Statistical and conceptual challenges in the comparative analysis of principal components

Josef C. Uyeda[1,*], Daniel S. Caetano[1], and Matthew W. Pennell[1]

[1] Department of Biological Sciences & Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID 83844, U.S.A.
[*] Email for correspondence: josef.uyeda@gmail.com

Keywords: Phylogenetic comparative methods, principal components analysis, Brownian motion, Ornstein-Uhlenbeck, Early Burst, multivariate evolution

Quantitative geneticists long ago recognized the value of studying evolution in a multivariate framework (Pearson, 1903). Due to linkage, pleiotropy, coordinated selection and mutational covariance, the evolutionary response in any phenotypic trait can only be properly understood in the context of other traits (Lande, 1979; Lynch and Walsh, 1998). This is of course also well–appreciated by comparative biologists. However, unlike in quantitative genetics, most of the statistical and conceptual tools for analyzing phylogenetic comparative data (recently reviewed in Pennell and Harmon, 2013) are designed for analyzing a single trait (but see, for example Revell and Harmon, 2008; Revell and Harrison, 2008; Hohenlohe and Arnold, 2008; Revell and Collar, 2009; Schmitz and Motani, 2011; Adams, 2014b). Indeed, even classical approaches for testing for correlated evolution between two traits (e.g., Felsenstein, 1985; Grafen, 1989; Harvey and Pagel, 1991) are not actually multivariate as each trait is assumed to have evolved under a process that is independent of the state of the other (Hansen and Orzack, 2005; Hansen and Bartoszek, 2012). As a result of this limitations, researchers with multivariate datasets are often faced with a choice: analyze each trait as if they were independent or else decompose the dataset into statistically independent set of traits, such that each set can be analyzed with the univariate methods.

Principal components analysis (PCA) is the most common method for reducing the dimensionality of the dataset prior to analyzing the data using phylogenetic comparative methods. The first PC axis is the eigenvector in the direction of greatest variance, the second PC axis, the second greatest variance, and so on. However, standard methods for calculating PC scores assume that the samples are independent of one another, which is hardly ever the case for comparative data. As a result of shared common ancestry, relatives are likely to share many traits and trait combinations. Performing comparative analyses without considering the species' evolutionary relationships is anathema to most evolutionary biologists but it is less well–appreciated that phylogeny should be considered when transforming data (Revell, 2009; Polly et al., 2013).

Standard PCA continues to be regularly used in comparative biology. This is applied to a variety of types of traits including geometric morphometric landmarks (e.g., Dornburg et al., 2011; Hunt, 2013), measurements of multiple morphological traits (e.g., Harmon et al., 2010; Bergmann and Irschick, 2012; Weir and Mursleen, 2013; Pienaar et al., 2013; Price et al., 2014), and climatic variables (e.g., Kozak and Wiens, 2010; Schnitzler et al., 2012). We stress that the papers that we have cited here are simply examples selected from a substantial number of papers where standard PCA was used.

The most frequently used approach for correcting for the non–independence of species is to assume a phylogenetic model for the evolution of measured

traits and incorporate the expected covariance in the calculation of the PC axes and scores (Revell, 2009). Revell's method (explained in detail below) assumes that the measured traits have evolved under a multivariate Brownian motion (BM; Edwards and Cavalli-Sforza, 1964) model of trait evolution. In a brief simulation study, Revell (2009) demonstrated that if the underlying model for the traits was indeed a multivariate BM model, performing standard PCA gave biased estimates of the eigenvalues, whereas pPCA did not.

In this paper, we first extend the argument of Revell (2009) and demonstrate how biased eigenvalues obtained from PCA systematically distort biological inference in a predictable manner. Performing comparative analyses on standard PC axes can therefore positively mislead inference. This point has been made in other fields that deal with auto–correlated data, such as population genetics (Novembre and Stephens, 2008), ecology (Podani and Miklós, 2002), climatology (Richman, 1986) and paleobiology (Bookstein, 2012). However, the connection between these previous results and phylogenetic comparative data has not been widely appreciated and standard PCs continue to be regularly used in the field. We hope that our paper helps change this practice.

Second, as stated above, Revell (2009) assumed that the measured traits had evolved under a multivariate BM process. As the pPC scores are not phylogenetically independent (Revell, 2009; Polly et al., 2013, see below), one must use comparative methods to analyze them which will in turn require selecting an evolutionary model for the scores. The choice of model for the traits and the pPC scores are separate steps in the analysis (Revell, 2009). This has the potential to introduce an odd circularity into the analysis: it seems probable that the choice of a model for the evolution of the traits will influence the apparent macroevolutionary dynamics of the resulting pPC scores. To our knowledge this effect has not been previously explored. Here we analyze simulated data to investigate whether assuming a BM model for the traits introduces systematic biases in the pPC scores when the generating model is different. Furthermore, we analyze two real comparative datasets — wherein the traits almost certainly have not evolved via a strict BM process – to understand the implications of these results for the types of data that researchers actually have.

Last, we consider the interpretation of evolutionary models fit to pPC axes and discuss the conceptual and statistical advantages and disadvantages of using pPCA compared to alternative approaches for studying multivariate evolution in a phylogenetic comparative framework. We argue that the statistical advantages of using pPC axes comes at a substantial conceptual cost and that alternative techniques are likely to be much more informative for addressing many evolutionary questions.

Methods

Overview of pPCA

Before describing our analyses, we briefly overview standard and phylogenetic PCA and highlight the differences between the two. In conventional PCA, a $m \times m$ covariance matrix $\mathbf{R}$ is computed from a matrix of trait values $\mathbf{X}$ for the n species and m traits

$$\mathbf{R} = (n-1)^{-1}(\mathbf{X} - \mathbf{1}\boldsymbol{\mu}^{\intercal})^{\intercal}(\mathbf{X} - \mathbf{1}\boldsymbol{\mu}^{\intercal}) \tag{1}$$

where $\boldsymbol{\mu}$ is a vector containing the means of all m traits and $\mathbf{1}$ is a column vector of ones. We note that in many applications $\mathbf{X}$ may not represent the raw trait values; in geometric morphometrics, for example, size, translation and rotation will often be removed from $\mathbf{X}$ prior to computing $\mathbf{R}$ (Rohlf and Slice, 1990; Bookstein, 1997). The eigenvalues $\mathbf{D}$ and eigenvectors $\mathbf{V}$ of $\mathbf{R}$ are then obtained using singular–value decomposition $\mathbf{R} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$ or some related technique. The scores $\mathbf{S}$, the trait values of the species along the PC axes, are computed as

$$\mathbf{S} = (\mathbf{X} - \mathbf{1}\boldsymbol{\mu}^{\intercal})\mathbf{V}. \tag{2}$$

Phylogenetic PCA differs from this procedure in two important ways (Revell, 2009; Polly et al., 2013) . First the covariance matrix is inversely weighted by the expected covariance of trait values between taxa under a given model $\boldsymbol{\Sigma}$. Under a BM model of trait evolution, $\boldsymbol{\Sigma}$ is simply proportional to the matrix representation of the phylogenetic tree $\mathbf{C}$, such that $\Sigma_{i,j}$ is the shared path length between lineages i and j. (We note that as only relative branch lengths matter, under a multivariate BM process, we can simply set $\boldsymbol{\Sigma} = \mathbf{C}$ without loss of generality.) Including the expected covariance between trait values essentially just re–orients the axes according to the phylogeny. Second, the space is centered on the "phylogenetic means" $\mathbf{a}$ of the traits rather than their arithmetic means, which can be computed following Revell and Harmon (2008):

$$\mathbf{a} = [(\mathbf{1}^{\intercal}\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}^{\intercal}\boldsymbol{\Sigma}^{-1}\mathbf{X}]^{\intercal}. \tag{3}$$

In pPCA, Equation 1 is therefore modified as

$$\mathbf{R} = (n-1)^{-1}(\mathbf{X} - \mathbf{1}\mathbf{a}^{\intercal})^{\intercal}\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \mathbf{1}\mathbf{a}^{\intercal}) \tag{4}$$

Similarly, $\mathbf{S}$ can be calculated for pPCA using Equation 2 but substituting the phylogenetic means for the arithmetic means

$$\mathbf{S} = (\mathbf{X} - \mathbf{1}\mathbf{a}^{\intercal})\mathbf{V} \tag{5}$$

where again, $\mathbf{V}$ is a matrix containing the eigenvectors of $\mathbf{R}$ (in this case obtained from Equation 4).

The effect of weighting the covariance and centering the space using phylogeny has an important statistical consequence (Revell, 2009; Polly et al., 2013). In PCA, each PC score is independent of all other scores from the same PC axis and from scores on other axes. Due to the phylogenetic structure of the data, this property of independence does not hold when using pPCA. Therefore it is necessary to analyze pPC scores using phylogenetic comparative methods, just as one would for any other trait (Revell, 2009; Polly et al., 2013).

Effect of PCA on model selection under multivariate Brownian motion

We simulated 100 replicate datasets under multivariate Brownian motion to evaluate the effect of using standard versus phylogenetic PCA to infer the mode of evolution. For each dataset, we used `TreeSim` (Stadler, 2011) to simulate a phylogeny of 50 terminal taxa under a pure–birth process and scaled each tree to unit height. We then simulated a 20–trait dataset under multivariate Brownian motion. When analyzing phylogenetic comparative data, $\mathbf{R}$ is estimated from the data; here, we set $\mathbf{R}$ to be a known quantity and use it to simulate phylogenetically structured trait data. For each simulation, we generated a positive definite covariance matrix for $\mathbf{R}$, by drawing eigenvalues from an exponential distribution with a rate $\lambda = 1/100$ and randomly oriented orthogonal eigenvectors. We then used this matrix to sample a covariance matrix for the tip states $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{R} \otimes \mathbf{C})$ where $\otimes$ is the Kronecker product. For each of the 100 simulated datasets, we computed PC scores using both standard methods and pPCA (using the `phytools` package; Revell, 2012). We used `phylolm` (Ho and Ané, 2014) to fit models of trait evolution to the original data and to all PC scores obtained by both methods. Following Harmon et al. (2010), we considered three models of trait evolution: 1) BM; 2) Ornstein–Uhlenbeck with a fixed root (OU: Hansen, 1997); and 3) Early Burst (EB: Blomberg et al., 2003; Harmon et al., 2010). We then calculated the Akaike weights (AICw) for each model/transformation/trait combination.

