11-2019

Phylogenetic Comparative Methods and the Evolution of Multivariate Phenotypes

Dean C. Adams
Iowa State University, dcadams@iastate.edu

Michael L. Collyer
Chatham University

Phylogenetic Comparative Methods and the Evolution of Multivariate Phenotypes

Abstract
Evolutionary biology is multivariate, and advances in phylogenetic comparative methods for multivariate phenotypes have surged to accommodate this fact. Evolutionary trends in multivariate phenotypes are derived from distances and directions between species in a multivariate phenotype space. For these patterns to be interpretable, phenotypes should be characterized by traits in commensurate units and scale. Visualizing such trends, as is achieved with phylomorphospaces, should continue to play a prominent role in macroevolutionary analyses. Evaluating phylogenetic generalized least squares (PGLS) models (e.g., phylogenetic analysis of variance and regression) is valuable, but using parametric procedures is limited to only a few phenotypic variables. In contrast, nonparametric, permutation-based PGLS methods provide a flexible alternative and are thus preferred for high-dimensional multivariate phenotypes. Permutation-based methods for evaluating covariation within multivariate phenotypes are also well established and can test evolutionary trends in phenotypic integration. However, comparing evolutionary rates and modes in multivariate phenotypes remains an important area of future development.

Disciplines
Ecology and Evolutionary Biology | Genetics | Statistical Methodology

PHYLOGENETIC COMPARATIVE METHODS AND THE EVOLUTION OF MULTIVARIATE PHENOTYPES

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

[1]*Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames IA, Department of Statistics, Iowa State University, Ames IA, USA. email: dcadams@iastate.edu*

[2]*Department of Science, Chatham University, Pittsburgh, PA, USA. email: M.Collyer@Chatham.edu*

Running Title: Evolution in multivariate phenotypes

Corresponding author contact information: Dean C. Adams; Dept. Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, 50011. dcadams@iastate.edu

Keywords: multivariate, phylogenetic comparative methods, macroevolution, shape analysis, high-dimensional data

Abstract

Evolutionary biology is multivariate, and advances in phylogenetic comparative methods for multivariate phenotypes have surged to accommodate this fact. Evolutionary trends in multivariate phenotypes are derived from distances and directions between species in a multivariate phenotype space. For these patterns to be interpretable, phenotypes should be characterized by traits in commensurate units and scale. Visualizing such trends, as is achieved with phylomorphospaces, should continue to play a prominent role in macroevolutionary analyses. Evaluating phylogenetic generalized least squares (PGLS) models (e.g., phylogenetic ANOVA and regression) is valuable, but using parametric procedures is limited to only few phenotypic variables. In contrast, non-parametric, permutation-based PGLS methods provide a flexible alternative, and are thus preferred for high-dimensional multivariate phenotypes. Permutation-based methods for evaluating covariation within multivariate phenotypes are also well-established, and can test evolutionary trends in phenotypic integration. However, comparing evolutionary rates and modes in multivariate phenotypes remains an important area of future development.

1. Introduction

Characterizing patterns of phenotypic diversity at macroevolutionary scales requires a phylogenetic perspective. It is widely recognized that shared evolutionary history leads to phenotypic similarity between closely-related species, and thus statistical summaries must account for evolutionary non-independence during the analysis (Felsenstein 1985, Harvey and Pagel 1991). The mathematical tools used to accomplish this task are known as phylogenetic comparative methods (Harmon 2018). Modern phylogenetic comparative biology began with the seminal work of Felsenstein (1985), whose phylogenetically independent contrasts revolutionized the way in which cross-species analyses are performed (for a recent review see Huey, et al. 2019). Subsequently, it was revealed that this approach is mathematically related to a broader class of statistical models (phylogenetic generalized least squares: Grafen 1989, Martins and Hansen 1997; see Blomberg, et al. 2012, Garland and Ives 2000, Rohlf 2001), thereby linking several approaches in one conceptual analytical framework. Thus, the current incarnation of the phylogenetic comparative toolkit was born.

Since these initial advances, there has been an explosion of analytical methods contributing to the phylogenetic comparative toolkit, enabling evolutionary biologists to quantify phenotypic trends that inform on a wide array of biological hypotheses. For instance, phylogenetic comparative methods may be utilized to evaluate trends of evolutionary covariation between traits (Felsenstein 1985, Garland, et al. 1993, Grafen 1989, Revell and Collar 2009), to quantify the degree of phylogenetic signal in phenotypes (Blomberg, et al. 2003, Pagel 1999), to compare rates of phenotypic evolution among clades or between traits (Adams 2013, Garland 1992, O'Meara, et al. 2006, Revell and Harmon 2008, Thomas, et al. 2006), and to evaluate the fit of differing models of trait evolution (Beaulieu, et al. 2012, Butler and King 2004, Hansen 1997). Unfortunately, while

such approaches yield considerable power for characterizing patterns of phenotypic diversity across the tree of life, the biological insights derived from them have been largely restricted to univariate traits, as most comparative methods were developed for only a single column of phenotypic variables (e.g., body size). This is regrettable, as evolutionary biology is inherently multivariate (Blows 2007, Collyer, et al. 2015), and processes such as natural selection can act on more than one trait simultaneously (Lande 1979, Lande and Arnold 1983). Furthermore, it has become common in evolutionary biology to characterize phenotypes using more than one trait (Harmon, et al. 2008, Losos 1992, Price, et al. 2010), or by using complex, multi-dimensional traits that require a vector of values to encode (Adams 2010, Kirkpatrick and Meyer 2004, McPeek, et al. 2008). Thus, the ability to evaluate multivariate phenotypic trends across the phylogeny has become a pressing need.

Theorists have recently endeavored to develop phylogenetic comparative methods capable of evaluating phylogenetic patterns in multivariate datasets (e.g., Adams 2014b, Adams 2014c, Adams and Collyer 2015, Bartoszek, et al. 2012, Bastide, et al. 2018, Goolsby 2015, Klingenberg and Marugán-Lobón 2013, Revell and Harmon 2008, among others). These methods are gaining prominence in the field, and are increasingly used to address evolutionary hypotheses in multivariate phenotypic datasets in a manner analogous to what has long been possible for univariate traits (e.g., Chira, et al. 2018, Felice and Goswami 2017, Grunstra, et al. 2018, Martinez, et al. 2018, Serb, et al. 2017, Zelditch, et al. 2015). In this review, we survey the recent advances for evaluating evolutionary trends in multivariate phenotypes, highlight some biological insights discovered through use of multivariate phylogenetic comparative approaches, and identify several areas for future analytical development. We describe the various types of datasets that biologists use to characterize multivariate phenotypes and relate these to the properties displayed by the

resulting multivariate dataspaces. We contend that visualizing patterns in multivariate phenotypic spaces plays an important role in macroevolutionary analyses, and argue for the importance of such methods in complementing quantitative macroevolutionary hypothesis testing. We then summarize multivariate phylogenetic hypothesis testing approaches, discuss their utility, and identify some current limitations for evaluating patterns in multivariate phenotypic datasets. Finally, we provide pertinent suggestions for empiricists to guide them in their analytical studies of multivariate phenotypes, as well as point to areas in need of future theoretical development.

