Morphological integration during postnatal ontogeny: implications for evolutionary biology

Alex Hubbe[1,*]

Guilherme Garcia[2]

Harley Sebastião[2]

Arthur Porto[3]

Fábio Machado[4,5]

James Cheverud[6]

Gabriel Marroig[2]

1. Departamento de Oceanografia, Instituto de Geociências, Universidade Federal da Bahia, Bahia 40170-020, Brazil;

2. Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo 05508-900, Brazil;

3. Seattle Children's Research Institute, Center for Developmental Biology and Regenerative Medicine, Washington 98101, USA;

4. Department of Biology, University of Massachusetts, Massachusetts 01003, USA;

5. División Mastozoología, Museo Argentino de Ciencias Naturales "Bernardino Rivadavia" - CONICET, Buenos Aires C1405 DJR, Argentina;

6. Department of Biology, Loyola University Chicago, Illinois 60660, USA;

∗ Corresponding author; e-mail: alexhubbe@yahoo.com

Manuscript elements: Figure 1, Figure 2, Figure 3, Figure 4, Table 1, online appendix A (including Figure A1, Figure A2, Figure A3, Figure A4, Figure A5, Figure A6, Figure A7, Figure A8,

Figure A9).

Keywords: Development, **G**-matrix, **P**-matrix, Marsupialia, Placentalia.

Manuscript type: Article.

Prepared using the suggested LaTeX template for *Am. Nat.*

Abstract

Understanding how development changes the genetic covariance of complex phenotypes is fundamental for the study of evolution. If the genetic covariance changes dramatically during postnatal ontogeny, one cannot infer confidently evolutionary responses based on the genetic covariance estimated from a single postnatal ontogenetic stage. Mammalian skull morphology is a common model system for studying the evolution of complex structures. These studies often involve estimating covariance between traits based on adult individuals. There is robust evidence that covariances changes during ontogeny. However, it is unknown whether differences in age-specific covariances can, in fact, bias evolutionary analyses made at subadult ages. To explore this issue, we sampled two marsupials from the order Didelphimorphia, and one precocial and one altricial placental at different stages of postnatal ontogeny. We calculated the phenotypic variance-covariance matrix (**P**-matrix) for each genera at these postnatal ontogenetic stages. Then, we compared within genus **P**-matrices and also **P**-matrices with available congeneric additive genetic variance-covariance matrices (**G**-matrices) using Random Skewers and the Krzanowsky projection methods. Our results show that the structural similarity between matrices are in general high (> 0.7). Our study supports that the **G**-matrix in therian mammals is conserved during most of postnatal ontogeny. Thus it is feasible to study life-history changes and evolutionary responses based on the covariance estimated from a single ontogenetic stage. Our results also suggest that at least for some marsupials the **G**-matrix varies considerably prior to weaning, which does not invalidate our previous conclusion because specimens at this stage would experience striking differences in selective regimes than during later ontogenetic stages.

Introduction

Development is an essential component of evolution. Although natural selection acts upon phenotypes (Lande and Arnold, 1983), the response to selection depends on how development shapes the genetic variation into variable adult phenotypes (Hendrikse et al., 2007; Hallgrímsson et al., 2009; Alberch, 1982; Hall, 1999). By changing the availability of genetic variation at different life stages, development can alter the capacity of a species to change along specific directions in morphospace (Hallgrímsson et al., 2009; Mitteroecker and Bookstein, 2009).

If changes in the patterns of genetic variation across ontogeny are substantial, earlier life stages might produce responses to selection that are remarkably different from later stages. For example, consider a case in which a set of traits present overall high intensity of association [i.e. high integration sensu Olson and Miller (1958)] in juvenile specimens, but in adults these traits are much less integrated. Selection operating on juveniles would likely lead to responses aligned with the major direction of variation, while the same selection operating on adults could lead to highly divergent evolution. In this scenario, the reconstruction of selection patterns is problematic, because using only the additive genetic covariance from a specific population's ontogenetic stage (e.g. adults) can result in inaccurate conclusions if direct selective pressures occurred on both ontogenetic stages. On the other hand, if the additive genetic covariance is relatively stable throughout ontogeny, the net impact of evolutionary forces acting on populations can be reconstructed satisfactorily from individuals belonging to a single ontogenetic stage (i.e. adult stage). Thus, to which extent development causes variation in the additive genetic covariance has profound implications for evolutionary, genetic, and ecological studies limited to one or few ontogenetic stages of a population.

The mammalian skull is a common model system for investigating the evolution of complex structures, and studies often focus on adult specimens [e.g. Machado et al. (2018); Goswami (2006); Haber (2014)]. If we accept that the phenotypic covariance can be used as surrogate for the additive genetic covariance (Cheverud, 1988, 1996; Roff, 1995; Akesson et al., 2007; Porto et al., 2009; Reusch and Blanckenhorn, 1998; House and Simmons, 2005; Leamy and Cheverud,

1984), there is compelling evidence that the additive genetic covariance changes during ontogeny (Hallgrímsson et al., 2009; Mitteroecker and Bookstein, 2009; Atchley, 1984; Coleman et al., 1994; Zelditch, 1988; Zelditch and Carmichael, 1989; Zelditch et al., 2006, 1992; Mitteroecker et al., 2012; Nonaka and Nakata, 1984; Goswami et al., 2012). However, it is not well known to what extent this developmental variation in the additive genetic covariance can have a meaningful impact on evolutionary responses.

To address this issue we developed a cross-sectional study of cranial trait covariances based on a comprehensive sample of different mammalian developmental strategies. We sampled the precocial platyrrhine primate *Sapajus*, the altricial sigmodontinae rodent *Calomys*, and the Didelphimorphia marsupials *Didelphis* and *Monodelphis* in different age classes encompassing from the first months of life after birth to adulthood. We calculated the phenotypic variance-covariance matrix (**P**-matrix) for each genus and age class and also used the adult additive genetic variance-covariance matrices (**G**-matrix) available for *Calomys* (Garcia et al., 2014) and *Monodelphis* (Porto et al., 2015). We then compared within genus **P**- and **G**-matrices using different analytical methods.

Methods

Sample

Our sample is composed of 2098 specimens belonging to *Didelphis* and *Monodelphis* (Marsupialia, Mammalia), *Sapajus* (Primates, Placentalia, Mammalia), and *Calomys expulsus* (Rodentia, Placentalia, Mammalia; Table 1). Studied specimens are deposited in the following institutions: American Museum of Natural History (New York, USA); Field Museum of Natural History (Chicago, USA); Museu Nacional (Rio de Janeiro, Brazil); Museu Paraense Emilio Goeldi (Belém, Brazil), Museu de Zoologia da Universidade de São Paulo (São Paulo, Brazil); Museum of Vertebrate Zoology (Berkeley, USA), National Museum of Natural History (Washington D.C, USA), and Texas Biomedical Research Institute (San Antonio, USA).

The age of specimens was determined either by dental eruption and wear (i.e. dental age

class) or by days after birth (i.e. birth age class; Figure 1). Dental age class for *Didelphis* and *Monodelphis* were determined based on maxillary dental eruption and wear (Tyndale-biscoe and Mackenzie, 1976; Tribe, 1990). We added an extra class (zero), composed of specimens with no erupted teeth. For *Didelphis*, the ontogenetic stages sampled most likely include lactation (dental age class zero), the start of solid food ingestion (dental age class one), end of the weaning (between dental age classes one and two), to adulthood [dental age class above four; van Nievelt and Smith (2005); Tyndale-biscoe and Mackenzie (1976); McManus (1974); Abdala et al. (2001); Sebastião and Marroig (2013)]. For *Monodelphis*, the ontogenetic stages sampled most likely include the end of the weaning process (during dental age class two) to adulthood [dental age class above four; van Nievelt and Smith (2005); Sebastião and Marroig (2013)].

An independent set of *Monodelphis* specimens was classified according to birth age classes 20, 40, 70, 150, 300, and 400 days after birth. These classes encompass exclusive lactation (20 and 40), to the very end of the weaning process (70), to adulthood [300 and 400; Nievelt and Smith (2005)]. The birth age classes are analogue to the following dental age classes : 20 = zero; 40 = zero to one; 70 = two; $\geq$ 150 = $\geq$ four (van Nievelt and Smith, 2005; Nievelt and Smith, 2005). Throughout our paper, the dental age classes will be labeled *Monodelphis* (D), while the birth age classes will be labeled *Monodelphis* (B).

The dental age classes for *Sapajus* were determined based on the premaxillary and maxillary dental eruption (Richtsmeier et al., 1993). They span from weaning (dental age class one) to adulthood [dental age class six; Marroig and Cheverud (2001); Fragazy et al. (2004)]. Almost all specimens of *Sapajus* we studied are assigned to *S. apella* (n = 200), however 24 specimens are of uncertain classification. These specimens belong to the age classes one (n = 6), two (n = 11) and three (n = 7). Therefore, we cannot rule out an interspecific bias in our results. However, inspections of Principal Component Analysis plots showed no particular partitioning of the morphospace between specimens that were and were not identified to the species level (results not shown).

The *Calomys* were classified according to the birth age classes 20, 30, 50, 100, 200, 300 and 400 days after birth. Specimens comprise from around weaning (20 days) to adulthood [100 days

onwards, which is when specimens reached sexual maturity; Hingst-Zaher et al. (2000)].

Didelphis, *Monodelphis* (D) and *Sapajus* specimens were wild caught while *Calomys* (Garcia et al., 2014) and *Monodelphis* (B) (Porto et al., 2015) specimens were derived from captive-bred colonies.

Landmarks and Measurements

Three-dimensional coordinates were collected for every specimen for 32 skull landmarks using a 3D digitizer. These landmarks are homologous among the species. All specimens were measured with the same instrument (Microscribe MX; Immersion Corporation, San Jose, California) with the exception of adult *Sapajus* that were measured with a 3Draw digitizer (Polhemus Inc., Colchester, Vermont). Tests performed with adult specimens measured with both devices indicated that that source of variation is negligible (results not shown). Based on these landmarks, 35 linear distances were calculated in millimeters [for details related to the landmarks and distances acquisition refer to Cheverud (1995); Porto et al. (2009); Shirai and Marroig (2010)]. The set of distances calculated aimed to represent the whole cranium morphology and important developmental and functional relationships among cranial regions while avoiding redundancy (Marroig and Cheverud, 2001; Cheverud, 1982).

Didelphis and *Monodelphis* (D) crania were measured by HS. *Monodelphis* (B) crania were measured by AP. *Calomys* crania were measured by GG, and *Sapajus* crania were measured by GM. All specimens were measured following the same protocol. Even if interobserver measurement error exists, it should not bias our conclusions because we are studying within-genus variation and for each genus only one observer took the measurements.

Didelphis, *Monodelphis* (D), and *Calomys* specimens were measured twice to account for measurement errors (Lessells and Boag, 1987). Measurement errors were calculated for each genus at each age class with more than 14 specimens for each trait independently. In general, repeatabilities were high (> 0.9; Figure A 1). Lower repeatabilities were associated with traits exhibiting low total variation between replicates. It is worth noting there are three distances in *Calomys* (MTPNS, APETBA, APETTS) that presented low repeatability values (< 0.75) for certain ages. Besides the low variance observed for these distances, they were also very short, approaching the

spatial resolution of the digitizer. Therefore, measurement errors are negligible in most cases and specimens' traits were represented in further analyses by the mean of replicated measurements. This ensures that the actual repeatability values of the traits we used are underestimated (Falconer and MacKay, 1996), further limiting the impact of measurement error. If you have the repeatability of single measurements, you can calculate the repeatability of the mean of two replicates as

$$r(n) = r/(r + (1 - r)/n)/r) \tag{1}$$

where n is the number of replicates and r is repeatability. So the repeatability of the mean of two repeated measures is 0.888 if the repeatability of single measurements is 0.8.

Although the *Sapajus* and the *Monodelphis* (B) specimens were not measured twice, the measurement errors observed for other Platyrrhini measured by GM (Marroig and Cheverud, 2001), and for smaller Didelphimorphia measured by AP (Porto et al., 2009) were negligible. Therefore, we assumed the *Sapajus* and *Monodelphis* (B) specimens also presented negligible measurement errors.

*Estimating variance-covariance and correlation **P**-matrices*

The **P**-matrix is the phenotypic variance-covariance matrix (hereafter covariance matrix) and its standardized counterpart, the phenotypic ccorrelation matrix. Covariance and correlation **P**-matrices are estimated with error, and the smaller the ratio between sample size and number of traits analyzed the more pronounced the influence of error on the matrix estimates (Marroig et al., 2012). For this reason, age class and genus specific **P**-matrices were estimated only in cases with at least 15 specimens (Table 1).

