doi:10.1111/evo.13864

# Measuring the magnitude of morphological integration: The effect of differences in morphometric representations and the inclusion of size

**Fabio A. Machado,[1,2,3,4] Alex Hubbe,[5] Diogo Melo,[6] Arthur Porto,[7] and Gabriel Marroig[6]**

[1]*Department of Biology, University of Massachusetts, Boston, Massachusetts 02125*

[2]*División Mastozoología, Museo Argentino de Ciencias Naturales "Bernardino Rivadavia," Av. Ángel Gallardo 470 (C1405DJR) Buenos Aires, Argentina*

[3]*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina*

[4]*E-mail: macfabio@gmail.com*

[5]*Departamento de Oceanografia, Instituto de Geociências, Universidade Federal da Bahia R. Barão de Jeremoabo, S/N - Ondina Salvador, Bahia 40170-110, Brazil*

[6]*Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Rua do Matão, 277 Universidade de São Paulo, São Paulo São Paulo 05508-090, Brazil*

[7]*Department of Biosciences, Centre for Ecological and Evolutionary Synthesis (CEES), University of Oslo, 0315, Oslo Norway*

Received March 4, 2019
Accepted October 3, 2019

**The magnitude of morphological integration is a major aspect of multivariate evolution, providing a simple measure of the intensity of association between morphological traits. Studies concerned with morphological integration usually translate phenotypes into morphometric representations to quantify how different morphological elements covary. Geometric and classic morphometric representations translate biological form in different ways, raising the question if magnitudes of morphological integration estimates obtained from different morphometric representations are compatible. Here we sought to answer this question using the relative eigenvalue variance of the covariance matrix obtained for both geometric and classical representations of empirical and simulated datasets. We quantified the magnitude of morphological integration for both shape and form and compared results between representations. Furthermore, we compared integration values between shape and form to evaluate the effect of the inclusion or not of size on the quantification of the magnitude of morphological integration. Results show that the choice of morphological representation has significant impact on the integration magnitude estimate, either for shape or form. Despite this, ordination of the integration values within representations is relatively the same, allowing for similar conclusions to be reached using different methods. However, the inclusion of size in the dataset significantly changes the estimates of magnitude of morphological integration, hindering the comparison of this statistic obtained from different spaces. Morphometricians should be aware of these differences and must consider how biological hypothesis translate into predictions about integration in each particular choice of representation.**

**KEY WORDS:** **Covariance matrix, canidae, eigenvalue variance, P matrix, skull.**

Morphological integration describes the association between continuous morphological traits and is a key component in understanding multivariate evolution (Lande 1979). We expect functionally and/or developmentally related traits to be more associated among themselves than with others and, consequently, to evolve in a coordinated fashion (Riedl 1978; Cheverud 1984). A fundamental aspect of morphological integration is the magnitude of morphological integration (sensu Olson and Miller 1958), which measures the overall intensity of the association between traits. A high magnitude of morphological integration means that the available variation is restricted to relatively few dimensions in phenotypic space. In these cases, evolutionary change will be strongly influenced by the interaction between selection and available variation, and we expect evolutionary change to proceed preferentially along these few dimensions in which variation is available (Felsenstein 1988; Goswami et al. 2014; Melo et al. 2016; Felice et al. 2018, but see Schluter 1996). Conversely, lower magnitudes of morphological integration might indicate relative independence between the traits, allowing different aspects of the phenotype to evolve without interference imposed by other parts of the organism. Thus, populations with similar covariation patterns but with different magnitudes of morphological integration can present different evolutionary responses when subjected to the same selective pressure (Lande 1979; Hansen and Houle 2008; Pavlicev et al. 2009).

In the last few decades, with the advent of modern morphometric techniques, several authors have measured the magnitude of morphological integration in multivariate phenotypes using geometric morphometrics (GM). At the same time, morphometric studies based on classic linear measures, such as interlandmark distances (ILDs), are still popular, making up a significant portion of all papers published on the topic (Esteve-Altava 2017). Using different morphometric representations of the same phenotype goes beyond a mere stylistic choice because using different representation affects how biological processes are accounted for. Take size, for example. GM representations usually involve a size-scaling procedure, which leads many researchers to put aside isometric variation in the investigation of morphological integration, focusing on the morphological integration in shape alone (e.g., Jamniczky and Hallgrímsson 2009; Goswami et al. 2015; Curth et al. 2017; Randau et al. 2019). ILD analyses, in turn, measure traits on a ratio scale (sensu Houle et al. 2011), which leads to isometric size variation being embedded in the variation of the traits, thus leading to the joint evaluation of the overall form (size plus shape) of biological structures (e.g., Meiri et al. 2005; Young et al. 2010; Haber 2015). Furthermore, different morphometric representations can vary in how they quantify shape changes, which can lead to incompatible interpretations regarding magnitude of morphological integration. For example, applying purely homoscedastic (and noncorrelated) error on landmark data can lead to nonzero correlations among ILD data drawn from the same configurations (Rohlf 2000). Conversely, the application of affine, global deformations on landmark configurations will not necessarily lead to equally coordinated changes on ILD variables (Mitteroecker and Bookstein 2007). Nevertheless, if we expect to make general statements about the evolutionary properties of integrated structures, we need to understand how measures of the magnitude of morphological integration might differ between these popular approaches and if and when we can directly compare them.

Here, we investigate if measures of magnitudes of morphological integration obtained using different morphometric representations can be directly compared. We approach this objective in two ways. First, using simulations we alter a fixed GM covariance matrix to produce matrices with different magnitudes of morphological integration. From these altered covariance matrices, we sampled sets of shapes and sizes and calculated the magnitude of morphological integration for both GM and a set of linear distances. Second, we validate these comparisons by evaluating magnitudes of morphological integration on the skull of a diverse group of carnivores.

## Material and Methods

### SAMPLE AND MORPHOMETRICS

Morphometric data were obtained from 3440 skulls of adult specimens from 67 species of Carnivora from the Caniformes suborder, which comprises very distinct forms such as wolf (Canidae), walrus (Odobenidae), bear (Ursidae), badger (Mustelidae), and raccoon (Procyonidae). In addition to morphological variation, the family Canidae was recently found to have a diverging pattern of morphological integration when compared to other carnivoran families (Machado et al. 2018). Therefore, this sample comprises not only considerable shape disparity but also differences in patterns of association among cranial structures. For each specimen, a set of 32 landmarks (8 mid-line and 12 symmetric) was measured by the first author using a Microscribe MLX system (Immersion Corporation, San Jose, CA). To produce GM shape variables the samples were subjected to a generalized Procrustes alignment (GPA, Rohlf and Slice 1990). Because GPA scales configurations to have the same size, data were evaluated both with and without size by removing or including the logarithm of the centroid size as an additional variable (Mitteroecker et al. 2004).