To explore the effect of trait correlation on inference, we conducted an additional set of simulations where $\mathbf{R}$ was varied from the above simulations to result in more or less correlated sets of phenotypic traits. We drew eigenvalues $\mathbf{m}$ from an exponential distribution and scaled these so that the leading eigenvalue m_1 was equal to 1. We then exponentiated this vector across a sequence of exponents ranging for $\ll 1$ to $\gg 1$; this gave us a series of covariance matrices ranging from highly correlated ($m_1 = 1; m_2, \ldots, m_{20} \approx 0$) to nearly independent ($\mathbf{m} \approx \mathbf{1}$), respectively. We

chose the series of exponents to obtain a regular sequence of $m_1 / \sum_{i=1}^{20} m_i$ from 0.05 to 1. For each set of eigenvalues, we simulated 25 datasets and estimated the slope of the relationship between the absolute size of phylogenetically independent contrasts (Felsenstein, 1985) and the height of the node at which they were calculated (i.e., the "node height test" of Freckleton and Harvey, 2006). Under OU models, this relationship is expected to be positive, while under EB models this relationship is negative. BM models are expected not to show correlation between contrasts and height of the nodes.

Effect of using PCA when traits are not Brownian

We then simulated datasets under alternative models of trait evolution. Because of difficulties in efficiently simulating large multivariate datasets of covarying traits under OU or EB models, we instead simulated 20 independent traits under BM, OU and EB for 50 taxa trees (as above, but setting all eigenvalues equal to one another). Of course, this is certainly not representative of the process that have shaped real multivariate data, but considering this simple case allowed us to investigate how misspecifying the model of trait evolution can impact analyses under the simplest scenario.

For the BM simulations, we set $\sigma^2 = 1$. For OU, we set $\sigma^2 = 1$ and $\alpha = 2$, such that the phylogenetic half–life $\log(2)/\alpha$ (Hansen et al., 2008) was equal to ~ 0.35 of the total tree depth. For EB, we again set $\sigma^2 = 1$ and set a, the exponential rate of deceleration, to be $\log(0.02)$.

As above, we fit BM, OU and EB models to the original data, PC scores and pPC scores for each simulated dataset and estimated parameters and AICw. In addition to the model–fitting and comparison, for every transformation, we applied two common diagnostic tests for deviation from BM–like evolution to all trait/PC axes. First, we calculated the slope of the node height test as described in the preceding section. Second, we characterized the disparity through time (Harmon et al., 2003) using the `geiger` package (Pennell et al., in press).

Finally, we examined the scenario in which a set of traits each follow a model of evolution with unique evolutionary parameters. In particular, we use the accelerating-decelerating (ACDC) model of Blomberg et al. (2003) to generate independent trait datasets. This model is a general case of the EB model which allows both accelerating or decelerating rates of phenotypic evolution. Accelerating rates of evolution result in identical likelihoods as the OU model (assuming the root state is at the optimal trait value), and thus are equivalent for our purposes. We simulated 100 datasets with 50 taxa and 20 traits. Trees were generated as in previous simulations. Each trait was simulated along the phylogeny with an ACDC parameter (a) drawn from a normal distribution with mean 0 and standard deviation

of 5. Values of a above 0 correspond to accelerating evolutionary rates, while those below 0 correspond to decelerating, or Early-Burst models of evolution. For each dataset, we conducted both standard and phylogenetic PCA in which the traits are standardized to unit variance (i.e., using correlation matrices, which ensured traits across parameter values had equal expected variances). For each PC or pPC, we regressed the magnitude of the trait loadings against the trait's ACDC parameter value. We then visualized whether there were systematic trends in the relationship between the ACDC parameter value, and the weight given to a particular trait across PC axes. Such systematic trends would indicate that either PCA or pPCA "sorts" traits into PC axes according to the particular evolutionary model each trait follows.

Empirical examples

We analyzed two comparative datasets assembled from the literature, allowing us to investigate the effects of principal components analyses on realistically structured data. First, we analyzed phenotypic evolution across the family Felidae (cats) using measurements from two independent sources — five cranial measurements from Slater and Van Valkenburgh (2009) and body mass and skull width from Sakamoto et al. (2010). For the analysis, we used the supertree compiled by Nyakatura and Bininda-Emonds (2012). Second, we analyzed 23 morphometric traits in *Anolis* lizards and phylogeny from Mahler et al. (2010). In both datasets, all measurements were linear measurements on the logarithmic scale. We conducted standard and phylogenetic PCA and examined the effect of each on model–fitting, the slope of the node–height test, and the average disparity through time. All simulations and analyses were conducting using R v3.0.2 (R Development Core Team, 2013). Scripts to reproduce all analyses are available at `https://github.com/mwpennell/phyloPCA`.

Results

Effect of PCA on model selection under multivariate Brownian motion

Standard PCA introduces a systematic bias in the favored model across principal components. In our simulations where the traits evolved under a multivariate BM model, EB models had systematically elevated support as measured by Akaike weights for the first few components, for which it generally exceeded support for the BM model (Figure 1). Fitting models sequentially across PC axes 1–20 revealed a regular pattern of increas-

ing support for BM models moving from the first toward the intermediate components, followed by increasing support for OU models among later components (which generally approached an AICw of 1). This regular pattern across trait axes was not present for either the original datasets, or for phylogenetic principal components, which found strong support for the BM model regardless of which trait was analyzed. We note that the theoretical maximum AICw for the BM model in the three–model comparison is $1/(2e^{-1}+1) \approx 0.576$, as BM is a special case of both OU and EB and therefore the ΔAICw for these models cannot be greater than 2.

Multivariate datasets simulated with high correlations (low effective dimensionality) showed increased support for BM across PC axes. When the leading eigenvalue explained a large proportion of the variance, the slope of the node height test converged toward 0, indicating no systematic distortion of the contrasts through time (Figure 2). However, when the eigenvalues of the rate matrix were more even, standard PCA resulted in a negative slope in the node height test among the first few PCs, which in turn provides elevated support for EB models. This pattern is reversed among higher PC axes, which have positive slopes between node-height and absolute contrast size, which provides elevated support for OU models (Figures 2 and 4).

Effect of using PCA when traits are not Brownian

If the underlying model was either OU or EB rather than BM, then phylogenetic PCA tended to increase support for the true model relative to the original trait variables for the first few component axes (Figures 3, S1, and S2). For example, when each of the original trait variables were simulated under an OU process, support for the OU model increased for pPC1 relative to the original trait variables. Higher principal component axes inferred a regular pattern of decreasing support for the OU model, while the last few PCs have equivocal support for either a BM or OU model (Figure 3). Furthermore, parameter estimation was affected by phylogenetic PCA. The α parameter of the OU model was estimated to be stronger than the value simulated for individual traits for the first few pPC scores and lower for the higher components (Figure S3).

Examining the outcomes of the node–height tests (Figure 4) and the disparity through time analyses (Figure 5) help clarify the results we observed from model comparison and parameter estimation. Under OU models, traits are expected to have the highest contrasts near the tips, whereas under EB models, traits will have the highest contrasts near the root of the tree. Under multivariate BM, standard PCA maximizes the overall variance explained across the entire dataset, thereby tending to select linear combinations of traits that maximize the contrasts at the root of the tree.

Thus, the first few PCs are skewed toward resembling EB models, while the last few PCs are skewed toward resembling OU models. By contrast, the effect of pPCA on the node–height relationship depends on the generating model. When traits are evolved under an OU model, the first few pPC axes show an exaggerated pattern of high variance towards the tips. Likewise, when traits are evolved under an EB model, the first few pPC axes show an exaggerated pattern of high variance towards the root of the tree. For traits generated under both OU and EB models, the higher components resemble BM–like patterns.

When the data includes traits evolved under different parameters, both PCA and pPCA resulted in the systematic assignment of traits with particular parameter values to each PCA axes. Traits which follow EB models are preferentially given higher loadings under both PCA and pPCA for the first few PCs as well as the last few PCs (Figure 6). Furthermore, both PCA and pPCA assign fewer PCs with higher loadings to traits that follow EB models, whereas the intermediate PCs have more even loadings slightly skewed toward accelerating rates (i.e., OU–like models). This asymmetry may reflect the fact that EB models are more variable in their outcomes to the phylogeny, owing to the fewer independent branches among which divergence can occur closer to the root of the tree. Our results indicate that both pPCA and PCA can be biased in the selection of PC axes with respect to the generating evolutionary model.

Empirical examples

In the Felid dataset, the seven morphometric traits were extremely highly correlated, with the first PC explaining 96.9% and 93.7% of the total variation in the dataset for standard PCA and phylogenetic PCA, respectively. All raw traits and the first PC axis of both standard and phylogenetic PCA support a BM model of evolution (PC and pPC axes have AICw's of 0.574, which is near the theoretical maximum for BM). The last four standard PC axes show strong support for a OU model (AICw $\approx$ 1) whereas under phylogenetic PCA the last axes have mixed support favouring BM or OU (Figure S4). Both the node-height test and the disparity through time plots show this same pattern. The node-height slope of the first axis is approximately zero while the slope of the remaining axes are slightly positive under standard and phylogenetic PCA. The first axis show the same disparity through time pattern of the untransformed data in both standard and phylogenetic PCA. However, the last PC axes show disparity accumulated toward the tips under standard PC, while phylogenetic PCA produced a less clear pattern (Figure S5).

For the morphometric traits in the *Anolis* dataset, the first PC also explained a large proportion of the variation (92.6% and 90.0% for standard

and phylogenetic PCA, respectively). Most of the untransformed traits had equivocal support for either a BM or EB model (Figure 7). While PC1 of both PCA and pPCA mirrored this pattern, the remaining PCs for both PCA and pPCA show a general pattern of decreasing support for an EB model (Figure 7). Collectively PCs 2-4 had higher support for the EB model than any other PC in both standard PCA (AICw $PC2 = 1.0$, $PC3 = 0.47$, $PC4 = 0.28$) and phylogenetic PCA (AICw $pPC2 = 1.0$, $pPC3 = 0.43$, $pPC4 = 0.27$). Similarly, these early PC axes tended to have more negative slopes from the node-height test relative to the average trait in the dataset (Figure S6).

Discussion

Different ways of representing the same set of data can change the meaning of measurements and alter the interpretations of subsequent statistical analyses (Houle et al., 2011). PCA is often considered to be a simple linear transformation of a multivariate dataset and the potential consequences of performing phylogenetic comparative analyses on PC scores has received very little attention. In this paper, we sought to highlight the fact that fitting macroevolutionary models to a handful of PC axes may positively mislead inference — what appears like the signal of an interesting biological process may simply be an artifact stemming from how PCA is computed. By focusing analyses exclusively on the first few PC axes, as is commonly done in comparative studies, researchers are, in effect, taking a biased sample of a multivariate distribution. We demonstrate how these biased samples can influence inference in both PCA and pPCA — particularly toward inferring decreasing rates of evolution in highly dimensional datasets.