Before estimating covariance and correlation **P**-matrices, sources of variation within each genus that were not of interest for this study (species, sex and their interaction) were identified and controlled. To identify the effects, we used the following approaches. First we used pairwise nonparametric multivariate analyses of variance (NP-MANOVA; significance at $P(\alpha) < 0.05$). In cases of insufficient sample sizes for NP-MANOVA we used pairwise nonparametric univariate

analysis of variance (NP-ANOVA) (McArdle and Anderson, 2001; Anderson, 2001; Collyer and Adams, 2018), considering the tested effects as influencing traits means whenever two or more traits had significant covariances at $P(\alpha) < 0.01$ for that effect. Second, effects were also evaluated with Principal Component Analysis. Lastly, we assessed the impact that controlling for different combinations of effects and their interactions has on the covariance estimates by evaluating the extent to which the covariance structure is sensitive to their inclusion in the analyses. To do that, first we calculated all possible matrices resulting from controlling all possible combinations of effects and their interactions. For example, for *Didelphis* at age class one, we calculated one matrix after controlling for sex, another matrix after controlling for sex and species, and so on. After that, we compared these matrices using both the Random Skewers method (for details on the method see below) and the difference between the trace of the matrices. For the adult age class, we also evaluated the effect of the age classes involved, as described above. Whenever an effect or interaction was considered a source of variation (Table 1) in at least one age class, residual pooled within-group phenotypic covariance and correlation matrices were estimated for each genus and age class using the general linear model approach (Marroig and Cheverud, 2004).

After estimating the age class and genus specific **P**-matrices, we examined two fundamental aspects of morphometric mammalian skull covariance/correlation matrices: their structure (i.e. relationship between the morphological elements), and overall magnitude of morphological integration [i.e. the intensity of the relationships between the morphological elements; (Porto et al., 2009).

Matrix structural similarity

The structural similarity between age class and genus specific **P**-matrices was evaluated for both covariance and correlation matrices. Additionally, all **P**-matrices from *Calomys* and *Monodelphis* (B and D) were compared to the adult **G**-matrices of *Calomys* (Garcia et al., 2014) and *Monodelphis* (B) (Porto et al., 2015), respectively. Matrices were compared pairwise using Random Skewers [RS; (Cheverud and Marroig, 2007)], and the Krzanowski Projection [KP; (Krzanowski, 1979, 2000; Blows et al., 2004).

We chose the RS and KP because these statistics provide simple and intuitive measurements of overall similarity between two matrices, and have high correspondence with the Riemaniann Distance (Figure A 2; (Machado et al., 2018)), which is a statistic used in a similar context to our study (Mitteroecker and Bookstein, 2009) Moreover, RS is particularly relevant in our study because it goes to the core of the evolutionary question we are addressing. RS assesses whether two **G**- or **P**-matrices have, on average, similar evolutionary responses to the same selection gradients (Cheverud and Marroig, 2007).

The RS is based on the multivariate response to selection equation and measures the mean correlation between the evolutionary response vectors of each pair of **P**-matrices obtained by multiplying these **P**-matrices by the same standardized 1,000 random selection vectors. RS measures how similar, on average, two populations (**P**-matrices) will respond to the same 1,000 selection regimes. If these populations have **P**-matrices with the same structure, they will have the same orientation in their responses and the RS equals one. If they have unrelated structures the RS equals zero. Statistical significance was evaluated by comparing the observed RS value to an empirically-derived correlation distribution for 1,000 random vectors. If the observed value exceeds 99% of this empirically-derived distribution, it was considered different from zero, therefore rejecting the null hypothesis of structural dissimilarity (Cheverud and Marroig, 2007; Cheverud, 1996).

The KP is based on the projection **S**-matrix:

$$S = A^t B B^t A \tag{2}$$

where **A**- and **B**-matrices are sets of orthogonal axes (Principal components – PCs) from each of the compared matrices in the same k-dimensonal subspace, and ^{t} denotes transposed matrix. The sum of the eigenvalues of the **S**-matrix divided by k represents the similarity of the two subspaces. The KP measures the similarity between the two k-dimensional subspaces. If the two subspaces are completely dissimilar the KP equals zero and if the two subspaces are strictly similar the KP equals one. The k dimensions were determined by the first k PCs of the integer part of half the original dimensions subtracted from another dimension. In this study $k = 16$ [integer part of

$\frac{35}{2} - 1$; Krzanowski (1979, 2000); Blows et al. (2004)].

The concept of low versus high structural similarity is subjective to some extent, since there is no consensus among researchers about the minimum amount of structural similarity between **G**-matrices to ensure that they will not bias evolution in a meaningful way (Cheverud 1988; Arnold and Phillips 1999; Marroig and Cheverud 2001). With respect to the Random Skewers, it has been argued that values > 0.7 can be interpreted as high similarity between two matrices (Marroig Cheverud, 2001; Prôa et al., 2012), even though a value as low as 0.43 would reject the null hypothesis of structural dissimilarity, at $P(\alpha) < 0.01$ for 35 traits. Therefore, rather than focusing on significance tests, it is more interesting to evaluate these metrics within the context in which they are measured (Haber, 2014). In the present case, the different taxonomic scales compared and the marked differences in life-history phases and development implied by these different scales provide the proper context to interpret these matrix structural similarity metrics. For these reasons, we consider that matrix structural similarities > 0.7 should imply a highly conserved covariance structure.

Due to the sampling error associated with the estimation of covariance and correlation **P**-matrices, the maximum structural similarity (r_{max}) between two matrices will never be one. It will be the geometric mean of repeatabilities (t) of the pair of matrices. Therefore, an adjusted similarity (r_{adj}) can be obtained as the proportion of r_{max} and the observed similarity [r_{obs}; Cheverud (1996)], as follows:

$$r_{adj} = r_{obs} / r_{max} \tag{3}$$

The t value for both correlation and covariance matrices was determined using a Monte Carlo resampling procedure of self-correlation (Porto et al., 2009; Marroig and Cheverud, 2001). For every **P**-matrix, 1,000 bootstrap resamplings of the original data were made after removing sources of variation (Table 1), and keeping sample size constant. Correlation and covariance matrices were estimated for each of the resamples and RS and KP were used to compare the original and the resampled matrices. The t value for both RS and KP was then obtained as the mean comparison value between original and resampled matrices. Please notice this procedure provides

inflated estimates of repeatabilities for poorly estimated **P**-matrices, thus providing conservative corrections of matrices structural similarities when at least one poorly estimated **P**-matrix is considered in the analysis.

Overall magnitude of morphological integration

The overall magnitude of morphological integration was defined for age class and genus specific covariance and correlation **P**-matrices by calculating the relative eigenvalue variance (Pavlicev et al., 2009):

$$Var_{rel}(\lambda) = \frac{\frac{\sum_{i=1}^{N}(\lambda_i - \bar{\lambda})^2}{N}}{N-1} \tag{4}$$

where λ represents the eigenvalues of each **P**-matrix and N represents the number of traits (Pavlicev et al., 2009). The $Var_{rel}(\lambda)$ determines the overall level of correlation among all traits and may vary between zero (all traits are independent) and one (all traits are fully correlated).

Due to differences in sample sizes and its known effect on $Var_{rel}(\lambda)$ estimations (Haber, 2011), we calculated the $Var_{rel}(\lambda)$ as the difference between the observed and the expected values for the specific sample size (Wagner, 1984), which was verified to have virtually no bias for the present sample (Figure A 3).

Statistical analyses

All analyses were done in the R Core Team (2019) programming environment using the package EvolQG (Melo et al., 2015). The only exceptions are the pairwise NP-ANOVAs and MANOVAs that were performed with the package RRPP (Collyer and Adams, 2018, 2019).

Results

For the sake of brevity, we chose to present in the main text only the covariance matrix results, which are very similar to the correlation matrix results (Figure 2, 3, 4 and Figure A 4, 5, 6).

Matrix structural similarity

In the analysis comparing **P**-matrices within genera and between age classes, 77%, and 49% of the observed similarities were above 0.7 and 0.8, respectively, and 83%, and 30% of the adjusted similarities were above 0.8 and 0.9, respectively (Figure 2). For the adjusted similarities, the only values < 0.7 are the comparisons involving *Monodelphis* (B) birth age classes 20 and 40. All RS similarities involving these age classes ranged between 0.24 and 0.45 for observed similarities and between 0.27 and 0.52 for adjusted similarities. The observed KP similarities clustered between 0.54 and 0.62, with the exception of two values that were > 0.7. Adjusted KP similarities were > 0.68.

The comparisons within *Monodelphis* (D), *Monodelphis* (B) and *Calomys* between age class and genus specific **P**-matrices and congeneric **G**-matrices showed very similar results to the comparison of generic **P**-matrices (Figure 3).

Overall magnitude of morphological integration

For *Sapajus* and *Calomys* the difference between the largest and smallest values was slightly below 0.1, and the 95% confidence interval overlaps for all age classes. For *Monodelphis* (B) and *Monodelphis* (D) this difference is ~ 0.19, and the confidence interval do not overlap for all age classes. However, there is no clear trend in the fluctuation of the $Var_{rel}(\lambda)$ values between age classes. In the comparison between corresponding age classes between *Monodelphis* (B) and *Monodelphis* (D), there is overlap between the birth age class 70 and the dental age class two, while for the other classes in some cases there is overlap and in others not. For *Didelphis*, the difference between the largest and the smallest values was ~ 0.22 and there is a trend for the first three dental age classes to have higher $Var_{rel}(\lambda)$ than the other classes (Figure 6).

Discussion

Developmental changes on the additive genetic covariance

Pre- and postnatal development differ drastically between marsupials and placentals and within placentals (Smith, 1997). Moreover, mammalian postnatal ontogeny occurs over major life-history phases (McManus, 1974; Abdala et al., 2001; Nievelt and Smith, 2005) that may have influence over development (Atchley, 1984; Zelditch, 1988; Hallgrímsson et al., 2009; Sibly et al., 2014), such as lactation, beginning of solid food ingestion, and weaning. For instance, these life stages translate into different functional demands in terms of food acquisition (Rayner, 1985), implying that the optimal trait association might change during ontogeny. The postnatal stage is also characterized by several skull morphological changes (Zelditch et al., 1992; Abdala et al., 2001; Flores et al., 2006), such as the faster growth of the viscerocranium in relation to the neurocranium, and the development of muscles of the masticatory apparatus to cope with the ingestion of solid food, which influences the growth of underlying bones (Enlow and Hans, 1996; Moss, 1997). Thus, it is not surprising that our results support previous findings that skull trait covariances changes during postnatal ontogeny (Hallgrímsson et al., 2009; Mitteroecker and Bookstein, 2009; Atchley, 1984; Coleman et al., 1994; Zelditch, 1988; Zelditch and Carmichael, 1989; Zelditch et al., 2006, 1992; Mitteroecker et al., 2012; Nonaka and Nakata, 1984; Goswami et al., 2012).

However, it is remarkable that for marsupials (*Didelphis* and *Monodelphis*), an altricial placental (*Calomys*), and a precocial placental (*Sapajus*) the adjusted matrix structural similarities from weaning to adulthood were high (> 0.7, often > 0.8). This means that the populations covariance structure for skull traits during this part of ontogeny respond similarly to different selection gradients (RS results) and share aligned morphological subspaces (KP results). However *Monodelphis* (B) during lactation (age classes 20 and 40) showed a different pattern, presenting systematically low RS adjusted matrix structural similarities (< 0.7). This suggest that at least in some marsupials the skull trait covariance is subject to pronounced changes during early stages of postnatal development, which will stabilize only around weaning. Unfortunately, our data is insufficient to explore this issue in depth (Figure 1). First, *Didelphis* during lactation (dental age

class zero) presented high adjusted structural similarity with all other dental age classes. This may be a real biological phenomenon, but it is likely the consequence of sampling specimens based on dental age classes (see further discussion in the next section). Second, we did not sample placental mammals before weaning (Figure 1). Consequently, more studies focusing on this ontogenetic stage will be required to better understand this phenomenon in therian mammals.

Regarding the overall magnitude of morphological integration, our results are hard to interpret in most cases due to confounding factors related to sampling and biological aspects (see next section for further discussion). Nevertheless, even in the case of *Didelphis* and *Monodelphis* (D) that have noticeable differences in the overall magnitude of morphological integration among classes, this variation does not affect how these **P**-matrices respond to different selection gradients (i.e. RS results are similar).