2. Characterizing Multivariate Phenotypes

Before summarizing multivariate phylogenetic comparative methods, it is useful to review what is meant by a multivariate phenotype (note: evaluating evolutionary patterns in ecological data can also be performed multivariately [e.g., Pie, et al. 2017], but will not be discussed here). Typically, a multivariate phenotype is a set of continuously measured trait values, which may be correlated with one another (Collyer and Adams 2007, Collyer, et al. 2015, Huttegger and Mitteroecker 2011). Various datatypes are used to characterize multivariate phenotypes. For example, Catlett, et al. (2010) measured gestation length, age at weaning, and other variables to represent multivariate life history phenotypes in lemurs. Patterns of multivariate gene expression have also been used (Valenzuela 2010). Likewise, sets of performance measures, including out-lever to in-lever ratios (Carroll, et al. 2004), force and power estimates (Friedman, et al. 2016), and empirical measures of locomotor performance (Moen, et al. 2013), can represent multivariate phenotypes. Function-valued traits representing an ordered sequence of phenotypic values, such as data describing a growth curve, are also examples of multivariate phenotypes (Goolsby 2015, Kingsolver, et al. 2001). However, most frequently, multivariate phenotypes describe

morphological traits. Some common data types include sets of individual traits such as the lengths, ratios, and angles between structures (multivariate morphometrics: Blackith and Reyment 1971), sets of shape variables obtained from the coordinates of anatomical points (landmark-based geometric morphometrics: Adams, et al. 2013, Mitteroecker and Gunz 2009), or variables derived from anatomical curves or surfaces, often obtained from CT scans or other representations of anatomical objects (semilandmark methods: Gunz and Mitteroecker 2013, or spherical harmonics: McPeek, et al. 2008). **Figure 1** presents a visual summary of some common multivariate phenotypic data types.

2.1. Phenotypic Dataspaces: Mathematical Considerations

Mathematically, multivariate phenotypes are represented by a vector of trait values. The vectors for a set of species are then assembled into a $N \times p$ matrix (**Y**), whose rows contain the phenotypes of the N species. The columns of **Y** contain the trait values for each of the p trait dimensions, and correspond to the axes of a p-dimensional phenotypic dataspace. Therefore, each row of **Y** describes the location of the N^{th} species as a point in this multivariate dataspace (**Figure 1)**. The axes of the phenotypic dataspace are typically considered to be orthogonal; thereby assuming that the dataspace displays Euclidean geometry. This corresponds to the commonsense notion of dataspaces, where similar phenotypes are close together in the dataspace and dissimilar phenotypes are further apart. Assuming that Euclidean geometry is appropriate for the span of phenotypic values in the dataspace also implies that distances and directions between specimens confer biological meaning, and comparisons of such measures are interpretable (see discussion in Huttegger and Mitteroecker 2011, Mitteroecker and Huttegger 2009). However, it is important to recognize that not all multivariate phenotypic datasets display these crucial properties.

For phenotypic dataspaces to be Euclidean, their trait dimensions must minimally be in commensurate units and be of the similar scale. Otherwise, the mathematical definition of similarity and difference is not the concordant across the trait dimensions (see Legendre and Legendre 2012). For example, a phenotypic dataspace could theoretically be constructed from a combination of continuous measurements, count variables (e.g., the number of scales), and the presence or absence of particular structures. However, relationships among specimens in this space are uninterpretable, because the notion of distance differs between the axes representing continuous variables (Euclidean distance), counts (Gower's distance), and binary traits (Hamming distance). Importantly, this concern persists even for traits that are all continuously valued, if those traits are measured in different units (e.g., a dataset of dimensionless ratios, masses, and angular extents). The reason is that the deviations between specimens in each trait dimension are represented in differing units, and combining these deviations across traits to estimate distances between species, or covariances between traits, results in values that are uninterpretable. To construct a valid Euclidean space from such data first requires mathematical transformation or standardization of the trait dimensions, so the variables are expressed in similar units (see: Legendre and Legendre 2012 for a related discussion). In such cases however, some downstream phylogenetic comparative analyses may no longer be useful. For instance, estimates of disparity or evolutionary rates are not meaningful when performed on standard normal deviates, as standardizing data in this manner alters the original trait variances upon which disparity measures and evolutionary rates are based.

Fortunately, many phenotypic datasets are comprised of variables measured in similar units and scale, thereby preserving these important properties. Examples include sets of linear measurements quantified in similar units and expressed in the same scale (but see Huttegger and

Mitteroecker 2011), parameters summarizing function-valued traits or other curves, and shape variables from landmark-based geometric morphometric methods. In these cases, it is reasonable to assert Euclidean geometry, and empirical comparisons based on distances and directions in the multivariate phenotype space can be interpreted biologically. We recommend that empiricists carefully consider whether the phenotypic traits under investigation are of commensurate units and scale, so that downstream analyses of macroevolutionary patterns are interpretable.

Finally, multivariate phenotype spaces displaying Euclidean geometry also exhibit other useful properties, including rotation-invariance. This means that the dispersion of species in the dataspace, and statistical summaries based on them, remain unchanged when the dataspace is viewed from a different orientation (see Adams and Collyer 2018a). Rigid rotations of data spaces are an essential component of many ordination methods, such as principal components analysis. As described in the next section, ordination methods are important tools for visualizing patterns of phenotypic evolution in multivariate phenotypes. And as explained below, rotation-invariance is an essential property that multivariate phylogenetic comparative methods should retain.

3. Visualizing Evolutionary Patterns in Multivariate Phenotypes

Because multivariate phenotypes are represented by many trait dimensions, visualizing patterns of phenotypic dispersion is often challenging. One solution is to utilize ordination methods that provide a low-dimensional view of a high-dimensional dataspaces. Presently, two ordination methods incorporate phylogenetic information into their plots: phylomorphospaces and phylogenetic principal components analysis.

3.1. Phylomorphospace

Phylomorphospaces are ordination plots with the phylogeny superimposed (Klingenberg and Ekau 1996, Rohlf 2002). For multivariate phenotypes, they provide a low-dimensional view of the phenotypic variation among the extant species, while including hypothesized trait evolution along the branches of the phylogeny. To obtain a phylomorphospace, principal component (PC) axes are first obtained in the usual manner (i.e., from the $p \times p$ trait covariance matrix (**S**) calculated from **Y**). Next, PC scores are obtained (via matrix projection) for all species, as well as for estimated ancestral values at the nodes of the phylogeny. Scores on the first few PC axes are then plotted to provide a graphical visualization of phenotypic dispersion relative to the phylogeny. Because this procedure is based on a decomposition of the trait covariance matrix (**S**), the axes of the phylomorphospace are orthogonal (Polly, et al. 2013), and the method is thus a rigid rotation of the original phenotype space. Additionally, scores on the PC axes are uncorrelated with one another, and patterns of dispersion among species are retained. This means that the approach preserves the total variation in the dataset throughout the analysis, and that distances and directions among species may be interpreted biologically. Note, however, that the phylomorphospace plot is a projection of the full dataspace into a sub-space of fewer dimensions (typically two). Thus, exploring patterns in higher dimensions may yield additional revelations. Additionally, any downstream statistical analyses should be performed on the full set of trait dimensions to ensure that 100% of the phenotypic variation is included in the analysis. When subsets of trait dimensions are evaluated (e.g., the first few principal components), evolutionary inferences from them can be misleading (Adams and Collyer 2018a, Uyeda, et al. 2015; for a related discussion see Bookstein 2013).

Phylomorphospaces provide powerful tools for visualizing evolutionary patterns in multivariate phenotypes, which can lead to significant biological insights. For example, Aristide,

et al. (2018) discovered that New World monkeys diversified into distinct regions of morphospace with little phenotypic overlap among genera; a pattern interpreted as being consistent with an adaptive radiation. Likewise, Davis and Betancur-R (2017) found that herbivorous and carnivorous fish species occupied distinct regions of morphospace, and did so consistently across lineages, demonstrating strong phylogenetic convergence of ecotypes in the group. In other cases, phylomorphospaces have revealed that one lineage displays considerably greater phenotypic disparity as compared to another lineage (e.g., Sidlauskas 2008, Zelditch, et al. 2015), implying possible selective release, differing rates of morphological evolution between lineages, or that one lineage repeatedly evolves similar phenotypes throughout its evolutionary history. Finally, phylomorphospaces can be used to identify groups where phenotypes appear accentuated in a consistent manner over time (Sherratt, et al. 2016), providing evidence of a directional trend in multivariate phenotypic evolution. Several hypothetical examples of patterns commonly observed in phylomorphospaces are shown in **Figure 2.**

Importantly, phylomorphospaces provide a visual means of examining evolutionary patterns that may be evaluated quantitatively using statistical hypothesis testing approaches. For example, visual patterns of phenotypic distinctness among ecotypes may be formally evaluated using phylogenetic ANOVA (*sensu* Adams and Collyer 2018b), while patterns of recurrent phenotypic evolution may be confirmed via statistical tests of evolutionary convergence (see Stayton 2015). Likewise, apparent differences in phenotypic variance may be evaluated formally via disparity comparisons among clades (Serb, et al. 2017, Zelditch, et al. 2015), and by comparisons of rates of phenotypic evolution (Sherratt, et al. 2017). Finally, patterns revealed in phylomorphospaces may precipitate simulation-based approaches to explore more complex macroevolutionary hypotheses and evolutionary scenarios (e.g., Sherratt, et al. 2016, Sidlauskas

2008). We assert that phylomorphospaces are an important component of the multivariate phylogenetic comparative toolkit and recommend that empiricists employ them as a regular part of their statistical arsenal for evaluating trends in multivariate phenotypes.

