



A case study of extant and extinct *Xenarthra* cranium covariance structure: implications and applications to paleontology

Alex Hubbe, Diogo Melo, and Gabriel Marroig

Abstract.—Most of the mammalian diversity is known only from fossils, and only a few of these fossils are well preserved or abundant. This undersampling poses serious problems for understanding mammalian phenotypic evolution under a quantitative genetics framework, since this framework requires estimation of a group's additive genetic variance–covariance matrix (**G** matrix), which is impossible, and estimating a phenotypic variance–covariance matrix (**P** matrix) requires larger sample sizes than what is often available for extinct species. One alternative is to use **G** or **P** matrices from extant taxa as surrogates for the extinct ones. Although there are reasons to believe this approach is usually safe, it has not been fully explored. By thoroughly determining the extant and some extinct *Xenarthra* (Mammalia) cranium **P** matrices, this study aims to explore the feasibility of using extant **G** or **P** matrices as surrogates for the extinct ones and to provide guidelines regarding the reliability of this strategy and the necessary sample sizes. Variance–covariance and correlation **P** matrices for 35 cranium traits from 16 xenarthran genera (12 extant and 4 extinct) were estimated and compared between genera. Results show xenarthran **P**-matrix structures are usually very similar if sample sizes are reasonable. This study and others developed with extant therian mammals suggest, in general, that using extant **G** or **P** matrices as an approximation to extinct ones is a valid approach. Nevertheless, the accuracy of this approach depends on sample size, selected traits, and the type of matrix being considered.

Alex Hubbe. Departamento de Oceanografia, Instituto de Geociências, Universidade Federal da Bahia, Salvador, BA, 40170-020, Brazil. E-mail: alexhubbe@yahoo.com

Diogo Melo and Gabriel Marroig. Laboratório de Evolução de Mamíferos, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, Rua do Matão 277, São Paulo, SP, 05508-090, Brazil. E-mail: diogro@usp.br, gmarroig@usp.br

Accepted: 12 November 2015

Published online: 13 May 2016

Supplemental materials deposited at Dryad: doi:10.5061/dryad.6br16

Introduction

The evolution of continuous traits is critically dependent on the genetic variance available in populations. Moreover, since traits in multicellular complex organisms often are not genetically independent (due to pleiotropy, epistasis, and linkage disequilibrium), these organisms cannot be regarded as a collection of independent parts being changed by evolutionary processes. Instead, an organism must be understood as a coherent whole, with relationships described by a covariance structure. Thus, traits usually evolve in a correlated way, and to fully understand the evolution of complex structures (like the mammalian skull) we need to deal with the inheritance of such multidimensional phenotypes (Fisher 1930; Wright 1931; Lande 1979; Lande and Arnold 1983; Falconer and MacKay 1996). Quantitative genetics provides a framework with which to understand multivariate

phenotypic character evolution, and this framework has been used to study many evolutionary questions in several extant and some extinct species, from plants to vertebrates (e.g., Lande 1979; Cheverud 1984; Lofsvold 1986; Arnold 1992; Stepan 1997; Goswami 2006; Goswami et al. 2014; Hansen and Houle 2008; Webster and Zelditch 2011; Porto et al. 2013; Armbruster et al. 2014; Haber 2015). Under this framework it is fundamental to determine the additive genetic variance–covariance matrix (hereafter **G** matrix), since it interacts with evolutionary processes to determine the rate and direction of evolution (Lande 1979; Lande and Arnold 1983; Cheverud 1984).

The **G** matrix is a symmetric square matrix in which each row/column represents a phenotypic trait measured in a population. The **G** matrix describes the additive genetic variance of each trait (i.e., unstandardized measure of heritability) on the diagonal and

the additive genetic covariance between traits on the off-diagonal elements (Lande 1979; Arnold 1981; Cheverud 1988; Steppan et al. 2002). Therefore, the **G** matrix describes the population's variance and covariance that are linearly inheritable for a given set of traits and is relevant for the prediction of evolution on these traits (Falconer and MacKay 1996; McGuigan 2006). There are two aspects of the **G** matrix that should be conjointly studied to understand a population's mean phenotypic evolution (Marroig and Cheverud 2001; Marroig et al. 2009; Porto et al. 2009): one is the matrix structure, which is the relationship (covariance or correlation) between traits; the other is the matrix's overall magnitude of integration, which is related to the intensity of the relationships between traits (Porto et al. 2009).