The high adjusted structural similarities for *Monodelphis* and *Calomys* between **P**-matrices at different ontogenetic stages and their congeneric **G**-matrices suggest that at least from weaning onwards the **P**-matrix is an appropriate surrogate for the **G**-matrix. Our results expand the age range for which covariance patterns are similar across post-weaning age points, as they have been found to correspond to patterns exhibited by adult specimens. (Cheverud, 1988, 1996; Roff, 1995; Akesson et al., 2007; Porto et al., 2009; Reusch and Blanckenhorn, 1998; House and Simmons, 2005).

We have sampled species that are distantly related and that show profoundly different pre- and postnatal developmental strategies. We suggest that our results can be extended to all therian mammals because there is no a priori reason to believe evolutionary lineages with more similar developmental patterns, such as within placentals or marsupials, will have a higher influence of developmental changes over **G**-matrices. Future studies with comparable methods will no doubt be important to evaluate this conclusion.

Our findings have important implications for evolutionary, genetic, and ecological studies. We showed that during lactation **G**-matrices may vary considerably, but that from weaning onwards **G**-matrices become relatively stable. Thus, for specimens spanning from weaning to adulthood, selection operating on different postnatal ontogenetic stages will have similar consequences in

terms of the responses produced. Since specimens during lactation would be subjected to very different selective regimes when compared to specimens from weaning onwards, we consider that our results are reassuring that working with a single ontogenetic stage will not lead to misleading conclusions.

At this point we can only speculate on the reasons for this conservation of the **G**-matrix during most of the postnatal ontogeny. The **G**-matrices observed in adult populations are the consequence of several hierarchical developmental processes (Hallgrímsson et al., 2009, 2007). The later the developmental input, the smaller may be its effects on **G**-matrix determination (Atchley, 1984). Most postnatal development is shaped by growth in size and muscle and bone interactions, which are considered relatively later developmental inputs (Hallgrímsson et al., 2009, 2007), that may have limited influence over the **G**-matrix determination. In addition, growth and muscle and bone interactions may exert a buffering effect on the skull form as a whole, while allowing for localized variation (Mitteroecker et al., 2020; Marroig et al., 2011). Alternatively, the **G**-matrix may be maintained by stabilizing selection, which selects against postnatal developmental processes that could generate disruptions in the covariance structure (Arnold et al., 2008; Melo and Marroig, 2014).

Influence of sample size and dental age class

Our results should be interpreted with caution. Low ratios between sample size and number of traits introduce considerable error in the estimation of matrices (Marroig et al., 2012). In terms of matrix structural similarity, poor sampling will most likely decrease the similarity between two matrices (Hubbe et al., 2016; Machado et al., 2018). This probably explains most of the lower similarities observed (note the clustering of warmer colors on the left of Figure 2, and 3). Even adjusted matrix structural similarities are probably understamited for comparisons involving poorly estimated matrices, since these similarities were determined based on a conservative approach.

However, sampling error is probably not the sole explanation for the discrepant low values for the observed matrix structural similarity involving the birth age classes 20 and 40 for *Monodelphis*

(B), which varied between 0.24 and 0.45. First, all **P**-matrices of *Monodelphis* (B) can be considered poorly estimated (Table 1), and it is suggestive that the only observed matrix structural similarities lower than < 0.6 were the ones involving at least one of the two youngest age classes (Figure 2 and 3). To further explore whether this result could be explained mainly by sampling, we did a rarefaction analysis comparing the **P**-matrix from *Monodelphis* (D) from dental age classes two and seven, which show high structural similarity (observed RS of 0.88). We did not use any **P**-matrix from *Monodelphis* (B) due to small sample sizes. We used RS to compare 1,000 bootstrap resamplings of the original data after removing sources of variation (Table 1), and keeping sample size constant at 13. The 95% confidence interval ranged from 0.51 to 0.80. Thus, it is likely that at least part of the observed results for the structural similarity involving *Monodelphis* (B) birth age classes 20 and 40 are due to developmental changes and not sampling artifacts.

Regarding the overall magnitude of morphological integration, low ratios between sample size and number of traits has unpredictable consequences on the statistics (Figure A 7). In other words, error may result in the overall magnitude of morphological integration among age classes to be more similar or dissimilar. As a consequence, it is impossible to determine if the differences observed among the $Var_{rel}(\lambda)$ within *Calomys*, *Sapajus*, and *Monodelphis* (B) are due to biology or sampling error. On the other hand, sampling error is likely negligible to interpret results for *Monodelphis* (D) and *Didelphis*. In addition, the 95% confidence interval determined based on the bootstrap procedures should be interpreted with caution for cases with relatively low $Var_{rel}(\lambda)$ (Figure A 8).

Another source of potential bias in our results is the strategy of sampling dental age classes, which combine in a single age class specimens with different absolute ages (Richtsmeier et al., 1993; van Nievelt and Smith, 2005). This may inflate the total amount of morphological variation explained by size, particularly for dental age classes in which specimens are experiencing higher growth rates (Figure A 9), since relatively small variation in absolute ages within the dental age class will result in relatively large variation in size. The implication of this sampling strategy is that when the influence of size is overestimated in both matrices, the matrix structural similarity will be inflated, and will decrease when comparing one matrix with an inflated influence of size

and another that was not. This, for instance, may explain why *Didelphis* during lactation (dental age class zero) showed high structural matrix similarities with other dental age classes, while *Monodelphis* (B) during lactation (birth age classes 20 and 40) did not. The overall magnitude of morphological integration will inflate. This, for instance, may explain the larger overall magnitude of morphological integration observed for younger age classes for *Didelphis*.

However, it is remarkable how results from dental age classes and birth age classes show the same general pattern of high structural similarities between matrices from different ontogenetic stages. Thus, in spite of sampling limitations, we argue that our results reflect a real biological phenomenon.

Conclusion

Our findings suggest that at least for therian mammals, one of the major groups in ecological systems, the study of the life-history changes and evolutionary consequences under selection (or genetic drift) is much facilitated by shared and common patterns of additive genetic covariance among traits during most of postnatal ontogeny. Thus, even though selection might be operating in different directions during most of postnatal ontogeny, due to differences in life-history phases and fitness components, the net response to such pressures will not be biased by differences in the **G**-matrix during the postnatal ontogeny.

While our findings support the investigation of selective pressures and evolutionary responses in natural systems based on a single weaning-onward ontogenetic stage, they also highlight the need for a more comprehensive understanding of how the **G**-matrix changes between birth and weaning. In addition, it is important to better understand species life histories to evaluate when and how selection is operating. This is particularly relevant if species show relatively drastic change in some ecological aspect during ontogeny (e.g. Drago et al. (2009); Tanner et al. (2010)).

Acknowledgment

The authors are thankful to people and institutions that provided generous help and access to collections: A. Fleming, R. MacPhee, R. Voss, and E. Westwig (American Museum of Natural History); B. Patterson, L. Heaney, and B. Stanley (Field Museum of Natural History); S. Costa, and J. de Queiros (Museu Paraense Emilio Goeldi); F. Barbosa, S. Franco, J. A. de Oliveira, and L. Salles (Museu Nacional); C. Conroy, E. Lacey and J. Patton (Museum of Vertebrate Zoology); M. Brett-Surman, R. Chapman, L. Gordon, and D. Lunde (National Museum of Natural History); J. Gualda, and M. de Vivo (Museu de Zoologia da Universidade de São Paulo); and J. VandeBerg (Texas Biomedical Research Institute). This research was supported by grants and fellowships from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Museum of Vertebrate Zoology (UC-Berkeley), American Museum of Natural History, and Field Museum of Natural History.

References

Abdala, F., Flores, D. A., and Giannini, N. P. (2001). Postweaning ontogeny of the skull of Didenphis albiventris. *Journal of Mammalogy*, 82:190–200.

Akesson, M., Bensch, S., and Hasselquist, D. (2007). Genetic and phenotypic associations in morphological traits: a long term study of great reed warblers Acrocephalus arundinaceus. *Journal of Avian Biology*, 38:58–72.

Alberch, P. (1982). Developmental constraints in evolutionary processes. In Bonner, J. T., editor, *Development in Evolution*, pages 313–332. Springer, Berlin and New York.

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1):32–46.

Arnold, S. J., Bürger, R., Hohenlohe, P. A., Ajie, B. C., and Jones, A. G. (2008). Understanding the evolution and stability of the G-matrix. *Evolution*, 62(10):2451–2461.

Atchley, W. R. (1984). Ontogeny, Timing of Development, and Genetic Variance-Covariances Structure. *The American Naturalist*, 123(4):519–540.

Blows, M. W., Chenoweth, S. F., and Hine, E. (2004). Orientation of the genetic variance-covariance matrix and the fitness surface for multiple male sexually selected traits. *American Naturalist*, 163(3):329–340.

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

Cheverud, J. M. (1988). A comparison of genetic and phenotypic correlations. *Evolution*, 42(5):958–968.

Cheverud, J. M. (1995). Morphological integration in the Saddle-back Tamarin (Saguinus-fuscicollis) cranium. *American Naturalist*, 145(1):63–89.

Cheverud, J. M. (1996). Quantitative genetic analysis of cranial morphology in the cotton-top (Saguinus oedipus) and saddle-back (S. fuscicollis) tamarins. *Journal of Evolutionary Biology*, 9(1):5–42.

Cheverud, J. M. and Marroig, G. (2007). Comparing covariance matrices: Random skewers method compared to the common principal components model. *Genetics and Molecular Biology*, 30(2):461–469.

Coleman, J. S., McConnaughay, K. D., and Ackerly, D. D. (1994). Interpreting phenotypic variation in plants. *Trends in Ecology & Evolution*, 9(5):187–191.

Collyer, M. L. and Adams, D. C. (2018). RRPP: An r package for fitting linear models to high-dimensional data using residual randomization.

Collyer, M. L. and Adams, D. C. (2019). RRPP: Linear model evaluation with randomized residuals in a permutation procedure. r package version 0.4.0.

Drago, M., Cardona, L., Crespo, E. A., and Aguilar, A. (2009). Ontogenic dietary changes in South American sea lions. *Journal of Zoology*, 279(3):251–261.

Eisenberg, J. F. (1989). *Mammals of the Neotropics, Volume 1: The Northern Neotropics: Panama, Colombia, Venezuela, Guyana, Suriname, French Guiana*. The University of Chicago Press, Chicago.

Eisenberg, J. F. and Redford, K. H. (1999). *Mammals of the Neotropics, Volume 3: The Central Neotropics: Ecuador, Peru, Bolivia, Brazil*. The University of Chicago Press, Chicago.

Falconer, D. S. and MacKay, T. F. C. (1996). *Introduction to quantitative genetics*. Longman, New York.

Flores, D. A., Giannini, N., and Abdala, F. (2006). Comparative postnatal ontogeny of the skull in the australidelphian metatherian Dasyurus albopunctatus (Marsupialia: Dasyuromorpha: Dasyuridae). *Journal of Morphology*, 267(4):426–440.

Fragazy, D. M., Visalberghi, E., and Fedigan, L. M. (2004). *The Complete Capuchin: The Biology of the Genus Cebus*. Cambridge University Press, Cambridge.

Garcia, 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. (2006). Morphological integration in the carnivoran skull. *Evolution*, 60(1):169–183.

Goswami, A., Polly, P. D., Mock, O. B., and SÁNchez-Villagra, M. R. (2012). Shape, variance and integration during craniogenesis: contrasting marsupial and placental mammals. *Journal of Evolutionary Biology*, 25(5):862–872.

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

Haber, A. (2014). The Evolution of Morphological Integration in the Ruminant Skull. *Evolutionary Biology*.

Hall, B. K. (1999). *Evolutionary developmental biology*. Dordrecht: Kluwer.

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

Hallgrímsson, B., Lieberman, D. E., Young, N. M., Parsons, T., and Wat, S. (2007). Evolution of Covariance in the Mammalian Skull. In Bock, G. and Goode, J., editors, *Tinkering: The Microevolution of Development*, pages 164–190. John Wiley & Sons, Ltd.

Hendrikse, J. L., Parsons, T. E., and Hallgrímsson, B. (2007). Evolvability as the proper focus of evolutionary developmental biology. *Evolution & Development*, 9(4):393–401.

Hingst-Zaher, E., Marcus, L. F., and Cerqueira, R. (2000). Application of geometric morphometrtcs to the study of postnatal size and shape changes in the skull of Calomys expdsus. *Hystrix*, 11:99–113.

House, C. M. and Simmons, L. W. (2005). The evolution of male genitalia: patterns of genetic variation and covariation in the genital sclerites of the dung beetle Onthophagus taurus. *Journal of Evolutionary Biology*, 18:1281–1292.

