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

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

[1]Department of Biology, University of Massachusetts (Boston), USA.

[2]División Mastozoología, Museo Argentino de Ciencias Naturales "Bernardino Rivadavia". Buenos Aires, Argentina.

[3]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

[4]Departamento de Oceanografia, Instituto de Geociências, Universidade Federal da Bahia, Brazil.

[5]Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, Brazil.

[6]Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Norway.

* Department of Biology, University of Massachusetts (Boston), USA

100 Morrissey Boulevard

ZIP Code: 02125

E-mail: macfabio@gmail.com

tel: (617) 543 8826

Running Title: Morphometrics and Morphological Integration

Keywords: Skull, Carnivora, Canidae, Eigenvalue variance, Covariance

Abstract Word Count: 237 words Total Word Count: 4743 words

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/evo.13864

Acknowledgments

We thank Mirian Zelditch, P. David Polly, Anjali Goswami, Philipp Mitteroecker and 2 anonymous reviews for their insightful comments that helped us to stir this manuscript into the right direction. We also thank Valentina Segura for lending the mandible outline to ilustrate 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 offi cial views of the National Institutes of Health.

Authors contributions

Conceptualization, all authors; Data gathering and code writing, FAM; Resources, GM; First draft, FAM and AH; Writing, all authors; Funding Acquisition, all authors.

Conflict of Interest Statement

The authors declare no conflict of interest.

Data Accessibility

Data necessary to replicate these results will be publicly available as a supplementary material and at Dryad online repository (http://datadryad.org/).

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

ABSTRACT:

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 by 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 (shape or form). Morphometricians should be aware of these differ-

ences and must consider how biological hypothesis translate into predictions about integration in each particular choice of representation.

Keywords: Eigenvalue variance, Covariance matrix, P matrix, Skull, Canidae

Introduction

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 (Cheverud, 1984; Riedl, 1978). 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 inter-landmark 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). ILDs analyses, in turn, measures 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 an 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 non-correlated) error on landmark data can lead to non-zero 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 geometric morphometrics 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 was 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 FM using a Microscribe MLX system (Immersion Corporation - San Jose, California). 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 was 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 inter-landmark distances. These distances have been used on a large set of evolutionary studies (e.g. Marroig and Cheverud, 2004; Porto et al., 2016; Assis et al., 2016; Hubbe 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 integration values using a Euclidean Distance Matrix Approach (EDMA), in which all possible ILDs are calculated. The relationship between integration measures obtained for the 35 ILD and the EDMA was slightly non-linear and highly correlated (Spearman rank Order correlation $r_s = 0.965$). Since the use of both ILD datasets provide the same general patterns, we focus on the results of the 35 ILD, henceforth called ILD for simplicity.

Following the analysis of the GM dataset, ILD 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 (inter-landmark distance) 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 non-allometric 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) in order 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 (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. Inter-landmark distances 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 P_W was calculated as the eigenvalue variance

$$Var(\lambda) = \frac{\sum_{i=1}^{N}(\lambda_i - \bar{\lambda})^2}{N} \tag{1}$$

where λ are the eigenvalues of 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

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

where $tr(\mathrm{P}_W)$ is the total amount of variation on the matrix P_W. The relative eigenvalue variance is calculated then as $Var_{rel}(\lambda) = Var(\lambda)/Var_{max}(\lambda)$. Values closer to 0 imply that eigenvalues have similar scales, implying a lack of covariance among the original traits. If $Var_{rel}(\lambda)$ is closer to 1, then $Var(\lambda)$ is close to the theoretical maximum $Var_{max}(\lambda)$, with all variation concentrated along a single axis, suggesting a highly integrated structure. The use of $Var_{rel}(\lambda)$ 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 $Var_{rel}(\lambda)$ 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 inter-landmark distances 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 $Var_{rel}(\lambda)$ 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

Simulations were based on the pooled within-groups GM covariance matrix for the full form data-set. In other words, it is the average covariance matrix for the full dataset weighted by within-group degrees of freedom. 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, we used the form P_W for the full sample of our empirical dataset, as described above, 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 in order 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 λ^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). In order 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

$$P_W^p = V\Lambda^p V^t \quad (3)$$

where V is a matrix of the first 35 eigenvectors of the baseline matrix and Λ^p is a diagonal matrix with the modified eigenvalues λ^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 higher than one, 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 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 inter-landmark distances 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 the 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.