For the ILD analysis, we obtained a set of 35 linear ILDs. These distances have been used on a large set of evolutionary studies (e.g., Marroig and Cheverud 2004; Assis et al. 2016; Hubbe et al. 2016; Porto et al. 2016; Machado et al. 2018), and are particularly interesting because not only they measure specific localized dimensions of the skull bones and structures (Pearson and Davin 1924), but also because the covariance matrix obtained on

phenotypic data alone has been shown to be an accurate approximation of genetic patterns of covariation among the same traits (Garcia et al. 2014; Porto et al. 2015; Penna et al. 2017). We also calculated the magnitude of integration values using an Euclidean distance matrix approach (EDMA) in which all possible ILDs are calculated. The relationship between integration measures obtained for 35 ILDs and EDMA was slightly nonlinear and highly correlated (Spearman rank order correlation, $r_s = 0.965$). As the use of both ILD datasets provide the same general patterns, we focus on the results of 35 ILDs, henceforth called ILD for simplicity.

Following the analysis of the GM dataset, ILDs were analyzed both with and without the influence of size. To control the effect of size, we obtained the distances on the centroid-scaled configurations produced by GPA. This is equivalent to using Mosimann's shape ratios (Mosimann 1970), where each variable (ILD) is divided by a size factor (centroid size). Neither the GM nor ILD "size-corrected" data are free of shape changes associated with size (i.e., allometry). They are merely scaled to have a common size, thus reflecting variation on allometric and nonallometric shape changes. Henceforth, the analysis without isometric size will be referred to as "Shape" and the ones with isometric size as "Form."

To evaluate if different morphometric methods contain the same information of the biological variation on the sample, we performed a Procrustes correlation test (Peres-Neto and Jackson 2001). This test evaluates the distribution of observations on two multivariate spaces by applying the Procrustes transformations (translation, scale, and rotation) to reduce the residual sum of squares between the same observations on both spaces. A correlation-like statistic can be obtained as the square root of 1 minus the residual sum of squares, with values closer to 1 representing a closer match between spaces. Tests were performed for each species, evaluating the empirical distribution of values on the first 35 principal components of each sample (Peres-Neto and Jackson 2001). Average correlations were 0.939 (SD = 0.014) for Form and a 0.904 (SD = 0.012) for Shape, never reaching values lower than 0.878 for any comparison. These high correlations suggest that both ILD and GM are quantifying the same general patterns of morphological variation in our sample.

Covariance matrices were obtained for each species individually and for the full sample. Both species matrices and the full sample matrix were calculated after controlling for intraspecific sources of variation. For individual species, these sources of nuisance variation were subspecies and sex. For the full sample, in addition to subspecies and sex, species differences were also controlled for. This control procedure was done by fitting a linear model on the data using the nuisance variables as fixed effects, and using the residuals from these regressions to produce pooled within-group phenotypic covariance matrices ($\mathbf{P}_W$). These regressions were done using the landmark and log-centroid size data. After the regression, residuals were added to the average shape and size. To obtain configurations on the original scale, each resulting configuration was multiplied by its corresponding centroid size. ILDs were then calculated on these rescaled configurations, reducing possible sources of differences between GM and ILD datasets. Further details on data processing, landmarks, measurements, and nuisance sources of variation are described elsewhere (Machado et al. 2018).

## MAGNITUDE OF MORPHOLOGICAL INTEGRATION

The magnitude of morphological integration of each matrix $\mathbf{P}_W$ was calculated as the eigenvalue variance:

$$\mathrm{var}(\lambda) = \frac{\sum_{i=1}^{N}(\lambda_i - \bar{\lambda})^2}{N}, \quad (1)$$

where $\lambda$ are the eigenvalues of $\mathbf{P}_W$ and $N$ are the number of traits (Pavlicev et al. 2009). Pavlicev et al. (2009) suggested to scale the observed eigenvalue variance by the theoretical maximum to produce values that range between 0 and 1, which is compatible with the squared correlation coefficient among traits ($r^2$). For covariance matrices, the maximum theoretical maximum is achieved when all variation is concentrated on the first principal component:

$$\begin{aligned}\mathrm{var}_{max}(\lambda) &= \frac{(tr(\mathbf{P}_W) - \bar{\lambda})^2 + \sum_{i=2}^{N}\bar{\lambda}^2}{N} \\ &= \frac{(tr(\mathbf{P}_W) - (tr(\mathbf{P}_W)/N))^2 + \sum_{i=2}^{N}(tr(\mathbf{P}_W)/N)^2}{N} \\ &= \frac{tr(\mathbf{P}_W)^2(N-1)}{N^2}, \end{aligned} \quad (2)$$

where $tr(\mathbf{P}_W)$ is the total amount of variation on the matrix $\mathbf{P}_W$. The relative eigenvalue variance is calculated then as $\mathrm{var}_{rel}(\lambda) = \mathrm{var}(\lambda)/\mathrm{var}_{max}(\lambda)$. Values closer to 0 imply that eigenvalues have similar scales, implying a lack of covariance among the original traits. If $\mathrm{var}_{rel}(\lambda)$ is closer to 1, then $\mathrm{var}(\lambda)$ is close to the theoretical maximum $\mathrm{var}_{max}(\lambda)$, with all variation concentrated along a single axis, suggesting a highly integrated structure. The use of $\mathrm{var}_{rel}(\lambda)$ is convenient in the present study because the statistic is invariant to arbitrary rotations of the data, making it ideal for the use on both GM and ILD datasets.

Even though $\mathrm{var}_{rel}(\lambda)$ is thought to be comparable between systems with different dimensionality (Pavlicev et al. 2009), we only used the leading 35 eigenvalues for all analyses (the number of ILDs chosen for the ILD analysis) to standardize the number of traits ($N = 35$). This is equivalent to analyzing the datasets after performing a principal component analysis for dimensionality reduction. Despite this reduction, values of $\mathrm{var}_{rel}(\lambda)$ obtained with the reduced number of eigenvalues and the full sets were nearly identical for both simulated and empirical datasets, suggesting

that this procedure did not alter the $var_{rel}(\lambda)$ estimates for GM data significantly.

## SIMULATION

Even though it is possible to generate completely random GM matrices (e.g., Walker 2000), it can be difficult to generate meaningful alterations on those matrices and produce biologically plausible shapes due to the arbitrary nature of the values of covariances and correlations. For this reason, simulations were based on the pooled within-group form GM covariance matrix for the full sample of our empirical dataset. In other words, this $\mathbf{P}_W$ is the average covariance matrix for the full dataset weighted by within-group degrees of freedom, which included the covariances among Procrustes-superimposed coordinates plus the logarithm of the centroid size. From this baseline matrix, we altered its magnitude of morphological integration to produce a wide range of $var_{rel}(\lambda)$ values. This was done by first performing an eigen decomposition. The first 35 eigenvalues were kept and then raised to different powers $\lambda^p$, where $p$ was drawn from 20 equally spaced values between 0.5 and 2.5. This value range for $p$ was chosen because it produced values of $var_{rel}(\lambda)$ that are comparable to the ones in our empirical sample (Machado et al. 2018), which spans the values found for all major mammalian lineages (Porto et al. 2009). To maintain the total amount of variation constant in the altered matrices, eigenvalues were scaled to have the same total amount of variation as the original first 35 eigenvalues. Modified matrices were then reconstructed as follows:

$$\mathbf{P}_W^p = \mathbf{V}\Lambda^p\mathbf{V}^t, \quad (3)$$

where $\mathbf{V}$ is a matrix of the first 35 eigenvectors of the baseline matrix and $\Lambda^p$ is a diagonal matrix with the modified eigenvalues $\lambda^p$. When $p$ is closer to 1, the resulting matrix possesses $var_{rel}(\lambda)$ that is closer to the one of the original baseline matrix. When $p$ is greater than 1, the leading eigenvalues are proportionally larger than the last ones, increasing the disparity between eigenvalues and, therefore, increasing the $var_{rel}(\lambda)$.