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.

Krzanowski, W. (2000). *Principles of multivariate analysis: a user's perspective. Revised edition*. Oxford, Clarenton.

Krzanowski, W. J. (1979). Between-groups comparison of principal components. *Journal of the American Statistical Association*, 74(367):703–707.

Lande, R. and Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution*, 37(6):1210–1226.

Leamy, L. and Cheverud, J. M. (1984). Quantitative genetics and the evolution of ontogeny. II. Genetic and environmental correlations among age-specific characters in randombred house mice. *Growth*, 48(3):339–353.

Lessells, C. M. and Boag, P. T. (1987). Unrepeatable repeatabilities: a common mistake. *Auk*, 104(1):116–121.

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

Marroig, G. and Cheverud, J. M. (2001). A comparison of phenotypic variation and covariation patterns and the role of phylogeny. Ecology, and ontogeny during cranial evolution of new world monkeys. *Evolution*, 55(12):2576–2600.

Marroig, G. and Cheverud, J. M. (2004). Cranial evolution in sakis (Pithecia, platyrrhini) I: Interspecific differentiation and allometric patterns. *American Journal of Physical Anthropology*, 125(3):266–278.

Marroig, G., Melo, D., Porto, A., Sebastiao, H., and Garcia, G. (2011). Selection Response Decomposition (SRD): A New Tool for Dissecting Differences and Similarities Between Matrices. *Evolutionary Biology*, 38(2):225–241.

Marroig, G., Melo, D. A. R., and Garcia, G. (2012). Modularity, noise, and natural selection. *Evolution*, 66(5):1506–1524.

McArdle, B. H. and Anderson, M. J. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology*, 82(1):290–297. Publisher: Wiley Online Library.

McManus, J. (1974). Didelphis virginiana. *Mammalian Species*, 40:1–6.

Melo, D., Garcia, G., Hubbe, A., Assis, A., and Marroig, G. (2015). EvolQG - An R package for evolutionary quantitative genetics [version 1; referees: awaiting peer review]. *F1000Research*, 4(925).

Melo, D. and Marroig, G. (2014). Directional selection can drive the evolution of modularity in complex traits. *Proceedings of the National Academy of Sciences*, page 201322632.

Mitteroecker, P., Bartsch, S., Erkinger, C., Grunstra, N. D., Le Maître, A., and Bookstein, F. L. (2020). Morphometric variation at different spatial scales: coordination and compensation in the emergence of organismal form. *Systematic Biology*.

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., Gunz, P., Neubauer, S., and Mueller, G. (2012). How to Explore Morphological Integration in Human Evolution and Development? *Evolutionary Biology*, 39(4):536–553.

Nievelt, A. F. H. v. and Smith, K. K. (2005). To Replace or Not to Replace: The Significance of Reduced Functional Tooth Replacement in Marsupial and Placental Mammals. *Paleobiology*, 31(2):324–346. Publisher: Paleontological Society.

Nonaka, K. and Nakata, M. (1984). Genetic variation and craniofacial growth in inbred rats. *Journal of Craniofacial Genetics and Developmental Biology*, 4:271–302.

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.

Porto, A., Oliveira, F. B. D., Shirai, L. T., Conto, V. D., and Marroig, G. (2009). The evolution of modularity in the mammalian skull I : Morphological integration patterns and magnitudes. *Evolutionary Biology*, 36:118–135.

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

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rayner, J. M. V. (1985). Linear relations in biomechanics: the statistics of scaling functions. *Journal of Zoology*, 206(3):415–439. _eprint: https://zslpublications.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-7998.1985.tb05668.x.

Redford, K. H. and Eisenberg, J. F. (1992). *Mammals of the Neotropics, Volume 2: The Southern Cone: Chile, Argentina, Uruguay, Paraguay*. University of Chicago Press, Chicago.

Reusch, T. and Blanckenhorn, W. U. (1998). Quantitative genetics of the dung fly Sepsis cynipsea: Cheverud's conjecture revisited. *Heredity*, 81(1):111–119.

Richtsmeier, J. T., Corner, B. D., Grausz, H. M., Cheverud, J. M., and Danahey, S. E. (1993). The Role of Postnatal Growth Pattern in the Production of Facial Morphology. *Systematic Biology*, 42(3):307–330.

Roff, D. A. (1995). The estimation of genetic correlations from phenotypic correlations: a test of Cheverud's conjecture. *Heredity*, 74:481–490.

Sebastião, H. and Marroig, G. (2013). Size and shape in cranial evolution of 2 marsupial genera: Didelphis and Philander (Didelphimorphia, Didelphidae). *Journal of Mammalogy*, 94(6):1424–1437.

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.

Sibly, R. M., Grady, J. M., Venditti, C., and Brown, J. H. (2014). How body mass and lifestyle affect juvenile biomass production in placental mammals. *Proceedings of the Royal Society B: Biological Sciences*, 281(1777):20132818.

Smith, K. K. (1997). Comparative patterns of craniofacial development in Eutherian and Metatherian mammals. *Evolution*, 51:1663–1678.

Tanner, J. B., Zelditch, M. L., Lundrigan, B. L., and Holekamp, K. E. (2010). Ontogenetic change in skull morphology and mechanical advantage in the spotted hyena (Crocuta crocuta). *Journal of Morphology*, 271(3):353–365.

Tribe, C. J. (1990). Dental age classes in Marmosa incana and other didelphoids. *Journal of Mammalogy*, 71(4):566–569.

Tyndale-biscoe, C. H. and Mackenzie, R. B. (1976). Reproduction in Didelphis marsupialis and Didelphis albiventris in Colombia. *Journal of Mammalogy*, 57(2):249–265.

van Nievelt, A. F. H. and Smith, K. K. (2005). Tooth eruption in Monodelphis domestica and its significance for phylogeny and natural history. *Journal of Mammalogy*, 86:333–341.

Wagner, G. P. (1984). On the eigenvalue distribution of genetic and phenotypic dispersion matrices: evidence for a nonrandom organization of quantitative character variation. *Journal of Mathematical Biology*, 21(1):77–95. Publisher: Springer.

Zelditch, M. L. (1988). Ontogenetic Variation in Patterns of Phenotypic Integration in the Laboratory Rat. *Evolution*, 42(1):28–41.

Zelditch, M. L., Bookstein, F. L., and Lundrigan, B. L. (1992). Ontogeny of Integrated Skull Growth in the Cotton Rat Sigmodon fulviventer. *Evolution*, 46(4):1164–1180.

Zelditch, M. L. and Carmichael, A. C. (1989). Ontogenetic Variation in Patterns of Developmental and Functional Integration in Skulls of Sigmodon fulviventer. *Evolution*, 43(4):814–824.

Zelditch, M. L., Mezey, J., Sheets, H. D., Lundrigan, B. L., and Garland, T. (2006). Developmental regulation of skull morphology II: ontogenetic dynamics of covariance. *Evolution & Development*, 8(1):46–60.

Tables

Table 1: Sample sizes (Females/Males) for each age class and species and controlled fixed effects for each genus. Boldfaced sample sizes indicate that covariance/correlation matrices were estimated for that particular age class and genus combination. S: sex; SP: species.

			Dental Age Class											**Sources of variation controlled**
Order	**Genus**	**Species**	**Zero**	**One**	**Two**	**Three**	**Four**	**Five**	**Six**	**Seven**			**Adults**[a]	
Didelphimorphia	*Didelphis*		**19/29**	**72/89**	**91/86**	**53/63**	**72/70**	**74/65**	**59/69**	**78/79**			**137/148**	S, SP, S x SP
		D. aurita	0/0	7/9	20/22	8/4	9/16	13/10	8/9	27/22				
		D. marsupialis	2/9	28/34	35/33	20/20	27/18	22/24	21/19	17/14				
		D. virginiana	17/20	37/46	36/31	25/39	36/36	39/31	30/41	34/43				
	Monodelphis (D)				**21/27**	**36/40**	**19/21**	**44/36**	**40/45**	**39/63**			**85/102**	S, SP
		M. brevicaudata			2/7	3/8	0/1	8/7	8/9	15/11				
		M. domestica			15/18	29/25	18/20	15/14	16/17	10/16				
		M. glirina			0/2	1/1	1/0	13/10	13/8	9/23				
		M. palliolata			4/0	3/6	0/0	8/5	3/11	5/13				
Primates	*Sapajus*			**5/13**	**8/9**	**15/16**		**10/21**	**57/70**				**57/70**	S
		S. apella		2/10	1/5	13/11		10/21	57/70					
		sp indet		3/3	7/4	2/5		0/0	0/0					
			Birth Age Class											
			20	**30**	**40**	**50**	**70**	**100**	**150**	**200**	**300**	**400**	**Adults**[a]	
Didelphimorphia	*Monodelphis* (B)	*M. domestica*	**13/8**		**9/8**		**6/9**		**12/12**		8/5	**10/9**	**18/14**	S
Rodentia	*Calomys*	*C. expulsus*	**14/7**	**14/11**		**25/27**		**35/25**		**29/18**	**16/14**	**12/19**	**168**	S

[a] refer to text for adult definition.

601

Figure legends

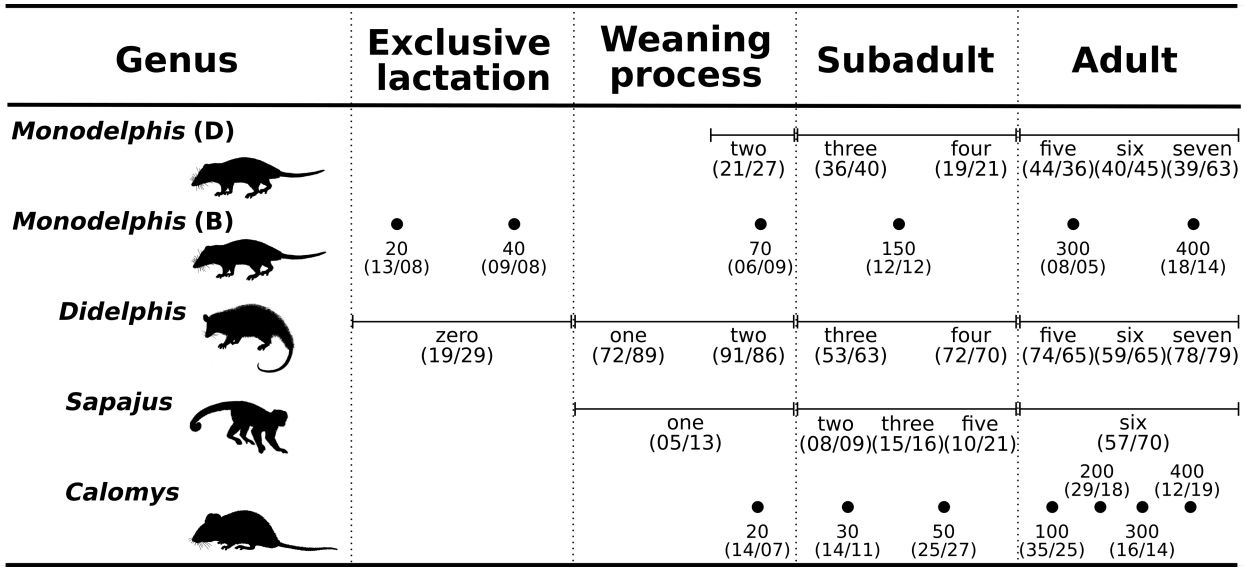

Figure 1: Schematic representation of the age classes in relation to major life-history phases for the genera studied. Birth age classes are represented by dots and dental age classes by error bars. The position of these symbols and the length of the error bars are not precise representations and are intended to show the broad distribution of data over the major life-history phases and differences in both sampling strategies (dental and birth age classes). Values in parentheses are female and male sample sizes. Figures for each genus are adapted from (Eisenberg, 1989; Eisenberg and Redford, 1999; Redford and Eisenberg, 1992). Those figures are not in scale.