We can obtain an intuitive understanding of how inference can be affected by PCA by considering data simulated under a multivariate BM model. Despite a constant rate of evolution across each dimension of trait space, stochasticity will ensure that some dimensions will diverge more rapidly than expected early in the phylogeny, while others will diverge less. All else being equal, dimensions that happen to diverge early are expected to have the greatest variance across species, and standard PCA will identify these axes as the primary axes of variation. However, the trait combinations that are most divergent early, will have apparent slowdowns towards the present simply due to "regression toward the mean", resulting in the characteristic "early burst" pattern of evolution for the first few principal components (for a related point in the context of models of lineage diversification, see Pennell et al., 2012). An analogous process will result in the last few PCs following an OU process, in which the amount of divergence will appear to increase toward the present. Standard PCA thus "sorts" orthogonal trait dimensions by whether they follow EB, BM

and finally, OU like patterns of trait divergence. Thus, traits studied using PCA may often be biased to reflect particular evolutionary models, merely as a statistical artifact.

These problems, which ultimately stem from using PCA on auto–correlated data, are not limited to phylogenetic comparative studies (see Richman, 1986; Podani and Miklós, 2002; Novembre and Stephens, 2008; Jolliffe, 2002; Bookstein, 2012). For example, Novembre and Stephens (2008) demonstrated that apparent waves of human migration in Europe obtained from PCA of genetic data (e.g., Cavalli-Sforza et al., 1994) could be attributed to artifacts similar to those we document here (in their case, the auto–correlation was the result of geography rather than phylogeny). While the bias introduced by applying standard PC to comparative data has been documented previously (Revell, 2009; Polly et al., 2013), we sought to clarify precisely how inferences of macroevolutionary processes and patterns can be impacted by this bias.

Revell (2009) recognized this biased selection of eigenvectors and introduced pPCA as a means for accounting for the phylogenetic correlation of data points. Our simulations verify that when the underlying model is multivariate BM, phylogenetic PCA mitigates the biased selection of PC axes by scaling divergence by the expected divergence given the branch lengths of the phylogeny. However, BM is often a poor descriptor of the macroevolutionary dynamics of trait evolution (for example, see Harmon et al., 2010; Pennell et al., 2014) and assuming this model when performing pPCA is less than ideal. Revell (2009) suggested that alternative covariance structures could be used to estimate phylogenetically independent PCs for different models. For example, one could first optimize the λ model (Pagel, 1999) across all traits simultaneously and then rescale the branch lengths of the tree according to the estimated parameter in order to obtain $\mathbf{\Sigma}$ for use in Equation 4. However, one cannot compare model fits across alternative linear combinations of traits, so the data transformation must occur separately from modeling the evolution of the PC axes. As Revell noted, parameters estimated to construct the covariance structure for the pPCA will likely be different from the same parameters estimated using the PC scores themselves. Furthermore, this procedure is restricted to models that assume a shared mean and variance structure across traits (see Hansen et al., 2008; Bartoszek et al., 2012, for examples where this does not apply). As such, if the question of interest relies on model–based inference, transformation of trait data using pPCA will require simplifying and rather ad–hoc assumptions, and researchers must hope that the resulting inferences are generally robust to these decisions.

We show that when the trait model is misspecified, systematic and predictable distortions occur across pPC axes — similar to those that were observed when the phylogeny was ignored altogether. In some scenar-

ios, such distortions may not substantially alter inference. For example, when all traits evolve under a OU model (or when all traits evolve under a EB model), we find that these distortions primarily serve to inflate the support for the true model. Even so, interpretation of parameter estimates for pPC scores becomes much more challenging (Figures 4, 5, and S3). However, more complex scenarios can produce more worrying distortions. When evolutionary rates vary through time and across traits, both PCA and pPCA will sort traits into PC axes according to which evolutionary model they follow, despite all traits being evolutionarily independent. Under the conditions we examined, this resulted in both PC1 and pPC1 being heavily-weighted toward EB-type models, despite simulating an even distribution of accelerating and decelerating rates across traits. Suggestively, we observe similar patterns for both PCA and pPCA in the *Anolis* morphometric dataset (Figures 7 and S6). Focusing on the first few axes of variation identified by pPCA alone may skew our view of evolutionary processes in nature, and bias researchers toward finding particular patterns of evolution.

When employed as a descriptive tool, PCA can be broadly used even when assumptions regarding statistical non–independence or multivariate normality are violated (Jolliffe, 2002). There is nothing wrong with using standard PCA or pPCA on comparative data to describe axes of maximal variation across species or for visualizing divergence across phylomorphospace (Sidlauskas, 2008). Furthermore, our simulations and empirical analyses suggest that strong correlations among traits (i.e., when the leading eigenvector explained a majority of the variation, e.g., $> 75\%$), PC scores may not be appreciably distorted (Figure 2). The statistical artifacts we describe are more likely to appear when matrices have high effective dimensionality (see Bookstein, 2012). Given that many morphometric datasets may be highly correlated, the overall effect of using PCA or of misspecifying the model in phylogenetic PCA may in some cases be relatively benign. And we certainly do not mean to imply that the biological inferences that have been made from analyzing standard or phylogenetic PC scores in a comparative framework are necessarily incorrect. When Harmon et al. (2010) analyzed the evolution of PC2 (what they referred to as "shape") obtained using standard PCA, they found very little support for the EB model across their 39 datasets. The magnitude of the bias introduced by using standard PCA is difficult to assess but any bias that did exist would be towards finding EB–like patterns. This only serves to strengthen their overall conclusion that such slowdowns are indeed rare (but see Slater and Pennell, 2014). However, our results do suggest that in some cases analyses conducted with PC axes should be revisited to ensure that results are robust.

The broader question raised by our study is how one should draw evolutionary inferences from multivariate data. The first principal component

axis from pPCA is the major axis of divergence across the sampled lineages in the clade (also known as the "line of divergence"). This axis is of considerable interest in evolutionary biology. The direction of this line of divergence may be affected by the orientation of within–population additive genetic (co)variance $\mathbf{G}$, such that evolutionary trajectories may be biased along the "genetic line of least resistance" (i.e., divergence occurs primarily along the leading eigenvector of $\mathbf{G}$, $\mathbf{g}_{max}$; Schluter, 1996). Alternatively, the line of divergence may align with $\boldsymbol{\omega}_{max}$, the "selective line of least resistance", due to the structure of phenotypic adaptive landscapes (Arnold et al., 2001; Jones et al., 2007; Arnold et al., 2008), or else may be driven by patterns of gene flow between populations (Guillaume and Whitlock, 2007) or the pleiotropic effects of new mutations (Jones et al., 2007; Hether and Hohenlohe, 2014).

While it is perfectly sensible and interesting to compare the orientation of pPC1 to that of $\mathbf{g}_{max}$, $\boldsymbol{\omega}_{max}$ or other within–population parameters, making explicit connections between macro– and microevolution requires a truly multivariate approach. Quantitative genetic theory (e.g., Lande, 1979; Lynch and Walsh, 1998) makes predictions about the overall pattern of evolution in multivariate space. By fitting evolutionary models to pPC scores, we are only considering evolution along these axes independently and are therefore missing most of what is going on. In contrast, multivariate tests for the correspondence of axes of trait variation within and between species can provide meaningful insights into the long–term determinants of evolutionary change, and the process by which traits evolve (Hohenlohe and Arnold, 2008; Bolstad et al., 2014).

The most conceptually straightforward multivariate approach for analyzing comparative data is to construct models in which there is a covariance in trait values between species (which is done in univariate models) and a covariance between different traits. Such multivariate extensions of common quantitative trait models have been developed (Butler and King, 2004; Revell and Harmon, 2008; Hohenlohe and Arnold, 2008; Revell and Collar, 2009; Thomas and Freckleton, 2012). These allow researchers to investigate the connections between lines of divergence and within–population evolutionary parameters (Hohenlohe and Arnold, 2008) as well as to study how the correlation structure between traits itself changes across the phylogeny (Revell and Collar, 2009).

However, these approaches also have substantial drawbacks. First, the number of free parameters of the models rapidly increases as more traits are added (Revell and Harmon, 2008), making them impractical for large multivariate datasets. This issue may be addressed by constraining the model in meaningful ways (Butler and King, 2004) or by assuming that all traits (or a set of traits) share the same covariance structure (Klingenberg and Marugán-Lobón, 2013; Adams, 2014b,a). Such restrictions of pa-

rameter space are especially appropriate for truly high–dimensional traits, such as shape inferred from geometric morphometric landmark data. For such traits, we are primarily interested in the evolution of the aggregate trait and not necessarily the individual components. The second drawback is that these models allow for inference of the covariance between traits but the cause of this covariance is usually not tied to specific evolutionary processes. This difficulty can be addressed by explicitly modeling the evolution of some traits as a response to evolution of others. Hansen and colleagues have developed a number of models in which a predictor variable evolves via some process and a response variable tracks the evolution of the first as OU process (Hansen et al., 2008; Hansen and Bartoszek, 2012; Bartoszek et al., 2012). This has been a particularly useful way of modeling the evolution of allometries (e.g. Hansen and Bartoszek, 2012; Voje et al., 2013; Bolstad et al., 2014). But, as with the "full covariance" models discussed above, increasing the number of traits makes the model much more complex and parameter estimation difficult.

As we can only estimate a limited number of parameters from most comparative datasets — and even when we consider large datasets, most existing comparative methods have only been developed for the univariate case — it often remains necessary to reduce the dimensionality of a multivariate dataset to one or a few compound traits. We believe that although PCA can be potentially quite usefully applied to this problem, it may be in ways that are statistically and conceptually distinct from how it is conventionally applied to comparative data.

First, we argue that reducing multivariate problems to more easily managed, lower-dimensionality analyses should be approached with the specific goal of maintaining biological meaning and interpretability (Houle et al., 2011). The common practice of examining only the first few PCs carries with it the implicit assumption that PCA ranks traits by their evolutionary importance and biological interest — a conclusion that may not necessarily be the case (Polly et al., 2013). If a certain PC axis is of sufficient biological interest in its own right, it may not matter if it is a biased subset of a multivariate distribution. The fact that a vast majority of the traits studied in adaptive radiations likely represent very biased axes of variation across the multivariate process of evolution does not diminish the importance of the inferences made from studying these traits (Schluter, 2000).