The theory involving the **G** matrix and its consequences for mean phenotype evolution was originally developed to explain changes occurring within a few generations (microevolution; Lande 1979; Marroig and Cheverud 2001). However, this theory can be extended to account for the evolution among distinct species (macroevolution) if one major condition is satisfied: **G** matrices from different lineages share a similar structure (Lande 1979; Turelli 1988; Jones et al. 2004). Matrices are similar when they predict similar responses to directional selection, or if they suggest similar partitions of traits in variational modules, that is, similar patterns of high and low correlations between traits in both matrices. Theoretical models (Lande 1979; Turelli 1988; Barton and Turelli 1989) and empirical studies (Arnold 1981; Lofsvold 1986; Cheverud 1988; Kohn and Atchley 1988; Wilkinson et al. 1990; Shaw et al. 1995; Roff 1996; Bégin and Roff 2003) do not agree in their assessment of the degree of structural similarity between **G** matrices from different evolutionary lineages. Therefore, we must empirically assess the structural similarity between the **G** matrices of the evolutionary lineages of interest before conducting any macroevolutionary study with them (Turelli 1988; Arnold 1992; Steppan 1997; Arnold et al. 2008).

However, the accurate estimation of the **G** matrix, even for a single population, is a

complex enterprise, because **G**-matrix estimation requires a large number of related individuals, from hundreds to thousands, in known genealogies (Steppan et al. 2002; McGuigan 2006). This is especially limiting when studying organisms that are difficult to breed in the lab or that have long life spans. In some cases, such extinct species, it is impossible to estimate **G** matrices. Thus, obtaining **G** matrices and testing their structural similarity is often unfeasible, and the understanding of mean phenotypic micro- or macroevolution using quantitative genetics is hampered.

Fortunately, there is consistent evidence, at least for morphological traits, and particularly for mammals, that the variance-covariance phenotypic matrix (henceforth **P** matrix) can be used as a surrogate for the **G** matrix (the Cheverud conjecture; Cheverud 1988, 1996; Roff 1995; Reusch and Blanckenhorn 1998; House and Simmons 2005; Akesson et al. 2007; Porto et al. 2009). Working with **P** matrices is easier, because they are simpler to estimate with any confidence than **G** matrices, since **P** matrices can be determined without genealogical information (Cheverud 1988). Besides, estimating **P** matrices is also relevant by itself, since "natural selection acts on phenotypes, regardless of their genetic basis" (Lande and Arnold 1983: 1210). The most challenging aspect in properly estimating **P** matrices is obtaining adequate sample sizes, because the lower the ratio between sample size and number of traits analyzed, the greater the influence of error on the **P**-matrix estimation (Marroig et al. 2012).

However, the Cheverud conjecture should not be taken as an axiom. Instead, it should be empirically evaluated for the evolutionary lineages under study (Marroig and Cheverud 2001, 2010). To do so, it is always preferable to have at least one **G** matrix for comparison, but this is not strictly necessary. If high structural similarity between the **P** matrices is observed, this is most probably due to a common **G**-matrix structure. Since the phenotype is mainly determined by the additive genetic effect (**G**) and an environmental variation (**E**; $\mathbf{P} = \mathbf{G} + \mathbf{E}$; Falconer and MacKay 1996), there are two possibilities to explain the structural similarity between **P** matrices.

First, the **G** matrices share a similar structure, and the resemblance between **P** matrices is related to a common genetic background. Second, the **G** matrices have different structures, and the environmental matrices of each lineage would accurately compensate for the effects of **G** matrices to always generate **P** matrices with similar structures. The second option is highly unlikely, especially in a context that compares a large number of taxa that have a long evolutionary history and diverse ecology (Marroig and Cheverud 2001).

