Phylogenetic Comparative Methods for Evaluating the Evolutionary History of Function-Valued Traits

Eric W. Goolsby

Interdisciplinary Toxicology Program, Department of Plant Biology, University of Georgia, Athens GA, 30602, USA
Correspondence to be sent to: Department of Plant Biology, 2502 Miller Plant Sciences, University of Georgia, Athens, GA 30602-7271, USA
E-mail: eric.goolsby.evolution@gmail.com

Abstract.—Phylogenetic comparative methods offer a suite of tools for studying trait evolution. However, most models inherently assume fixed trait values within species. Although some methods can incorporate error around species means, few are capable of accounting for variation driven by environmental or temporal gradients, such as trait responses to abiotic stress or ontogenetic trajectories. Such traits, often referred to as function-valued or infinite-dimensional, are typically expressed as reaction norms, dose–response curves, or time plots and are described by mathematical functions linking independent predictor variables to the trait of interest. Here, I introduce a method for extending ancestral state reconstruction to incorporate function-valued traits in a phylogenetic generalized least squares (PGLS) framework, as well as extensions of this method for testing phylogenetic signal, performing phylogenetic analysis of variance (ANOVA), and testing for correlated trait evolution using recently proposed multivariate PGLS methods. Statistical power of function-valued comparative methods is compared to univariate approaches using data simulations, and the assumptions and challenges of each are discussed in detail. [ancestral state reconstruction; dose–response curves; evolution of tolerance; function-valued traits; infinite-dimensional; phenotypic plasticity; PGLS; phylogenetic comparative method; phylogenetic signal; reaction norms.]

Phylogenetic comparative methods such as ancestral state reconstruction, phylogenetically independent contrasts (PICs), and phylogenetic generalized least squares (PGLS) comprise a suite of tools for inferring the evolutionary history of traits and for testing hypotheses of trait evolution (Felsenstein 1985; Grafen 1989; Martins and Hansen 1997). Although intraspecific differences due to measurement error, sampling error, or natural variation can be incorporated into error structures, these models generally operate under the assumption that individual species possess a fixed mean for any given trait (Martins and Hansen 1997; Ives et al. 2007). Many traits relevant to ecology and evolutionary biology are inherently function-valued, including developmentally varying traits, niche preference, thermal performance, life history patterns, plant ecophysiological strategies, responses to environmental hazards such as heavy metals and anthropogenic contaminants, and tolerance to abiotic stressors such as salinity, drought, and nutrient limitation. Rather than a single mean value, function-valued traits are best represented by curves defined by mathematical functions. Accordingly, univariate analytical approaches to such traits, which fail to account for trait dependence on independent exogenous variables such as time or environment, are statistically inappropriate (Kingsolver et al. 2001; Rocha and Klaczko 2012; Stinchcombe et al. 2012; Donovan et al. 2014).

Until recently, standard phylogenetic comparative methods have lacked tools for direct incorporation of function-valued analyses. Instead, many researchers have approached evolutionary questions regarding function-valued traits by summarizing trait curves with univariate mean values and applying these values directly to comparative methods. This approach typically utilizes one of two metrics: mean trait values

evaluated at specific environmental levels (Pollard et al. 2001; Pigliucci et al. 2003), or inverse functions evaluated at a standardized biological response (Guenard et al. 2011; Strachan et al. 2011; Larras et al. 2014). The inverse of a function $f(x)=y$ is the function $f^{-1}(y)=x$ that gives the value of x that is expected to produce a response y. For example, in toxicological studies, statistical responses to chemicals are typically expressed as inverse functions rather than response at a specific dose (the 50% lethal response dose [LD50] of a chemical is the dose expected to cause mortality in 50% of exposed subjects). For linear functions with similar slopes, both approaches typically yield consistent results; for nonlinear functions, approaches using mean trait values across treatment levels are unreliable. Conversely, inverse evaluation of monotonic functions at analogous reference trait values across species (e.g., LD50) is a valid approach regardless of linearity or nonlinearity. This is because inverse function evaluation at a fixed biological endpoint returns the value of the independent variable that produces said response, and is thus not subject to weighting by curve thresholds or other nonlinear features. However, summary values such as LD50 only capture a single dimension of function-valued traits and are likely to mask important trait variation across species. Consider two aquatic species exhibiting identical LD50 values for salinity toxicity, but the slope of the dose–response curve for one species is much steeper than the dose–response curve of the other species. In this example, interpretation of the LD50s in isolation might lead to the incorrect conclusion that salinity tolerance in the two species is identical. Similarly, other univariate summaries of function-valued traits, such as area under the curve, fail to distinguish biological differences such as generalist–specialist shifts in which the shape of the curve changes

but the total area under the curve remains constant (Izem and Kingsolver 2005; Stinchcombe et al. 2012). For these reasons, evolution of function-valued traits should be studied using methods that incorporate entire species trait curves rather than univariate summaries.

One particularly important challenge related to function-valued traits is the study of plasticity, which is the focus of the next section. However, it should be noted that the principles of function-valued traits discussed in the next section, in addition to trait plasticity, apply broadly to other classes of function-valued traits as well (e.g., environmental tolerance, ontogenetic variation, etc.).

Plastic Traits: Motivation for Study, Common Approaches, and Challenges

A central question implicit to the study of phenotypic plasticity is how plasticity affects evolutionary processes, such as gene flow, speciation, diversification, and persistence in novel environments (Wund 2012). Questions concerning the evolutionary dynamics of phenotypic plasticity, plasticity costs, and tradeoffs are also of particular interest (Callahan et al. 2008). Such questions may be applied to exploring evolvability and rates of evolutionary change of both plastic and fixed traits to better understand or anticipate species responses to processes such as environmental pollution (Guenard et al. 2011; Guenard et al. 2013; Larras et al. 2014), climate change (Bradshaw and Holzapfel 2006), and species invasions (Richards et al. 2006). By inferring the evolutionary history of trait responses to environmental conditions, evolutionary hypotheses relating to phenotypic plasticity, such as adaptive plasticity, plasticity costs, and tradeoffs in plastically varying traits, may be tested. Additionally, by taking into account phylogenetic variation in reaction norms, certain patterns previously obscured or falsely detected, either due to sampling traits in too few or too narrowly varying environments, assuming linearity of nonlinear reaction norms, or failing to consider interspecific differences in phenotypic minimum and maximum values, may be clarified.

Traits are often sampled under natural field conditions and assumed to be fixed (nonplastic), despite clear evidence in many traits of substantial variation driven by environment and development (Whitman and Agrawal 2009; Mason et al. 2013; Donovan et al. 2014). By failing to account for environmental or ontogenetic variation, or by assuming that environmental differences can be treated as covariates that uniformly affect all species of interest, statistical and biological inferences may be unreliable (Fig. 1a) (Rocha and Klaczko 2012). Additionally, environmental variation may obscure evolutionary signatures, leading to the incorrect conclusion that a trait is too labile for phylogenetic analysis (Revell 2010).