3.2. Phylogenetic Principal Component Analysis

Phylogenetic principal component analysis (pPCA: Revell 2009) is another approach for obtaining ordination plots while accounting for phylogenetic non-independence. Statistically, the method conditions the ordination on the phylogeny via a decomposition of the $p \times p$ evolutionary rate matrix (**R**: Revell and Harmon 2008), rather than using the original trait covariance matrix (**S**). The difference between the two is that the rate matrix (**R**) is simply the trait covariance matrix (**S**) standardized by the phylogeny, and is therefore weighted inversely by the evolutionary relationships among taxa. PC scores are then obtained for the extant species as well as for the estimated ancestral taxa, and these scores are used to generate the ordination plot (for details see Revell 2009).

While phylogenetic PCA has the appeal of incorporating the phylogeny into its computations, the approach does have some unintuitive properties. For instance, the axes of phylogenetic PCA are orthogonal, yet the species' scores on those axes are correlated with one another (see also Polly, et al. 2013). And unlike standard PCA, the eigenvalues of the phylogenetically corrected PC axes do not sum to the total phenotypic variation in the dataset (though summing variation in the pPC scores does: Polly, et al. 2013). Additionally, the statistical rationale for incorporating the phylogeny into the computations is unclear, as principal components analysis does not assume independence among observations, in contrast to hypothesis testing approaches that often require this assumption (e.g., ordinary least squares ANOVA and regression

methods: see below). On the other hand, the axes of phylogenetic PCA have been rotated to account for the effects of phylogeny, and when all axes are used, the distances between species in the phylogenetically-rotated space is identical to that of the original multivariate dataspace. Overall we agree with Polly, et al. (2013) that results from phylogenetic PCA can be difficult to interpret, and instead recommend phylomorphospaces as a means of visualizing dispersion in multivariate phenotypes. Nonetheless, we acknowledge that obtaining a visualization of multivariate phenotype spaces that aligns with the phylogenetic relatedness among species is an important goal worth pursuing. We therefore recommend that future theoretical work should explore alternative algebraic formulations to produce ordinations that maximize the covariation between the phenotypic data and the phylogeny but do so while preserving the desirable properties displayed by classical ordination approaches.

4. Phylogenetic Signal in Multivariate Phenotypes

Phylogenetic signal is the tendency for related species to be more phenotypically similar than species selected at random from a phylogeny (Blomberg, et al. 2003, Munkemuller, et al. 2012). Phylogenetic signal is expected under many macroevolutionary scenarios of trait evolution and is therefore frequently examined in phylogenetic comparative studies. For multivariate phenotypes, several analytical approaches are available. One method (Pagel 1999) evaluates the fit of the data to the phylogeny while including a scaling parameter (λ) that describes the degree of phylogenetic signal, but this approach is limited to cases where the number of variables is less than the number of species (see Adams 2014a). Another approach quantifies the sum-of-squared changes in multivariate phenotypes across branches of the phylogeny (Klingenberg and Gidaszewski 2010). However, this method is sensitive to both the number of variables (p) and the

number of species (N), complicating comparisons across datasets (Adams 2014a). Additionally, because this approach is based on ancestral state estimation, it will provide inaccurate estimates of phylogenetic signal when directional phenotypic evolution has occurred, as ancestral states are not faithfully estimated under scenarios of directional evolution (Royer-Carenzi and Didier 2016). A third approach (Adams 2014a) is a multivariate generalization of the Kappa statistic (Blomberg, et al. 2003), which measures phylogenetic signal as a ratio of observed to expected phenotypic variation obtained with and without considering phylogenetic non-independence. This approach (K_{mult}) has a known and constant expected value, per variable, under Brownian motion (1.0), and holds considerable promise for characterizing the degree of phylogenetic signal in multivariate datasets.

To provide a sense of the degree to which multivariate phenotypes display phylogenetic signal, we surveyed the literature for empirical studies of phylogenetic signal in multivariate phenotypes; obtaining over 330 K_{mult} estimates from nearly 100 published studies. The vast majority of these datasets (80%) described morphological phenotypes, while the remainder represented multivariate life history traits, behavior, physiology, and other measures. Overall, the degree of phylogenetic signal was not significantly different among these datatypes (R^2 = 0.038; F = 1.867, P = 0.09), though behavioral data did exhibit lower values as compared to the other phenotypic datasets (K_{mult} = 0.47 versus K_{mult} = 0.65). K_{mult} ranged from 0.031 to 2.130 in this sample (**Figure 3**), with a mean phylogentic signal of 0.65. K_{mult} did not vary with the number of species in the phylogeny (R^2 = 0.002; F = 0.7557, P = 0.368), and we observed that most of the values were less than 1.0. Interestingly, all of these findings were generally consistent with patterns observed in an earlier survey of phylogenetic signal in univariate phenotypes (Blomberg, et al. 2003). This suggests that broad-scale patterns of phylogenetic signal may be concordant across

both single-valued and multivariate phenotypes. Additionally, approximately 75% of the K_{mult} values displayed significant phylogenetic signal, though there was no difference in the distributions of K_{mult} for significant and non-significant datasets (t = 0.5498, P = 0.5832; $D_{KS.test}$ = 0.0698, P = 0.9251). This paradoxical result makes sense in light of phylogenetic signal latency, which we discuss below.

One interesting observation in this sample was that most estimates of multivariate phylogenetic signal were considerably less than 1.0, indicating that there was less phylogenetic signal than expected under Brownian motion. Yet the majority of these datasets displayed statistically significant phylogenetic signal when compared to a random association of phenotypes to the tips of the phylogeny. Indeed, a similar pattern was also observed in univariate datasets (Blomberg, et al. 2003), where it was attributed to either selection and phenotypic adaptation (which would reduce variation across taxa) or to various sources of measurement error. While such explanations remain a possibility here, for multivariate phenotypes we suggest a third possibility. Specifically, when phylogenetic signal is concentrated in one or a few phenotypic dimensions (phylogenetic signal latency), it is possible to observe significant multivariate phylogenetic signal whose summary measure (K_{mult}) is less than 1.0. This possibility was explored briefly through a simulation study, in which a subset of trait dimensions were obtained via phylogenetic simulation under Brownian motion, while the remaining trait dimensions contained variation that was not phylogenetically associated. Results of these simulations (**Figure 3**) confirmed that as the percentage of trait dimensions with non-phylogenetically associated variation increases, K_{mult} decreases (but the ratio of p to N does not impact K_{mult}). Yet because there was still phylogenetic signal in some trait dimensions, many such datasets still display significant phylogenetic signal. Biologists who observe this pattern in their empirical datasets should consider whether

phylogenetic signal is concentrated in a subset of trait dimensions in their data, as a possible explanation of this pattern. Unfortunately, it is currently not straightforward how to discern between a weak signal for many variables and a strong signal for few variables for small values of K_{mult}, other than to consider the significance of the signal based on the *P*-value. We propose that future research should consider evaluations of the effect size of K_{mult}, calculated as a standard deviate from its sampling distribution (values produced from resampling permutations: see related statistics derived in Adams and Collyer 2016, Adams and Collyer 2018b, Collyer, et al. 2015), as a way to resolve this conundrum.