The approach outlined above has been successfully used as a first fundamental step to understanding mammalian cranium morphological evolution (Cheverud 1982, 1988; Lofsvold 1986; Steppan 1997; Marroig and Cheverud 2001; Ackermann and Cheverud 2004; Oliveira et al. 2009; Porto et al. 2009; Shirai and Marroig 2010). The mammalian cranium is an appealing model for studying morphological macroevolution, because it has a complex development and is involved in several of the organism's vital functions (e.g., brain protection, food processing) that allow the exploration of a wide range of questions (e.g., Cheverud 1995; Ackermann and Cheverud 2004; Hallgrímsson et al. 2009; Marroig and Cheverud 2010; Cardini and Polly 2013; Porto et al. 2013; Haber 2015).

Unfortunately, there are cases in which even **P** matrices cannot be properly estimated, such as for fossils lacking adequate sample sizes. This is frustrating, because most of world's biological diversity is represented by fossils (Benton and Harper 2009). For mammalian cranium morphological evolutionary studies covering extant and extinct taxa, one approach is to use the parameters determined for the extant groups as proxies for the extinct ones (Ackermann 2003, 2005; Ackermann and Cheverud 2004). This approach relies on the uniformitarianism paradigm, which is not necessarily true (Ackermann 2005). However, empirical studies have shown high structural similarity between the **P** matrices of extant taxa representing several lineages of the therian mammals (Lofsvold 1986; Steppan 1997; Marroig and Cheverud 2001; Ackermann and Cheverud 2004; Goswami 2006, 2007; Oliveira

et al. 2009; Porto et al. 2009; Shirai and Marroig 2010; Haber 2015), suggesting the **P** matrices from extinct taxa could be replaced by **P** matrices from extant taxa.

Within this context, we aim to deepen our knowledge regarding Xenarthra, which is one of the major Placentalia group and has not been thoroughly studied under an evolutionary quantitative genetics framework. More specifically, one of our goals is to present the first broad-scale test of the Cheverud conjecture for xenarthran cranium **P** matrices, taking into account 12 extant and 4 extinct lineages. By testing the Cheverud conjecture within this magnaorder, we automatically achieve another goal: assessing the structural similarity of **P** matrices within Xenarthra. We also aim to explore the range of overall magnitude of integration within Xenarthra. By doing so, we characterize both fundamental aspects of the **P** matrix that are relevant to evolutionary studies. Complementarily, we explore whether phylogenetic proximity and morphological similarity influence both the structure of the **P** matrices and their overall magnitude of integration. Finally, since we found particularities in some xenarthran **P** matrices, we use the case study with Xenarthra to provide guidelines to researchers interested in performing evolutionary quantitative genetic analyses involving extinct lineages, especially when one cannot properly estimate **P** matrices for these lineages.

To accomplish these goals, we estimated **P** matrices for every evolutionary lineage at different taxonomic and phylogenetic levels (genera, subfamily, family, suborder, and order). The Cheverud conjecture was tested through the comparison of the structural similarity of the **P** matrices among genera to orders. Overall magnitude of integration was calculated for the **P** matrix of each genus. Distance matrices of the phylogeny, morphology, and overall magnitude of integration and a similarity matrix of **P**-matrices' structures were determined at the genus level. These matrices were compared to evaluate the morphological differentiation (on average) and the influence of phylogeny on the **P**-matrices' structures and their overall magnitude of integration.

Brief Contextualization of the Xenarthra (Mammalia)

Although the relationship of the Xenarthra with other mammals is not well resolved (Delsuc et al. 2001; Murphy et al. 2001; Bininda-Emonds et al. 2007; Goloboff et al. 2009; O'Leary et al. 2013), their monophyly is well supported both by morphological (Engelmann 1985; Gaudin 2004) and molecular data (Bininda-Emonds et al. 2007; Delsuc et al. 2012; O'Leary et al. 2013). The Xenarthra are represented by three major groups, markedly distinct in their morphology, ecology, and biology: Folivora (sloths), Vermilingua (anteaters), and Cingulata (armadillos; Eisenberg 1989; Redford and Eisenberg 1992; Eisenberg and Redford 1999; Gardner 2007).