As an alternative to field measurements, studies may be performed in homogenous laboratory settings to minimize plastic intraspecific variation, with the intention of "removing" the effect of environment and allowing "baseline" phenotypic expression. However, uniform environments fail to capture the full extent of trait combinations possible within a species; instead, only a snapshot of possible trait combinations is identified, which is potentially problematic when making evolutionary inferences on putatively adaptively plastic traits. Another strategy involves estimating the difference in field trait values with controlled common environment trait values as a measure of total trait plasticity (Knight and Ackerly 2003). Although an improvement, this method has several limitations: (i) the ability to isolate specific drivers of trait plasticity is limited; (ii) providing a single ideal environment for multiple species native to diverse habitats may not be possible, hence some species might be stressed under common environment conditions; and (iii) it is unlikely that the full range of trait plasticity is captured by sampling traits under field and single-treatment laboratory conditions (Fig. 1b) (Jacobs and Latimer 2012).

Researchers can capture a more comprehensive representation of the capacity for plasticity by exposing species to a controlled common environment while manipulating a single environmental variable. Such environmental manipulations often involve treating continuous variables as arbitrarily categorized treatments (e.g., "low" and "high"), which inherently impose the assumption of linearity and normality of data. Accordingly, problems are likely to arise if the slope, strength, or shape of reaction norms or plastically induced shifts in correlated traits differs across species (Fig. 1c) (Rocha and Klaczko 2012). Furthermore, categorical environmental manipulations pose several analytical challenges. In particular, trait measurements must be made under comparable treatment levels, and sampling differences and missing data are incompatible with many statistical methods. Additionally, the magnitude of arbitrarily categorized or binned continuous treatment values is ignored, thus sacrificing statistical power and limiting the ability to predict trait values at unobserved levels (Griswold et al. 2008; Stinchcombe et al. 2012).

Evaluating traits at multiple levels along a continuous environmental gradient offers an opportunity to more fully assess the function-valued relationship between environment and plastically varying traits (Fig. 1d) (Kingsolver et al. 2001; Rocha and Klaczko 2012; Stinchcombe et al. 2012; Murren et al. 2014). In general, function-valued analyses offer several advantages to other approaches: (i) they are robust to sampling differences; (ii) there is no need for measurements to be conducted at identical treatment levels (provided measurements provide sufficient coverage of the range of interest to accurately estimate function parameters); (iii) they have higher statistical power; (iv) they incorporate the magnitude and trend from continuous predictor variables; and (v) they allow prediction of trait responses in extant and ancestral taxa at unobserved levels (Kingsolver et al. 2001).
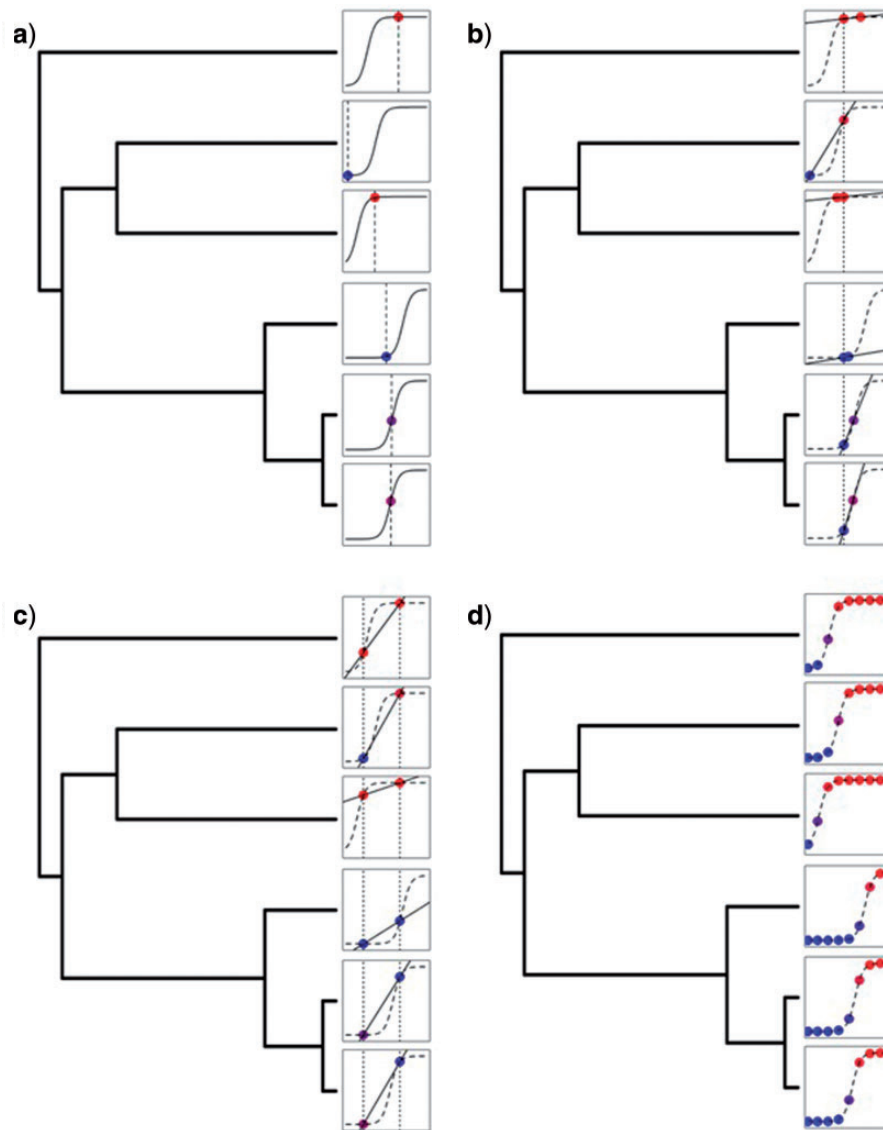
FIGURE 1.   Hypothetical result of performing phylogenetic comparative analysis on plastic traits from species traits (a) measured in nature; (b) measured in the field and in a controlled common environment; (c) measured in a controlled common environment at two different treatment levels; and (d) measured in a controlled common environment at multiple treatment levels. The *x*-axis represents an environmental gradient; the *y*-axis represents the trait value. In (a), the vertical dotted line represents the value of the environmental variable for each species collected in nature; in (b), the vertical dotted line represents the value of the environmental variable for the controlled common environment, and the second point represents the trait value measured in nature; and in (c), the vertical lines represent the two treatment levels of the environmental variable in the controlled common environment. Solid lines in (b and c) represent inferred reaction norms, several of which deviate considerably from true reaction norms. Depending on specific assumptions regarding the method used, contrasting conclusions regarding trait phylogenetic signal, model of evolution, and evolutionary history may be drawn.