The danger occurs when the biological significance of the set of traits is poorly understood, and when the source of the statistical signal may be either artifactual or biological. If a trait was not of interest *a priori*, then this essentially turns into a multiple comparisons problem in which PCA searches multivariate trait space for an unusual axis of variation, which tend to suggest models inconsistent with the generating multivariate pro-

cess as a whole. *A posteriori* interpretation of the PC axes by their loadings is something of an art — one must "read the tea leaves" to understand what these axes mean biologically. Even when a particular axis is correlated with a biological interpretation, it can be unclear whether the statistical signal supporting a particular inference results from the evolutionary dynamics of the trait of interest or if it is the result of statistical artifacts introduced by the imperfect representation of that trait by a PC axis. More rigorous algorithms can be applied to identify subsets of the original variables that best approximate the principal components, which — although still biased — are frequently more interpretable (Cadima and Jolliffe, 2001; Somers, 1986, 1989; Hausman, 1982; Vines, 2000; Jolliffe, 2002; Zou et al., 2006). Another potential approach is to use principal components computed from within–population data, rather than comparative data. For example, if **G** (or failing that, the phenotypic variance–covariance matrix **P**) is available for a focal species, then the traits associated to the principal axes of variation in that species can be measured across all species in the phylogeny. In other words, across species trait measurements can be projected along $\mathbf{g}_{\max}$. This alleviates the issues we discuss in this paper by estimating PCs from within–population data that is independent from the comparative data used for model–based inference.

Of course, components defined by within–population variance structure or by approximating principal components with interpretable linear combinations will not explain as much variance across taxa as standard PCA and will not necessarily be statistically independent of one another. However, the extra variance explained by the principal components of comparative data may in fact include a sizeable amount of stochastic noise, rather than interesting biological trait variation (as we have shown in our simulations). Furthermore, while the traits identified by pPCA will be statistically orthogonal, this is only true in the particular snapshot captured by comparative data and does not imply that they are evolving independently. The distinction between statistical and evolutionary independence is crucial (Hansen and Houle, 2008) but it is easy to conflate these concepts when the data has been abstracted from its original form. We argue that the added intepretability of carefully chosen and biologically meaningful trait combinations far outweighs the cost of some trait correlations or explaining less–than–maximal variation.

Concluding remarks

In this note, we sought to clarify some statistical and conceptual issues regarding the use of principal components in comparative biology. We have shown that from a statistical standpoint, failing to consider the phylogeny when performing PCA can be positively misleading. And despite the de-

velopment of methods to correct for this, in our reading of the empirical literature, we have found this to be a common oversight. We have also demonstrated that misspecifying the model of trait evolution when conducting pPCA may influence the inferences we make from the pPC scores. We show that in some scenarios, pPCA may sort traits according to the particular evolutionary models they follow in a similar manner as standard PCA — ignoring phylogeny altogether is, of course, a form of model misspecification. Consequently, we caution that the use of pPCA may bias inference toward identifying particular evolutionary patterns, which may or may not be representative of the true multivariate process shaping trait diversification as a whole.

We hope that our paper provokes discussion about how we should go about analyzing multivariate comparative data. We certainly do not have the answers but argue there are some major theoretical limitations inherent in using PCA (phylogenetic or not) to study macroevolutionary patterns and processes.

Acknowledgements

We would like to thank our advisor, Luke Harmon, for encouraging us to pursue this project and for providing insightful comments on the work and manuscript. We thank Luke Mahler for providing data for the *Anolis* empirical example. JCU was supported by NSF DEB 1208912 and DBI 0939454. DSC was supported by a fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES — 1093/12–6). MWP was supported by a NSERC postgraduate fellowship.

References

Adams, D. C. 2014a. A generalized k statistic for estimating phylogenetic signal from shape and other high-dimentional multivariate data. Systematic Biology DOI:10.1093/sysbio/syu030.

Adams, D. C. 2014b. Quantifying and comparing phylogenetic evolutionary rates for shape and other high-dimensional phenotypic data. Systematic Biology 63:166–177.

Arnold, S., M. Pfrender, and A. Jones. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. Genetica 112:9–32.

Arnold, S. J., R. Brger, P. A. Hohenlohe, B. C. Ajie, and A. G. Jones. 2008. Understanding the evolution and stability of the G-matrix. Evolution 62:2451–2461.

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

Bergmann, P. J. and D. J. Irschick. 2012. Vertebral evolution and the diversification of squamate reptiles. Evolution 66:1044–1058.

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

Bolstad, G. H., T. F. Hansen, C. Pélabon, M. Falahati-Anbaran, R. Pérez-Barrales, and W. S. Armbruster. 2014. Genetic constraints predict evolutionary divergence in dalechampia blossoms. Philosophical Transactions of the Royal Society B 369:1–15.

Bookstein, F. L. 1997. Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge University Press.

Bookstein, F. L. 2012. Random walk as a null model for high-dimensional morphometrics of fossil series: geometrical considerations. Paleobiology 39:52–74.

Butler, M. A. and A. A. King. 2004. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. The American Naturalist 164:683–695.

Cadima, J. F. and I. T. Jolliffe. 2001. Variable selection and the interpretation of principal subspaces. Journal of Agricultural, Biological, and Environmental Statistics 6:62–79.

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza. 1994. The History and Geography of Human Genes. Princeton University Press.

Dornburg, A., B. Sidlauskas, F. Santini, L. Sorenson, T. J. Near, and M. E. Alfaro. 2011. The influence of an innovative locomotor strategy on the phenotypic diversification of triggerfish (family: Ballistidae). Evolution 65:1912–1926.

Edwards, A. W. F. and L. L. Cavalli-Sforza. 1964. Reconstruction of evolutionary trees. Pages 67–76 *in* Phenetic and phylogenetic classification (W. Heywood and J. McNeill, eds.). Systematics Association, London.

Felsenstein, J. 1985. Phylogenies and the comparative method. The American Naturalist 125:1–15.

Freckleton, R. P. and P. H. Harvey. 2006. Detecting non-brownian trait evolution in adaptive radiations. PLoS Biol 4:e373.

Grafen, A. 1989. The phylogenetic regression. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 326:119–157.

Guillaume, F. and M. C. Whitlock. 2007. Effects of migration on the genetic covariance matrix. Evolution 61:2398–2409.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351.

Hansen, T. F. and K. Bartoszek. 2012. Interpreting the evolutionary regression: The interplay between observational and biological errors in phylogenetic comparative studies. Systematic Biology 61:413–425.

Hansen, T. F. and D. Houle. 2008. Measuring and comparing evolvability and constraint in multivariate characters. Journal of Evolutionary Biology 21:1201–1219.

Hansen, T. F. and S. H. Orzack. 2005. Assessing current adaptation and phylogenetic inertia as explanations of trait evolution: the need for controlled comparisons. Evolution 59:2063–2072.

Hansen, T. F., J. Pienaar, and S. H. Orzack. 2008. A comparative method for stuyding adaptation to a randomly evolving environment. Evolution 62:1965–1977.

Harmon, L. J., J. B. Losos, T. Jonathan Davies, R. G. Gillespie, J. L. Gittleman, W. Bryan Jennings, K. H. Kozak, M. A. McPeek, F. Moreno-Roark, T. J. Near, A. Purvis, R. E. Ricklefs, D. Schluter, J. A. Schulte II, O. Seehausen, B. L. Sidlauskas, O. Torres-Carvajal, J. T. Weir, and A. Ø. Mooers. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64:2385–2396.

Harmon, L. J., J. A. Schulte, A. Larson, and J. B. Losos. 2003. Tempo and mode of evolutionary radiation in iguanian lizards. Science 301:961–964.

Harvey, P. H. and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press.

Hausman, R. 1982. Constrained multivariate analysis. Optimisation in Statistics Pages 137–151.

Hether, T. D. and P. A. Hohenlohe. 2014. Genetic regulatory network motifs constrain adaptation through curvature in the landscape of mutational (co)variance. Evolution 68:950–964.

Ho, L. S. T. and C. Ané. 2014. A linear-time algorithm for gaussian and non-gaussian trait evolution models. Systematic biology 63:397–408.

Hohenlohe, P. A. and S. J. Arnold. 2008. Mipod: A hypothesistesting framework for microevolutionary inference from patterns of divergence. The American Naturalist 171:366–385.

Houle, D., C. Pélabon, G. P. Wagner, and T. F. Hansen. 2011. Measurement and meaning in biology. The Quarterly Review of Biology 86:3–34.

Hunt, G. 2013. Testing the link between phenotypic evolution and speciation: an integrated palaeontological and phylogenetic analysis. Methods in Ecology and Evolution 4:714–723.

Jolliffe, I. 2002. Principal component analysis. Springer.

Jones, A. G., S. J. Arnold, and R. Bürger. 2007. The mutation matrix and the evolution of evolvability. Evolution 61:727–745.

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

Kozak, K. H. and J. J. Wiens. 2010. Accelerated rates of climatic-niche evolution underlie rapid species diversification. Ecology Letters 13:1378–1389.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33:402–416.

Lynch, M. and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer.

Mahler, D. L., L. J. Revell, R. E. Glor, and J. B. Losos. 2010. Ecological opportunity and the rate of morphological evolution in the diversification of greater antillean anoles. Evolution 64:2731–2745.

Novembre, J. and M. Stephens. 2008. Interpreting principal componenet analyses of spatial population genetic variation. Nature Genetics 40:646–649.

Nyakatura, K. and O. R. Bininda-Emonds. 2012. Updating the evolutionary history of carnivora (mammalia): a new species-level supertree complete with divergence time estimates. BMC biology 10:12.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Pearson, K. 1903. Mathematical contributions to the theory of evolution. xi. on the influence of natural selection on the variability and correlation of organs. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 200:1–66.

Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E. Alfaro, and L. J. Harmon. in press. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics DOI: 10.1093/bioinformatics/btu181.

Pennell, M. W., R. G. FitzJohn, W. K. Cornwell, and L. J. Harmon. 2014. Model adequacy and the macroevolution of angiosperm functional traits. *BioRxiv*. DOI: http://dx.doi.org/10.1101/004002.

Pennell, M. W. and L. J. Harmon. 2013. An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. Annals of the New York Academy of Sciences 1289:90–105.

Pennell, M. W., B. A. J. Sarver, and L. J. Harmon. 2012. Trees of unusual size: Biased inference of early bursts from large molecular phylogenies. PLoS ONE 7:e43348.

Pienaar, J., A. Ilany, E. Geffen, and Y. Yom-Tov. 2013. Macroevolution of life-history traits in passerine birds: adaptation and phylogenetic inertia. Ecology Letters 16:571–576.

Podani, J. and I. Miklós. 2002. Resemblance coefficients and the horseshoe effect in pricipal coordinates analysis. Ecology 83:3331–3343.

Polly, D. P., A. M. Lawing, A. Fabre, and A. Goswami. 2013. Phylogenetic prinicpal components analysis and geometric morphometrics. Hystrix 24:33–41.

Price, T. D., D. M. Hooper, C. D. Buchanan, U. S. Johansson, D. T. Tietze, P. Alström, U. Olsson, M. Ghosh-Harihar, F. Ishtiaq, J. Gupta, Martens, B. Harr, P. Singh, and D. Mohan. 2014. Niche filling slows the diversification of himalayan songbirds. Nature DOI:10.1038/nature13272.