The origin of the Xenarthra precedes 60 Ma (Simpson 1980; Bergqvist et al. 2004; Bininda-Emonds et al. 2007; Delsuc et al. 2012). Most of the group diversification occurred in South America, but some lineages evolved in Central and North America (Simpson 1980). The current distribution of xenarthrans is from southern North America into Patagonia (Gardner 2007). Evidence suggests their occurrence in the Northern Hemisphere was broader in the past, with several fossils found throughout North America (Stock 1942; McDonald et al. 2000; McDonald and Pelikan 2006; Gardner 2007; Hoganson and McDonald 2007). The xenarthrans are represented by 31 extant species in 14 genera (21 armadillos, 6 sloths, and 4 anteaters) that vary between 100 g and 40 kg (Eisenberg and Redford 1999; Delsuc et al. 2012), but in the recent past (between the late Pleistocene and middle Holocene), they were far more speciose, and some species could weigh more than 1 ton (McDonald 2005; Cartelle et al. 2008; Vizcaino et al. 2008).

Material and Methods

Samples.—A total of 1139 adult specimens belonging to 16 genera (12 extant and 4 extinct; Fig. 1) were measured and used in this study. Specimens were considered adult when the supraoccipito-exoccipital and the basioccipito-exoccipital sutures were closed (i.e., completely ossified), the basioccipito-basisphenoid suture

was fused or closed, texture of external bones was uniformly smooth, and viscerocranium and neurocranium exhibited adult proportions (Cheverud 1995; Anderson and Handley 2001; Elbroch 2006; Porto et al. 2009). The process of determining whether a specimen is adult is to some extent iterative, in the sense that we need to examine several specimens of a species to develop the parameters that would allow determining adulthood. We would like to clarify that the definition of adulthood may vary according to research interest, and in our case adults are specimens in which development had no more influence, or drastically decreased influence, over phenotype. Xenarthra have, in relation to other mammals, morphological and developmental particularities (Eisenberg 1989; Redford and Eisenberg 1992; Eisenberg and Redford 1999; Gardner 2007; Hautier et al. 2011; Rager et al. 2014), and some classical adulthood parametrization (e.g., tooth eruption and wearing or basioccipito-basisphenoid closure) may not be applicable to Xenarthra; but with the several parameters adopted, we most likely sampled specimens whose phenotypes were no longer under major developmental influence. Using the specimens selected based on the above criteria, we further observed the within-genera specimen distribution in principal component analysis (PCA) scatter plots based on the first and second PCs for genera with at least 20 specimens. In all cases, there were no outliers. All specimens were examined and identified by A.H. Only specimens identified to the species level were included in our sample. All extant specimens were wild caught.

The fossil record is patchy, and paleontological analyses often rely on a limited number of fossils per species. The adequate number of specimens per species may vary according to the kind of analyses performed, and in our case less than five specimens is prohibitive, and as will be shown throughout our paper, samples that are smaller than the number of traits can be considered inappropriate, since matrix estimation will probably be dominated by sampling error (Marroig et al. 2012). Xenarthra has a rich fossil record, and ideally we would have sampled several extinct lineages within Cingulata, Vermilingua, and Folivora.