A Novel Approach for Function-Valued Phylogenetic Comparative Methods

Although several studies have implemented function-valued analyses for studying and predicting the evolution of function-valued traits within a single species (Stinchcombe et al. 2012), methods for analyzing function-valued traits in an explicitly phylogenetic context have only recently begun to emerge. Recently, a method called phylogenetic Gaussian process regression (PGPR), an extremely flexible method for reconstructing the evolutionary history of function-valued traits, was proposed (Aston et al. 2012; Hadjipantelis et al. 2013; Jones and Moriarty 2013). In PGPR, curves of any shape may be analyzed without *a priori* assumptions of function structure, multiple sources of uncertainty may be incorporated, and a variety of evolutionary models may be tested (Hadjipantelis et al. 2013). Despite its flexibility, PGPR currently lacks compatibility with several commonly used comparative methods such as phylogenetic regression and estimation of phylogenetic signal, as such methods inherently rely on PGLS-based

reconstruction of ancestral states (Martins and Hansen 1997; Blomberg et al. 2003).

Here, I present a function-valued extension of ancestral state reconstruction in a PGLS framework. This method inherits the flexibility of PGLS and is fully compatible with several recently proposed multivariate PGLS-based methods that can be used to test for phylogenetic signal (Adams 2014a), perform phylogenetic ANOVA (Adams 2014b), evaluate correlated trait evolution between functions and univariate traits (Adams 2014b) or between multiple functions (Adams and Felice 2014), and test for shifts in evolutionary rates of function-valued traits (Adams 2014c).

METHODS

Conceptually, a function-valued trait is composed of infinite discrete landmarks along a curve that evolve as a single multivariate trait (Adams 2014a). Representation of a function for a given species can be approximated with a finite-length sequence of landmarks described by x, y coordinates. However, before performing PGLS-based comparative analyses, functions must first be aligned to make landmarks analogous across species (Adams 2014a). Depending on specific function-valued attributes, landmark alignment may be accomplished using a variety of techniques, such as inverse function alignment, dynamic time warping (Myers et al. 1980; Giorgino 2009), and generalized time warping (Zhou and De la Torre 2012), as described in the following two sections.

Ancestral Curve Reconstruction of Sigmoidal Curves

Monotonic functions with constant minimum and maximum values, such as sigmoidal functions (e.g., as with binomial, proportional, or probabilistic data), are a special case of function-valued traits in which the y-axis represents absolute endpoints that are comparable across species. In such cases, the y-axis serves as a pre-aligned point of reference, and function-valued evolution of these types of curves need only be expressed along the x-axis using inverse functions evaluated at a constant vector of y-values spanning minimum and maximum function values (e.g., between 0 and 1) for each species. In this way, the trait being studied is the vector of x-values that corresponds to a fixed vector of reference y-values for a given species. For example, in dose–response curves, the LD50 is a comparable landmark across species that can be appropriately analyzed using conventional univariate phylogenetic comparative methods. Conversely, the response at a specific dose is not comparable across species due to the nonlinearity of sigmoidal curves, and inferences based on such approaches are incorrect.

Univariate ancestral state reconstruction can be easily extended to incorporate function-valued traits using a PGLS framework (Martins and Hansen 1997).

First, parameters for function-valued traits $f(x)$ are estimated for each of N species. The reconstructed vector of function-valued landmarks at the root of the tree, $\boldsymbol{\mu}_{\text{root}}$, is then calculated using the generalized least squares formula

$$\boldsymbol{\mu}_{\text{root}} = (\mathbf{1}^t \mathbf{C}^{-1} \mathbf{1})^{-1} \mathbf{1}^t C^{-1} \mathbf{Y} \qquad (1)$$

where $\mathbf{C}$ is the matrix of expected trait variance–covariance given by the phylogenetic tree and assuming a specific model of trait evolution (e.g., Brownian motion), $\mathbf{1}$ is an $N \times 1$ matrix of ones, y is a vector of n evenly spaced y-axis landmarks between a and b (0 and 1 for binomial data), and $\mathbf{Y}$ is the $N \times n$ matrix of inverse functions f_{tips}^{-1} evaluated at y. Root vector $\boldsymbol{\mu}_{\text{root}}$ is extended into an $N \times n$ matrix $\mathbf{R}$, and each row is filled with $\boldsymbol{\mu}_{\text{root}}$. Next, the matrix of M internal node landmark vectors is calculated

$$\boldsymbol{\mu}_{\text{nodes}} = (\mathbf{D}\mathbf{C}^{-1}(\mathbf{Y} - \mathbf{R}))^t + \mathbf{R}^t_{n \times M} \qquad (2)$$

where $\mathbf{D}$ is the $M \times N$ matrix of expected covariance between internal nodes and tips as specified by the phylogenetic tree and model of evolution, $\mathbf{R}^t_{n \times M}$ is an $n \times M$ matrix consisting of M columns each filled with $\boldsymbol{\mu}_{\text{root}}$. $\boldsymbol{\mu}_{\text{nodes}}$ represents $f_{\text{nodes}}^{-1}(y)$ in the set of landmark coordinates ($f_{\text{nodes}}^{-1}(y, y)$), which can then be used to estimate parameters for curves at each node. A hypothetical example of ancestral curve reconstruction of a sigmoidal trait is visualized in Fig. 2.

Variance and 95% confidence intervals may also be estimated for reconstructed ancestral curves (Rohlf 2001). First, the Brownian motion parameter σ^2 is estimated for each landmark i from 1 to n

$$\sigma_i^2 = \frac{((\mathbf{Y}_i - \boldsymbol{\mu}_{\text{root}_i})^t \mathbf{C}^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_{\text{root}_i}))^t}{N - 1} \qquad (3)$$

Next, variance is calculated for each reconstructed landmark at each internal node

$$\sigma_{i_{\text{nodes}}}^2 = \sigma_i^2 \text{diag}(\mathbf{D}^* - (\mathbf{D}\mathbf{C}^{-1}\mathbf{D}^t)$$
$$+ (\mathbf{1}^* - \mathbf{D}\mathbf{C}^{-1}\mathbf{1})(\mathbf{1}^t\mathbf{C}^{-1}\mathbf{1})^{-1}(\mathbf{1}^* - \mathbf{D}\mathbf{C}^{-1}\mathbf{1})) \quad (4)$$

where $\mathbf{D}^*$ is the $M \times M$ matrix of expected covariance among internal nodes due to phylogeny and the specified evolutionary model, and $\mathbf{1}^*$ is an M-length vector of ones (Cressie 1993). Finally, 95% confidence intervals for each curve can be calculated as $\boldsymbol{\mu}_{\text{nodes}} \pm 1.96(\sigma_{\text{nodes}})$.