R Development Core Team. 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, Austria ISBN 3-900051-07-0.

Revell, L. J. 2009. Size-correction and principal components for interspecific comparative studies. Evolution 63:3258–3268.

Revell, L. J. 2012. phytools: an r package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217–223.

Revell, L. J. and D. C. Collar. 2009. Phylogenetic analysis of the evolutionary correlation using likelihood. Evolution 64:1090–1100.

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

Revell, L. J. and A. S. Harrison. 2008. Pcca: a program for phylogenetic canonical correlation analysis. Bioinformatics 24:1018–1020.

Richman, M. B. 1986. Rotation of principal components. Journal of Climatology 6:293–335.

Rohlf, F. J. and D. Slice. 1990. Extensions of the procrustes method for the optimal superimposition of landmarks. Systematic Zoology 39:40–59.

Sakamoto, M., G. T. Lloyd, and M. J. Benton. 2010. Phylogenetically structured variance in felid bite force: the role of phylogeny in the evolution of biting performance. Journal of Evolutionary Biology 23:463–478.

Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50:1766–1774.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press.

Schmitz, L. and R. Motani. 2011. Nocturnality in dinosaurs inferred from scleral ring and orbit morphology. Science 332:705–708.

Schnitzler, J., C. H. Graham, C. F. Dormann, K. Schiffers, and H. Peter Linder. 2012. Climatic niche evolution and species diversification in the cape flora, south africa. Journal of Biogeography 39:2201–2211.

Sidlauskas, B. 2008. Continuous and arrested morphological diversification in sister clades of characiform fishes: a phylomorphospace approach. Evolution 62:3135–3156.

Slater, G. J. and M. W. Pennell. 2014. Robust regression and posterior predictive simulation increase power to detect early bursts of trait evolution. Systematic Biology 63:293–308.

Slater, G. J. and B. Van Valkenburgh. 2009. Allometry and performance: the evolution of skull form and function in felids. Journal of Evolutionary Biology 22:2278–2287.

Somers, K. M. 1986. Multivariate allometry and removal of size with principal components analysis. Systematic Biology 35:359–368.

Somers, K. M. 1989. Allometry, isometry and shape in principal components analysis. Systematic Biology 38:169–173.

Stadler, T. 2011. Simulating trees with a fixed number of extant species. Systematic Biology 60:676–684.

Thomas, G. H. and R. P. Freckleton. 2012. Motmot: models of trait macroevolution on trees. Methods in Ecology and Evolution 3:145–151.

Vines, S. 2000. Simple principal components. Journal of the Royal Statistical Society: Series C (Applied Statistics) 49:441–451.

Voje, K. L., A. B. Mazzarella, T. F. Hansen, K. Østbye, T. Klepaker, A. Bass, A. Herland, K. M. Bærum, F. Gregersen, and L. A. Vøllestad. 2013. Adaptation and constraint in a stickleback radiation. Journal of Evolutionary Biology 26:2396–2414.

Weir, J. T. and S. Mursleen. 2013. Diversity-dependent cladogenesis and trait evolution in the adaptive radiation of the auks (aves: Alcidae). Evolution 67:403–416.

Zou, H., T. Hastie, and R. Tibshirani. 2006. Sparse principal component analysis. Journal of computational and graphical statistics 15:265–286.

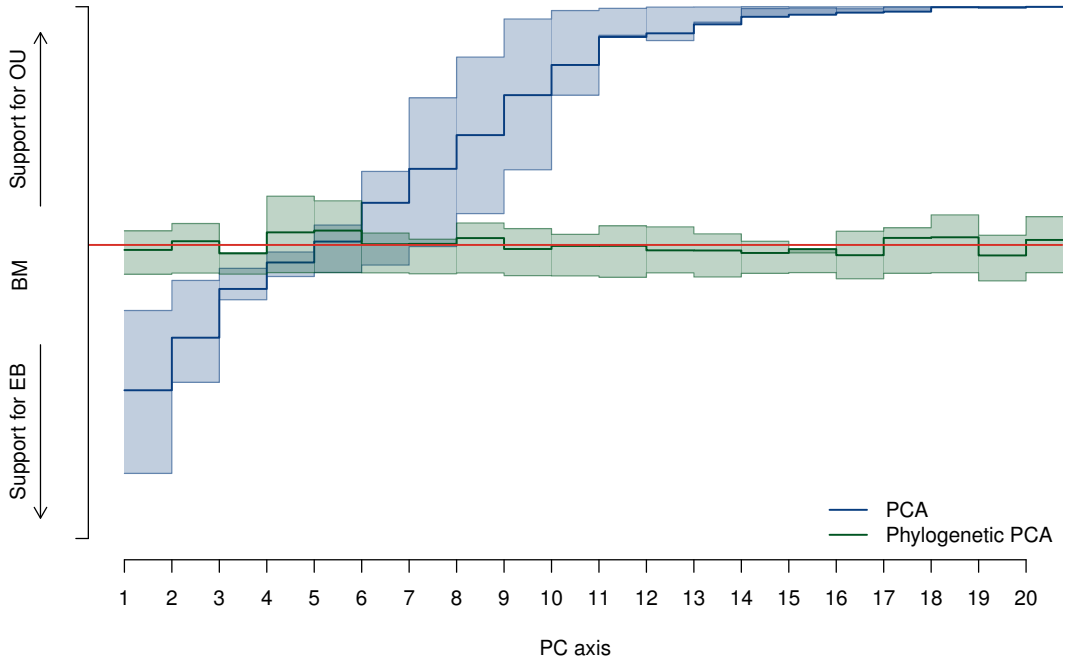

Figure 1: Distribution of support for BM, OU and EB models when the generating model is a correlated multivariate BM model. Support for models were transformed onto a linear scale by calculating an overall model support statistic: $AICw_{OU} - AICw_{EB}$. Thus high values support OU, low values support EB, and intermediate values near 0 indicate BM-like evolution. Models were fit to each replicated dataset for each of 20 different traits which were taken either from PC scores (blue line) or phylogenetic PC scores (green line). Shaded regions indicate the 25^{th} and 75^{th} quantiles of the mode–support statistic for 100 replicated datasets. The red line indicates the average model support statistic averaged over all 20 original trait variables. Note that EB models have higher Akaike weights for the first few PCs of standard PCA, and that later PCs subsequently favor BM and finally, OU models. No such bias is found across traits for either the original data or pPCA.

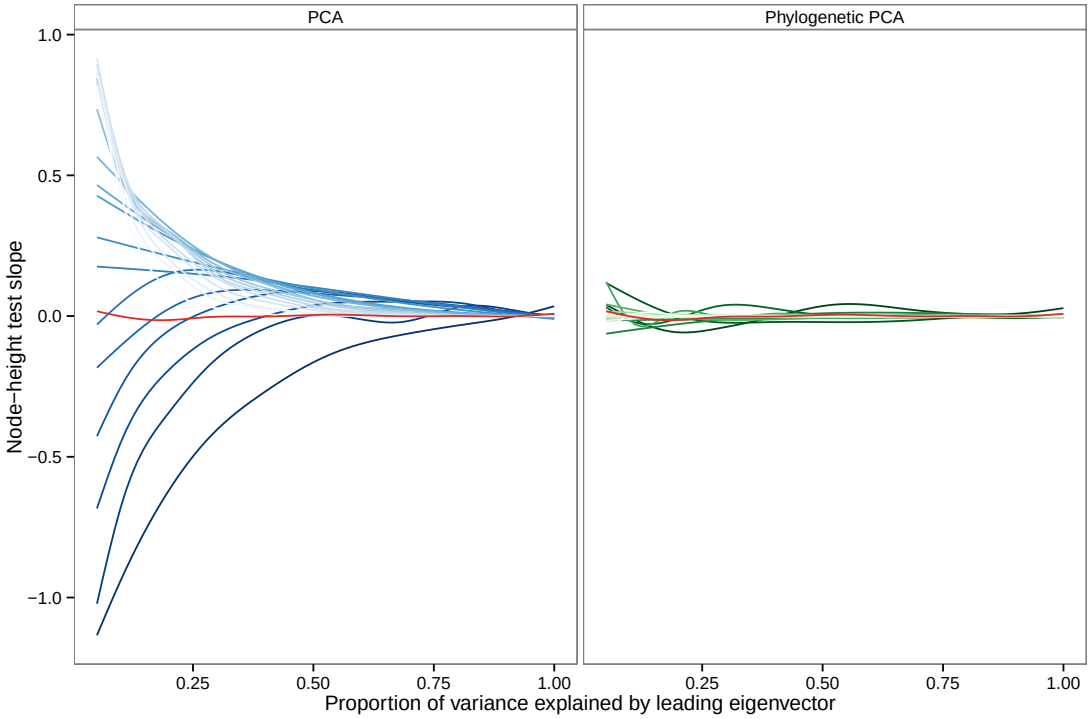

Figure 2: Effect of trait correlations on the slope of the node height test for PC scores (left) and pPC scores (right) under a multivariate BM model of evolution. The red line is the aggregated data for all 20 traits on the original (untransformed) scale. The intensity of the colors are proportional to the ranking of the PC or pPC axes, stronger lines represent the first axes. When the leading eigenvector explains very little variation in the data and the effective dimensionality is high, the slope of node height test increases from negative to positive across PC axes. This indicates that under standard PCA, PC1 has higher contrasts near the root of the tree, while later PCs have higher contrasts near the tips (resulting in the pattern of model support observed in Figure 1). As the amount of variance explained by the principal eigenvector increases, the slope of the node height test approaches 0. No such effect is found for phylogenetic PCA.