In order to evaluate the effect of amount of variance on integration, we calculated the r_p between $Var_{rel}(\lambda)$ and $tr(\mathrm{P}_W)$ 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 semi-partial correlations between $tr(\mathrm{P}_W)$ and $Var_{rel}(\lambda)$, with variation of $tr(\mathrm{P}_W)$ conditional on the percentage of variation on PC1.

Comparison between GM and ILD

Values of $Var_{rel}(\lambda)$ were compared between GM and ILD representations using both Reduced Major Axis regressions (RMA) and Correlation analyses. For RMA an slope equal to 1 and a intercept equal to 0 means that integration values are the same for

both morphometric representations being compared. This was done both statistically by estimating confidence intervals for the RMA statistics, but also graphically, by plotting values of integration obtained on both representations. Comparisons were made between shape (GM x ILD) and form (GM x ILD) representations and between shape and form for all representations (GM shape x GM form, ILD shape x ILD form, ILD shape x GM form, GM shape x ILD form). Because all shape x form comparisons yielded very similar results (not shown), we focus only on the GM shape x ILD form comparison, as these are among the most popular forms of morphological quantification in magnitude of morphological integration studies (Esteve-Altava, 2017). Because of the strong nonlinear relationship between representations observed on the simulations (see below) we refrain from using the RMA on this dataset. For the same reason we employed a Spearman Rank Order correlation analysis (r_s) on the simulated dataset. For the observed dataset we report the r_p, even though r_s produces the same overall results.

Additionally, in order 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 approach (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. Geometric morphometric 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, specially for form (Fig. 1A). All correlations between GM and ILD were strong ($r_s > 0.978$) and non-linear. The correlation between form ILD and shape GM was strongly non-linear (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 ± 0.003). Correlations between $Var_{rel}(\lambda)$ and $tr(\mathrm{P}_W)$ changed broadly between different morphometric representations. While 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, semi-partial correlations are vastly smaller ($\rho = -0.030 - -0.070$ for all comparisons), suggesting that most of the signal 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 0.1–0.7 and shape values ranging from 0.05–0.2. 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), while

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, while 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, specially 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, while 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 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 non-zero 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 a 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 (Marroig et al., 2009; Hallgrímsson 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 ± 0.116) than on ILD (0.498 ± 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 (Figure 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.

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 x shape GM comparison (Figure 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 on the 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, since 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 (Figure 1A, last panel), because the amount of variation explained by the PC1 decreases without size. However, since 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 clusters 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 bio-mechanical 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 on different aspects of morphological integration and evolution, and we should make choices that are appropriate to 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 kinds 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), i.e. 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 are 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, i.e. $Var_{rel}(\lambda)$ of different species will be similarly sorted in both ILD and GM, especially if we are analyzing form data. Furthermore, large-scale 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.

References

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

Adams, D. C. and Collyer, M. L. (2016). On the comparison of the strength of morphological integration across morphometric datasets. Evolution, 70(11):2623–2631.

Assis, A. P. A., Patton, J. L., Hubbe, A., and Marroig, G. (2016). Directional selection

effects on patterns of phenotypic (co)variation in wild populations. Proceedings of the Royal Society of London. Series B, Biological Sciences, 283(1843):20161615.

Bookstein, F. L. (1989). " Size and shape": a comment on semantics. Systematic Zoology, 38(2):173–180.

Bookstein, F. L. (2009). Measurement, explanation, and biology: Lessons from a long century. Biological Theory, 4(1):6–20.

Bookstein, F. L., Chernoff, B., Elder, R. L., Humphries, J., Smith, G. R., and Strauss, R. E. (1985). Morphometrics in evolutionary biology: the geometry of size and shape change, with examples from fishes. The Academy of Natural Sciences of Philadelphia, Special Publication No. 15.

Calder, W. A. (1984). Size, function, and life history. Courier Corporation.