Ancestral Curve Reconstruction of Other Function Types

Many function-valued traits cannot be appropriately expressed with y-bounded sigmoidal functions. In such cases, inverse function alignment does not properly align curves across species. For instance, unimodal and multimodal functions lack a unique inverse solution for any given y-value. Additionally, certain function-valued traits require complex curve specification such as with Gaussian process regression. For such traits,
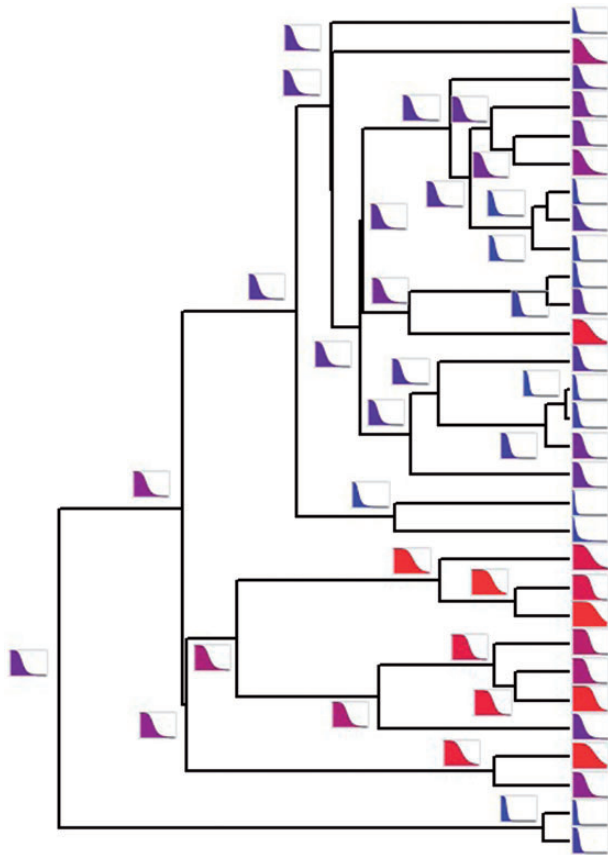
FIGURE 2. Ancestral curve reconstruction of a simulated data set for a proportion-based trait response (y-axis) in response to an environmental gradient (x-axis).

ancestral curve reconstruction must be performed on both the x- and y-axes, both of which must first be aligned across species. This can be accomplished for most function-valued relationships using a nonlinear alignment method called generalized time warping (Zhou and De la Torre 2012), a multisequence extension of dynamic time warping (Myers et al. 1980; Giorgino 2009). Although applicable to a wide variety of complex curve shapes, the appropriateness of time warping methods to specific function types should be assessed, and individual time warping parameters may require specification. For a more in-depth discussion, see Giorgino (2009).

For ancestral state reconstruction of aligned curves, Equations (1) and (2) must be applied individually to both the x-axis and the y-axis. For x-axis ancestral state reconstruction, Y is the $N \times n$ matrix of aligned x-values; for y-axis ancestral state reconstruction, Y is the $N \times n$ matrix of aligned y-values. The resulting set of μ_{nodes} forms coordinates $(x_{\text{nodes}}, y_{\text{nodes}})$ which are then used to estimate parameters for f_{nodes}. Similarly, variance and 95% confidence intervals for reconstructed curves must be applied on both the x and y axes for curves aligned with time warping. For monotonic sigmoidal function-valued traits, inferences from inverse function approaches and time warping alignment are equivalent

(R code is provided as Supplementary Material available on Dryad at http://dx.doi.org/10.5061/dryad.5nd50).

Compatibility with Distance-Based Multivariate Phylogenetic Comparative Methods

Recently, several multivariate extensions of phylogenetic comparative methods have been proposed for high-dimensional traits (Adams 2014a, 2014b, 2014c; Adams and Felice 2014). These methods allow quantification of metrics, such as multivariate phylogenetic signal, correlated trait evolution, phylogenetic ANOVA, and evolutionary rates. Distance-based comparative methods rely on removal of phylogenetic covariance from data by projecting phenotypic data onto the phylogenetic matrix **E** (Garland and Ives 2000):

$$\mathbf{E} = (\mathbf{U}\mathbf{W}^{1/2}\mathbf{U}^t)^{-1} \tag{5}$$

where **U** and **W** are eigenvectors and eigenvalues of the phylogenetic covariance matrix **C**. Although primarily presented as tools for analyzing Procrustes-aligned morphometric data (Rohlf and Slice 1990), these methods are not limited to the number of supplied trait dimensions and can handle hundreds of nonisotropic covarying dimensions. In particular, unlike covariance-based approaches, distance-based multivariate comparative methods maintain statistical power as the number of trait dimensions increases, and uniquely allow the number of dimensions to far exceed the number of species in the phylogeny. Thus, the evolution of function-valued traits can be appropriately analyzed using distance-based multivariate comparative methods (Adams 2014a).

As with Procrustes-aligned landmark coordinates (Adams 2014a, 2014b), function-valued traits are supplied to multivariate methods in the form of coordinate sequences from generalized time warping alignment (for curves of any form), or in the case of inverse function evaluations (for sigmoidal curves), only $f_{\text{nodes}}^{-1}(y)$ is supplied. Estimation of the phylogenetic mean is inherent to distance-based multivariate phylogenetic methods. Importantly, the estimated phylogenetic mean for distance-based methods is equivalent to the root estimate μ_{root} obtained from Equation (1) when applied to properly aligned landmarks of function-valued data. Specific applications and methodological details of applying distance-based multivariate comparative methods to function-valued traits are discussed in the following sections.

Phylogenetic Signal of Function-Valued Traits

Phylogenetic signal is a measure of the extent to which species exhibit phenotypic similarity due to phylogenetic relatedness. Various methods of signal quantification have been proposed, including Blomberg's K (Blomberg et al. 2003), Pagel's λ (Pagel 1999), autocorrelation

techniques (Pavoine et al. 2008), and correlation-based methods (Zheng et al. 2009). Blomberg's K is based on the ratio of observed variation relative to the variation expected under Brownian motion:

$$K = \frac{(\mathbf{Y} - E(\mathbf{Y}))^t(\mathbf{Y} - E(\mathbf{Y}))}{(\mathbf{Y} - E(\mathbf{Y}))^t\mathbf{C}^{-1}(\mathbf{Y} - E(\mathbf{Y}))} \Big/ \frac{\text{tr}(\mathbf{C}) - N(\mathbf{1}^t\mathbf{C}^{-1}\mathbf{1})^{-1}}{N - 1}$$

(6)

where $\mathbf{Y}$ is an $N \times 1$ vector of phenotypic values for a univariate trait and $E(\mathbf{Y})$ is the phylogenetic mean (estimated phenotypic value at the root of the phylogeny). Thus, the expectation of K under Brownian motion is 1. Phylogenetic permutation can be used to statistically test if K represents significant phylogenetic signal (Blomberg et al. 2003).