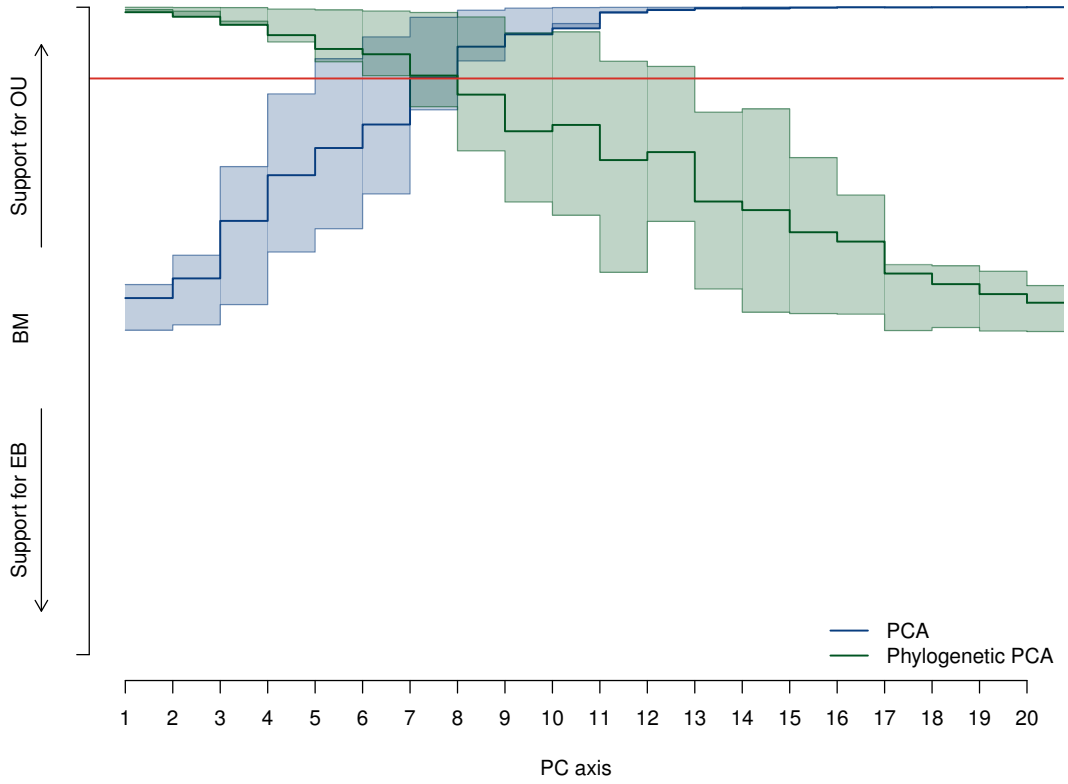

Figure 3: Distribution of support for BM, OU and EB models when the generating model is a uncorrelated multivariate OU model. Support for models were transformed into a linear scale by calculating an overall model support statistic: $AICw_{OU} - AICw_{EB}$. Thus high values support OU, low values support EB, and intermediate values near 0 indicate BM-like evolution. Models were fit to each replicated dataset for each of 20 different traits which were taken either from PC scores (blue line) or phylogenetic PC scores (green line). Shaded regions indicate the 25^{th} and 75^{th} quantiles of the mode–support statistic for 100 replicated datasets. The red line indicates the average model support statistic averaged over all 20 original trait variables. Note that EB models have higher Akaike weights for the first few PCs of standard PCA, and that later PCs subsequently favor BM and finally, OU models. No such bias is found across traits for either the original data or pPCA.

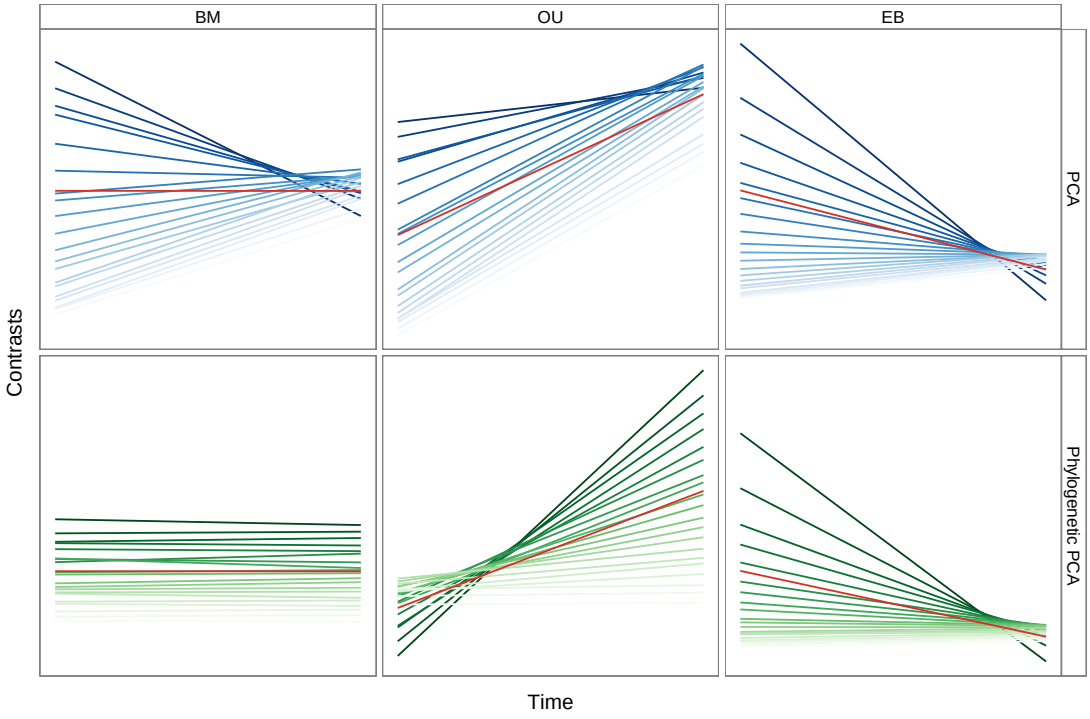

Figure 4: Relationship between the average phylogenetic independent contrasts and the height of the node across 100 datasets simulated under either a BM (left), OU (middle) or EB (right) model of evolution. Contrasts were calculated for each of the 20 traits corresponding to either PC scores (top row) or pPPC scores (bottom row). Each line represents a best–fit linear model to the aggregated data across all 100 replicate simulations. Red lines are aggregated over all 20 traits on the original data. The plots are oriented so that the left side of each panel corresponds to the root of the phylogeny, with time increasing tipward to the right. The intensity of the colors are proportional to the ranking of the PC or pPC axes, stronger lines represent the first axes. PCA results in a predictable pattern of increasing slope in the contrasts across PCs. By contrast, pPCA only has systematic distortions across pPC axes when the underlying model is not multivariate BM. When this occurs, the first few pPC axes tend to have more extreme slopes than the original data (but in the correct direction).

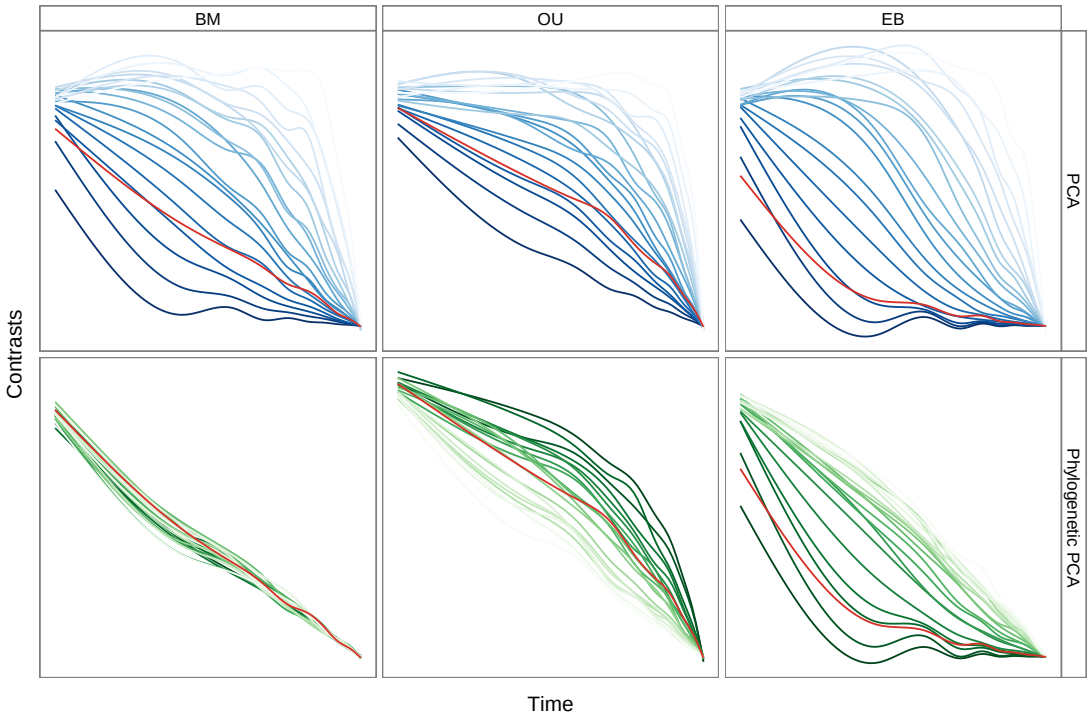

Figure 5: Disparity through time plots averaged across the 100 simulated datasets. The datasets were simulated under BM (left), OU (middle) or EB (right). The analyses were then performed on PC scores (top row) and pPPC scores (bottom row). The average disparity through time of all 20 original trait variables is indicated by the red line. We fit a loess curve through the relative disparities for each trait/transformation/model combination. The plots are oriented so that the left side of each panel corresponds to the root of the phylogeny, with time increasing tipward to the right. The intensity of the colors are proportional to the ranking of the PC or pPC axes, stronger lines represent the first axes. As in Fig. 4, the first few axes from the PCA show a strong pattern of high disparity early in the clades' histories with the higher components showing seemingly higher disparity towards the present. PPCA corrects the distortion if the generating model is multivariate BM. However, if the generating model was not BM, the first few pPC axes tend to show an exaggerated pattern of disparity relative to the original traits.

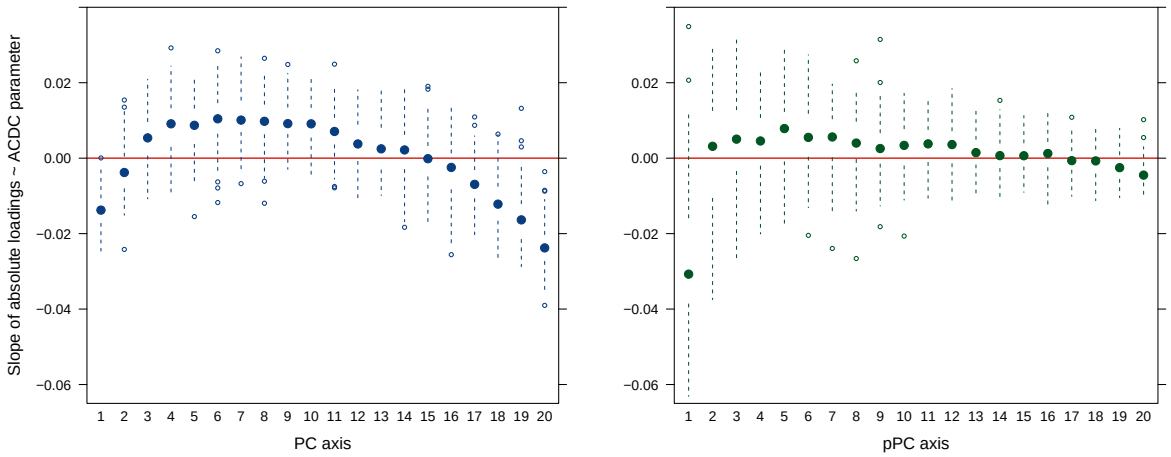

Figure 6: Relationship between factor loadings and ACDC parameter across PCA (left) and pPCA (right) across 100 simulated 20 trait datasets with ACDC parameters drawn from a Normal distribution with mean = 0 and sd = 5. Boxplots indicate the distribution of the slope of a linear model describing the relationship between the absolute factor loadings for a given PC and the magnitude of the ACDC parameter. A negative slope indicates that traits with decelerating rates of evolution tend to have higher loadings in that particular PC. Conversely, positive slopes indicate that traits with accelerating rates tend to have higher loadings.