Finally, for both univariate and multivariate phenotypes, current statistical tests do not evaluate whether the observed pattern differs from what is expected under Brownian motion. Instead, they evaluate the observed phylogenetic signal relative to a sampling distribution obtained by permuting phenotypic values across the tips of the phylogeny (i.e., a random association of phenotypes with species: for conceptual motivation of the original test see: Blomberg, et al. 2003). To fill this void we propose an additional approach, where the observed K_{mult} is compared to a distribution of values obtained from data simulated on the phylogeny under Brownian motion. This simulated distribution will have an expected value of 1.0 (see Adams 2014a), and thus comparisons to this distribution will evaluate whether the observed phylogenetic signal differs from what is expected under Brownian motion. Future theoretical work should formally develop the test procedure proposed here and evaluate its statistical properties.

5. Multivariate Phenotypes and Patterns of Covariation

For many evolutionary hypotheses, understanding the degree to which phenotypic traits covary is of paramount importance. There are many biological reasons to expect that multivariate

phenotypes would display correlations. For instance, common selective pressures and other mechanisms can generate covariation between phenotypic traits within an organism (Klingenberg 2014). Likewise, selection and adaptation can generate correlations between phenotypes and other parameters such as diet, climate, the presence or absence of competing species, or other ecological variables (e.g., Baab, et al. 2014, Mahler, et al. 2010, Martin and Wainwright 2011). Several analytical methods have been developed for evaluating patterns of trait covariation in multivariate phenotypes. Determining which method should be utilized depends in part upon whether the evolutionary correlations of interest are between phenotypic traits, or whether they describe the evolutionary covariation between phenotypes and other variables. Below we highlight methods for evaluating both types of patterns, and identify several biological hypotheses that can be addressed with these approaches.

5.1 Phenotypic Integration: Evolutionary Correlations Within Phenotypes

Phenotypic integration describes a pattern where phenotypic traits are correlated with one another (Olson and Miller 1958). Such patterns are expected when selection acts upon multiple, functionally related traits (Arnold 2005), or when traits display genetic linkages, exhibit pleiotropy, or have shared developmental pathways (see Cheverud 1996, Mitteroecker and Bookstein 2007). Several analytical methods have been developed to evaluate evolutionary correlations among traits in a phylogenetic context. For example, Revell and Collar (2009) used a likelihood framework to evaluate changes in evolutionary correlations between traits across the phylogeny. For instance, this method revealed that suction-feeding eel species exhibited higher evolutionary correlations between anatomical units as compared to bite-feeding species (Collar, et al. 2014). Similarly, a Bayesian approach can be used to evaluate shifts in evolutionary correlations

between traits across the phylogeny (Caetano and Harmon 2019). Both of these methods have potential to yield insights that may inform on how evolutionary correlations evolve. Nonetheless, it should be recognized that they only consider pairwise correlations between individual traits, because both methods evaluate the elements of the evolutionary rate matrix (**R**), which contains pairwise evolutionary trait correlations. When broader patterns of evolutionary correlations across traits are of interest, other analytical approaches are required.

Sometimes it is of interest to determine whether *sets* of traits (or modules) display evolutionary correlations with one another. For example, one may wish to determine whether distinct anatomical modules, such as the skull and mandible, correlate across the phylogeny (e.g., Adams and Felice 2014, Figueirido, et al. 2010). To evaluate such patterns, a multivariate equivalent of evolutionary correlation is required. Here, phylogenetic partial least squares (Adams and Felice 2014, Klingenberg and Marugán-Lobón 2013) may be used to evaluate covariation in multivariate phenotypes. As with pairwise evolutionary correlations, the approach starts with the evolutionary rate matrix (**R**: see above). However, rather than evaluating each evolutionary correlation individually, the overall covariation between blocks of variables is quantified and is evaluated. In one example, high levels of evolutionary integration between the basicranium and facial regions of primates was identified; a pattern that has impacted the evolution of phenotypic disparity in this group (Neaux, et al. 2018). In another study, Evans, et al. (2017) found that phenotypic integration between the braincase and facial regions of some teleost fishes was greater than that exhibited in carnivore mammals, and that the lower levels of integration in carnivores allowed for greater opportunity to evolve phenotypic disparity. Similarly, phenotypic integration between claw and toepad traits in *Anolis* lizards was found to enhance microhabitat specialization among species (Yuan, et al. 2019). These, and other examples, are a testament to the power of

phylogenetic partial least squares for identifying evolutionary correlations between sets of variables.

Finally, it may be of interest to determine whether all traits in a multivariate dataset display phenotypic integration with one another. Unfortunately, methods for evaluating such patterns in a phylogenetic context are less well developed. For patterns across individuals within species, summary measures such as eigenvalue variance may be used to characterize patterns of global integration (e.g., Pavlicev, et al. 2009: for an alternative approach see Bookstein 2015). However, while the phylogenetic distribution of variation across principal component dimensions has been visually examined (e.g., Klingenberg and Marugán-Lobón 2013), to our knowledge neither a phylogenetic equivalent of the above summary measures, nor a statistical test of such measures, has been proposed. We recommend that future theoretical work investigate this possibility, and develop formal statistical tests of global integration in a phylogenetic context.

5.2. Phylogenetic Linear Models: ANOVA and Regression

Many evolutionary hypotheses strive to evaluate the relationship between multivariate phenotypes and one or more independent variables. Such hypotheses are best characterized by phylogenetic generalized least squares (PGLS) models (Grafen 1989, Martins and Hansen 1997). These models are defined mathematically as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E} \qquad 1$$

where $\mathbf{Y}$ is a $N \times p$ matrix of multivariate phenotypic trait values, $\mathbf{X}$ is a $N \times k$ design matrix containing one or more independent (predictor) variables, $\boldsymbol{\beta}$ is a $k \times p$ matrix containing the model

coefficients, and **E** is a $N \times p$ matrix of residuals (see Adams 2014b, Adams and Collyer 2015, Adams and Collyer 2018b, Clavel, et al. 2015). Unlike ordinary least squares models where the residual error (**E**) is assumed to be independent, the residuals of PGLS are not independent, but instead contain the expected covariation between species as described by the phylogenetic covariance matrix under a specified model of evolutionary change (typically Brownian motion: Rohlf 2001). Thus, the analysis is tantamount to a weighted least squares model, where the weights are the inverse of phylogenetic relatedness.

Implementing phylogenetic generalized least squares models using parametric statistical techniques based on maximum likelihood and other formulations has long been the favored approach for evaluating evolutionary trends in univariate data (e.g., Grafen 1989, Martins and Hansen 1997). However, multivariate phenotypes present numerous challenges to this paradigm that have only recently come to light (see Adams 2014b, Adams and Collyer 2018a). For instance, approaches that evaluate multivariate PGLS models using maximum likelihood, or through standard multivariate test measures (e.g., Wilks' Λ), display increasing type I error as phenotypes become more highly multivariate (Adams 2014b, Adams and Collyer 2018a). The reason is that standard parametric implementations require finding the determinant and the inverse of the $p \times p$ trait covariance matrix, which becomes more challenging as the number of traits (p) approaches the number of species (N), and is not possible when $p > N$ (see Adams 2014b, Adams and Collyer 2018a, Adams and Collyer 2018b). To circumvent these issues, pairwise composite likelihood measures combined with phylogenetic simulations were proposed (Goolsby 2016). However, this method is not rotation-invariant, and statistical conclusions from it differ for the same phenotypic dataset when viewed in different orientations (for additional issues see: Adams and Collyer 2018a). This observation emphasizes why properties such as rotation-invariance are important in

multivariate comparative analyses, and demonstrates that analytical methods for describing evolutionary patterns in multivariate phenotypes should retain this important property if they are to provide useful biological inferences.