Adams (2014a) proposed a multivariate generalization of Blomberg's K (K_{mult}) which allows phylogenetic signal to be estimated in high-dimensional traits. K_{mult} utilizes a distance-based approach which produces identical estimates of K when applied to univariate traits but also allows for calculations of K_{mult} for multivariate traits. K_{mult} is calculated as follows:

$$K_{\text{mult}} = \frac{\mathbf{D}_{Y,\hat{a}}^t\mathbf{D}_{Y,\hat{a}}}{\mathbf{PD}_{U,0}^t\mathbf{PD}_{U,0}} \Big/ \frac{\text{tr}(\mathbf{C}) - N(\mathbf{1}^t\mathbf{C}^{-1}\mathbf{1})^{-1}}{N - 1}$$

(7)

where $\mathbf{D}_{Y,\hat{a}}$ is an $N \times 1$ vector of Euclidian distances between species means and the root of the phylogeny, as expressed by

$$\mathbf{D}_{i,\hat{a}} = \sqrt{(\mathbf{Y}_i - E(\mathbf{Y}))(\mathbf{Y}_i - E(\mathbf{Y}))^t}$$

(8)

and $\mathbf{PD}_{U,0}$ is an $N \times 1$ vector of Euclidean distances of species means to the origin projected onto the phylogenetic transformation matrix $\mathbf{E}$ from Equation (5). Current software implementations of K_{mult} in the R (R Core Team 2014) library geomorph 2.0 (physignal function) (Adams and Otárola-Castillo 2013; Adams et al. 2014) can readily incorporate the PGLS-based function-valued approaches presented here. For function-valued trait analysis utilizing inverse functions for alignment, $\mathbf{Y}$ is the $N \times n$ matrix of inverse functions f_{tips}^{-1} evaluated at y, and $E(\mathbf{Y})$ calculated by the physignal function is equivalent to the $1 \times n$ vector $\mathbf{\mu}_{\text{root}}$ obtained from Equation (1). For function-valued landmarks aligned using generalized time warping, x, y coordinates are analyzed as an $N \times 2n$ matrix of function-valued landmark coordinates. Again, the resulting $E(\mathbf{Y})$ is equivalent to results obtained from $\mathbf{\mu}_{\text{root}}$ in Equation (1).

Correlated Evolution Between Function-Valued Traits and Categorical (ANOVA) or Continuous (PGLS) Univariate Traits

Distance-based comparative approaches can be implemented to apply phylogenetic ANOVA or assess for correlated evolution between univariate and function-valued traits using a multivariate extension of PGLS called D-PGLS (Adams 2014b). In D-PGLS, independent variable $\mathbf{X}$ (a column of ones followed by one or more columns of values for independent variables) and dependent variable $\mathbf{Y}$ (an $N \times n$ or $N \times 2n$ matrix, depending on whether inverse functions or generalized time warping alignment is used) are both projected onto $\mathbf{E}$ to remove phylogenetic covariance from trait data, resulting in $\mathbf{X}_{\text{phy}}$ and $\mathbf{Y}_{\text{phy}}$. $\mathbf{Y}_{\text{phy}}$ is regressed on $\mathbf{X}_{\text{phy}}$ to obtain predicted values $\hat{\mathbf{Y}}_X$, and $\mathbf{Y}_{\text{phy}}$ is regressed on $\mathbf{1}_{\text{phy}}$ to obtain predicted values for $\hat{\mathbf{Y}}_1$. Next, the trace of the outer product of predicted values is used to calculate variation explained by the model:

$$SS_X = \text{tr}((\hat{\mathbf{Y}}_X - \hat{\mathbf{Y}}_1)(\hat{\mathbf{Y}}_X - \hat{\mathbf{Y}}_1)^t)$$

(9)

SS_X and residuals of ($\mathbf{Y}_{\text{phy}} \sim \mathbf{X}_{\text{phy}}$) are then used to calculate F-ratios and R^2, and phylogenetic permutation is used to determine the significance of the correlation. D-PGLS is implemented in the procD.pgls function in the R package geomorph (Adams and Otárola-Castillo 2013; Adams et al. 2014, 2014b).

Additional Distance-Based Multivariate Phylogenetic Comparative Methods

Although not discussed in detail here, similar methods are available to address questions such as correlated evolution between two separate function-valued traits (or between a function-valued trait and some other multivariate trait) using phylogenetic partial least squares (Adams and Felice 2014), as implemented in the phylo.pls function in geomorph (Adams and Otárola-Castillo 2013; Adams et al. 2014). Additionally, multivariate rates of evolution may be quantified, and evolutionary rates of function-valued traits for subgroups within a phylogeny may be statistically compared (Adams 2014c), as implemented in the compare.evol.rates function in geomorph (Adams and Otárola-Castillo 2013; Adams et al. 2014). For example, one could test the hypothesis that a function-valued plastic response to a particular environmental variable evolves more rapidly in a clade exposed to high levels of the environmental variable than a nonexposed clade. As distance-based multivariate methods continue to be developed, additional phylogenetic comparative methods, such as testing alternative models of evolution (e.g., Ornstein–Uhlenbeck) are expected to be available for further study of function-valued trait evolution (Adams 2014b).

Statistical Performance

Function-valued trait evolution was simulated under various conditions to evaluate the performance of function-valued ancestral curve reconstruction and to assess the statistical power of K_{mult} and D-PGLS when applied to function-valued traits. All curve simulations were based on evolution of a generalized linear model

using a logit link function:

$$f(x|\theta) = \frac{\exp(\theta_1 + \theta_2)}{1 + \exp(\theta_1 + \theta_2)} \qquad (10)$$

Curve evolution was simulated by evolving a root curve defined by two independently simulated function features, the median response level ($-\theta_1/\theta_2$) and the slope at the median response level ($\theta_2/4$), using the fastBM function in the R library phytools (Revell 2012). Internal node values were preserved and "true" regression parameters for each node were subsequently back calculated. To simulate correlated evolution between curves and univariate traits, sim.corrs was used to generate correlated evolution of a univariate trait with the curve median response level. Root curve parameters were set to $\theta_1 = -10.0$ and $\theta_2 = 2.0$, and to maintain positively sloped curves in the first quadrant (as with a typical dose–response curve), median response slopes of evolved curves were bounded between 0.2 and 1.0, and median response levels were bounded between 2.0 and 20.0.

For ancestral curve reconstruction, 1000 simulations were performed on randomly generated 128-taxa trees as described above. Ancestral curve reconstruction was performed on inverse tip functions (Equations (1) and (2)) and regression parameters were estimated for each node. For comparison with non-function-valued approaches (in which a single arbitrary x-value is chosen for evaluation of $f(x|\theta)$ and the function-valued relationship is ignored), univariate ancestral state reconstruction was conducted individually at 100 evenly spaced tip evaluations of $f(x|\theta)$ where x ranges from 0.0 to 15.0 (spanning the entire sigmoid portion of curves for all species). Coefficients of the known root state and the root state estimated from ancestral curve reconstruction were evaluated at $f(x|\theta)$ at the same 100 points for which univariate ancestral state reconstruction was conducted. The difference between the true value of $f(x|\theta_{\text{root}})$ and values obtained from both ancestral curve reconstruction and individual univariate ancestral state reconstruction were averaged across 1000 simulations and compared. PGPR ancestral curve reconstruction was also conducted for these simulated data sets using the R implementation provided in the supplement of Hadjipantelis (2013).