Genus	Subfamily	Family	Suborder	Order
<i>Bradypus</i> (49, 136, 136, 136)	> Bradypodinae (49, 136, 136, 136)	> Bradypodidae (49, 136, 136, 136)	Folivora (80, 181, 213, 218)	Pilosa (252, 477, 509, 514)
<i>Paramylodon</i> † (NA, NA, 31, 37)	> Mylodontinae (NA, NA, 31, 37)	Mylodontidae (-, -, 32, 37)		
<i>Scelidotherium</i> †* (NA, NA, NA, 16)	> Scelidotheriinae* (NA, NA, NA, 16)			
<i>Choloepus</i> (31, 45, 45, 45)	> Choloepodinae (31, 45, 45, 45)	Megalonychidae (31, 45, 45, 45)		
<i>Acratocnus</i> †* (NA, NA, NA, 8)	> Choloepodinae* (NA, NA, NA, 8)			
<i>Neocnus</i> †* (NA, NA, NA, 6)	> Choloepodinae* (NA, NA, NA, 6)			
<i>Cyclopes</i> (12, 91, 91, 91)	> Cyclopedinae (12, 91, 91, 91)	> Cyclopedidae (12, 91, 91, 91)	Vermilingua (172, 296, 296, 296)	
<i>Myrmecophaga</i> (22, 67, 67, 67)	Myrmecophaginae (160, 205, 205, 205)	> Myrmecophagidae (160, 205, 205, 205)		
<i>Tamandua</i> (138, 138, 138, 138)				
<i>Dasyus</i> (219, 219, 219, 219)	> Dasypodinae (219, 219, 219, 219)	Dasypodidae (535, 535, 535, 535)	> Cingulata (535, 535, 535, 535)	
<i>Euphractus</i> (71, 71, 71, 71)	Euphractinae (203, 203, 203, 203)			
<i>Chaetophractus</i> (92, 92, 92, 92)				
<i>Zaedyus</i> (40, 40, 40, 40)				
<i>Priodontes</i> (24, 24, 24, 24)	Tolypeutinae (113, 113, 113, 113)			
<i>Cabassous</i> (47, 47, 47, 47)				
<i>Tolypeutes</i> (42, 42, 42, 42)				

FIGURE 1. Phylogenetic/taxonomy hierarchy adopted. Within parentheses are the sample sizes for the different data sets of distances (35D, 32D, 28D, and 25D). See section “Sets of distances” for clarifications. † denotes an extinct genus. * denotes taxa not included in the analyses above the genus level. NA means no available sample for the specified set of distances.

However, we were able to measure at least five skulls containing the appropriate number of landmarks (see subsections “Landmarks and Measurements” and “Sets of Distances”) and not presenting morphological distortion only for four Folivora genera. Among the four extinct sloths, only *Paramylodon* was reasonably sampled (Fig. 1). Consequently, for the other three genera, sampling error probably contributed substantially to their estimated **P** matrices (Marroig et al. 2012). To avoid biases in the results due to potentially

high sampling errors, data from *Acratocnus*, *Neocnus*, and *Scelidotherium* were not included in the analyses above the genus level or in the determination of the similarity/distance matrices.

Taxonomy and Phylogeny.—The adopted taxonomy reflects the phylogenetic relationships, and the analyses were therefore both taxonomically and phylogenetically structured. The least inclusive taxonomic level analyzed was genus. Species were not considered due to the absence of a well-resolved phylogeny for the

Xenarthra at this level and the lack of adequate sample sizes for several species. The taxonomy of extant Xenarthra is in accordance with Gardner (2007). The taxonomy of the extinct genera followed the proposals of MacPhee et al. (2000), McAfee (2009), and Miño-Boilini and Carlini (2009; see Fig. 1 for details regarding the taxonomic/phylogenetic hierarchy adopted).

Here, we follow the fully resolved extant genus phylogeny proposed by Möller-Krull et al. (2007). When including the extinct *Paramylodon*, the Folivora clade was considered a polytomy due to the absence of a molecular phylogeny that includes this extinct genus and the major incongruence observed among morphological and molecular analyses involving extant and extinct taxa (see discussions in Gaudin 2004; Clark 2010).

Landmarks and Measurements.—As noted earlier, all specimens were measured by

A.H. For each specimen, three-dimensional coordinates were recorded for 32 landmarks (Fig. 2; Supplementary Table 1; Porto et al. 2009) using an MX or MLX Microscribe digitizer. The determination of some landmarks varied according to the taxon and should be considered only partially homologous among the Xenarthra (see Supplementary Table 1 for details). Worth mentioning is that ZS and ZI in the anteater *Cyclopes* and ZYGO in Pilosa (anteaters and sloths; see Fig. 2 for details regarding these landmarks) could not be determined by the meeting of the sutures present in the other taxa due to the absence (or reduction) of the jugal and squamosal, respectively, and were determined as extremity points of anatomical regions determined by one or two bones (Supplementary Table 1).