An alternative implementation to multivariate PGLS utilizes phylogenetic transformation, and test statistics derived from traces of covariance matrices rather than determinants, making them more robust to the challenges described above (Adams 2014b, Adams and Collyer 2015, Adams and Collyer 2018b). Statistical evaluation of these measures is then accomplished using residual randomization permutation procedures (RRPP: Adams and Collyer 2018b, Collyer and Adams 2018), where residuals from a reduced model are permuted to generate empirical sampling distributions against which the observed test statistics are compared. This approach is rotation-invariant, yields identical model parameters to standard implementations, displays appropriate type I error and high statistical power, and sampling distributions generated from it align with statistical distributions derived from statistical theory (see Adams and Collyer 2018a, Adams and Collyer 2018b). Finally, the method can be used for various statistical designs, including phylogenetic ANOVA, phylogenetic regression, phylogenetic factorial models, and phylogenetic analyses of covariance.

An increasing number of empirical studies use permutation-based PGLS to evaluate patterns in multivariate phenotypes. For example, Paluh and Bauer (2018) examined the evolution of quadrate shape in geckos, revealing distinct allometric trajectories across genera, suggesting that disparate functional pressures resulted in shifts in the direction of evolutionary allometry among lineages. Likewise, patterns of jaw shape variation were found to covary with dietary preferences in both tree-dwelling and ground-dwelling squirrel species (Zelditch, et al. 2017). In fact, ecomorphological associations between microhabitat use and multivariate phenotypes have

been identified in numerous lineages using this approach, including butterflies (Chazot, et al. 2015), lacertid lizards (Hipsley and Muller 2017), marine scallops (Serb, et al. 2017, Sherratt, et al. 2016), and other taxa. In general, permutation-based PGLS approaches provide a powerful tool for understanding patterns of covariation in multivariate phenotypes. We recommend that permutation-based PGLS be used in future studies to evaluate covariation in multivariate phenotypes as described by phylogenetic linear models (regression, ANOVA, etc.).

6. Evolutionary Tempo and Mode in Multivariate Phenotypes

Phylogenetic comparative methods describe the accumulation of phenotypic variation across the phylogeny under some process of evolutionary change. Typically, Brownian motion is used, where trait variation accumulates proportional to time under random (neutral) trait perturbations (Felsenstein 1973, Felsenstein 1981). However, other models of evolutionary change could be envisioned, such as Ornstein-Uhlenbeck (OU) models that incorporate selection into the variance-generating process (Butler and King 2004, Hansen 1997). Recent years have seen the development of analytical methods for comparing the fit of phenotypic data to the phylogeny under alternative evolutionary models. These methods have the advantage of providing empiricists with a means of modeling different evolutionary scenarios and evaluating the fit of the data to the phylogeny under models describing these hypothesized processes (e.g., Beaulieu, et al. 2012, Butler and King 2004). Typically, such 'evolutionary model fitting' approaches are described from a likelihood perspective; however, it is important to recognize that algebraically these methods can also be described using the PGLS model:

$$\mathbf{Y} = \mathbf{1}\mathbf{B} + \mathbf{E} \qquad 2$$

where **Y** is the matrix of multivariate phenotypic trait values, **1** is a column of ones (indicating a single-mean model), $\bar{\mathbf{B}}$ is a vector of model coefficients, and **E** is a matrix of residuals. In this case, the residuals of the model are normally distributed, but only under the specific model of evolutionary change (e.g., Brownian motion, Ornstein-Uhlenbeck, etc.). Viewed from this framework, different evolutionary models can be described by using different evolutionary covariance matrices embodied by **E** (see Adams and Collyer 2018a, Clavel, et al. 2015). Thus, evolutionary model comparisons are accomplished by obtaining summary statistics (e.g., logL or AIC) describing the fit of the data to the phylogeny under differing models of trait evolution, and selecting the preferred model based on these statistics (e.g., Butler and King 2004). While most analytical methods for comparing evolutionary models were developed for univariate traits (Beaulieu, et al. 2012, Butler and King 2004, O'Meara, et al. 2006, Thomas, et al. 2006), several approaches can now accommodate model comparisons for multivariate phenotypes.

6.1. The Tempo of Evolution in Multivariate Phenotypes

One class of models facilitates comparisons of rates of phenotypic evolution across lineages. Here the fit of the data to the phylogeny is obtained under a model containing a single rate of evolutionary change for all species, and then under a second model where rates of evolution differ between two or more groups. For multivariate phenotypes, this is tantamount to fitting the data to the phylogeny using one or more evolutionary rate matrices (**R**), where the evolutionary rates (σ^2) for each phenotypic trait dimension are found along the diagonal of the $p \times p$ evolutionary rate matrix (**R**). One approach compares the fit of one or more evolutionary rate matrices to the phylogeny using likelihood ratio tests and AIC values (Clavel, et al. 2015, Revell

and Harmon 2008: for a related Bayesian approach see: Caetano and Harmon 2019). Such likelihood methods can be appropriate when one is evaluating rates of phenotypic evolution from many taxa and just a few phenotypic trait variables (i.e., when $N >> p$: see Revell and Harmon 2008). However, when the $N:p$ ratio decreases, these methods suffer from high levels of model misspecification; favoring multi-rate models when data were generated under a single-rate model. Additionally, as the phenotypic data become more highly multivariate, model misspecification errors increase precipitously (Adams 2014c, Adams and Collyer 2018a). Therefore, likelihood-based methods for evaluating shifts in evolutionary rate matrices (**R**) are not a general solution for evaluating rate-shifts in high-dimensional phenotypic data, and are only appropriate when the dataset comprises a few trait dimensions and many species (i.e., when $N >> p$).

An alternative approach estimates a single multivariate rate of evolution (σ^2_{mult}) for all traits simultaneously (Adams 2014c), and uses simulations or permutations to evaluate the fit of the data under a single-rate versus a multi-rate model. Similarly, evolutionary rates among several multivariate phenotypes for the same taxa can be compared using σ^2_{mult} (see Denton and Adams 2015). Both approaches are robust to the challenges described above, and tests based on them display appropriate statistical properties. For instance, rates of evolution were found to differ in distinct regions of the avian skull, and these patterns were negatively correlated with levels of integration within modules (Felice and Goswami 2017). A similar pattern was discovered in ray-finned fishes, where lower levels of modularity in some traits was suggested to promote phenotypic diversification (Larouche, et al. 2018). Rates of multivariate phenotypic evolution have also been shown to differ in shell shapes among ecotypes of marine scallops (Sherratt, et al. 2017), in the pectoral fins of acanthomorph fishes (Du, et al. 2019), and in body shape evolution between several endangered freshwater fish lineages (Foster and Piller 2018), among other examples.

One limitation with these methods is that σ^2_{mult} describes the net evolutionary rate of change across the entire multivariate phenotype space. Therefore, if rates of evolution along individual trait axes are of interest, methods more akin to those developed using the evolutionary rate matrix **R** would be appropriate; though current implementations display high rates of model misspecification when phenotypes are highly multivariate. We consider having robust methods based on **R** for high-dimensional phenotypes as a current analytical need, and thus recommend that future theoretical work focus on the development of robust methods that can evaluate sets of phenotypic traits individually, but do so in a manner that minimizes misspecification rates and maximizes statistical power.

6.2. Multivariate Phenotypes and Evolutionary Mode

Because Brownian motion may not be the most appropriate model for describing patterns of phenotypic evolution, models that incorporate selection and other evolutionary processes have been developed. For instance, likelihood-based analytical approaches can characterize selection-based models in multivariate phenotypes, such as those embodied by Ornstein-Uhlenbeck (OU) processes. One method attempts to discover how many adaptive peaks are observed in a multivariate dataset, by fitting a series of Brownian motion and OU models to each trait dimension separately, and combining the optimal fitting models across trait dimensions to arrive at an estimate of the adaptive landscape for the dataset (Ingram and Mahler 2013). Unfortunately, the method suffers extremely high levels of model misspecification, and is unreliable. In a recent analysis, nearly 95% of datasets simulated under Brownian motion were incorrectly predicted to display two or more OU-adaptive peaks (see Adams and Collyer 2018a). Further, comparing predicted patterns from the observed dataset to those obtained from data simulated under Brownian

motion does not provide additional insight. With this approach, one could, for instance, determine that there were more predicted adaptive peaks in the observed dataset than in the simulated data. However, even under such circumstances, it is still unknown which of the predicted adaptive peaks represent the real peaks in the data (if any), and which describe the peaks generated by the method. Thus, while the method has considerable intuitive appeal, its current implementation is prohibitive for biological insight (for additional issues with the approach see: Adams and Collyer 2018a).