All simulations related to K_{mult} and D-PGLS were carried out on randomly generated pure-birth trees with 128 taxa, and all simulations were conducted 1000 times with randomly generated data. For K_{mult} and D-PGLS hypothesis testing, 999 iterations of phylogenetic permutation were conducted. Varying numbers of trait dimensions ($p = 2$, 5, 10, 25, and 50) were simulated to estimate the optimal number of dimensions for representing logit-type functions. Dimensions were input as sequences evaluated at $f^{-1}(y|\theta)$ where y is a sequence of evenly divided numbers from 0.001 to 0.999 of length p. Type I error and power were quantified for K_{mult} using the proportion of significant K_{mult} values for data simulated under λ tree transformations

(0, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, and 1.0). For D-PGLS Type I error and power analysis, functions and a continuous univariate trait were simulated with covariances of $\sigma = 0.0$, 0.1, 0.3, 0.5, 0.7, and 0.9 and the proportion of significant correlations was recorded. For comparison, all multivariate simulations were performed alongside analogous univariate approaches (Blomberg's K and PGLS) which were performed at 100 individual evaluations of $f(x|\theta)$ where x ranges from 0.0 to 15.0.

R code for all simulations is provided as Supplementary Material available on Dryad at http://dx.doi.org/10.5061/dryad.5nd50.

RESULTS AND DISCUSSION

Performance of Ancestral Curve Reconstruction

Root estimations of ancestral curve reconstruction were compared to simulated logit curve evolution using the inverse function approach described in previous sections. Root coefficients were estimated from reconstructed inverse functions (Equation (1)), and both the known and estimated functions were evaluated along 100 evenly spaced x-values ranging from 0.0 to 15.0. $f(x|\theta_{\text{estimated}})$ and $f(x|\theta_{\text{true}})$ were compared at each point. The average predicted curve for each method is plotted in Fig. 3a. The same approach was applied to reconstructed curves obtained from PGPR. For comparison with univariate methods in which the function-valued nature of the trait is ignored, ancestral state reconstruction of each observed tip y-value at each x-value is also plotted. Confidence intervals (95%) were calculated for PGLS (Fig. 3b), PGPR (Fig. 3c), and univariate (Fig. 3d) ancestral reconstruction methods using a single representative simulation (Rohlf 2001; Hadjipantelis et al. 2013). On average, ancestral curve reconstruction using PGLS- and PGPR-based methods yielded similar results (Fig. 3a). The small discrepancies that are present are likely attributable to differences such as the underlying assumptions of the Brownian motion simulations performed here and the assumption of an Ornstein–Uhlenbeck model of evolution inherent to PGPR. Although univariate ancestral state reconstruction visually appears to give similar results to both PGLS and PGPR ancestral curve reconstruction, in practice, the entire curve would not be reconstructed with this approach; rather, univariate ancestral state reconstruction is typically performed in isolation at a single x-level, and no knowledge of surrounding parts of the curve would be represented. Therefore, only instantaneous deviation from the true y-value at an individual x-value should be considered for univariate ancestral state reconstruction (i.e., not the sum of the resulting curve as a whole).

Even for x-values where the estimated univariate response is identical to the true response, a function-valued trait cannot be properly represented by a single value. Such attempts are likely to produce misleading
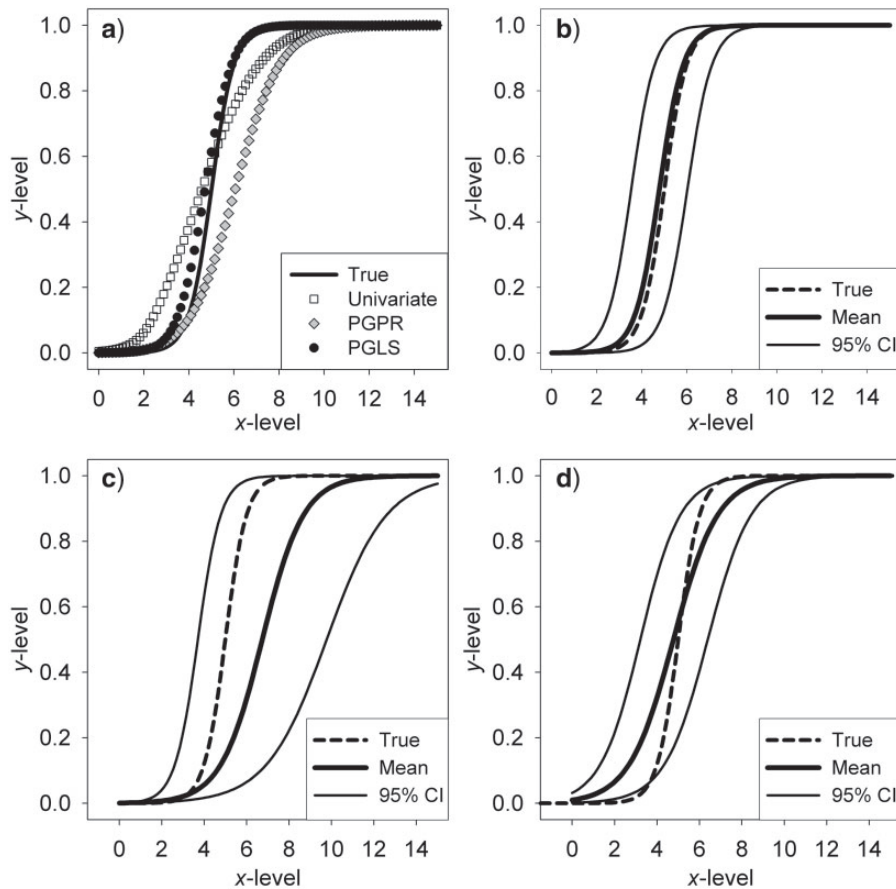
FIGURE 3.   Average result of ancestral root curve reconstruction (a) from 1000 simulated data sets ($N = 128$) for PGLS-based ancestral curve reconstruction (circles), PGPR (diamonds), and univariate ancestral state reconstruction (squares). 95% confidence intervals of root reconstructions are presented from a single representative simulation for (b) PGLS, (c) PGPR, and (d) univariate curve reconstruction. All three methods produced resulting curves similar to the actual root curve (solid line). However, univariate methods are only evaluated at a single point (not in the context of an entire curve), hence only deviation from the true curve at individual x-levels should be considered for the univariate approach. Univariate representations of function-valued traits are inappropriate, regardless of the instantaneous deviation from the actual curve; and such approaches are likely to result in incorrect interpretations. Therefore, despite greater absolute deviation from the root at various x-levels than univariate approaches, the overall performance of PGPR was comparable to that of PGLS, with both methods providing root curve estimates close to actual root curve.

inferences and potentially contradictory conclusions based on the arbitrary x-value choice at which the trait is evaluated. These issues are further highlighted in simulations of phylogenetic signal and correlated trait evolution.

In contrast to univariate ancestral state reconstruction, PGPR ancestral curve reconstruction *should* be taken in context of the entire reconstructed curve. Accordingly, PGPR ancestral curve reconstruction provided similar results to PGLS-based curve reconstruction overall (Fig. 3a). Therefore, both methods provide vastly superior representations of ancestral function-valued traits than univariate attempts at ancestral state reconstruction.