For each of the 20 values of $p$, we sampled 1000 populations of 100 simulated individuals using a multivariate normal distribution with the mean shape and mean log-centroid size as averages, and the corresponding $\mathbf{P}_W^p$ as the covariance matrix. Therefore, each individual in the simulated populations is defined by a set of landmark coordinates and a centroid size. For each sampled population, a new covariance matrix was calculated from Procrustes-superimposed configurations, along with its corresponding $var_{rel}(\lambda)$. Next, for each simulated individual landmark coordinates, we calculated the 35 ILDs used on the ILD analysis, and obtained the $var_{rel}(\lambda)$ from the ILD covariance matrix. Both GM and ILD $var_{rel}(\lambda)$ estimates were obtained for shape and form, as described above.

## EMPIRICAL ANALYSIS

To evaluate if the simulations are good representations of what can happen in empirical datasets, we calculated $var_{rel}(\lambda)$ values for shape and form variables on both ILD and GM representation on our set of 67 carnivore species. Before doing so, we evaluated two possible sources of bias in empirical estimates of integration: sample size (Haber 2011; Fruciano et al. 2013; Adams and Collyer 2016; Grabowski and Porto 2017) and total amount of variance (Hallgrímsson et al. 2009; Young et al. 2010).

To evaluate the effect of sample sizes on magnitude of integration, we calculated Pearson's product moment correlation ($r_p$) between integration values and sample sizes. Furthermore, we produced rarefied estimates of $var_{rel}(\lambda)$ to remove the effect of sample size (Fruciano et al. 2013). This was done by resampling the original datasets with a fixed sample of 40 (the smaller sample size in our dataset) 100 times. The $var_{rel}(\lambda)$ is calculated for each iteration, and the average $var_{rel}(\lambda)$ across all iterations is taken as the rarefied $var_{rel}(\lambda)$ for that species. Rarefied and full sample $var_{rel}(\lambda)$ values were compared through Pearson's correlation.

To evaluate the effect of amount of variance on the magnitude of integration, we calculated the $r_p$ between $var_{rel}(\lambda)$ and $tr(\mathbf{P}_W)$ for shape and form variables on both ILD and GM. However, the simple correlation between these factors might not indicate bias in itself because variance might be preferentially concentrated on PC1 due to the interaction of life-history traits and aspects of the developmental system under study (Hallgrímsson et al. 2009; Porto et al. 2013). Therefore, we also performed semipartial correlations between $tr(\mathbf{P}_W)$ and $var_{rel}(\lambda)$, with variation of $tr(\mathbf{P}_W)$ conditional on the percentage of variation on PC1.

## COMPARISON BETWEEN GM AND ILD

For the empirical dataset, values of $var_{rel}(\lambda)$ were compared between GM and ILD representations using both reduced major axis (RMA) regressions and correlation analyses using Pearson's product moment ($r_p$) and Spearman rank order ($r_s$) correlations indexes. For RMA, a slope equal to 1 and an intercept equal to 0 means that integration values are same for both morphometric representations being compared. This was done both statistically, by estimating confidence intervals for the RMA statistics, and also graphically, by plotting values of integration obtained on both representations. Comparisons were made between shape (GM × ILD) and form (GM × ILD) representations, and between shape and form for all representations (GM shape × GM form, ILD shape × ILD form, ILD shape × GM form, GM shape × ILD form). Because all shape × form comparisons yielded very similar results (not shown), we focus only on the GM shape × ILD form comparison, as these are among the most popular forms of morphological quantification in magnitude of morphological integration studies (Esteve-Altava 2017). For the correlation analysis

we report only the $r_p$, because $r_s$ produces the same overall results. For the simulated dataset, because of the presence of the strong nonlinear relationship between representations (see Results), we refrain from using the RMA and report only the $r_s$.

Additionally, to visualize the evolution of magnitude of morphological integration, we mapped the $var_{rel}(\lambda)$ values for each morphometric representation on the phylogeny of the group. Ancestral character state were estimated through maximum likelihood (Schluter et al. 1997) using the same phylogeny used in Machado et al. (2018) trimmed to match the current sample.

All analyses were run under the R Core Team (2015) programming environment using the "tidyverse" set of packages. GM analyses were carried out using the geomorph package (Adams and Otarola-Castillo 2013). Procrustes correlation tests were done with the vegan package (Oksanen et al. 2017). Ancestral character state reconstruction was done with the phytools (Revell 2012) package.

## Results

Simulations were able to sample almost the full theoretical range of $var_{rel}(\lambda)$ values, especially for form (Fig. 1A). All correlations between GM and ILD were strong ($r_s > 0.978$) and nonlinear. The correlation between form ILD and shape GM was strongly nonlinear (Fig. 1A, third panel). Intermediate values are the most discrepant between GM and ILD for form and shape. For form, GM values tended to be greater than the ones for ILD values (Fig. 1A, first panel), while for shape, the opposite was true (Fig. 1A, second panel). Values of form ILD are much larger than values of shape GM, with the exception of values at the extremes of the distribution.

For the empirical datasets, $var_{rel}(\lambda)$ values were not correlated with sample size ($p > 0.41$ for all morphometric representations), and rarefied values were nearly identical with the ones obtained from the full dataset ($r_p > 0.989$ on all accounts, average difference $= 0.003 \pm 0.003$). Correlations between $var_{rel}(\lambda)$ and $tr(\mathbf{P}_W)$ changed broadly between different morphometric representations. Although correlations between these factors was low for shape variables (both GM and ILD had $r_p < 0.298$), they were higher for form variables, particularly for GM ($r_p = 0.721$ for GM and $r_p = 0.591$ for ILD). However, semipartial correlations are vastly smaller ($-0.030$ to 0.070 for all comparisons), suggesting that most of the correlation is given by PC1. Because both sources of bias are considered to be less relevant on the present dataset, we chose to focus empirical comparisons on the observed values of $var_{rel}(\lambda)$ for the full samples.