A set of 35 linear distances in millimeters were calculated based on the landmarks recorded (Table 1). Bilateral distances were averaged between both sides. If the distance from one side was missing, the distance from the other side was used to represent the bilateral distance. The set of distances calculated aimed to represent the whole cranium morphology and important developmental and functional relationships among cranial regions while avoiding redundancy (Cheverud 1982; Marroig and Cheverud 2001). Furthermore, measurements were designed to capture local developmental/functional aspects of the skull, which has often been accomplished by

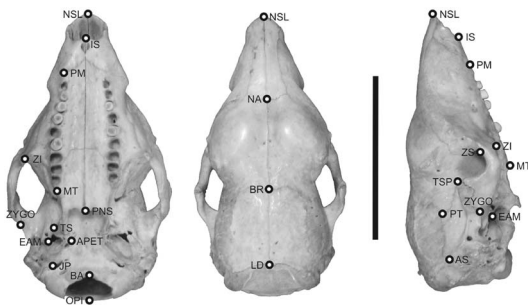


FIGURE 2. Landmarks on ventral, dorsal, and lateral view of an armadillo *Cabassous* skull. Scale bar, 5 cm.

TABLE 1. Linear distances and their relation to the cranial anatomical region.

Distance	Anatomical region	Position	Distance	Anatomical region	Position
ISPM	Face	Bilateral	ZSZI	Face	Bilateral
ISNSL	Face	Sagittal	ZIMT	Face	Bilateral
ISPNS	Face	Sagittal	ZIZYGO	Face	Bilateral
PMZS	Face	Bilateral	ZITSP	Face	Bilateral
PMZI	Face	Bilateral	MTPNS	Face	Bilateral
PMMT	Face	Bilateral	PNSAPET	Neurocranium	Bilateral
NSLNA	Face	Sagittal	APETBA	Neurocranium	Bilateral
NSLZS	Face	Bilateral	APETTS	Neurocranium	Bilateral
NSLZI	Face	Bilateral	BAEAM	Neurocranium	Bilateral
NABR	Neurocranium	Sagittal	EAMZYGO	Face	Bilateral
NAPNS	Face	Sagittal	ZYGOTSP	Face	Bilateral
BRPT	Neurocranium	Bilateral	LDAS	Neurocranium	Bilateral
PRAPET	Neurocranium	Bilateral	BRLD	Neurocranium	Sagittal
PTAPET	Neurocranium	Bilateral	OPILD	Neurocranium	Sagittal
PTBA	Neurocranium	Bilateral	PTAS	Neurocranium	Bilateral
PTEAM	Neurocranium	Bilateral	JPAS	Neurocranium	Bilateral
PTZYGO	Face	Bilateral	BAOPI	Neurocranium	Sagittal
PTTSP	Neurocranium and face	Bilateral			

measuring within bone interlandmark distances. We opt to use distances instead of landmark-based data under a generalized Procrustes analysis (GPA), because Van der Linde and Houle (2009) demonstrated that GPA tends to disperse local variation onto all landmarks, producing covariation structures that are difficult to interpret. We believe that while GPA is good at describing shape changes, studies interested in covariation should adopt different methods. Thus, we think our measurements are better suited to our goal of representing the phenotypic part of the genetic-phenotypic map (*sensu* Wagner and Altenberg 1996) than the common practice of using PCs of shape (Walker 2000; Van der Linde and Houle 2009; Berner et al. 2011; Marqu ez et al. 2012). Each specimen was measured twice to account for measurement errors (Lessells and Boag 1987). The measurement error was estimated for genera with more than 15 specimens as a proportion of total variance due to within-individual differences in repeated measurements (Fig. 1). The measurement error calculated individually for each genus and trait ranged between 0.71 and 1, with an average for all genera and traits of 0.98 and a standard deviation of 0.03. Only six cases presented values below 0.89, and they were associated with traits exhibiting low total variation and not larger differences between replicas. Therefore, measurement error has a negligible impact on the subsequent analyses. Specimen traits were represented by the average between both measurements.