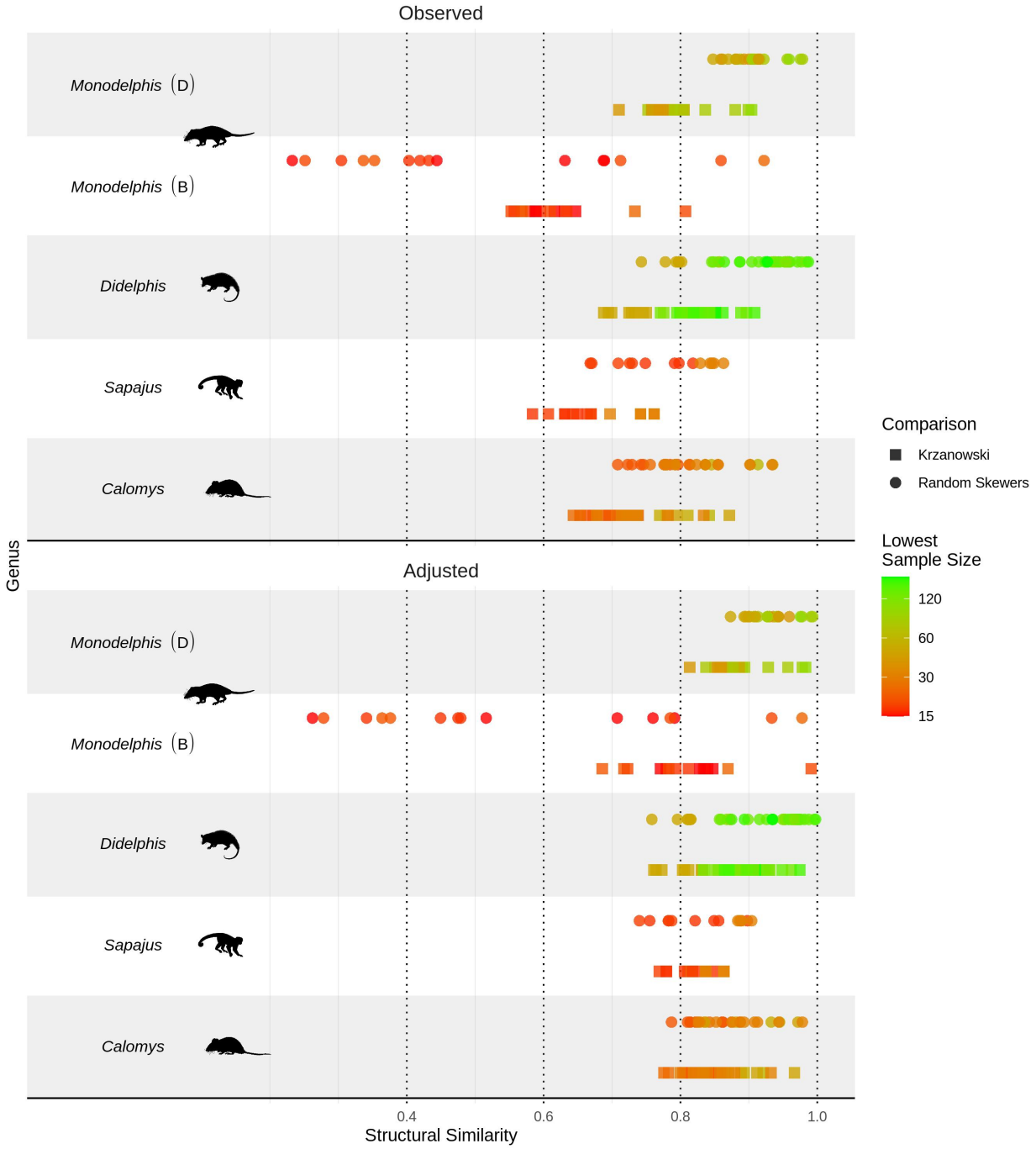

Figure 2: Distribution of Random Skewers and Krzanowski Projection values for age class and genus specific covariance matrix comparisons. The upper panel (Observed) depicts raw pairwise comparison values; in the lower panel (Adjusted), pairwise comparisons were normalized using the geometric mean of repeatabilities (t). The color scale represents the lowest sample size in each comparison. Figures for each genus are adapted from (Eisenberg, 1989; Eisenberg and Redford, 1999; Redford and Eisenberg, 1992). Those figures are not in scale.

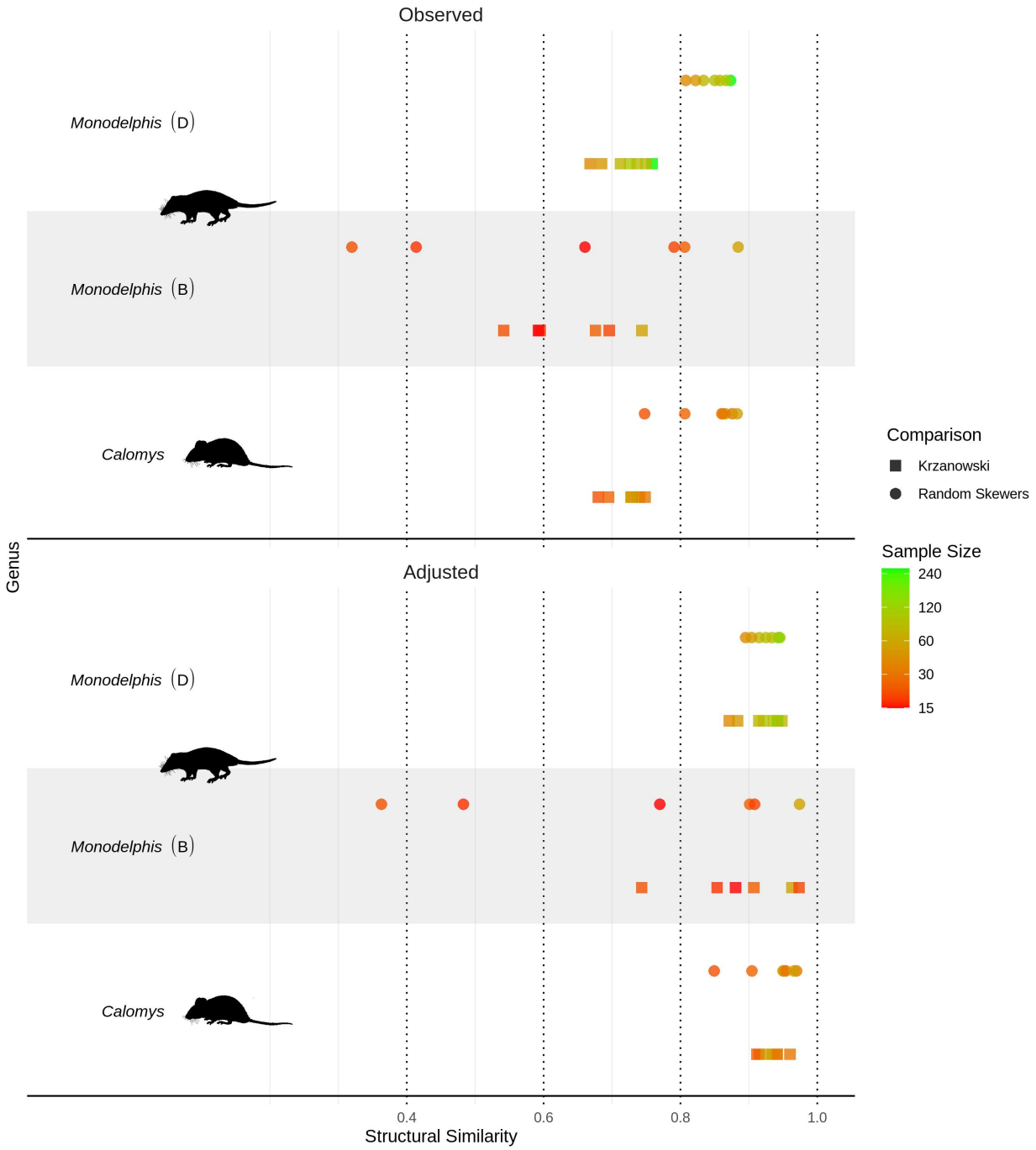

Figure 3: Distribution of Random Skewers and Krzanowski Projection values for covariance matrix comparisons between each **P**-matrix and the corresponding **G**-matrix for *Calomys* and *Monodelphis*. The upper panel (Observed) depicts raw pairwise comparison values; in the lower panel (Adjusted), pairwise comparisons were normalized using the geometric mean of repeatabilities (t). The color scale represents the sample size for the **P**-matrix in each comparison. Figures for each genus are adapted from (Eisenberg, 1989; Eisenberg and Redford, 1999; Redford and Eisenberg, 1992). Those figures are not in scale.

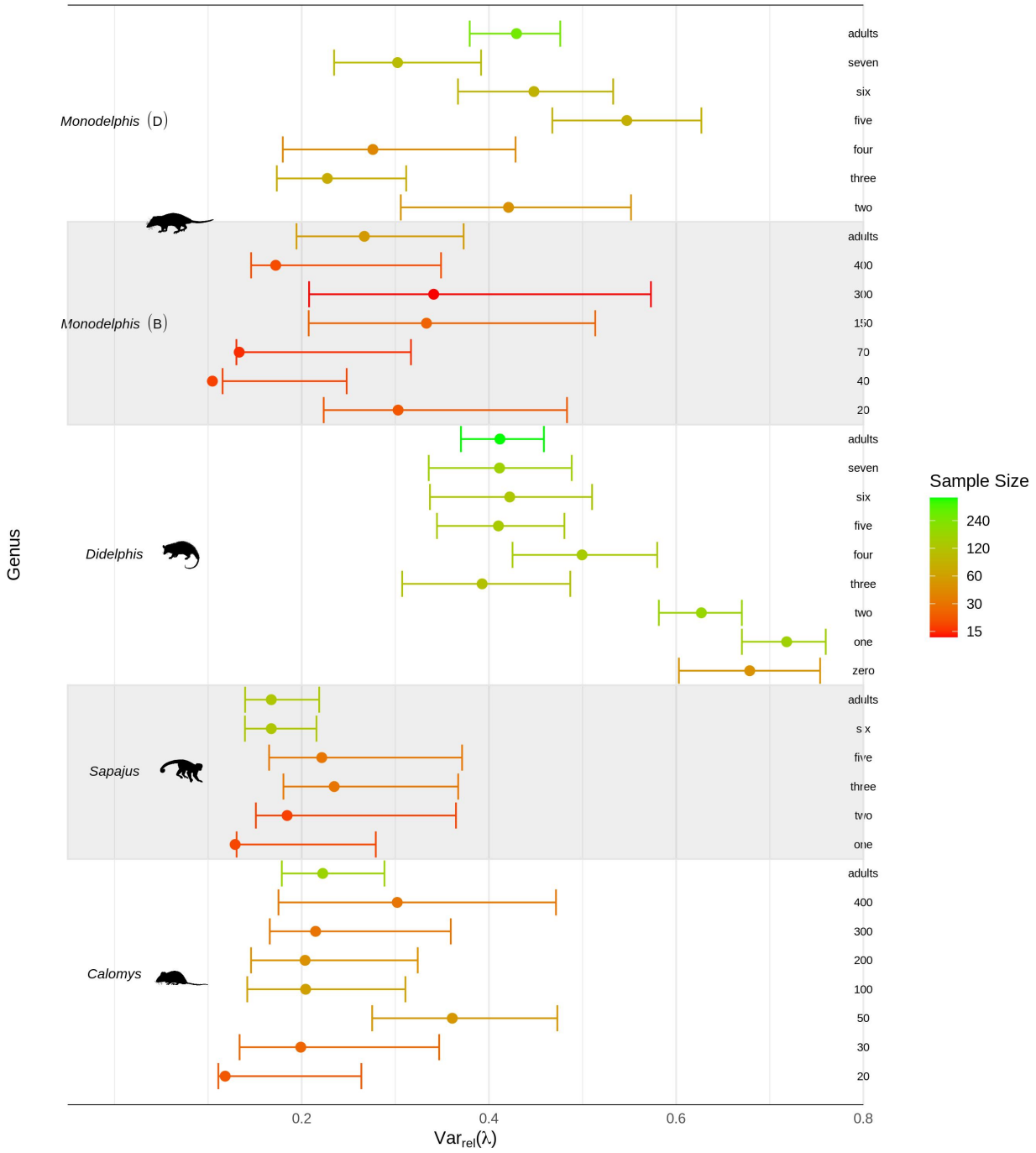

Figure 4: Distribution of age class and genus specific overall magnitude of morphological integration ($Var_{rel}(\lambda)$) values. Non-parametric bootstrap was used to estimate 95% confidence intervals. Figures for each genus are adapted from (Eisenberg, 1989; Eisenberg and Redford, 1999; Redford and Eisenberg, 1992). Those figures are not in scale.