Cheverud, J. M. (1982). Phenotypic, genetic, and environmental morphological integration in the cranium. Evolution, 36(3):499–516.

Cheverud, J. M. (1984). Quantitative genetics and developmental constraints on evolution by selection. Journal of Theoretical Biology, 110(2):155–171.

Collar, D. C., Wainwright, P. C., Alfaro, M. E., Revell, L. J., and Mehta, R. S. (2014). Biting disrupts integration to spur skull evolution in eels. Nature Communications, 5(1):430–9.

Curth, S., Fischer, M. S., and Kupczik, K. (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., Grosse, I. R., and Slater, G. J. (2009). Requirements for comparing the performance of finite element models of biological structures. Journal of Theoretical Biology, 256(1):96–103.

Esteve-Altava, B. (2017). In search of morphological modules: a systematic review. Biological reviews of the Cambridge Philosophical Society, 92(3):1332–1347.

Felice, R. N., Randau, M., and Goswami, A. (2018). A fly in a tube: Macroevolutionary expectations for integrated phenotypes. Evolution; international journal of organic evolution, 72(12):2580–2594.

Felsenstein, J. (1988). Phylogenies and quantitative characters. Annual Review of Ecology and Systematics, 19:445–471.

Fruciano, C., Franchini, P., and Meyer, A. (2013). Resampling-Based Approaches to Study Variation in Morphological Modularity. PLOS ONE, 8(7):e69376.

Garcia, G. R. G., Hingst-Zaher, E., Cerqueira, R., and Marroig, G. (2014). Quantitative Genetics and Modularity in Cranial and Mandibular Morphology of Calomys expulsus. Evolutionary Biology, 41(4):619–636.

Goswami, A., Binder, W. J., Meachen, J., and O'Keefe, F. R. (2015). The fossil record of phenotypic integration and modularity: A deep-time perspective on developmental and evolutionary dynamics. Proceedings of the National Academy of Sciences, 112(16):4891–4896.

Goswami, A., Smaers, J. B., Soligo, C., and Polly, P. D. (2014). The macroevolutionary consequences of phenotypic integration: from development to deep time. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1649):20130254.

Grabowski, M. and Porto, A. (2017). How many more? Sample size determination in studies of morphological integration and evolvability. Methods in ecology and evolution, 8(5):592–603.

Greaves, W. S. (1983). A functional analysis of carnassial biting. Biological Journal of the Linnean Society, 20:353–363.

Haber, A. (2011). A Comparative Analysis of Integration Indices. Evolutionary Biology, 38(4):476–488.

Haber, A. (2015). The Evolution of Morphological Integration in the Ruminant Skull. Evolutionary Biology, 42(1):99–114.

Hallgrímsson, B., Jamniczky, H., Young, N., Rolian, C., Parsons, T., Boughner, J., and Marcucio, R. S. (2009). Deciphering the palimpsest: studying the relationship between morphological integration and phenotypic covariation. Evolutionary Biology, 36(4):355–376.

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

Houle, D., Pélabon, C., Wagner, G. P., and Hansen, T. F. (2011). Measurement and Meaning in Biology. The Quarterly Review Of Biology, 86(1):3–34.

Hubbe, A., Melo, D., and Marroig, G. (2016). A case study of extant and extinct Xenarthra cranium covariance structure: implications and applications to paleontology. Paleobiology, 42(3):465–488.

Huxley, J. S. (1924). Constant differential growth-ratios and their significance. Nature, 114(2877):895–896.

514 Jamniczky, H. A. and Hallgrímsson, B. (2009). A comparison of covariance structure in
515 wild and laboratory muroid crania. Evolution, 63(6):1540–1556.

516 Jolicoeur, P. (1963). The Multivariate Generalization of the Allometry Equation. Bio-
517 metrics, 19(3):497–499.

518 Klingenberg, C. P. (2016). Size, shape, and form: concepts of allometry in geometric
519 morphometrics. Development genes and evolution, 226(3):113–137.

520 Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain:
521 Body size allometry. Evolution, 33(1):402–416.