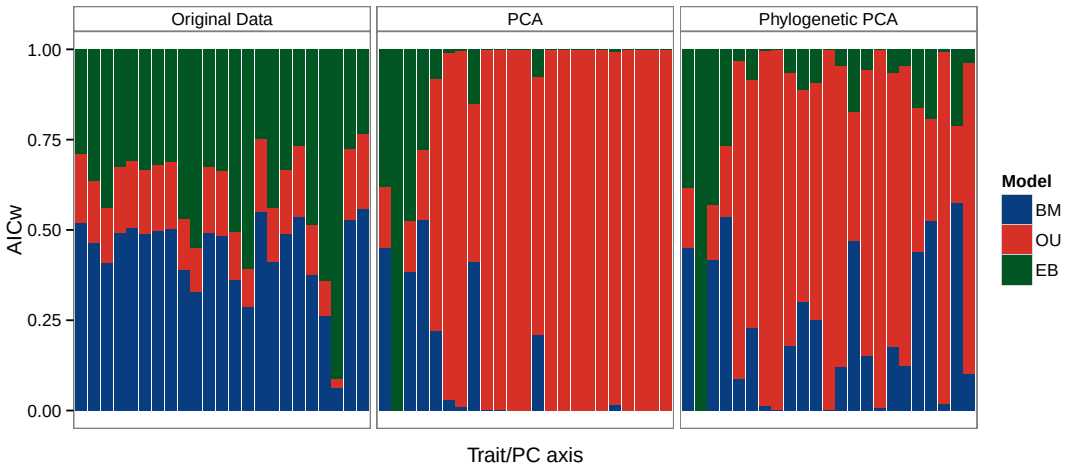

Figure 7: Distribution of support for BM, OU and EB models for a 23–trait morphometric dataset taken from Mahler et al. (2010). Support is measured in Akaike weights across all original trait variables (left), as well as standard PCA (middle) and pPCA (right). For both PCA and pPC, support for the EB model appears to be concentrated in PCs 1-4, with a suggestive pattern of decreasing support across PCs 2-4.

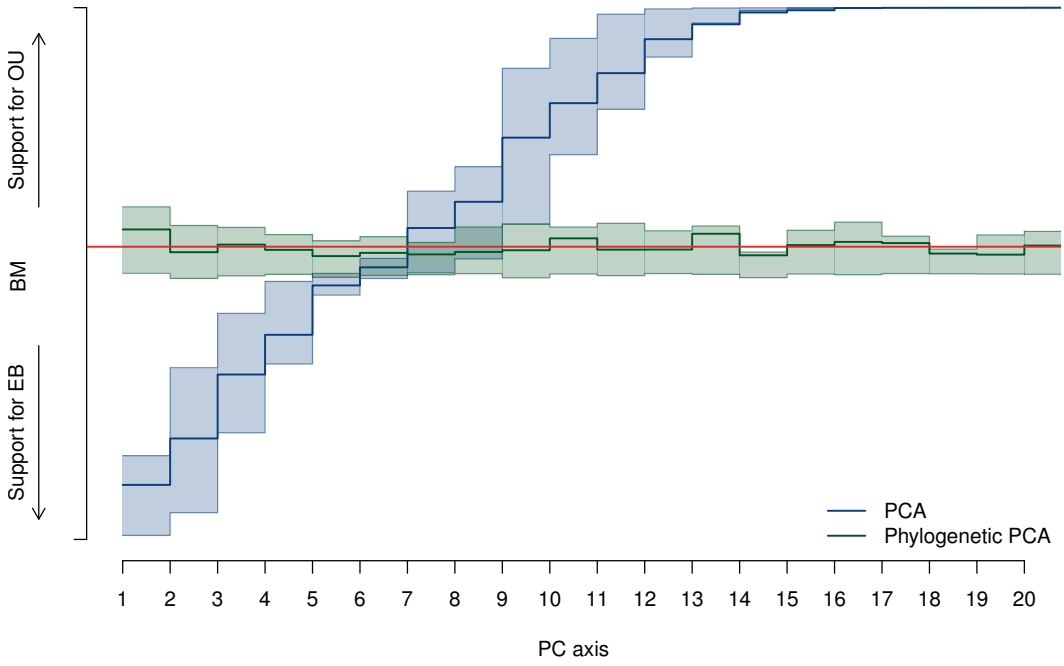

Figure S1: Distribution of support for BM, OU and EB models when the generating model is an uncorrelated multivariate BM model. Support for models were transformed into a linear scale by calculating an overall model support statistic: $AICw_{OU} - AICw_{EB}$. Thus high values support OU, low values support EB, and intermediate values near 0 indicate BM-like evolution. Models were fit to each replicated dataset for each of 20 different traits which were taken either from PC scores (blue line) or phylogenetic PC scores (green line). Shaded regions indicate the 25^{th} and 75^{th} quantiles of the mode–support statistic for 100 replicated datasets. The red line indicates the average model support statistic averaged over all 20 original trait variables. Note that EB models have higher Akaike weights for the first few PCs of standard PCA, and that later PCs subsequently favor BM and finally, OU models. No such bias is found across traits for either the original data or pPCA.

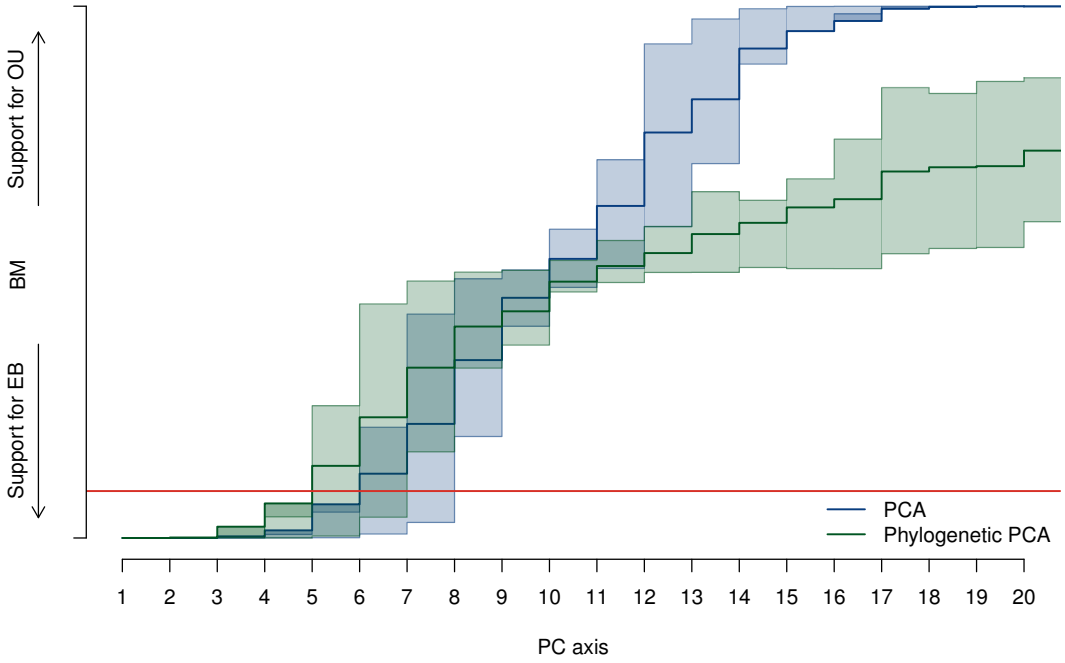

Figure S2: Distribution of support for BM, OU and EB models when the generating model is an uncorrelated multivariate EB model. Support for models were transformed into a linear scale by calculating an overall model support statistic: $AICw_{OU} - AICw_{EB}$. Thus high values support OU, low values support EB, and intermediate values near 0 indicate BM-like evolution. Models were fit to each replicated dataset for each of 20 different traits which were taken either from PC scores (blue line) or phylogenetic PC scores (green line). Shaded regions indicate the 25^{th} and 75^{th} quantiles of the model–support statistic for 100 replicated datasets. The red line indicates the average model support statistic averaged over all 20 original trait variables. Note that EB models have higher Akaike weights for the first few PCs of standard PCA, and that later PCs subsequently favor BM and finally, OU models. No such bias is found across traits for either the original data or pPCA.