Empirical values of $var_{rel}(\lambda)$ for carnivoran species were more restricted in range than for the simulated data, with form values ranging from 0.1 to 0.7 and shape values ranging from 0.05 to 0.2. The correlations between GM and ILD within shape and form were strong (shape: $r_p = 0.804$; form: $r_p > 0.964$; Fig. 1B, first and second panels, respectively), whereas the correlation between form ILD and shape GM was moderate ($r_p = 0.393$; Fig. 1B, third panel). RMA analyses showed that intercepts were very similar to 0 in all comparisons, and slopes diverged from the expected value of 1 (Fig. 1B), with the comparison between form ILD and shape GM showing slopes that differed greatly from 1.

Ancestral estimates of magnitude of morphological integration along the phylogeny shows little divergence in the general phylogenetic pattern observed for both representations (GM and ILD) in both form and shape variables (Fig. 2). For form representations (Fig. 2A), pinnipeds and ursids show the largest magnitudes of morphological integration, whereas procyonids are the ones with the lowest values, on average. For shape variables (Fig. 2B), despite the lower between-representation correlations in comparison to form variables (Fig. 1), the overall phylogenetic pattern is similar, especially the fact that canids present lower magnitude of morphological integration values.

## DISCUSSION

We show that the choice of morphometric representation can lead to important differences on the estimates of integration magnitudes. As a consequence, it is not possible to directly compare values of magnitudes of morphological integration obtained from different morphometric representations. Despite that, both simulation and empirical results show that magnitude of morphological integration values are sorted similarly in both representations within form and within shape. In other words, a species that is considered to have "low" (or "high") magnitude of morphological integration for form or shape in one representation will most likely also be considered as such in the other representation. The same is not true for comparisons between form ILD and shape GM, which shows that magnitudes of morphological integration measured for form and the ones measured for shape are not comparable either on the empirical or simulated datasets. Our ancestral state reconstruction reinforces both these ideas, as the observed patterns for the evolution of magnitude of morphological integration of form or shape were consistent within methods, but divergent between methods (Fig. 2). Therefore, although it is not advisable to combine results from studies using different morphometric representations, it is possible to understand general patterns of evolution and variation of form or shape magnitudes of morphological integration using different morphometric representations.

Both simulation and empirical results show that magnitude of morphological integration values are biased depending on the morphometric representation and space chosen. Specifically, values for shape magnitude of morphological integration were higher for ILD than for GM, and the opposite trend was observed for

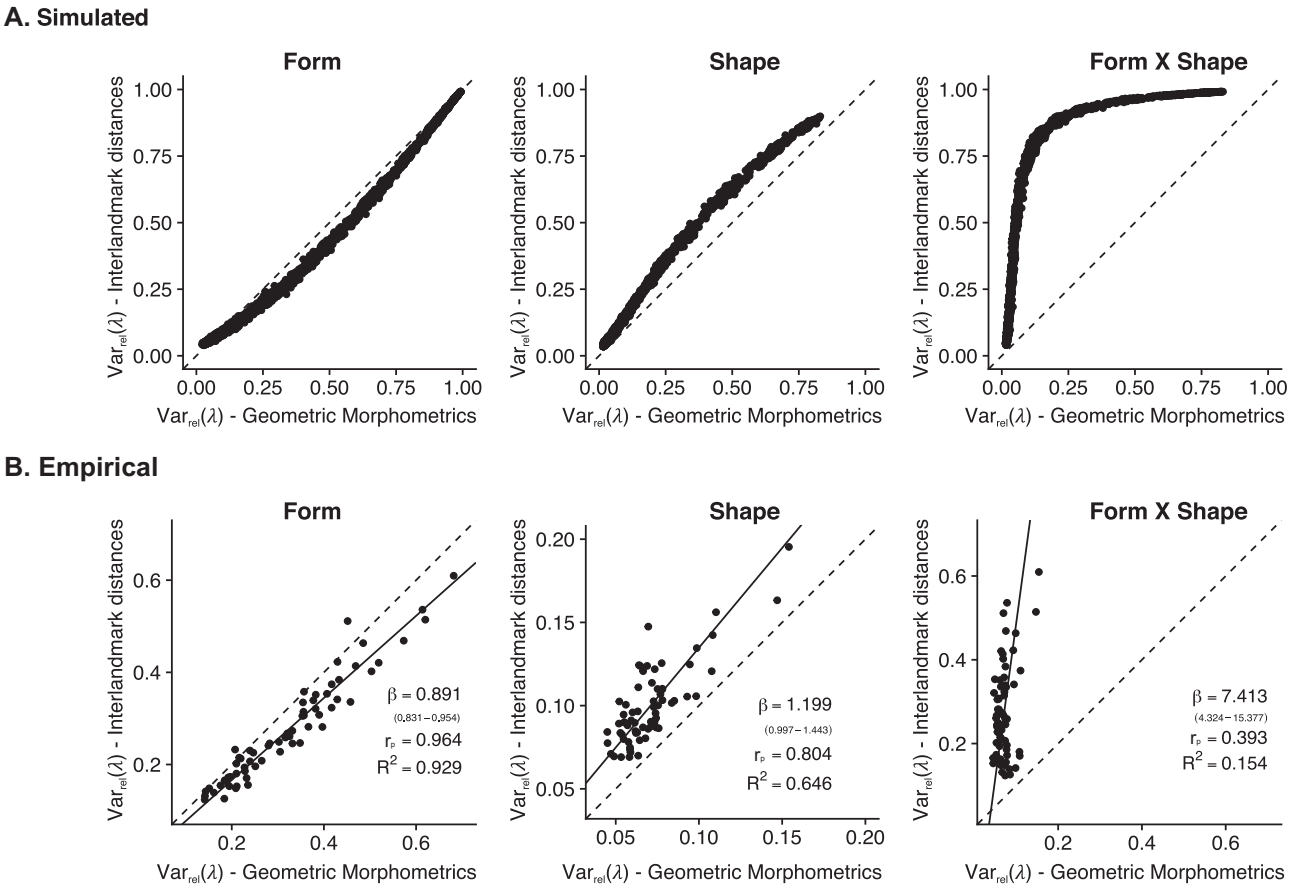


**Figure 1. Relationship between the $var_{rel}(\lambda)$ obtained on geometric morphometric and interlandmark distance for simulated (A) and empirical (B) datasets. SMA regression statistics: β-slope, values between parentheses are the confidence intervals; $r_p$-Pearson's product moment correlation index; $R^2$-coefficient of determination. Solid line—empirical SMA regression line. Dashed line—line where values are equal for both morphometric representations (slope = 1 and intercept = 0).**

form. These patterns are somewhat expected. It has long been argued that ILD data might add spurious correlations among traits depending on how traits are defined. For example, we expect traits that share landmarks to present nonzero covariances (Rohlf 2000), and if distances are defined across similar structures (i.e., mapping similar overall dimensions), they will most likely vary in a similar fashion (Zelditch et al. 2012). Despite the fact that the ILD used here are chosen to minimize those factors (Cheverud 1982), some small effect might be enough to slightly increase the magnitude on the ILD dataset. Thus, by adding covariances among traits we would expect the shape magnitude of morphological integration of ILD to be higher than the one for GM.