Statistical Power of K_{mult} and D-PGLS

Type I error and statistical power of K_{mult} and D-PGLS for function-valued traits were assessed using the methods described above. Type 1 error rates for all simulations of K_{mult} and D-PGLS, where $p \geq 5$ (as determined by data simulations on star trees and input covariance of zero, respectively) were ~ 0.05. Similarly, Type I error rates were ~ 0.05 for univariate Blomberg's K and PGLS evaluated at arbitrary x-levels. As expected, statistical power increased for K_{mult} and D-PGLS as the number of dimensions increased (Adams 2014a, 2014b) (Fig. 4). Dimensionality beyond $p = 25$ resulted in diminishing gains in statistical power, coupled with a substantially increased computational burden. K_{mult} power scaled approximately linearly with λ transformations up to 1.0, whereas D-PGLS power reached 1.0 with input covariation of 0.5 and higher (Fig. 5). For univariate methods, statistical power varied widely depending on the x-level at which the trait was assessed, although statistical power approached that of K_{mult} and D-PGLS when x-level was near the median response level of the root function and phylogenetic signal or input covariation was high. It should be
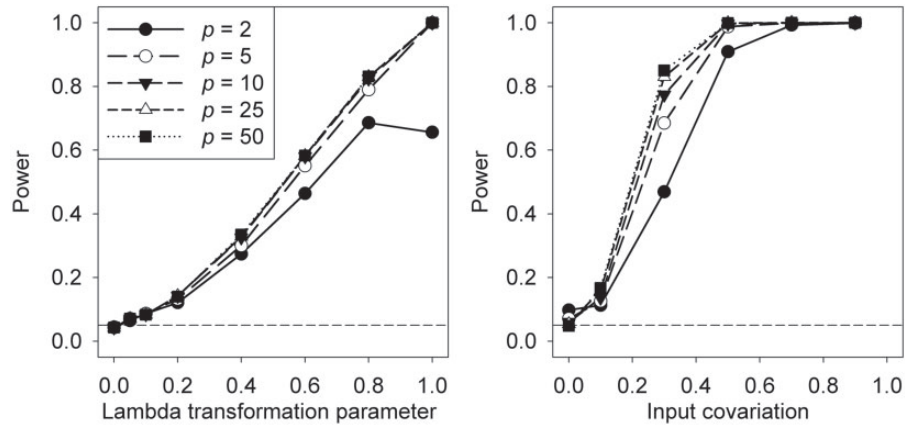
FIGURE 4. Results from K_{mult} (left) and D-PGLS (right) type I error and power from 1000 data simulations ($N=128$) with increasing trait dimensionality ($P=2, 5, 10, 25, 50$).
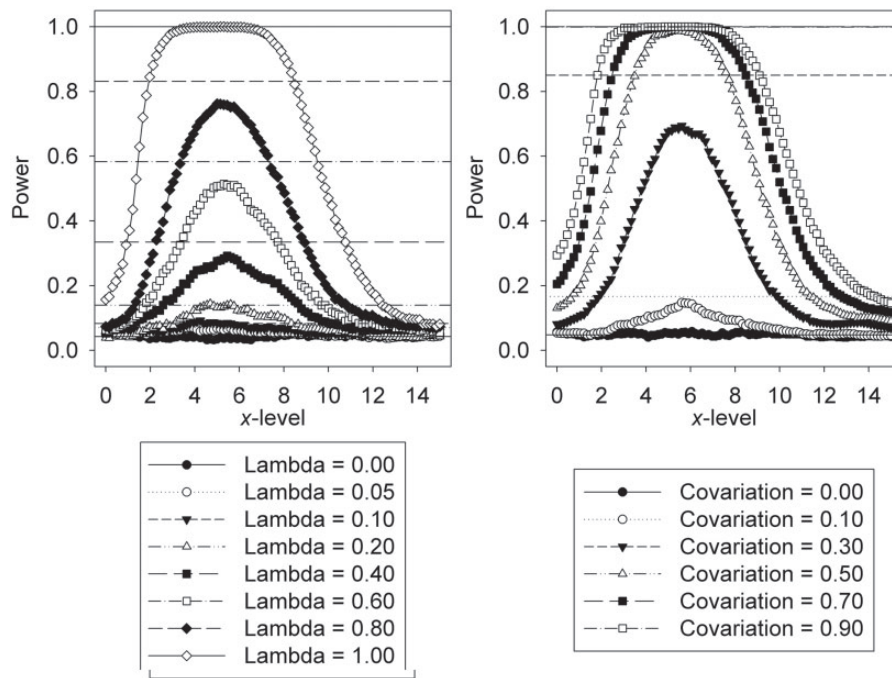

FIGURE 5. Statistical power of univariate Blomberg's K (left) and PGLS (right) at individual x-levels ranging from 0.00 to 15.00 ($N=128$). K_{mult} and D-PGLS power results ($P=50$) are plotted as horizontal lines for comparison. Statistical power for univariate approaches to function-valued traits varies widely for these tests.

noted that these results do *not* suggest that univariate approximations of phylogenetic signal or correlated trait evolution can be applied to function-valued traits. Indeed, phylogenetic comparative experiments evaluating species traits at a single level are blinded to true function-valued relationships, so the potential for erroneously drawn conclusions is likely to be overlooked, and such attempts provide no means for validation of results.

Overall, Type I error rates and statistical power of K_{mult} and D-PGLS suggest that distance-based multivariate phylogenetic comparative methods can be appropriately applied to the study of function-valued trait evolution,

and further highlight that univariate approaches to function-valued traits are inappropriate.

Limitations

Function-valued phylogenetic comparative methods offer promising improvements over conventional methods for many types of analyses. However, several assumptions and limitations must be accounted for to properly take advantage of these methods. First, a substantial amount of data is required for every species at multiple levels to accurately infer function parameters (Stinchcombe et al. 2012). Second, even

with large amounts of data, function parameters are estimated with error, but the methods described here assume species curves are estimated without error. Third, a thorough understanding of function-valued relationships is necessary to appropriately align function-valued landmarks and to express function-valued trait evolution appropriately. Finally, no methods for testing alternative models (e.g., Ornstein–Uhlenbeck) of high-dimensional multivariate trait evolution are currently available. Further development of distance-based multivariate phylogenetic comparative methods offers promising expansions of the possibilities of approaching function-valued trait evolution.

Conclusions

Phylogenetic comparative methods have transformed the field of comparative biology. While extensive work has contributed to the integration of evolution with inherently function-valued applications in developmental biology, toxicology, gene expression, and trait plasticity (particularly the prediction of species short-term evolutionary trajectories in response to selection gradients), few methods exist to unite function-valued approaches with phylogenetic comparative methods. Building on the methods presented here, future work should seek to address issues such as testing alternative evolutionary models, developing methods for dealing with sparse data, incorporating parameter estimation error into models, accounting for phylogenetic uncertainty, and applying the methods described here in alternative statistical frameworks, such as Bayesian approaches.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.5nd50.