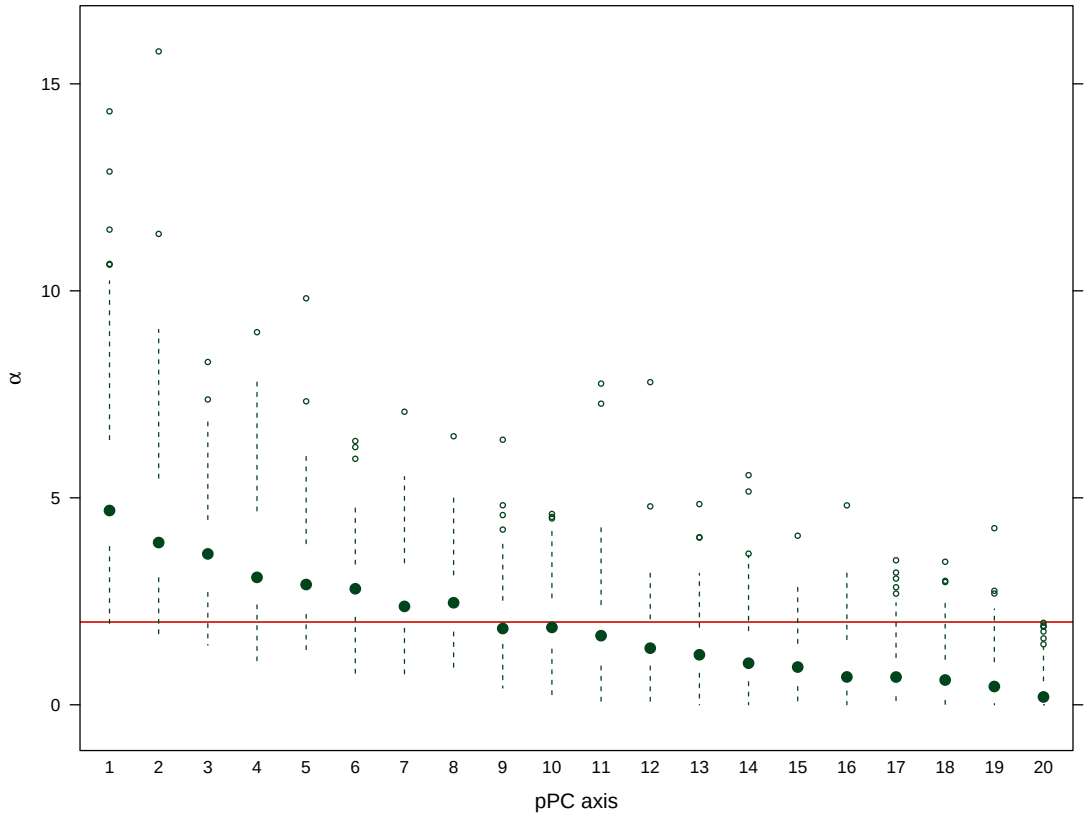

Figure S3: Estimated values of the α parameter from phylogenetic PCA when data is simulated under an uncorrelated multivariate OU model. The simulating value $\alpha = 2$ is depicted with the red line. The estimate of α is inflated in the first few pPC axes consistent with an exaggerated support for the OU model. In the last pPC axes, α is estimated to be very close to 0, such that the OU model is statistically indistinguishable from a BM model. These results mirror those depicted in Figure 3.

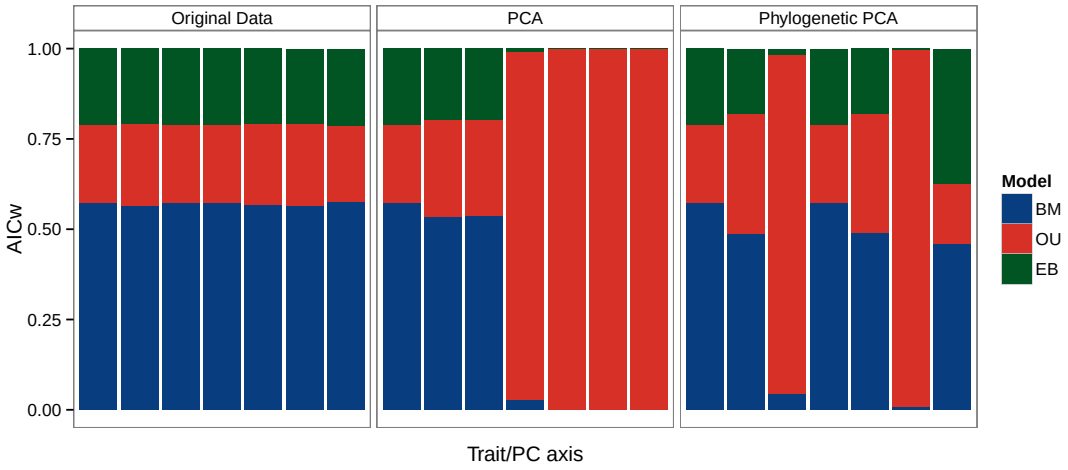

Figure S4: Proportion of support for BM, OU and EB models for each of the traits/PC axes from the morphological dataset of Felidae species. Traits were log transformed prior to analysis. Note that all original traits and the first axes under standard and phylogenetic PCA show strong support for a BM model.

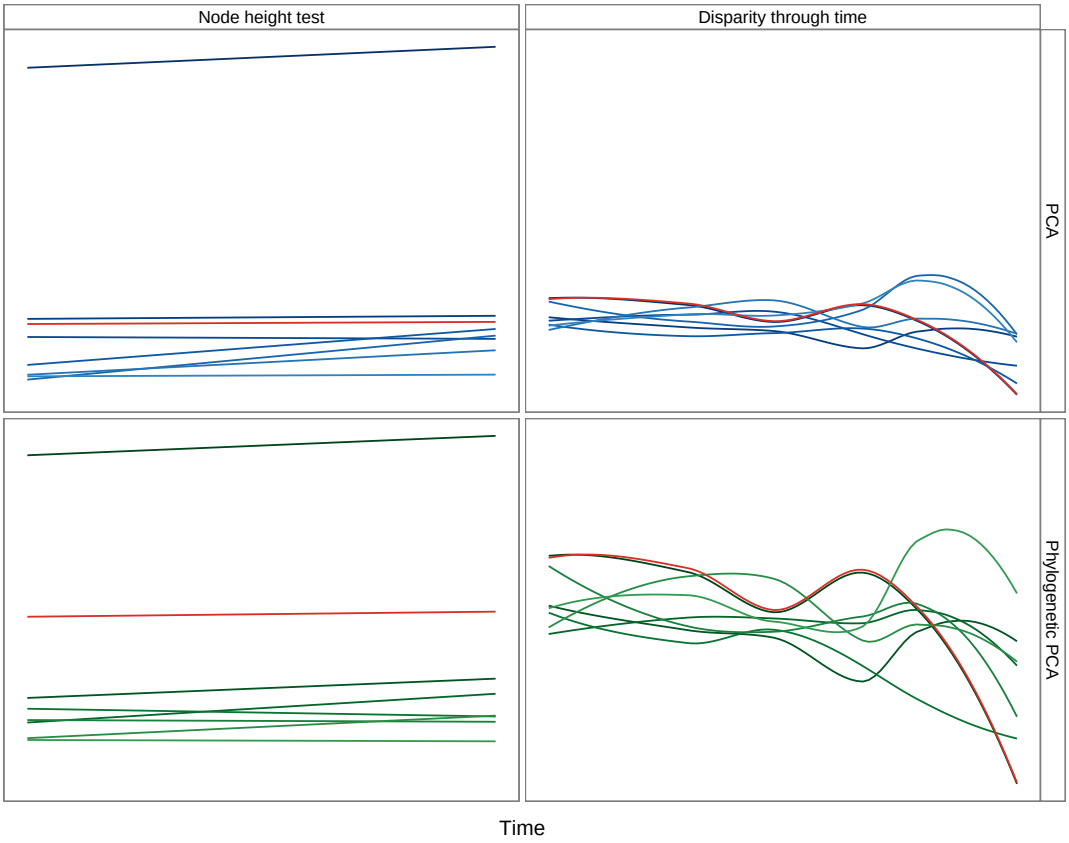

Figure S5: Node height test and disparity through time plots for the morphological dataset of Felidae species. Each line represents a best–fit linear model or loess curve fitted to the original traits, PC or pPC scores. All traits were log transformed prior to analysis. The intensity of color is proportional to the ranking of the PC or pPC axes, stronger lines represent the first axes. Order of original traits is arbitrary. Top panels show the relationship between the average phylogenetic independent contrasts and the height of the node. Bottom panels show disparity through time plots. The plots are oriented so that the left side of each panel corresponds to the root of the phylogeny, with time increasing tipward to the right. Compare this highly correlated dataset with only 7 traits to the larger, less correlated dataset of *Anolis* lizards (Figure S6).

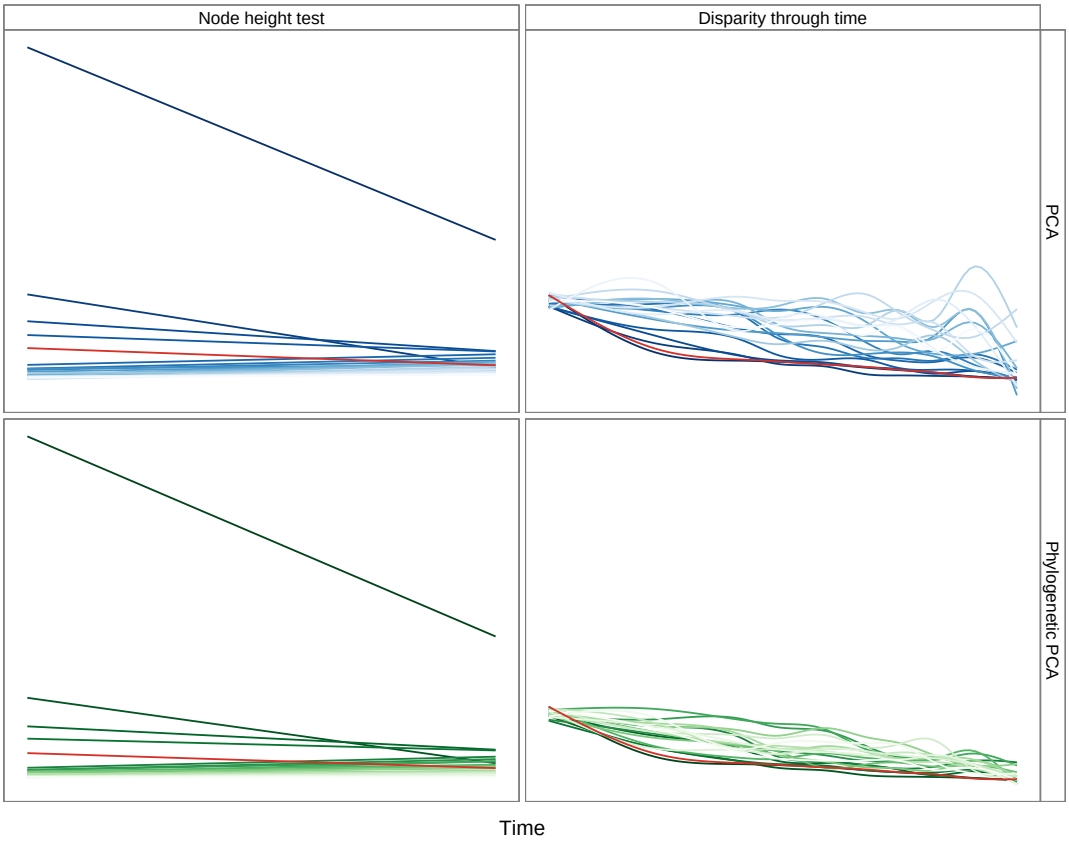

Figure S6: Node height test and disparity through time plots for the morphological dataset of *Anolis* lizards. Each line represents a best–fit linear model or loess curve fitted to the original traits, PC or pPC scores. All traits were log transformed prior to analysis. The intensity of color is proportional to the ranking of the PC or pPC axes, stronger lines represent the first axes. Order of original traits is arbitrary. Top panels show the relationship between the average phylogenetic independent contrasts and the height of the node. Bottom panels show disparity through time plots. The plots are oriented so that the left side of each panel corresponds to the root of the phylogeny, with time increasing tipward to the right.