Appendix A: Supplementary Figures

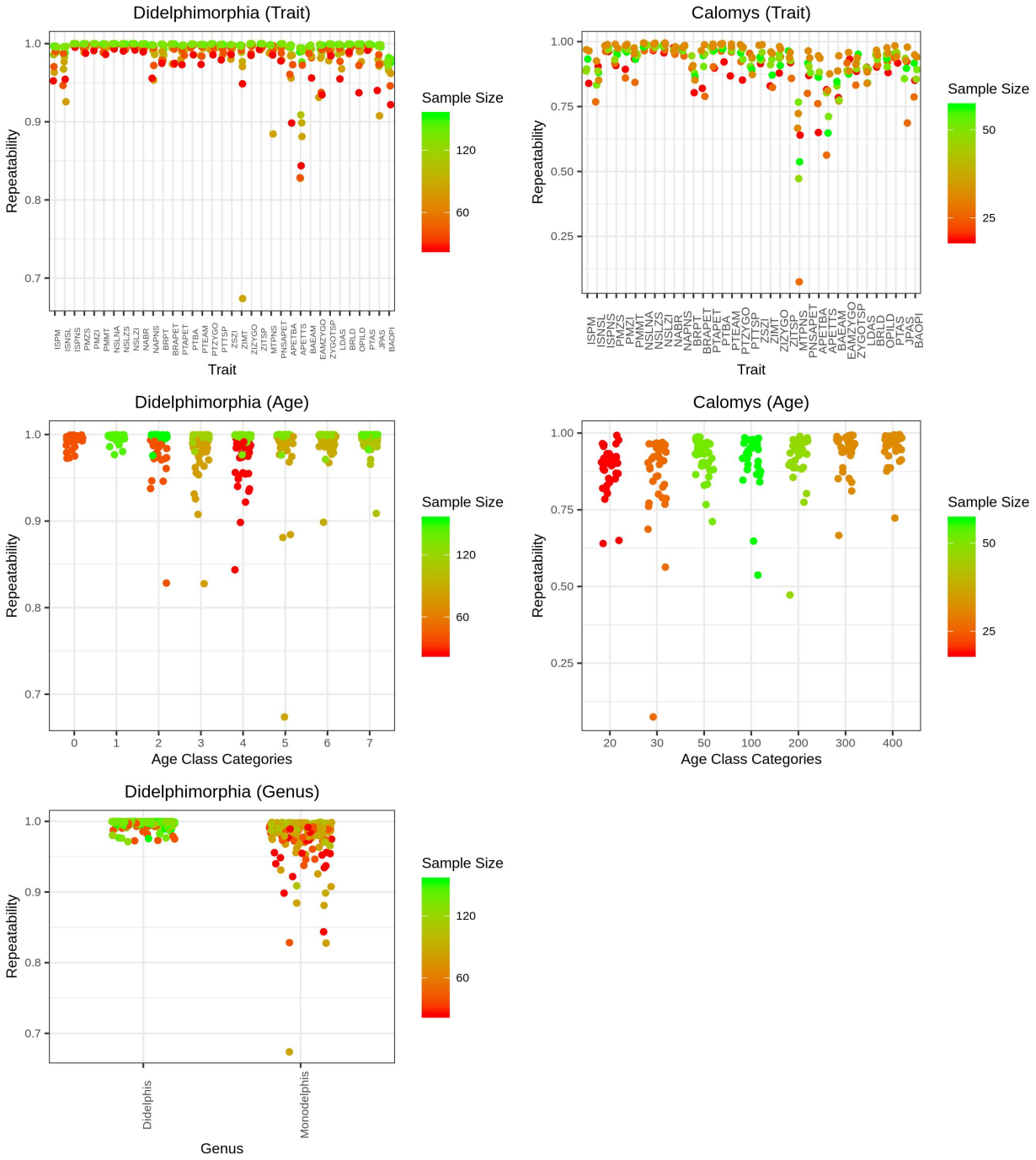

Figure 1: Repeatabilities for all measured traits for *Didelphis*, *Monodelphis* (D) and *Calomys* separated by trait, age classes and genus (for Didelphimorphia only).

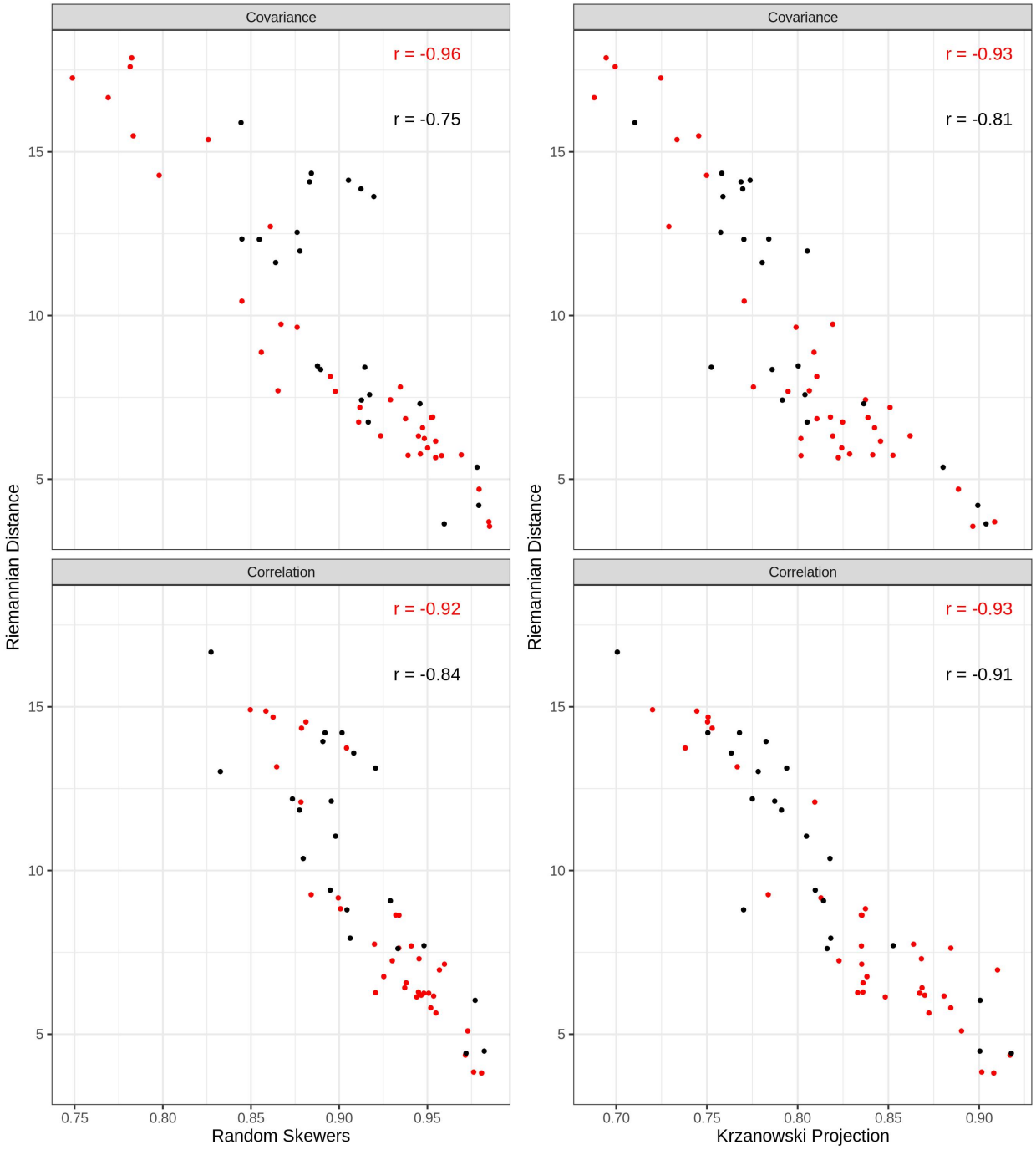

Figure 2: Riemannian Distance against Random Skewers or Krzanowski Projection for comparisons between age classes within *Didelphis* (red dots) and *Monodelphis* (D) (black dots), which were the genera with sufficient sample sizes in all age classes.

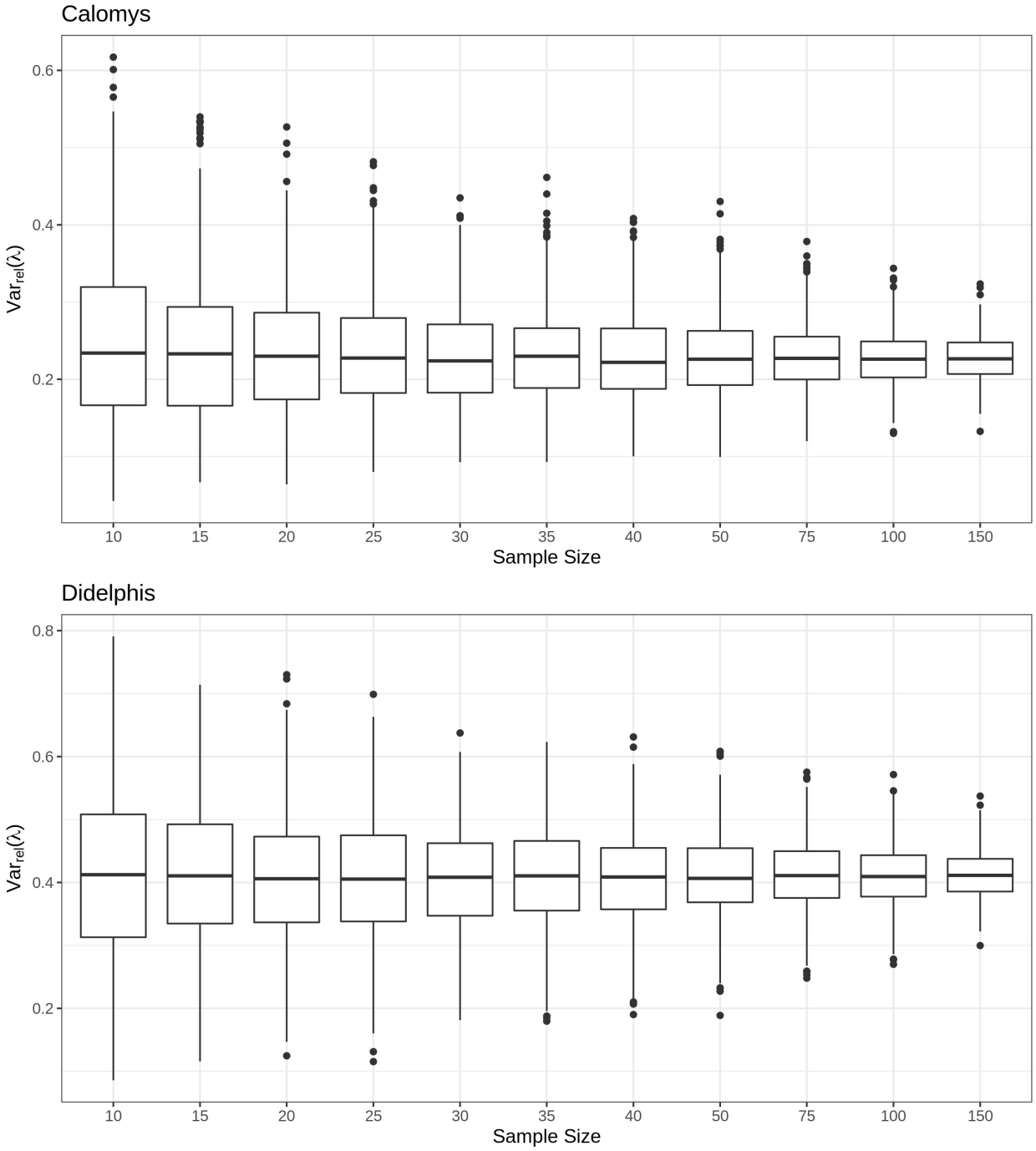

Figure 3: $Var_{rel}(\lambda)$ rarefaction analysis for *Calomys* and *Didelphis* adult covariance **P**-matrices based on 1000 bootstrap resamplings of the original data after removing sources of variation (Table 1 from the main text), and keeping sample size constant.