522 Machado, F. A., Zahn, T. M. G., and Marroig, G. (2018). Evolution of morphological
523 integration in the skull of Carnivora (Mammalia): Changes in Canidae lead to increased
524 evolutionary potential of facial traits. Evolution, 72(7):1399–1419.

525 Marroig, G. and Cheverud, J. M. (2004). Did natural selection or genetic drift produce the
526 cranial diversification of neotropical monkeys? The American Naturalist, 163(3):417–
527 428.

528 Marroig, G., Shirai, L. T., Porto, A., de Oliveira, F. B., and De Conto, V. (2009). The
529 evolution of modularity in the mammalian skull II: Evolutionary consequences. Evolu-
530 tionary Biology, 36(1):136–148.

531 Marroig, G., Vivo, M., and Cheverud, J. M. (2004). Cranial evolution in sakis (Pithe-
532 cia, Platyrrhini) II: evolutionary processes and morphological integration. Journal of
533 Evolutionary Biology, 17(1):144–155.

534 Meiri, S., Dayan, T., and Simberloff, D. (2005). Variability and correlations in carnivore
535 crania and dentition. Functional Ecology, 19(2):337–343.

Melo, D., Porto, A., Cheverud, J. M., and Marroig, G. (2016). Modularity: Genes, Development, and Evolution. Annual review of ecology, evolution, and systematics, 47(1):463–486.

Mitteroecker, P. and Bookstein, F. (2009). The ontogenetic trajectory of the phenotypic covariance matrix, with examples from craniofacial shape in rats and humans. Evolution, 63(3):727–737.

Mitteroecker, P. and Bookstein, F. L. (2007). The conceptual and statistical relationship between modularity and morphological integration. Systematic Zoology, 56(5):818–836.

Mitteroecker, P., Gunz, P., Bernhard, M., Schaefer, K., and Bookstein, F. L. (2004). Comparison of cranial ontogenetic trajectories among great apes and humans. Journal of Human Evolution, 46(6):679–698.

Morrissey, M. B., Kruuk, L. E. B., and Wilson, A. J. (2010). The danger of applying the breeder's equation in observational studies of natural populations. Journal of Evolutionary Biology, 23(11):2277–2288.

Mosimann, J. E. (1970). Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions. Journal of the American Statistical Association, 65(330):930–945.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H. (2017). vegan: Community Ecology Package.

Olson, E. C. and Miller, R. L. (1958). Morphological integration. University of Chicago Press.

Pavlicev, M., Cheverud, J. M., and Wagner, G. P. (2009). Measuring Morphological Integration Using Eigenvalue Variance. Evolutionary Biology, 36(1):157–170.

Pearson, K. and Davin, A. G. (1924). On the biometric constants of the human skull. Biometrika, 16(3/4):328–363.

Penna, A., Melo, D., Bernardi, S., Oyarzabal, M. I., and Marroig, G. (2017). The evolution of phenotypic integration: How directional selection reshapes covariation in mice. Evolution, 71(10):2370–2380.

Peres-Neto, P. and Jackson, D. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia, 129(2):169–178.

Polly, P. D., Stayton, C. T., Dumont, E. R., Pierce, S. E., Rayfield, E. J., and Angielczyk, K. D. (2016). Combining geometric morphometrics and finite element analysis with evolutionary modeling: towards a synthesis. Journal of Vertebrate Paleontology, 36(4):e1111225.

Porto, A., Oliveira, F. B. d., Shirai, L. T., Conto, V., and Marroig, G. (2009). The Evolution of Modularity in the Mammalian Skull I: Morphological Integration Patterns and Magnitudes. Evolutionary Biology, 36(1):118–135.

Porto, A., Schmelter, R., VandeBerg, J. L., Marroig, G., and Cheverud, J. M. (2016). Evolution of the Genotype-to-Phenotype Map and the Cost of Pleiotropy in Mammals. Genetics, 204:1601–1612.

Porto, A., Sebastião, H., Pavan, S. E., VandeBerg, J. L., Marroig, G., and Cheverud, J. M. (2015). Rate of evolutionary change in cranial morphology of the marsupial genus

Monodelphisis constrained by the availability of additive genetic variation. Journal of Evolutionary Biology, 28(4):973–985.