For form, however, this interpretation does not hold because the GM magnitude of morphological integration values are slightly inflated in relation to the ILD values. In this case, the amount of size variation may explain the observed pattern. Size is the main feature of biological form, and will usually dominate the variation in both ILD and GM (Jolicoeur 1963; Bookstein 1989; Mitteroecker et al. 2004). This means that size will load strongly on the first PC and therefore will be a major component in determining the magnitude of morphological integration (Hallgrímsson et al. 2009; Marroig et al. 2009; Porto et al. 2013). This phenomenon is simple to understand if we consider the magnitude of morphological integration in terms of the $var_{rel}(\lambda)$: everything else being equal if the relative contribution of the first eigenvalue increases, the standard deviation of eigenvalues will increase accordingly, leading to higher values of $var_{rel}(\lambda)$. Curiously, the amount of variation on the PC1 is higher on GM ($0.551 \pm 0.116$) than on ILD ($0.498 \pm 0.119$). Based on our discussion above, the inflation observed could be due to the large difference in the scale of size variation when compared to the scale of shape variation in GM. This could, in turn, explain why magnitude of morphological integration values for form tend to be higher on GM than on ILD (Fig. 1B, second panel). Because the difference between GM and ILD magnitudes of morphological integration values are relatively small, it is probable that size variation is similarly captured by ILD and GM, despite the difference in how that variation is measured.

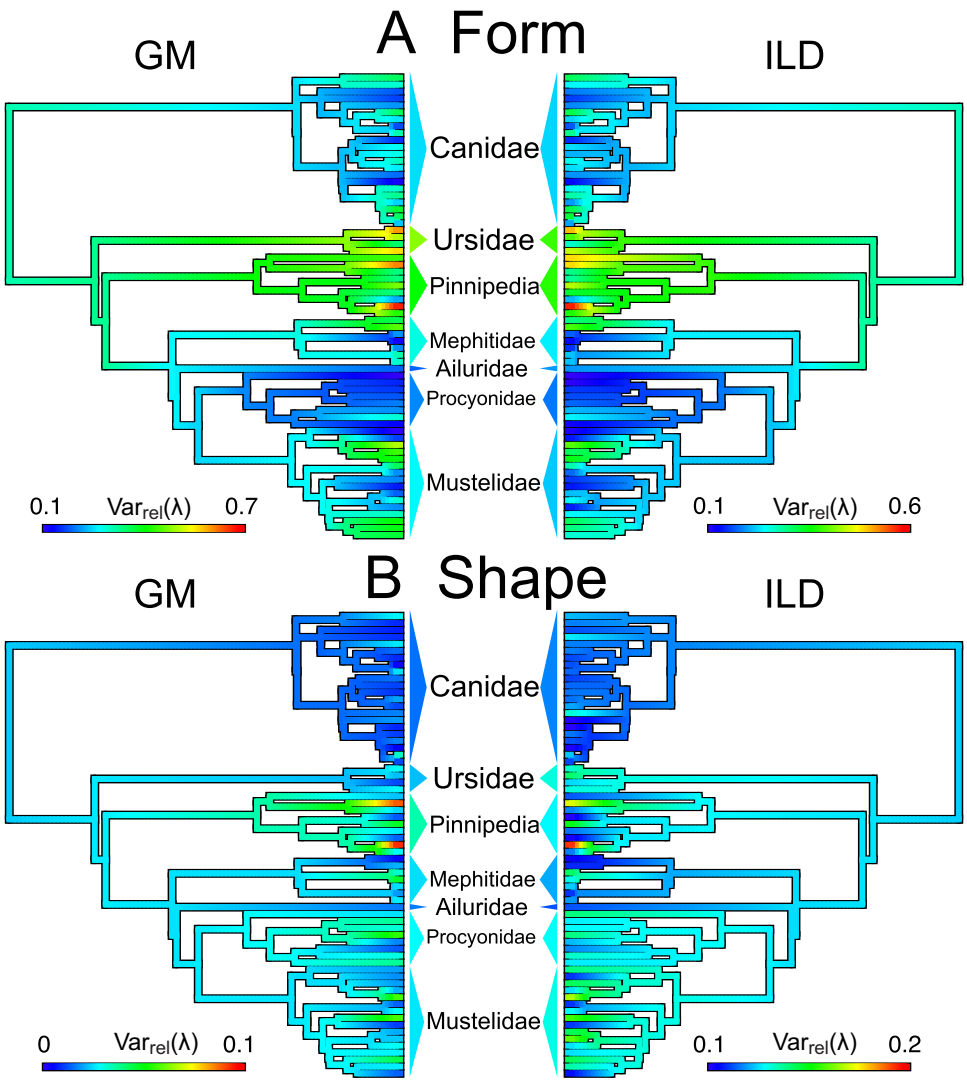


**Figure 2.** **Mapping of $var_{rel}(\lambda)$ as a continuous trait on the Caniform (Carnivora) phylogeny. Triangles associated with taxa names refer to the phylogenetic average of $var_{rel}(\lambda)$ for that group. GM, geometric morphometrics; ILD, interlandmark distances.**

The relationship between size and shape may also explain why the extreme values for the magnitude of morphological integration converge in both representations. This pattern was observed in all cases, but it is most apparent on the simulation dataset for the form ILD × shape GM comparison (Fig. 1A, last panel). In this case, matrices that present low magnitude of morphological integration values (i.e., near the theoretical minimum of $var_{rel}(\lambda) = 0$) were produced by applying low values of $p$ to equation (3). This means that the distribution of eigenvalues was homogenized and that the first PC explains a similar amount of variation in relation to other PCs. In other words, this matrix describes a mostly spherical distribution of traits. In this case, size variation will not dominate the covariance matrix structure because all dimensions contribute almost equally to the total amount of variation in the sample. Thus, matrices will present low magnitudes of morphological integration with or without size. Analogously, when $p$ is high, almost all variance in a sample will be accounted for by the first PC and the magnitude of morphological integration will be the strongest (will approach the $var_{rel}(\lambda) = 1$). The PC1 usually is a combination of size variation and allometry (Huxley 1924; Jolicoeur 1963; Bookstein et al. 1985; Klingenberg 2016) and our case is no exception. What we observe when $p$ is high is that the removal of size results in lower magnitudes of morphological integration in comparison to the form analysis (Fig. 1A, last panel) because the amount of variation explained by the PC1 decreases without size. However, as allometry is also an important contributor for the amount of variation explained by the PC1, this PC still presents considerably more variation than other axes. As a consequence, high magnitude of morphological integration values will still be observed for these matrices. Thus, unless ones' data cluster in the extremes of the magnitude of morphological integration distribution, which is

not the case for mammalian skulls at least, it is not advisable to conjointly discuss results obtained on different representations.