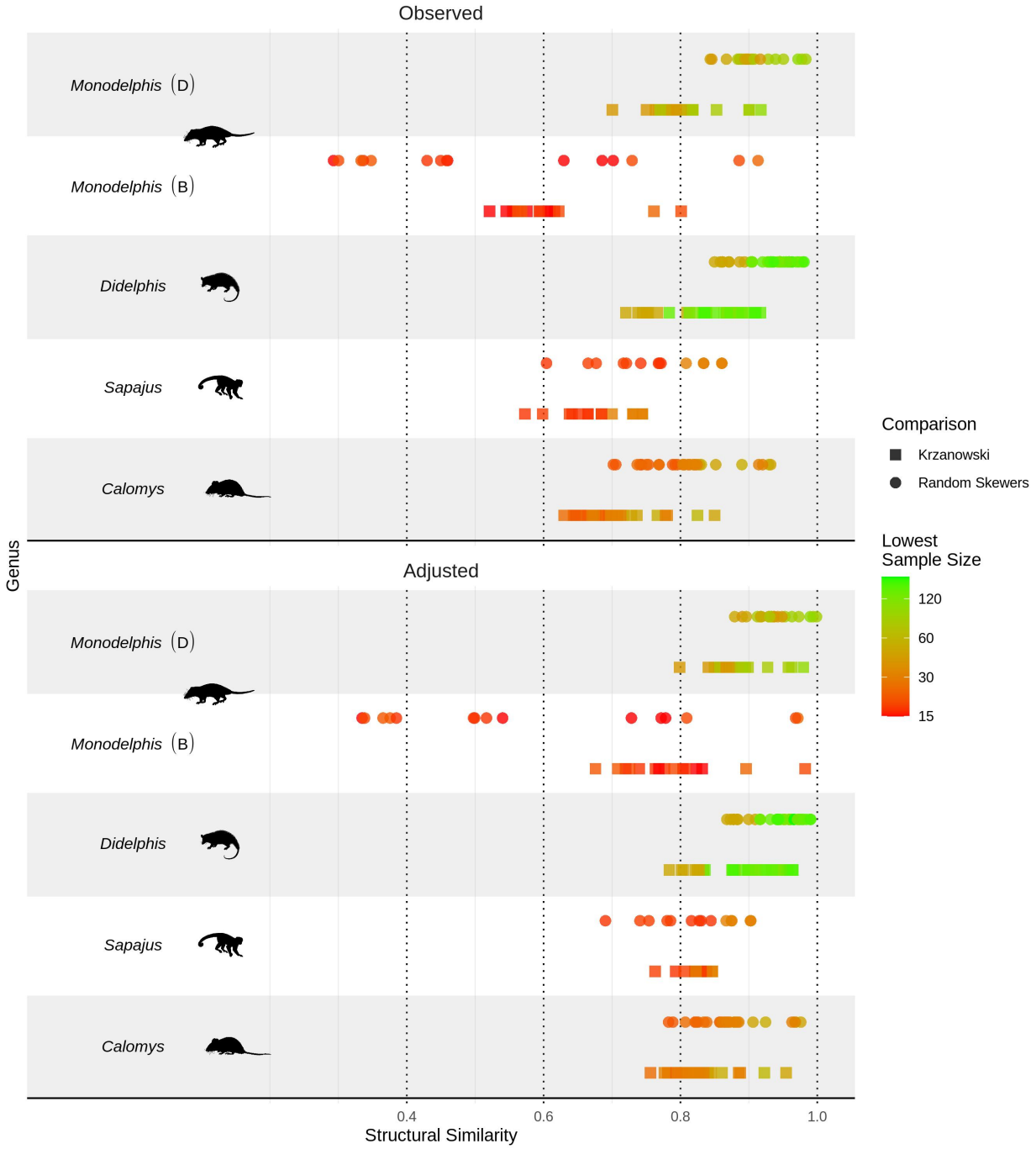

Figure 4: Distribution of Random Skewers and Krzanowski Projection values for correlation matrix comparisons between age classes within each genus. The upper panel (Observed) depicts raw pairwise comparison values; in the lower panel (Adjusted), pairwise comparisons were normalized using the geometric mean of repeatabilities (t). The color scale represents the lowest sample size in each comparison. Figures for each genus are adapted from (Eisenberg, 1989; Eisenberg and Redford, 1999; Redford and Eisenberg, 1992). Those figures are not in scale.

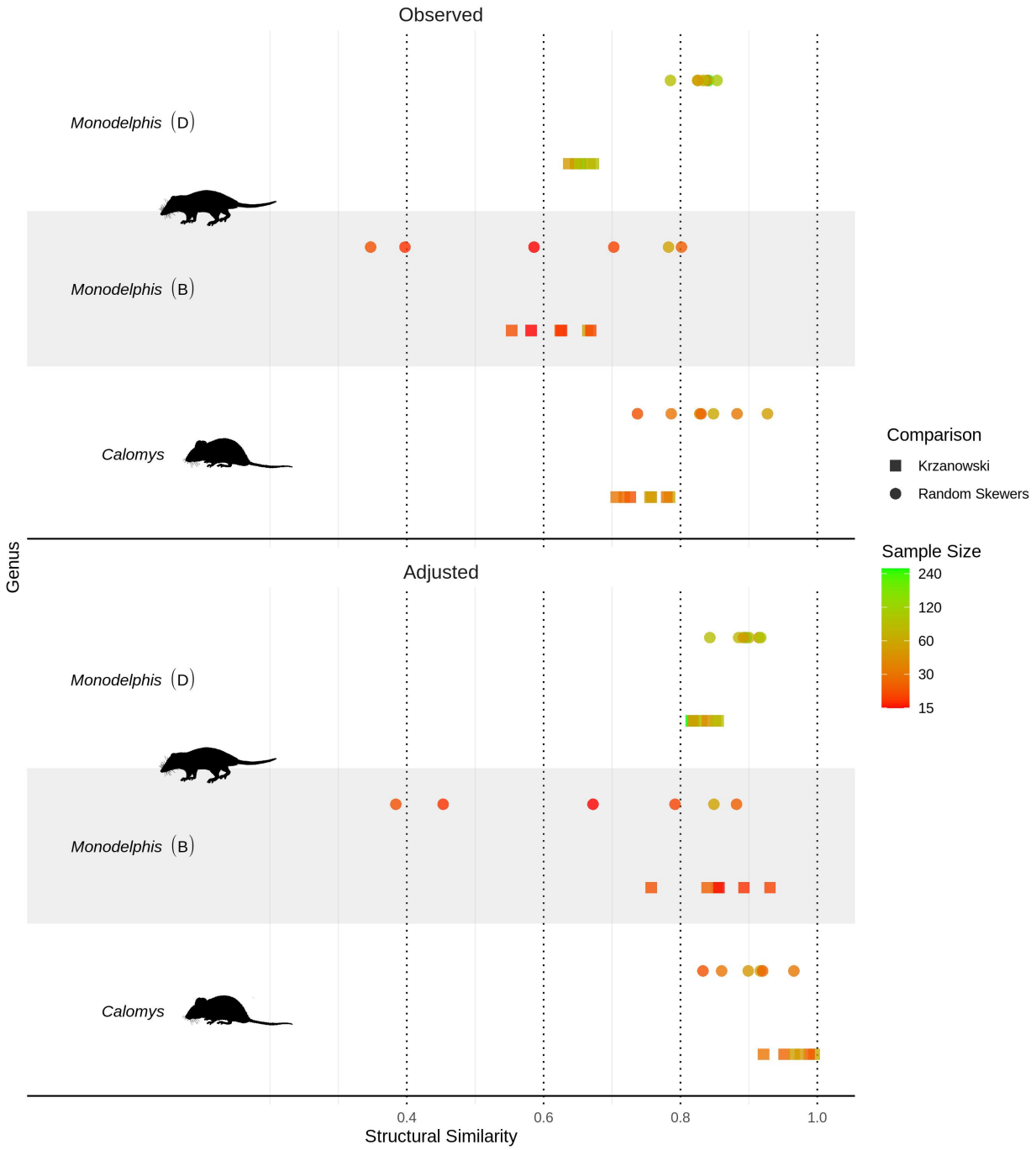

Figure 5: Distribution of Random Skewers and Krzanowski Projection values for correlation matrix comparisons between each **P**-matrix and the correspondent **G**-matrix for *Calomys* and *Monodelphis*. The upper panel (Observed) depicts raw pairwise comparison values; in the lower panel (Adjusted), pairwise comparisons were normalized using the geometric mean of repeatabilities (t). The color scale represents the sample size for the **P**-matrix in each comparison. Figures for each genus are adapted from (Eisenberg, 1989; Eisenberg and Redford, 1999; Redford and Eisenberg, 1992). Those figures are not in scale.

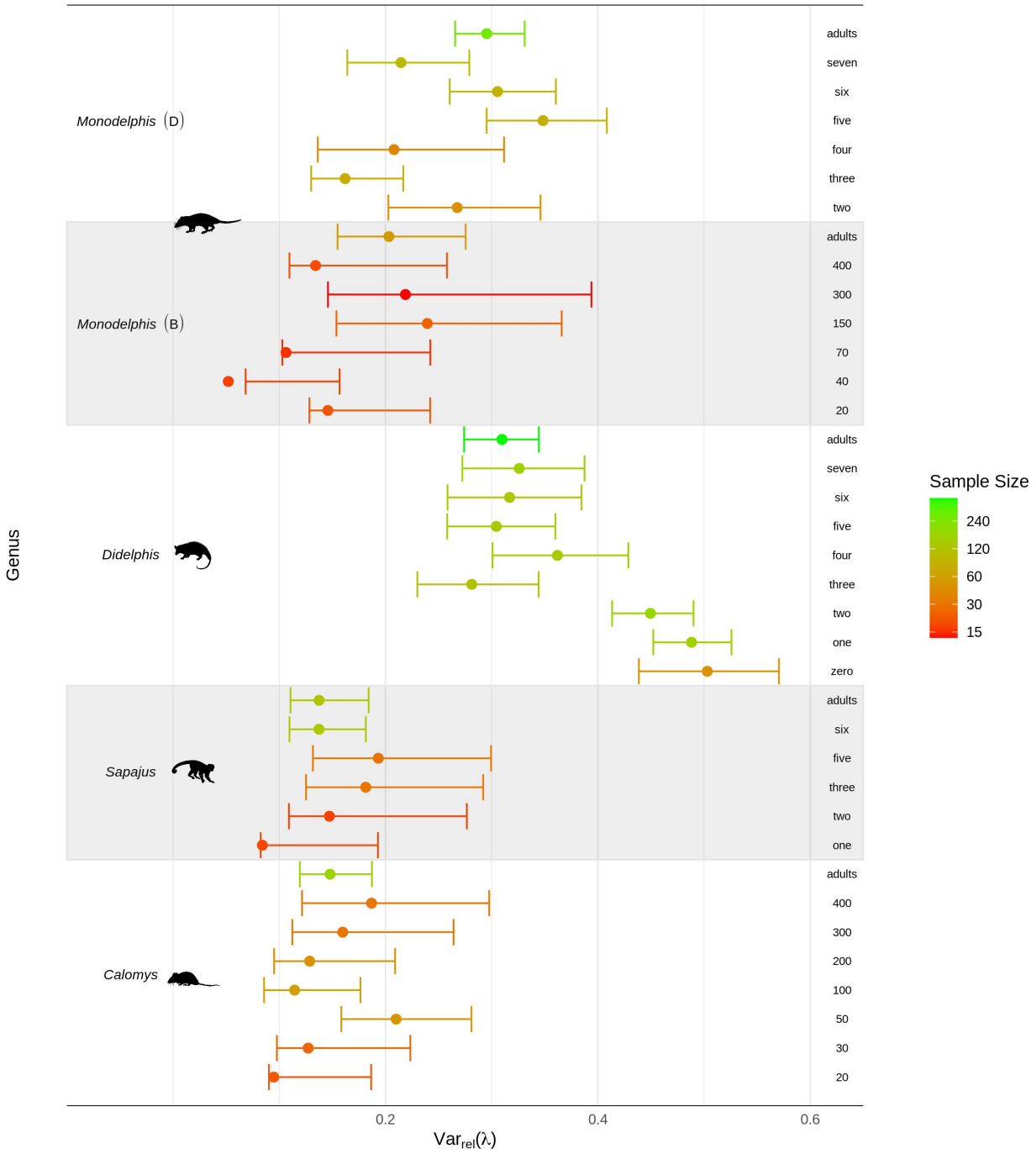

Figure 6: Distribution of overall magnitude of morphological integration ($Var_{rel}(\lambda)$) values for all genera and age classes based on correlation matrices. Non-parametric bootstrap was used to estimate 95% confidence intervals. Figures for each genus are adapted from (Eisenberg, 1989; Eisenberg and Redford, 1999; Redford and Eisenberg, 1992). Those figures are not in scale.