Porto, A., Shirai, L. T., Oliveira, F. B. d., and Marroig, G. (2013). Size Variation, Growth Strategies, And The Evolution Of Modularity In The Mammalian Skull. Evolution, 67(11):3305–3322.

R Core Team (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Randau, M., Sanfelice, D., and Goswami, A. (2019). Shifts in cranial integration associated with ecological specialization in pinnipeds (Mammalia, Carnivora). Royal Society Open Science, 6(3):190201–24.

Rayner, J. M. V. (1985). Linear relations in biomechanics: the statistics of scaling functions. Journal of zoology, 206(3):415–439.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution, 3(2):217–223.

Riedl, R. (1978). Order in Living Organisms: A Systems Analysis of Evolution. John Wiley.

Rohlf, F. and Slice, D. (1990). Extensions of the procrustes method for the optimal superimposition of landmarks. Systematic Zoology, 39(1):40–59.

Rohlf, F. J. (2000). On the use of shape spaces to compare morphometric methods. Hystrix-the Italian Journal of Mammalogy, 11(1):8–24.

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

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

Shirai, L. T. and Marroig, G. (2010). Skull modularity in neotropical marsupials and monkeys: size variation and evolutionary constraint and flexibility. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution, 314B(8):663–683.

Walker, J. A. (2000). Ability of geometric morphometric methods to estimate a known covariance matrix. Systematic Biology, 49(4):686–696.

Young, N. M., Wagner, G. P., and Hallgrímsson, B. (2010). Development and the evolvability of human limbs. Proceedings of the National Academy of Sciences of the United States of America, 107(8):3400–3405.