Several other approaches have been developed that fit OU models to multivariate phenotypic data (e.g., Bartoszek, et al. 2012, Bastide, et al. 2018, Clavel, et al. 2015, Goolsby 2016, Khabbazian, et al. 2016). With these methods, one fits the data to a series of *a priori* models (Brownian motion, OU 1-peak, OU 2-peaks, etc.), and summary measures such as AIC are used to identify the model with the highest support. However, at present, all current implementations of multivariate OU models suffer from one of three critical shortcomings (see Adams and Collyer 2018a). Either the methods display very high levels of model misspecification (preferring more complex models for data simulated under Brownian motion) that increases with trait dimensionality, or they are not rotation-invariant (meaning, different outcomes are obtained for the same data viewed in different orientations), or they only work if data-dimensionality is reduced to accommodate log-likelihood estimation. One reason for these issues is the large number of parameters that must be estimated (particularly covariance terms), even for relatively simple evolutionary models and just a few trait dimensions. For instance, whereas a two-peak OU model for univariate data is described by 4 parameters (1 alpha, 1 sigma, and 2 theta parameters), the same model for three-dimensional data requires up to 18 parameters (6 alpha parameters, 6 sigma parameters, 6 theta parameters). Likewise, a three-group rate shift model for three phenotypic trait dimensions requires up to 21 parameters to encode (18 sigma parameters, 3 phylogenetic mean

parameters). Clearly, evolutionary models can become very parameter rich, even for multivariate phenotypes described by only a few trait dimensions. Thus, comparing complex evolutionary models, even in those cases where $N >> p$, is still problematic, because of the large number of parameters that must be accurately estimated and evaluated. As such, at present we lack a reliable approach for fitting OU models to multivariate phenotypes, which is clearly a pressing need. Future theoretical work should explore alternative implementations to arrive at a robust approach to evaluating multivariate OU models.

7. Conclusions

The past decade has seen tremendous growth in the use of phylogenetic comparative methods to evaluate evolutionary trends in multivariate phenotypes. Methods for visualizing such trends in high-dimensional datasets are now available, the degree of phylogenetic signal can be reliably quantified, and patterns of covariation – both within multivariate phenotypes, and between multivariate phenotypes and other variables – can be characterized. We identified several current analytical challenges in the field where methods need to be developed, and remain confident that the scientific community is up to the task. As the field continues to mature, we are excited for the future discoveries empiricists will make, which will enrich our understanding of how evolutionary processes shape patterns of diversity in multivariate phenotypes across the tree of life.

Summary Points

1. Phylogenetic comparative analyses on multivariate phenotypes describe patterns of dispersion in multivariate dataspaces. For these analyses to have meaning, distances and directions in the phenotype space must be interpretable. To ensure this, the axes of

multivariate phenotypes should be in commensurate units and scale. Empiricists should carefully consider whether the phenotypic traits they use are in commensurate units and scale, so that downstream analyses of macroevolutionary patterns are interpretable.

2. Phylomorphospaces provide a low-dimensional visualization of a high-dimensional phenotypic dataspace. They are extremely useful for describing multivariate trait evolution and for generating hypotheses for future evaluation. Phylomorphospaces should be used as a regular part of any macroevolutionary analysis of multivariate phenotypes.
3. Phylogenetic signal in multivariate phenotypes is quantified in a manner analogous to what is accomplished for univariate traits. Patterns of phylogenetic signal in different types of multivariate phenotypes are similar to those observed in univariate traits. Most datasets display significant phylogenetic signal, yet those values are less than what is expected under Brownian motion.
4. Testing hypotheses of covariation in multivariate phenotypes becomes challenging as the number of trait dimensions increases. For this reason, standard (parametric) statistical hypothesis testing methods break down as phenotypes become more highly multivariate. Phylogenetic transformation, combined with the use of robust summary statistics and permutation methods (residual randomization), provides a solution to these challenges, so that patterns described by linear models can be evaluated in a phylogenetic framework. Permutation-based PGLS should be a regular part of the macroevolutionary toolkit for evaluating hypotheses of covariation in multivariate phenotypes while accounting for phylogenetic non-independence.
5. Characterizing the manner in which phenotypic diversity accumulates is accomplished by evaluating alternative evolutionary models. For multivariate phenotypes, the net rate of

evolutionary change under Brownian motion can be reliably compared between clades and between multivariate traits. However, characterizing non-Brownian evolution in multivariate phenotypes, such as processes described by Ornstein-Uhlenbeck models, represents a current challenge in the field.

Future Issues

1. New phylogenetic ordination approaches that maximize the covariation between the phenotypic data and the phylogeny should be developed.
2. Tests evaluating phylogenetic signal relative to what is expected under Brownian motion or other evolutionary models should be formalized for both univariate and multivariate phenotypes.
3. Robust evolutionary models for describing trait change in highly multivariate phenotypes, including Ornstein-Uhlenbeck models, are a critical need, and should be developed.

Acknowledgments

We thank E. Baken, E. Glynne, and B. Juarez for comments on the manuscript and assistance with data preparation, E. Glynne for image assistance, and A. Kaliontzopoulou for comment and discussion. This work was sponsored in part by National Science Foundation grants DEB-1556379 to D.C.A., and DEB-1737895 to M.L.C.

Literature Cited

Adams DC. 2010. Parallel evolution of character displacement driven by competitive selection in terrestrial salamanders. *BMC Evol. Biol.* 10:1-10

Adams DC. 2013. Comparing evolutionary rates for different phenotypic traits on a phylogeny using likelihood. *Syst. Biol.* 62:181-92

Adams DC. 2014a. A generalized Kappa statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. *Syst. Biol.* 63:685-97

Adams DC. 2014b. A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. *Evolution* 68:2675-88

Adams DC. 2014c. Quantifying and comparing phylogenetic evolutionary rates for shape and other high-dimensional phenotypic data. *Syst. Biol.* 63:166-77

Adams DC, Collyer ML. 2015. Permutation tests for phylogenetic comparative analyses of high-dimensional shape data: what you shuffle matters. *Evolution* 69:823-29

Adams DC, Collyer ML. 2016. On the comparison of the strength of morphological integration across morphometric datasets. *Evolution* 70:2623-31

Adams DC, Collyer ML. 2018a. Multivariate phylogenetic comparative methods: Evaluations, comparisons, and recommendations. *Syst. Biol.* 67:14-31

Adams DC, Collyer ML. 2018b. Phylogenetic ANOVA: Group-clade aggregation, biological challenges, and a refined permutation procedure. *Evolution* 72:1204-15

Adams DC, Felice RN. 2014. Assessing trait covariation and morphological integration on phylogenies using evolutionary covariance matrices. *PLoS ONE* 9:e94335

Adams DC, Rohlf FJ, Slice DE. 2013. A field comes of age: Geometric morphometrics in the 21st century. *Hystrix* 24:7-14

Aristide L, Bastide P, Reis SFd, Santos TMPd, Lopes RT, et al. 2018. Multiple factors behind early diversification of skull morphology in the continental radiation of New World monkeys. *Evolution* 72:2697-711

Arnold SJ. 2005. The ultimate causes of phenotypic integration: lost in translation. *Evolution* 59:2059-61

Baab KL, Perry JMG, Rohlf FJ, Jungers WL. 2014. Phylogenetic, ecological, and allometric correlates of cranial shape in Malagasy Lemuriforms. *Evolution* 68:1450-68

Bartoszek K, Pienaar J, Mostad P, Andersson S, Hansen TF. 2012. A phylogenetic comparative method for studying multivariate adaptation. *J. Theor. Biol.* 314:204-15

Bastide P, Ané C, Robin S, Mariadassou M. 2018. Inference of adaptive shifts for multivariate correlated traits. *Syst. Biol.* 67:662–80

Beaulieu JM, Jhwueng DC, Boettiger C, O'Meara BC. 2012. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. 66:2369-83