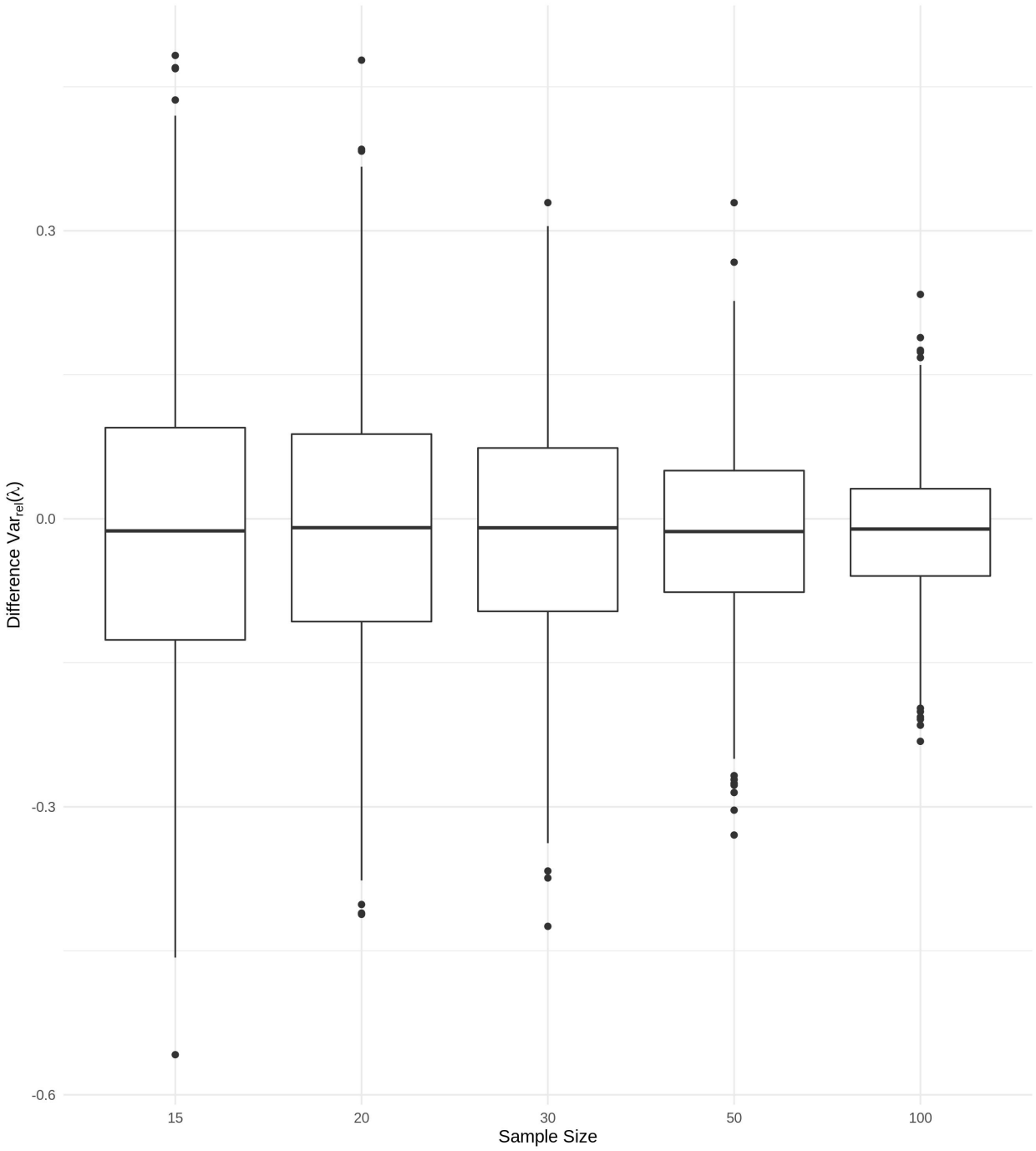

Figure 7: Difference between $Var_{rel}(\lambda)$ obtained on covariance matrices for the dental age class six and seven for *Didelphis* at different sample sizes. For each boxplot, 1,000 bootstrap resamplings of the original data for each dental age class were obtained after removing sources of variation (Table 1 from the main text).

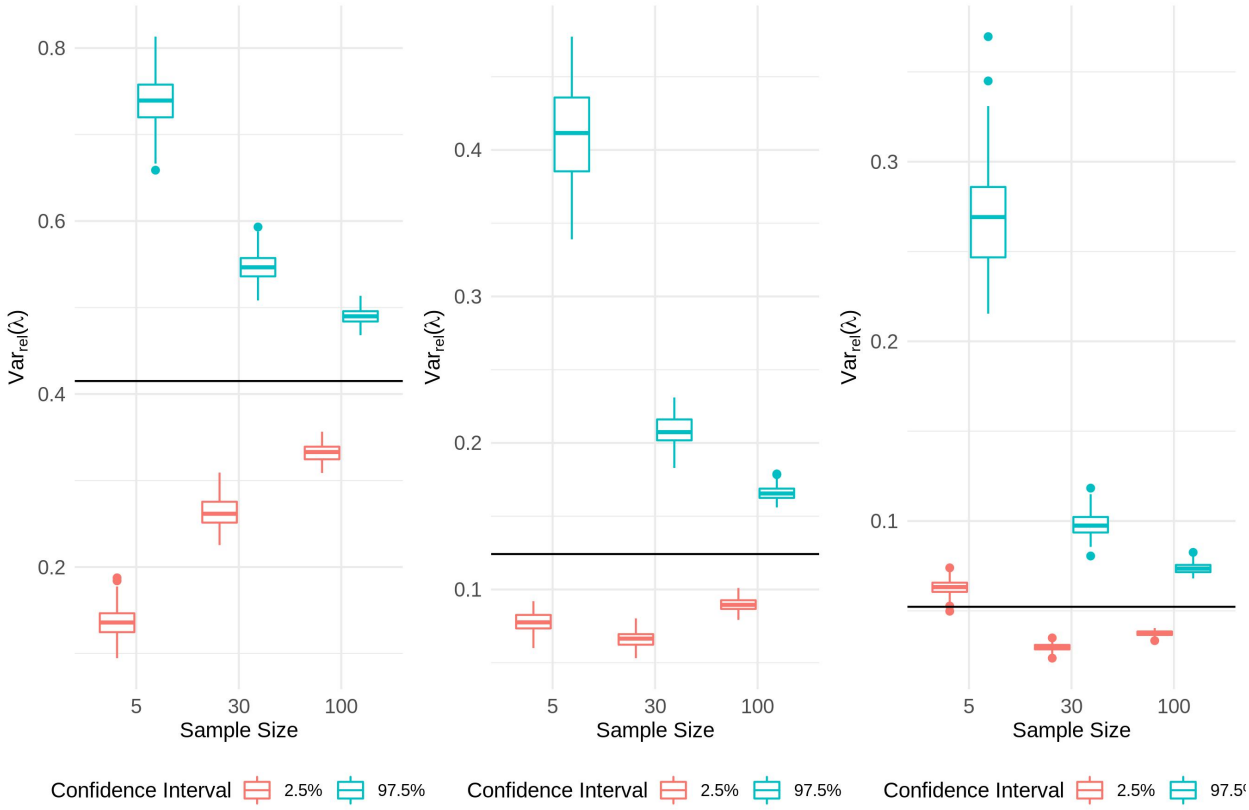

Figure 8: Boxplot of 100 95% confidence intervals for covariance matrices with different $Var_{rel}(\lambda)$ (horizontal lines) at different sample sizes (5, 30, 100). 95% confidence intervals were generated based on 100 bootstrap resamplings from the population. For each graph, a random population of 10,000 specimens was drawn from a multivariate normal distribution with traits' mean equal to zero and a specif covariance matrix. On the left, this matrix was the **P**-matrix from *Didelphis* at dental age class seven. On the center and on the right we constructed matrices with the same eigenvectors of the matrix above, but with different eigenvalues. Note that this figure replicates the pattern observed in our real data (Figure 4 from the main text). For higher $Var_{rel}(\lambda)$ (left graph), the observed value is around the center of the confidence interval, although for the sample size of 5 this does not hold true. For lower $Var_{rel}(\lambda)$ (graph on the center and on the right), the observed value is closer to the 2.5% confidence interval limit. This is probably do to the fact that the $Var_{rel}(\lambda)$ is constrained between zero and one, and irrespective of the matrix generated after the bootstrap procedure the $Var_{rel}(\lambda)$ cannot be inferior to zero (observe how the 2.5% confidence interval boxplot is narrower than the one for the 97.5 confidence interval). Lastly, for low $Var_{rel}(\lambda)$ and small sample sizes the observed value may fall outside the confidence interval.

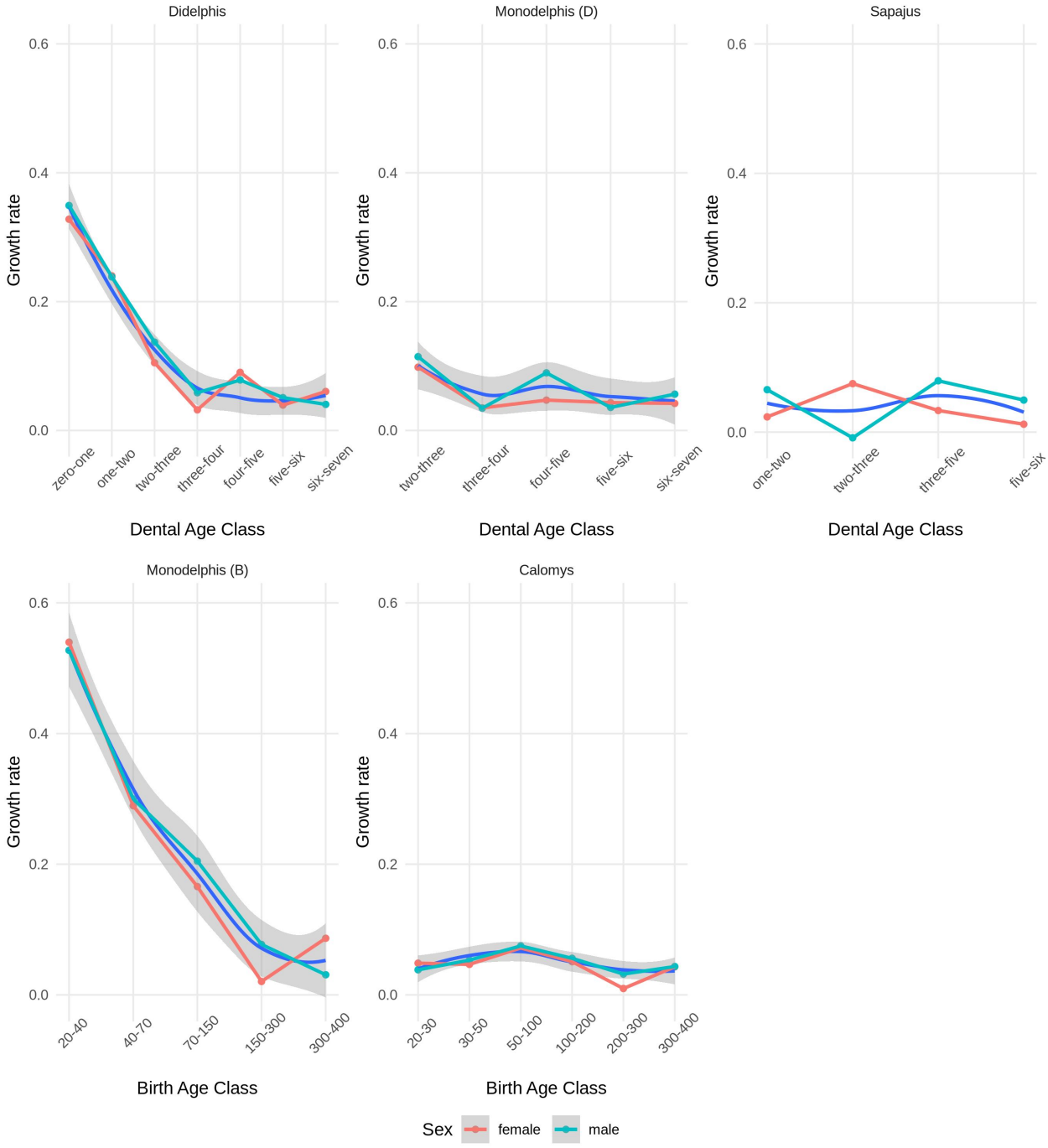

Figure 9: Growth rate for each genus between contiguous age classes. Growth rate was calculated as the percentage increase in centroid size from one age to the next. The shaded area represents the confidence interval of the LOESS regression line, representing the slowdown in the growth curve during ontogeny. Note that *Monodelphis* (B) and *Didelphis* both contain age classes where growth is occurring rapidly.