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

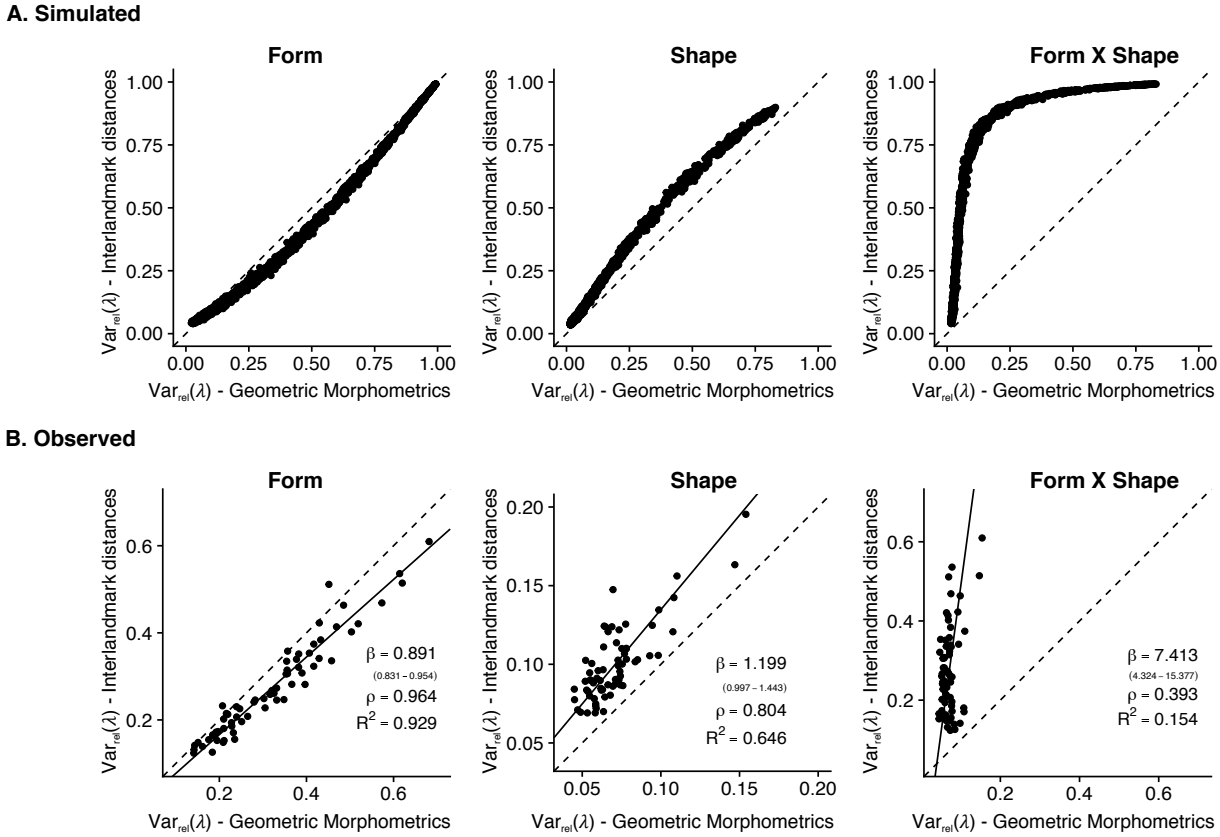

Figure 1: Relationship between the $Var_{rel}(\lambda)$ obtained on Geometric Morphometric and Inter-landmark Distance for simulated (A) and observed (B) datasets. SMA regression statistics: β-slope, values between parenthesis are the confidence intervals. ρ- Pearson's product moment correlation index. 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).