Blackith RE, Reyment RA. 1971. *Multivariate morphometrics*. London: Academic Press

Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717-45

Blomberg SP, Lefevre JG, Wells JA, Waterhouse M. 2012. Independent contrasts and PGLS regression estimators are equivalent. *Syst. Biol.* 61:382-91

Blows MW. 2007. A tale of two matrices: multivariate approaches in evolutionary biology. *J. Evol. Biol.* 20:1-8

Bookstein FL. 2013. Random walk as a null model for high-dimensional morphometrics of fossil series: geometrical considerations. *Paleobiology* 39:52-74

Bookstein FL. 2015. Integration, disintegration, and self-similarity: characterizing the scales of shape variation in landmark data. *Evol. Biol.* 42:395-426

Butler MA, King AA. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* 164:683-95

Caetano DS, Harmon LJ. 2019. Estimating correlated rates of trait evolution with uncertanty. *Syst. Biol.* 68:412-29

Carroll AM, Wainwright PC, Huskey SH, Collar DC, Turingan RG. 2004. Morphology predicts suction feeding performance in centrarchid fishes. *J. Exp. Biol.* 207:3873–81

Catlett KK, Schwartz GT, Godfrey LR, Jungers WL. 2010. "Life history space": A multivariate analysis of life history variation in extant and extinct Malagasy lemurs. *Am. J. Phys. Anthropol.* 142:391-404

Chazot N, Panara S, Zilbermann N, Blandin P, Y.L. Poul, et al. 2015. Morpho morphometrics: Shared ancestray and selection drive the evolution of wing size and shape in *Morpho* butterflies. *Evolution* 70:181-94

Cheverud JM. 1996. Developmental integration and the evolution of pleiotropy. *Am. Zool.* 36:44-50

Chira AM, Cooney CR, Bright JA, Capp EJR, Hughes EC, et al. 2018. Correlates of rate heterogeneity in avian ecomorphological traits. *Ecol. Lett.* 21:1505-14

Clavel J, Escarguel G, Merceron G. 2015. mvMORPH: an R package for fitting multivariate evolutionary models to morphometric data. *Methods Ecol. Evol.* 6:1311-19

Collar DC, Wainwright PC, Alfaro ME, Revell LJ, Mehta RS. 2014. Biting disrupts integration to spur skull evolution in eels. *Nature Comm.* 5:5505

Collyer ML, Adams DC. 2007. Analysis of two-state multivariate phenotypic change in ecological studies. *Ecology* 88:683-92

Collyer ML, Adams DC. 2018. RRPP: An R package for fitting linear models to high-dimensional data using residual randomization. *Methods Ecol. Evol.* 9:1772-79

Collyer ML, Sekora DJ, Adams DC. 2015. A method for analysis of phenotypic change for phenotypes described by high-dimensional data. *Heredity* 115:357-65

Davis AM, Betancur-R R. 2017. Widespread ecomorphological convergence in multiple fish families spanning the marine-freshwater interface. *Proc. B.* 284:20170565

Denton JSS, Adams DC. 2015. A new phylogenetic test for comparing multiple high-dimensional evolutionary rates suggests interplay of evolutionary rates and modularity in lanternfishes (Myctophiformes; Myctophidae). *Evolution* 69:2425-40

Du TY, Tissandier SC, Larsson HCE. 2019. Integration and modularity of teleostean pectoral fin shape and its role in the diversification of acanthomorph fishes. *Evolution* 73:In Press

Evans KM, Waltz BT, Tagliacollo VA, Sidlauskas BL, Albert JS. 2017. Fluctuations in evolutionary integration allow for big brains and disparate faces. *Sci. Reports* 7:40431

Felice RN, Goswami A. 2017. Developmental origins of mosaic evolution in the avian cranium. *Proc. Natl. Acad. Sci. U.S.A.* 115:555-60

Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Gen.* 25:471-92

Felsenstein J. 1981. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution* 35:1229-42

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

Figueirido B, Serrano-Alarcón FJ, Slater GJ, Palmqvist P. 2010. Shape at the cross-roads: homoplasy and history in the evolutoin of the carnivoran skull towards herbivory. *J. Evol. Biol.* 23:2579-94

Figueirido B, Tseng ZJ, Serrano-Alarcón FJ, Martín-Serra A, Pastor JF. 2014. A three-dimensional computer simulation of feeding behaviour in red and giant pandas relates skull biomechanics with dietary niche partitioning. *Biol. Letters* 10:20140196

Foster KL, Piller KR. 2018. Disentangling the drivers of diversification in an imperiled group of freshwater fishes (Cyprinodontiformes: Goodeidae). *BMC Evol. Biol.* 18:1-12

Friedman ST, Price SA, Hoey AS, Wainwright PC. 2016. Ecomorphological convergence in planktivorous surgeonfishes. *J. Evol. Biol.* 29:965-78

Garland T, Jr., Dickerman AW, Janis CM, Jones JA. 1993. Phylogenetic analysis of covariance by computer simulation. *Syst. Biol.* 43:265-92

Garland TJ. 1992. Rate tests for phenotypic evolution using phylogenetically independent contrasts. *Am. Nat.* 140:2104-11

Garland TJ, Ives AR. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *Am. Nat.* 155:346-64

Goolsby EW. 2015. Phylogenetic comparative methods for evaluating the evolutionary history of function-valued traits. *Syst. Biol.* 64:568–78

Goolsby EW. 2016. Likelihood-based parameter estimation for high-dimensional phylogenetic comparative models: Overcoming the limitations of "distance-based" methods. *Syst. Biol.* 65:852-70

Grafen A. 1989. The phylogenetic regression. *Phil. Trans. Roy. Soc. London B.* 326:119-57

Grunstra NDS, Mitteroecker P, Foley RA. 2018. A multivariate ecogeographic analysis of macaque craniodental variation. *Am. J. Phys. Anthropol.* 166:386-400

Gunz P, Mitteroecker P. 2013. Semilandmarks: a method for quantifying curves and surfaces. *Hystrix* 24:103-09

Hansen TF. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341-51

Harmon LJ. 2018. *Phylogenetic comparative methods: Learning from trees*

Harmon LJ, Melville J, Larson A, Losos JB. 2008. The role of geography and ecological opportunity in the diversification of day geckos (*Phelsuma*). *Syst. Biol.* 57:562-73

Harvey PH, Pagel MD. 1991. *The comparative method in evolutionary biology*. Oxford: Oxford University Press

Hipsley CA, Muller J. 2017. Developmental dynamics of ecomorphological convergence in a transcontinental lizard radiation. *Evolution* 71:936-48

Huey RB, Garland T, Turelli M. 2019. Revisiting a key innovation in evolutionary biology: Felsenstein's "Phylogenies and the Comparative Method". *Amer. Natur.* 193:In Press

Huttegger SM, Mitteroecker P. 2011. Invariance and meaningfulness in phenotype spaces. *Evol. Biol.* 38:335-51

Ingram T, Mahler DL. 2013. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods Ecol. Evol.* 4:416-25

Khabbazian M, Kriebel R, Rohe K, Ané C. 2016. Fast and accurate detection of evolutionary shifts in Ornstein-Uhlenbeck models. *Methods Ecol. Evol.* 7:811-24

Kingsolver JG, Gomulkiewicz R, Carter PA. 2001. Variation, selection and evolution of function-valued traits. *Genetica* 112:87-104

Kirkpatrick M, Meyer K. 2004. Direct estimation of genetic principal components: simplified analysis of complex phenotypes. *Genetics* 168:2295-306

Klingenberg CP. 2014. Studying morphological integration and modularity at multiple levels: concepts and analysis. *Phil. Trans. Roy. Soc. Lond. B.* 369:20130249

Klingenberg CP, Ekau W. 1996. A combined morphometric and phylogenetic analysis of an ecomorphological trend: pelagization in Antarctic fishes (Perciformes: Nototheniidae). *Biol. J. Linn. Soc.* 59:143-77