This then raises the question of what should be analyzed when making statements of magnitude of morphological integration: shape or form? One possible answer is that a structure should always be analyzed as a whole, and all factors should be taken into account (Bookstein 2009). From an evolutionary modeling point of view, the exclusion of important traits related to fitness variation could significantly mislead our understanding of the micro- and macroevolutionary dynamics (Morrissey et al. 2010). Given that size has major effects on fitness components (Calder 1984), excluding it from morphometric analysis might be ill advised. On the other hand, size is thought to affect all structure in a coordinated fashion, possibly obscuring localized genetic and epigenetic effects (Marroig et al. 2004; Mitteroecker and Bookstein 2009). Thus, its removal from the data might help us better understand details about morphological modularity that would not be evident otherwise (Marroig et al. 2004; Shirai and Marroig 2010; Porto et al. 2013). Also, it is conceivable that for some biological questions size truly is unimportant, such as the evolution of biomechanical proprieties, which might be fully independent of size (Rayner 1985; Dumont et al. 2009; Collar et al. 2014; Polly et al. 2016). These different choices for analyzing morphometric data could give insights into different aspects of morphological integration and evolution, and we should make choices that are appropriate for the questions we are trying to answer.

The standardized eigenvalue variance allows us to measure the magnitude of morphological integration in any numerical representation of shape or form, but we should always consider the actual theoretical implications of analyzing integration in these different morphospaces. The theory of morphological integration was originally developed to study measurements that are individually interpretable, such as lengths or weights (Olson and Miller 1958), and some predictions regarding the magnitude of integration only make sense in the context of those measurements. For example, the idea that functionally related traits will be more correlated, while true for these types of measurements, might not be so for shape variables such as ratios. Consider the case of the carnivoran mandible (Fig. 3). According to theory, the optimal position of the resultant force of the jaw closing muscles is 60% of the way from the mandibular-skull articulation to the carnassial tooth (Greaves 1983; Fig. 3A). The magnitude of morphological integration between the distance of the adductor muscle insertion (which influences the position of the resultant force) and of the carnassial tooth to the articulation is strong, as expected (Fig. 3B). However, if we measure the same features as a ratio of the total length of the mandible, we find that these variables are less integrated (Fig. 3C), a fact that could then lead to the misguided rejection of the hypothesis of functional association between the two traits. Interpretations of magnitude of morphological integration can be even less straightforward on shape–size morphospaces (sensu Mitteroecker et al. 2004), that is, when shape variables are analyzed along with a properly scaled size variable. On the mandible example above, if we add a size measure to the shape ratios we obtain a strong magnitude of morphological integration, despite the absence of association between ratios (Fig. 3C–E). In this case, the high magnitude of morphological integration is indeed the consequence of functional association among traits, as shape ratios and sizes stem directly from the original highly integrated variables. When interpreting morphological integration, we must take the meaning of the variables into account and consider how biological hypothesis translate into predictions about integration in the particular choice of representation.

Our results for carnivoran species are illustrative of how analyzing integration on different spaces might lead to different results and conclusions. Magnitude of morphological integration values for form have shown that large species, such as pinnipeds and bears, are among the most integrated ones (Fig. 2A). Larger sizes could be achieved by increased growth rates during ontogeny, a fact that could result into a greater variance in size and allometric variation in detriment of other aspects of shape (Porto et al. 2013). However, once size is removed, these groups are no longer among the most integrated ones. In fact, when we evaluate only the magnitude of integration of shape, Canidae is the group that stands out as being consistently less integrated than other taxa (Fig. 2B). As discussed above, interpreting shape integration can be problematic on its own. However, adding the log-centroid size on the shape data (as in the GM form dataset) do yield intermediary values of integration (the same is true if we add the log-centroid size to the ILD shape ratios, not shown), suggesting that the low shape integration in Canidae is not due to lack of morphological variation in shape variables. In fact, previous analysis have shown that canids have an increased evolutionary potential for some localized aspects of skull shape, namely those relating to the relative length of the face. Thus, it is likely that a lower shape integration observed for the group actually reflects a less constrained and more flexible (sensu Marroig et al. 2009) phenotype. This shows that a proper examination of morphological integration magnitudes cannot be taken out of the context of the morphospace under analysis.

In conclusion, our results show that the magnitude of morphological integration obtained from different morphometric representations is unlikely to be directly comparable. Even in the case of form, where absolute values are similar among representations, GM estimates tend to be larger than those for ILD (Fig. 1). Despite this, we observe a high correlation between values obtained from different representations, that is, $\mathrm{var}_{rel}(\lambda)$ of different species will be similarly sorted in both ILD and GM, especially if we are analyzing form data. Furthermore, large-scale