References

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.N. 2014. Assessing trait covariation and morphological integration on phylogenies using evolutionary covariance matrices. PLoS ONE 9:e94335.

Adams D.C., Otárola-Castillo E., Sherratt E. 2014. geomorph: software for geometric morphometric analyses. CRAN: The Comprehensive R Archive Network, Vienna, Austria. http://cran.r-project.org/package=geomorph (last accessed December 17, 2014).

Aston J.A.D, Buck D., Coleman J., Cotter C.J., Jones N.S., Macaulay V., MacLeod N., Moriarty J.M., Nevins A. 2012. Functional Phylogenies Group. Phylogenetic inference for function-valued traits: speech sound evolution. Trends Ecol. Evol. 27:160–166.

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.

Bradshaw W.E., Holzapfel C.M. 2006. Climate change – evolutionary response to rapid climate change. Science 312:1477–1478.

Callahan H.S., Maughan H., Steiner U.K. 2008. Phenotypic plasticity, costs of phenotypes, and costs of plasticity toward an integrative view. Ann. N. Y. Acad. Sci. 1133:44–66.

Cressie N.A.C. 1993. *Statistics for Spatial Data*. New rev. edn. New York: Wiley-Interscience.

Donovan L.A., Mason C.M., Bowsher A.W., Goolsby E.W., Ishibashi C.D.A. 2014. Ecological and evolutionary lability of plant traits affecting carbon and nutrient cycling. J. Ecol. 102:302–314.

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

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.

Garland T., Midford P.E., Ives A.R. 1999. An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral values. Am. Zool. 39:374–388.

Giorgino T. 2009. Computing and visualizing dynamic time warping alignments in R: the dtw package. J. Stat. Softw. 31:1–24.

Grafen A. 1989. The phylogenetic regression. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 326:119–157.

Griswold C.K., Gomulkiewicz R., Heckman N. 2008. Hypothesis testing in comparative and experimental studies of function-valued traits. Evolution 62:1229–1242.

Guenard G., Legendre P., Peres-Neto P. 2013. Phylogenetic eigenvector maps: a framework to model and predict species traits. Methods Ecol. Evol. 4:1120–1131.

Guenard G., von der Ohe P.C., de Zwart D., Legendre P., Lek S. 2011. Using phylogenetic information to predict species tolerances to toxic chemicals. Ecol. Appl. 21:3178–3190.

Hadjipantelis P.Z., Jones N.S., Moriarty J., Springate D.A., Knight C.G. 2013. Function-valued traits in evolution. J. R. Soc. Interface 10:20121032; doi:10.1098/rsif.2012.1032.

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.

Izem R., Kingsolver J.G. 2005. Variation in continuous reaction norms: quantifying directions of biological interest. Am. Nat. 166:277–289.

Jacobs B.S., Latimer A.M. 2012. Analyzing reaction norm variation in the field vs. greenhouse: comparing studies of plasticity and its adaptive value in two species of Erodium. Perspect. Plant. Ecol. 14:325–334.

Jones N.S., Moriarty J. 2013. Evolutionary inference for function-valued traits: Gaussian process regression on phylogenies. J. R. Soc. Interface 10:20120616; doi:10.1098/rsif.2012.0616.

Kingsolver J.G., Gomulkiewicz R., Carter P.A. 2001. Variation, selection and evolution of function-valued traits. Genetica 112:87–104.

Knight C.A., Ackerly D.D. 2003. Evolution and plasticity of photosynthetic thermal tolerance, specific leaf area and leaf size: congeneric species from desert and coastal environments. New Phytol. 160:337–347.

Larras F., Keck F., Montuelle B., Rimet F., Bouchez A. 2014. Linking diatom sensitivity to herbicides to phylogeny: a step forward for biomonitoring? Environ. Sci. Technol. 48:1921–1930.

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.

Mason C.M., McGaughey S.E., Donovan L.A. 2013. Ontogeny strongly and differentially alters leaf economic and other key traits in three diverse Helianthus species. J. Exp. Bot. 64:4089–4099.

Murren C.J., Maclean H.J., Diamond S.E., Steiner U.K., Heskel M.A., Handelsman C.A., Ghalambor C.K., Auld J.R., Callahan H.S., Pfennig D.W., Relyea R.A., Schlichting C.D., Kingsolver J. 2014. Evolutionary change in continuous reaction norms. Am. Nat. 183:453–467.

Myers C., Rabiner L.R., Rosenberg A.E. 1980. Performancy tradeoffs in dynamic time warping algorithms for isolated word recognition. IEEE Trans. Acoust. Speech Signal Process. 28:623–635.

Pagel M.D. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–844.

Pavoine S., Ollier S., Pontier D., Chessel D. 2008. Testing for phylogenetic signal in phenotypic traits: new matrices of phylogenetic proximities. Theor. Pop. Biol. 73:79–91.

Pigliucci M., Pollard H., Cruzan M.B. 2003. Comparative studies of evolutionary responses to light environments in Arabidopsis. Am. Nat. 161:68–82.

Pollard H., Cruzan M., Pigliucci M. 2001. Comparative studies of reaction norms in Arabidopsis. I. Evolution of response to daylength. Evol. Ecol. Res. 3:129–155.

R Core Team 2014. R: a language and environment for statistical computing. Vienna (Austria) R Foundation for Statistical Computing. Available from: URL http://www.R-project.org/ (last accessed December 17, 2014).

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.

Richards C.L., Bossdorf O., Muth N.Z., Gurevitch J., Pigliucci M. 2006. Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. Ecol. Lett. 9:981–993.

Rocha F.B., Klaczko L.B. 2012. Connecting the dots of nonlinear reaction norms unravels the threads of genotype-environment interaction in Drosophila. Evolution 66:3404–3416.

Rohlf F.J., Slice D.E. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Zool. 39: 40–59.

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

Stinchcombe J.R., Kirkpatrick M. 2012. Function-valued traits working group. Genetics and evolution of function-valued traits: understanding environmentally responsive phenotypes. Trends Ecol. Evol. 27:637–647.

Strachan L.A., Tarnowski-Garner H.E., Marshall K.E., Sinclair B.J. 2011. The evolution of cold tolerance in drosophila larvae. Physiol. Biochem. Zool. 84:43–53.

Whitman D.W., Agrawal A.A. 2009. What is phenotypic plasticity and why is it important? In: Whitman D.W., Ananthakrishnan T.N., editors. Phenoytpic plasticity of insects. Enfield (NH): Science Publishers. p. 1–63.

Wund M.A. 2012. Assessing the impacts of phenotypic plasticity on evolution. Integr. Comp. Biol. 52:5–15.

Zheng L., Ives A.R., Garland T., Larget B.R., Yu Y., Cao K. 2009. New multivariate tests for phylogenetic signal and trait correlations applied to ecophsyiological phenotypes of nine Manglietia species. Funct. Ecol. 23:1059–1069.

Zhou, F., De la Torre, F. 2012. Generalized time warping for multi-modal alignment of human motion. IEEE Comput Soc Conf Comput Vis Pattern Recogn. 1282–1289.