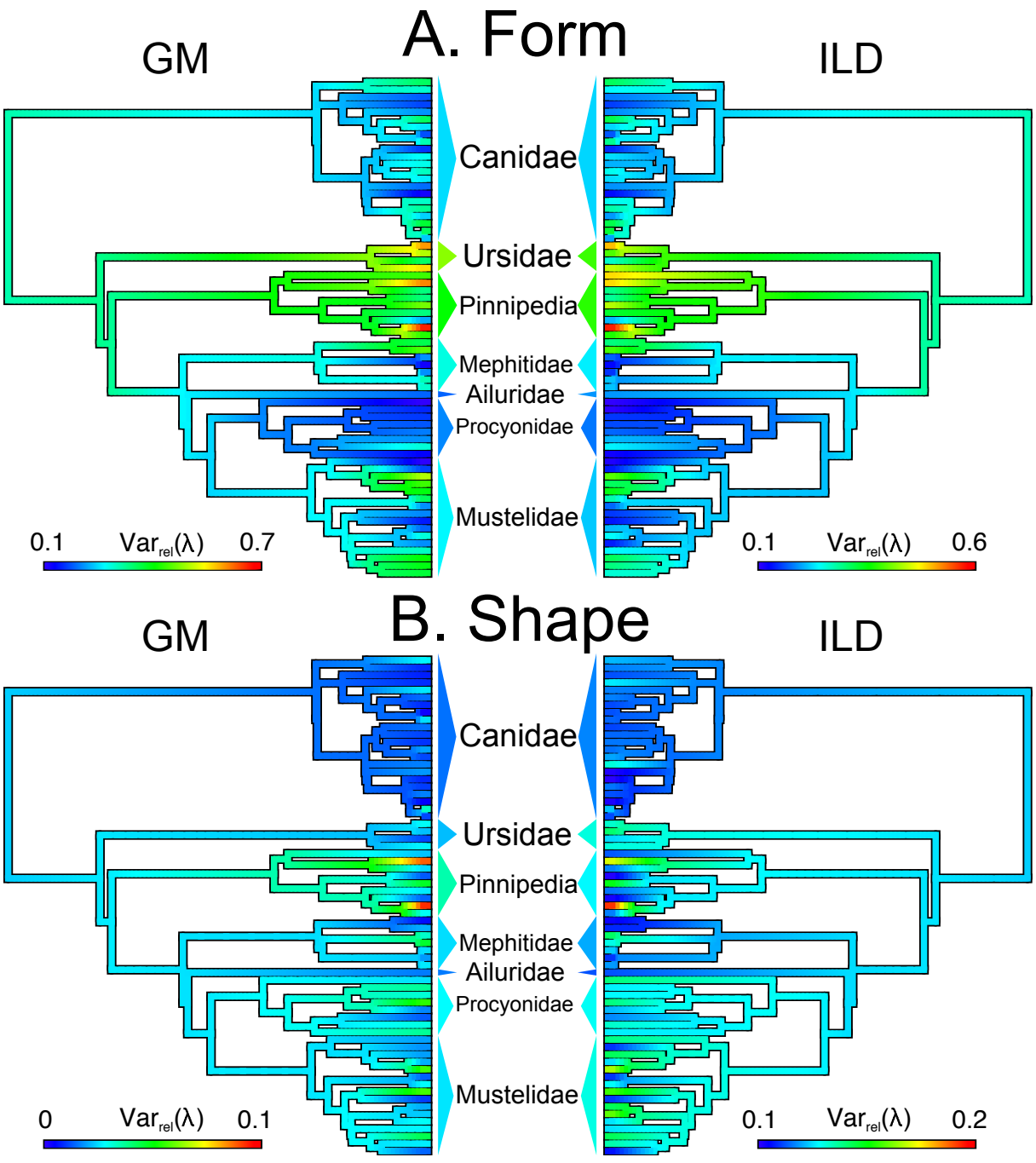

Figure 2: Mapping of $Var_{rel}(\lambda)$ as a continuous trait on the Caniform (Carnivora) phylogeny. Triangles associated with taxa names refer to the phylogenetic average for that group. GM- Geometric Morphometrics. ILD- Inter-landmark Distances.

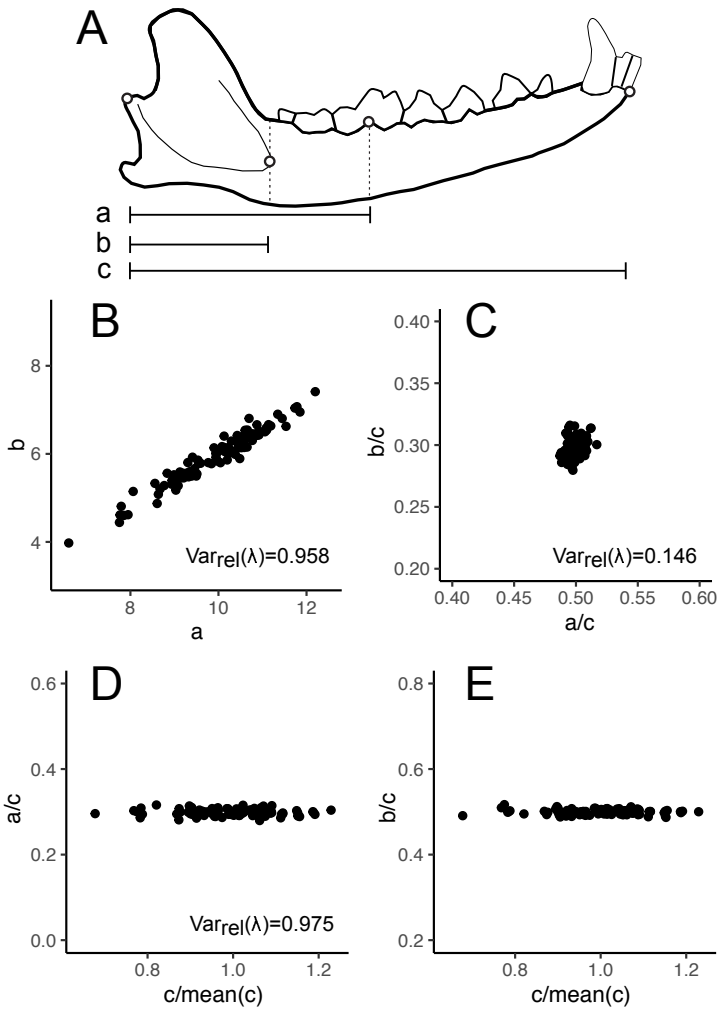

Figure 3: A- Example of functional traits on the carnivoran mandible: a- length between the articulation and abductor muscle insertion; b- length between the articulation and the carnassial tooth; c- dentary length. B- Relationship between the ratios b/c and a/c. C- Relationship between b and a as ratios of the c. D- Relationship between a/c and c standardized by its average. E- Relationship between a/c and c standardized by its average. $Var_{rel}(\lambda)$ 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).