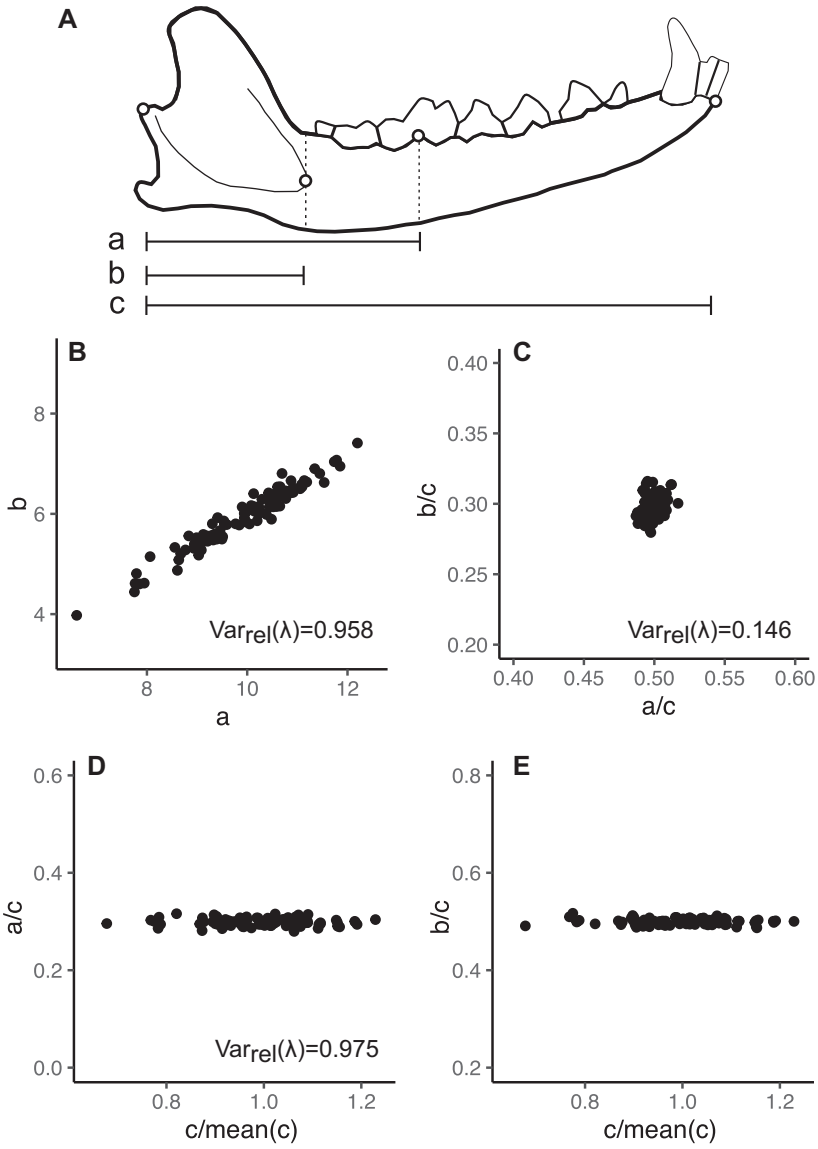


**Figure 3.** **(A) An example of functional traits on the carnivoran mandible: *a*—length between the articulation and the abductor muscle insertion; *b*—length between the mandibular-skull articulation and the carnassial tooth; *c*—dentary length. (B) Relationship between a and b. (C) Relationship between a and b as ratios of the *c* (a/c and b/c, respectively). (D) Relationship between *a*/*c* and *c* standardized by its average [c/mean(c)]. (E) Relationship between *a*/*c* and c/mean(c). var$_{rel}$(λ) was calculated for the association between linear variables *a* and *b* (B), shape ratios *a*/*c* and *b*/*c* (C), and both shape ratios, *a*/*c* and *b*/*c*, and the standardized size *c*/mean(*c*) (D).**

phylogenetic comparisons of form or shape can produce similar conclusions regardless of representation, even if species values are not equal. However, magnitudes of morphological integration obtained for form will not be compatible to those obtained for shape and vice versa, and care should be taken when evaluating conclusions reached by works focusing on these different representations of morphological variation.

## AUTHOR CONTRIBUTIONS

All authors conceptualized the study. FAM performed statistical analysis and wrote the software code GM gathered resources. FAM and AH provided the first draft of this manuscript. All authors were involved in writing and funding acquisition of this study.

## ACKNOWLEDGMENTS

We thank Mirian Zelditch, P. David Polly, Anjali Goswami, Philipp Mitteroecker, and two anonymous reviewers for their insightful comments that helped us to stir this manuscript to the right direction. We also thank Valentina Segura for lending the mandible outline to illustrate our paper. This work was partially supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) to FAM (2011/21674-4, 2013/22042-7), AH (2012/24937-9), DM (2014/26262-4), and GM (2011/14295-7). FAM was also partially supported by the NSF grant

(DEB 1350474 to L. Revell). AP was supported by the National Institute of Dental and Craniofacial Research of the National Institutes of Health, award number F31DE024944. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors declare no conflict of interest.

## DATA ARCHIVING

Data necessary to replicate these results will be publicly available as Supporting Information and at Dryad online repository: https://doi.org/10.5061/dryad.dv41ns1th.

## LITERATURE CITED

Adams, D., and E. Otarola-Castillo. 2013. geomorph: an r package for the collection and analysis of geometric morphometric shape data. Methods Ecol. Evol. 4:393–399.

Adams, D. C., and M. L. Collyer. 2016. On the comparison of the strength of morphological integration across morphometric datasets. Evolution 70:2623–2631.

Assis, A. P. A., J. L. Patton, A. Hubbe, and G. Marroig. 2016. Directional selection effects on patterns of phenotypic (co)variation in wild populations. Proc. Biol. Sci. 283:20161615.

Bookstein, F. L. 1989. "Size and shape": a comment on semantics. Syst. Zool. 38:173–180.

———. 2009. Measurement, explanation, and biology: lessons from a long century. Biol. Theory 4:6–20.

Bookstein, F. L., B. Chernoff, R. L. Elder, J. Humphries, G. R. Smith, and R. E. Strauss. 1985. Morphometrics in evolutionary biology: the geometry of size and shape change, with examples from fishes. Vol. 15, Special Publication. The Academy of Natural Sciences of Philadelphia, Philadelphia, PA.

Calder, W. A. 1984. Size, function, and life history. Courier Corporation, North Chelmsford, MA.

Cheverud, J. M. 1982. Phenotypic, genetic, and environmental morphological integration in the cranium. Evolution 36:499–516.

———. 1984. Quantitative genetics and developmental constraints on evolution by selection. J. Theor. Biol. 110:155–171.

Collar, D. C., P. C. Wainwright, M. E. Alfaro, L. J. Revell, and R. S. Mehta. 2014. Biting disrupts integration to spur skull evolution in eels. Nat. Commun. 5:430–439.

Curth, S., M. S. Fischer, and K. Kupczik. 2017. Patterns of integration in the canine skull: an inside view into the relationship of the skull modules of domestic dogs and wolves. Zoology 125:1–9.

Dumont, E. R., I. R. Grosse, and G. J. Slater. 2009. Requirements for comparing the performance of finite element models of biological structures. J. Theor. Biol. 256:96–103.

Esteve-Altava, B. 2017. In search of morphological modules: a systematic review. Biol. Rev. Cambr. Philos. Soc. 92:1332–1347.

Felice, R. N., M. Randau, and A. Goswami. 2018. A fly in a tube: macroevolutionary expectations for integrated phenotypes. Evolution 72:2580–2594.

Felsenstein, J. 1988. Phylogenies and quantitative characters. Annu. Rev. Ecol. Syst. 19:445–471.

Fruciano, C., P. Franchini, and A. Meyer. 2013. resampling-based approaches to study variation in morphological modularity. PLOS ONE 8: e69376.

Garcia, G. R. G., E. Hingst-Zaher, R. Cerqueira, and G. Marroig. 2014. Quantitative genetics and modularity in cranial and mandibular morphology of *Calomys expulsus*. Evol. Biol. 41:619–636.

Goswami, A., J. B. Smaers, C. Soligo, and P. D. Polly. 2014. The macroevolutionary consequences of phenotypic integration: from development to deep time. Philos. Trans. R. Soc. B 369:20130254.

Goswami, A., W. J. Binder, J. Meachen, and F. R. O'Keefe. 2015. The fossil record of phenotypic integration and modularity: a deep-time perspective on developmental and evolutionary dynamics. Proc. Natl. Acad. Sci. 112:4891–4896.

Grabowski, M., and A. Porto. 2017. How many more? Sample size determination in studies of morphological integration and evolvability. Methods Ecol. Evol. 8:592–603.

Greaves, W. S., 1983. A functional analysis of carnassial biting. Biol. J. Linn. Soc. 20:353–363.

Haber, A. 2011. A comparative analysis of integration indices. Evol. Biol. 38:476–488.

———. 2015. The evolution of morphological integration in the ruminant skull. Evol. Biol. 42:99–114.

