeger, R. G. 1970. Potential extinction through competition between two species of terrestrial salamanders. Evolution 24:632–642.

Johnston, J., and J. DiNardo. 1997. Econometric methods. McGraw Hill Publishing, New York.

Jöreskog, K. G., and H. Wold. 1982. Systems under indirect observation—causality structure prediction. North Holland Publishing Company, Amsterdam, The Netherland.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lutkepohl, and T.-C. Lee. 1985. The theory and practice of econometrics. John Wiley & Sons, New York.

Kariya, T., and H. Kurata. 2004. Generalized least squares. John Wiley & Sons, New York.

Klingenberg, C. P., and N. A. Gidaszewski. 2010. Testing and quantifying phylogenetic signals and homoplasy in morphometric data. Syst. Biol. 59:245–261.

Klingenberg, C. P., and J. Marugán-Lobón. 2013. Evolutionary covariation in geometric morphometric data: analyzing integration, modularity, and allometry in a phylogenetic context. Syst. Biol. 62:591–610.

Losos, J. B. 1992. The evolution of convergent structure in Caribbean *Anolis* communities. Syst. Biol. 41:403–420.

Maddison, W. P., and R. G. FitzJohn. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. Syst. Biol. 64: 127–136.

Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a binary character's effect on speciation and extinction. Syst. Biol. 56: 701–710.

Martins, E. P., and T. F. Hansen. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149:646–667.

Mazel, F., T. J. Davies, D. Georges, S. Lavergne, W. Thuiller, and P. R. Peres-Neto. 2016. Improving phylogenetic regression under complex evolutionary models. Ecology 97:286–293.

Moen, D. S., H. Morlon, and J. J. Wiens. 2016. Testing convergence versus history: convergence dominates phenotypic evolution for over 150 million years in frogs. Syst. Biol. 65:146–160.

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

Pennell, M. W., and L. J. Harmon. 2013. An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. Ann. NY Acad. Sci. 1289:90–105.

Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E. Alfaro, and L. J. Harmon. 2014. Geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics 30:2216–2218.

R Core Team. 2017. R: a language and environment for statistical computing. Version 3.4.1. R Foundation for Statistical Computing, Vienna. Available at http://cran.R-project.org.

Rabosky, D. L., and E. E. Goldberg. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Syst. Biol. 64:340–355.

Rencher, A. C. 2000. Linear models in statistics. John Wiley & Sons, New York.

Revell, L. J. 2010. Phylogenetic signal and linear regression on species data. Methods Ecol. Evol. 1:319–329.

Revell, L. J., and L. J. Harmon. 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. Evol. Ecol. Res. 10:311–331.

Reynolds, R. G., D. C. Collar, S. A. Pasachnik, M. L. Niemiller, A. R. Puente-Rolón, and L. J. Revell. 2016. Ecological specialization and morphological diversification in Greater Antillean boas. Evolution 70:1882–1895.

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

Rohlf, F. J., and M. Corti. 2000. The use of partial least-squares to study covariation in shape. Syst. Biol. 49:740–753.

Serb, J. M., E. E. Sherratt, A. Alejandrino, and D. C. Adams. 2017. Phylogenetic convergence and multiple shell shape optima for gliding scallops (Bivalvia: Pectinidae). J. Evol. Biol. 30:1736–1747.

Sherratt, E., A. Alejandrino, A. C. Kraemer, J. M. Serb, and D. C. Adams. 2016. Trends in the sand: directional evolution in the shell shape of recessing scallops (Bivalvia: Pectinidae). Evolution 70:2061–2073.

Stone, E. A. 2011. Why the phylogenetic regression appears robust to tree misspecification. Syst. Biol. 60:245–260.

Uyeda, J. C., D. S. Caetano, and M. W. Pennell. 2015. Comparative analysis of principal components can be misleading. Syst. Biol. 64:677–689.

Uyeda, J. C., R. Zenil-Ferguson, and M. W. Pennell. 2018. Rethinking phylogenetic comparative methods. Syst. Biol. https://doi.org/10.1101/222729.

Wilke, T., M. Benke, M. Brändle, C. Albrecht, and J. M. Bichain. 2010. The neglected side of the coin: non-adaptive radiations in spring snails (*Bythinella* spp.). Pp. 551–578 *in* M. Glaubrecht, ed. Evolution in action. Springer, Berlin.

Zelditch, M. L., J. Li, L. A. P. Tran, and D. L. Swiderski. 2015. Relationships of diversity, disparity, and their evolutionary rates in squirrels (Sciuridae). Evolution 69:1284–1300.

Associate Editor: G. Slater
Handling Editor: M. Noor

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Results from computer simulations comparing parameter estimates for phylogenetic regression for existing procedures and RRPP.

Figure S2. Results from computer simulations evaluating parameter bias in scenarios with (A) random (B) and aggregated groups on the phylogeny.

Figure S3. Results from computer simulations evaluating the variance in sampling distributions of βest. for random and aggregated groups on the phylogeny.

Figure S4. Empirical sampling distributions (black) and theoretical *F* distribution (red) for alternating and aggregated groups (*X*) for differing numbers of taxa ($N = 32, 64, 128$).

Figure S5. Empirical sampling distributions (black) and theoretical *F* distribution (red) for alternating and aggregated groups (*X*) for differing numbers of taxa ($N = 32, 64, 128$).

Figure S6. Empirical sampling distributions (black) and theoretical *F* distribution (red) for alternating and aggregated groups (*X*) for differing numbers of taxa ($N = 32, 64, 128$).

Figure S7. Correlation between significance levels from phylogenetic ANOVA for simulated datasets versus the same datasets rotated to their principal axes for: (a) random groups and (b) aggregated groups.

Figure S8. Power curves for RRPP for differing numbers of response (**Y**) variables ($p = 1, 5, 10$) for differing numbers of taxa ($N = 32, 64, 128$).