Klingenberg CP, Gidaszewski NA. 2010. Testing and quantifying phylogenetic signals and homoplasy in morphometric data. *Syst. Biol.* 59:245-61

Klingenberg CP, Marugán-Lobón J. 2013. Evolutionary covaration in geometric morphometric data: analyzing integration, modularity, and allometry in a phylogenetic context. *Syst. Biol.* 62:591-610

Lande R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* 33:402-16

Lande R, Arnold SJ. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210-26

Larouche O, Zelditch ML, Cloutier R. 2018. Modularity promotes morphological divergence in ray-finned fishes. *Sci. Reports* 8:7278

Legendre P, Legendre L. 2012. *Numerical Ecology*. Amsterdam: Elsevier

Losos JB. 1992. The evolution of convergent structure in Caribbean *Anolis* communities. *Syst. Biol.* 41:403-20

Mahler DL, Revell LJ, Glor RE, Losos JB. 2010. Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean Anoles. *Evolution* 64:2731-45

Martin CH, Wainwright PC. 2011. Trophic novelty is linked to exceptional rates of morphological diversification in two adaptive radiations of *Cyprinodon* pupfish. *Evolution* 65:2197-212

Martinez CM, McGee MD, Bornstein SR, Wainwright PC. 2018. Feeding ecology underlies the evolution of cichlid jaw mobility. *Evolution* 72:1645-55

Martins EP, Hansen TF. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* 149:646-67

McPeek MA, Shen L, Torrey JZ, Farid H. 2008. The tempo and mode of three-dimensional morphological evolution in male reproductive structures. *Am. Nat.* 171:E158-E78

Mitteroecker P, Bookstein FL. 2007. The conceptual and statistical relationship between modularity and morphological integration. *Syst. Biol.* 56:818–36

Mitteroecker P, Gunz P. 2009. Advances in geometric morphometrics. *Evol. Biol.* 36:235-47

Moen DS, Irschick DJ, Wiens JJ. 2013. Evolutionary conservation and convergence both lead to striking similarity in ecology, morphology and performance across continents in frogs. *Proc. Roy. Soc. B.* 280:20132156

Munkemuller T, Lavergne S, Bzeznik B, Dray S, Jombart T, et al. 2012. How to measure and test phylogenetic signal. *Methods Ecol. Evol.* 3:743-56

Neaux D, Sansalone G, Ledogar JA, Ledogar SH, Luk THY, et al. 2018. Basicranium and face: Assessing the impact of morphological integration in primate evolution. *J. Hum. Evol.* 118:43-55

O'Meara BC, C.Ane, Sanderson MJ, Wainwright PC. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60:922-33

Olson EC, Miller RL. 1958. *Morphological Integration*. Chicago: University of Chicago Press

Pagel MD. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877-84

Paluh DJ, Bauer AM. 2018. Phylogenetic history, allometry and disparate functional pressures influence the morphological diversification of the gekkotan quadrate, a keystone cranial element. *Biol. J. Linn. Soc.* 125:693-708

Pavlicev M, Cheverud JM, Wagner GP. 2009. Measuring morphological integration using eigenvalue variance. *Evol. Biol.* 36:157-70

Pie MR, Campos LLF, Meyer ALS, Duran A. 2017. The evolution of climatic niches in squamate reptiles. *Proc. Roy. Soc. B.* 284:20170268

Polly DP, Lawing AM, Fabre A, Goswami A. 2013. Phylogenetic principal components analysis and geometric morphometrics. *Hystrix* 24:33-41

Price SA, Wainwright PC, Bellwood DR, Kazancioglu E, Collar DC, et al. 2010. Functional innovations and morphological diversification in parrotfish. *Evolution* 64:3057-68

Revell LJ. 2009. Size-correction and principal components for interspecific comparative studies. *Evolution* 63:3258-68

Revell LJ, Collar DC. 2009. Phylogenetic analysis of the evolutionary correlation using likelihood. *Evolution* 63:1090-100

Revell LJ, Harmon LJ. 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. *Evol. Ecol. Res.* 10:311-31

Rohlf FJ. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution* 55:2143-60

Rohlf FJ. 2002. Geometric morphometrics and phylogeny. In *Morphology, shape and phylogeny*, ed. N MacLeod, PL Forey, pp. 175-93. London: Taylor & Francis

Royer-Carenzi M, Didier G. 2016. A comparison of ancestral state reconstruction methods for quantitative characters. *J. Theor. Biol.* 404:126-42

Serb JM, E. Sherratt E, Alejandrino A, Adams DC. 2017. Phylogenetic convergence and multiple shell shape optima for gliding scallops (Bivalvia: Pectinidae). *J. Evol. Biol.* 30:1736-47

Sherratt E, Alejandrino A, Kraemer AC, Serb JM, Adams DC. 2016. Trends in the sand: directional evolution in the shell shape of recessing scallops (Bivalvia: Pectinidae). *Evolution* 70:2061-73

Sherratt E, Serb JM, Adams DC. 2017. Rates of morphological evolution, asymmetry and morphological integration of shell shape in scallops. *BMC Evol. Biol.* 17:1-11

Sidlauskas B. 2008. Continuous and arrested morphological diversification in sister clades of characiform fishes: a phylomorphospace approach. *Evolution* 62:3135-56

Stayton CT. 2015. The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. *Evolution* 69:2140–53

Thomas GH, Freckleton RP, Székely. T. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. *Proc. Roy. Soc. B.* 273:1619-24

Uyeda JC, Caetano DS, Pennell MW. 2015. Comparative analysis of principal components can be misleading. *Syst. Biol.* 64:677-89

Valenzuela N. 2010. Multivariate expression analysis of the gene network underlying sexual development in turtle embryos with temperature-dependent and genotypic sex determination. *Sex. Develop.* 4:39-49

Yuan ML, Wake MH, Wang IJ. 2019. Phenotypic integration between claw and toepad traits promotes microhabitat specialization in the *Anolis* adaptive radiation. *Evolution* 73:In Press

Zelditch ML, Li J, Tran LAP, Swiderski DL. 2015. Relationships of diversity, disparity, and their evolutionary rates in squirrels (Sciuridae). *Evolution* 69:1284-300

Zelditch ML, Ye J, Mitchell JS, Swiderski DL. 2017. Rare ecomorphological convergence on a complex adaptive landscape: body size and diet mediate evolution of jaw shape in squirrels (Sciuridae). *Evolution* 71:633-49

Figure Captions

Figure 1 Examples of common multivariate phenotypic data. (*a*) Different types of measurements obtained from anatomical objects; including linear distances (orange), landmarks (red), and curves (blue). Image from the phenome10K project **www.phenome10k.org**, freely distributed under a Creative Commons license; scan contributed by Figueirido, et al. 2014). (*b*) A function-valued trait representing a growth curve. (*c*) Multivariate phenotypes represented mathematically from each of the four datasets (left to right): linear distances, landmarks, curves, and function-valued traits. (*d*) Example of a multivariate phenotype space for a particular multivariate dataset, with several species displayed. Species closer together in the space are similar in their multivariate phenotypes, while species further apart are less similar.

Figure 2 Hypothetical examples phylomorphospace plots for 32 species related by a phylogeny. (*a*) Clades for these species are denoted in orange and blue. Phylomorphospace patterns include: (*b*) Clade overlap in multivariate phenotype space. (*c*) Clade divergence of phenotypes. (*d*) Differences in clade disparity. (*e*) Directional phenotypic evolution in one clade.

Figure 3 Patterns of phylogenetic signal in multivariate phenotypes. (*a*) Frequency distribution of the degree of phylogenetic signal in multivariate phenotypes from 330 empirical datasets obtained from the literature. (*b*) Phylogenetic signal represented as the mean across five multivariate datasets simulated under Brownian motion evolution across a phylogeny of 400 species ($N = 400$). At each simulation level, the ratio of variables to species ($p:N$) was altered, and the number of trait dimensions simulated under Brownian motion, versus the number of dimensions whose traits were generated with no association to the phylogeny, were altered. Estimates of phylogenetic signal

(K_{mult}) and estimates of statistical significance are shown.