Hallgrímsson, B., H. Jamniczky, N. Young, C. Rolian, T. Parsons, J. Boughner, and R. S. Marcucio. 2009. Deciphering the palimpsest: studying the relationship between morphological integration and phenotypic covariation. Evol. Biol. 36:355–376.

Hansen, T. F., and D. Houle. 2008. Measuring and comparing evolvability and constraint in multivariate characters. J. Evol. Biol. 21:1201–1219.

Houle, D., C. Pélabon, G. P. Wagner, and T. F. Hansen. 2011. Measurement and meaning in biology. Q. Rev. Biol. 86:3–34.

Hubbe, A., D. Melo, and G. Marroig. 2016. A case study of extant and extinct *Xenarthra cranium* covariance structure: implications and applications to paleontology. Paleobiology 42:465–488.

Huxley, J. S. 1924. Constant differential growth-ratios and their significance. Nature 114:895–896.

Jamniczky, H. A., and B. Hallgrímsson. 2009. A comparison of covariance structure in wild and laboratory muroid crania. Evolution 63:1540–1556.

Jolicoeur, P. 1963. The multivariate generalization of the allometry equation. Biometrics 19:497–499.

Klingenberg, C. P. 2016. Size, shape, and form: concepts of allometry in geometric morphometrics. Dev. Genes Evol. 226:113–137.

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

Machado, F. A., T. M. G. Zahn, and G. Marroig. 2018. Evolution of morphological integration in the skull of Carnivora (Mammalia): changes in Canidae lead to increased evolutionary potential of facial traits. Evolution 72:1399–1419.

Marroig, G., and J. M. Cheverud. 2004. Did natural selection or genetic drift produce the cranial diversification of neotropical monkeys? Am. Nat. 163:417–428.

Marroig, G., M. Vivo, and J. M. Cheverud. 2004. Cranial evolution in sakis (Pithecia, Platyrrhini) II: evolutionary processes and morphological integration. J. Evol. Biol. 17:144–155.

Marroig, G., L. T. Shirai, A. Porto, F. B. de Oliveira, and V. De Conto. 2009. The evolution of modularity in the mammalian skull II: evolutionary consequences. Evol. Biol. 36:136–148.

Meiri, S., T. Dayan, and D. Simberloff. 2005. Variability and correlations in carnivore crania and dentition. Funct. Ecol. 19:337–343.

Melo, D., A. Porto, J. M. Cheverud, and G. Marroig. 2016. Modularity: genes, development, and evolution. Annu. Rev. Ecol. Evol. Syst. 47:463–486.

Mitteroecker, P., and F. L. Bookstein. 2007. The conceptual and statistical relationship between modularity and morphological integration. Syst. Zool. 56:818–836.

Mitteroecker, P., and F. Bookstein. 2009. The ontogenetic trajectory of the phenotypic covariance matrix, with examples from craniofacial shape in rats and humans. Evolution 63:727–737.

Mitteroecker, P., P. Gunz, M. Bernhard, K. Schaefer, and F. L. Bookstein. 2004. Comparison of cranial ontogenetic trajectories among great apes and humans. J. Hum. Evol. 46:679–698.

Morrissey, M. B., L. E. B. Kruuk, and A. J. Wilson. 2010. The danger of applying the breeder's equation in observational studies of natural populations. J. Evol. Biol. 23:2277–2288.

Mosimann, J. E. 1970. Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions. J. Am. Stat. Assoc. 65:930–945.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'hara, G. L. Simpson, P. Solymos, et al. 2017. vegan: community ecology package. R package version 2.4-3. Available at https://CRAN.R-project.org/package=vegan Accessed April 7, 2018.

Olson, E. C., and R. L. Miller. 1958. Morphological integration. Univ. of Chicago Press, Chicago, IL.

Pavlicev, M., J. M. Cheverud, and G. P. Wagner. 2009. Measuring morphological integration using eigenvalue variance. Evol. Biol. 36:157–170.

Pearson, K., and A. G. Davin. 1924. On the biometric constants of the human skull. Biometrika 16:328–363.

Penna, A., D. Melo, S. Bernardi, M. I. Oyarzabal, and G. Marroig. 2017. The evolution of phenotypic integration: how directional selection reshapes covariation in mice. Evolution 71:2370–2380.

Peres-Neto, P., and D. Jackson. 2001. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia 129:169–178.

Polly, P. D., C. T. Stayton, E. R. Dumont, S. E. Pierce, E. J. Rayfield, and K. D. Angielczyk. 2016. Combining geometric morphometrics and finite element analysis with evolutionary modeling: towards a synthesis. J. Vertebr. Paleontol. 36:e1111225.

Porto, A., F. B. D. Oliveira, L. T. Shirai, V. Conto, and G. Marroig. 2009. The evolution of modularity in the mammalian skull I: morphological integration patterns and magnitudes. Evol. Biol. 36:118–135.

Porto, A., L. T. Shirai, F. B. D. Oliveira, and G. Marroig. 2013. size variation, growth strategies, and the evolution of modularity in the mammalian skull. Evolution 67:3305–3322.

Porto, A., H. Sebastião, S. E. Pavan, J. L. VandeBerg, G. Marroig, and J. M. Cheverud. 2015. Rate of evolutionary change in cranial morphology of the marsupial genus *Monodelphisis* constrained by the availability of additive genetic variation. J. Evol. Biol. 28:973–985.

Porto, A., R. Schmelter, J. L. VandeBerg, G. Marroig, and J. M. Cheverud. 2016. Evolution of the genotype-to-phenotype map and the cost of pleiotropy in mammals. Genetics 204:1601–1612.

R Core Team. 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/.

Randau, M., D. Sanfelice, and A. Goswami. 2019. Shifts in cranial integration associated with ecological specialization in pinnipeds (Mammalia, Carnivora). R. Soc. Open Sci. 6:190201–24.

Rayner, J. M. V. 1985. Linear relations in biomechanics: the statistics of scaling functions. J. Zool. 206:415–439.

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.

Riedl, R. 1978. Order in living organisms: a systems analysis of evolution. John Wiley, Hoboken, NJ.

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

Rohlf, F. J. 2000. On the use of shape spaces to compare morphometric methods. Hystrix Ital. J. Mammal. 11:8–24.

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

Schluter, D., T. Price, A. Ø. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51:1699–1711.

Shirai, L. T., and G. Marroig. 2010. Skull modularity in neotropical marsupials and monkeys: size variation and evolutionary constraint and flexibility. J. Exp. Zool. B 314B:663–683.

Walker, J. A. 2000. Ability of geometric morphometric methods to estimate a known covariance matrix. Syst. Biol. 49:686–696.

Young, N. M., G. P. Wagner, and B. Hallgrímsson. 2010. Development and the evolvability of human limbs. Proc. Natl. Acad. Sci. USA 107:3400–3405.

Zelditch, M. L., D. L. Swiderski, and H. D. Sheets. 2012. Geometric morphometrics for biologists: a primer. Academic Press, Cambridge, MA.

Associate Editor: A. Kaliontzopoulou
Handling Editor: D. Hall