Missing Data.—Some of the fossil crania were incomplete and presented missing data. Fossil specimens that had missing data for sagittal distances were not considered in this study. For fossil specimens that had missing data for bilateral distances obtained from landmarks off the sagittal plane, a missing data-substitution procedure was adopted whenever the specimen had one of the landmarks that composes the distance at one side of the cranium and the other at the other side. The procedure consisted of reflecting one of the landmarks coordinates in relation to the cranium sagittal plane and then calculating the distance based on the original coordinates of one landmark and the reflected coordinates of

the other (see supplemental material for more information and tests of reliability and accuracy of this procedure).

Sets of Distances.—Pilosa has the premaxilla loosely attached to the maxilla, and several specimens analyzed did not have this bone, hindering the record of IS and preventing the calculation of all distances based on this landmark. Moreover, fossil specimens were consistently devoid of the cranium regions presenting IS, ZYGO, and NSL, preventing the calculation of the distances based on these landmarks. Due to these circumstances, four different sets of distances were used to progressively include more taxa and increase sample sizes in the analyses (Fig. 1): the original set of 35 distances (hereafter 35D); a set of 32 distances, excluding ISPM, ISNSL, and ISPNS (32D); a set of 28 distances, excluding ISPM, ISNSL, ISPNS, PTZYGO, ZIZYGO, EAMZYGO, and ZYGOTSP (28D); and a set of 25 distances, excluding ISPM, ISNSL, ISPNS, PTZYGO, ZIZYGO, EAMZYGO, ZYGOTSP, NSLNA, NSLZS, and NSLZI (25D).

On the one hand, this approach progressively allows the incorporation of new taxa in the analyses, increases sample sizes, and minimizes sampling error effect on the estimation of the **P** matrices (Marroig et al. 2012). On the other hand, the overall cranium shape representation is gradually simplified. The regions most affected were the zygomatic arch (28D and 25D) and the face (25D). All analyses were performed for the four sets of distances.

Estimating Variance–Covariance and Correlation P Matrices.—The **P** matrix can be represented as the variance–covariance matrix (hereafter covariance matrix) as well as its standardized form, the correlation matrix. Both forms were estimated, but before that, sources of potential variation in trait means that were not of immediate interest were statistically evaluated within each genus (Oliveira et al. 2009). The effect of sex, species, intraspecific variation, and their potential interactions, were tested through (1) MANOVA based on Wilk’s lambda statistic, considering the significance at $p < 0.05$; and (2) univariate ANOVA, considering the tested effect as influencing trait means whenever two or more variables (distances) had significance at $p < 0.01$.

Notice that in some cases low sample size prevented analysis with MANOVA or in the 35D data set (see details in Table 2).

When the uni- and multivariate tests were significant, the effects of the respective source of variation and/or interaction between them were controlled (Table 2). Residual pooled within-group phenotypic covariance and correlation matrices were estimated for each genus using the general linear model approach (Marroig et al. 2004). Above the genus level, P covariance, and correlation matrices were estimated as the pooled within-group phenotypic matrices of the taxa following the adopted phylogeny (Fig. 1).

Testing the Cheverud Conjecture.—The Cheverud conjecture was tested throughout different taxonomic/phylogenetic levels (genus, subfamily, family, suborder, and order). P matrices were calculated for each taxon at each level of analysis and compared pairwise. The structural similarity among covariance and correlation matrices was evaluated through the random skewers method (RS; Cheverud and Marroig 2007) and the Krzanowski projection method (KP; Krzanowski 1979, 2000; Blows et al. 2004). Although there are other methods suitable for matrix structure comparisons (e.g., Cheverud 1988; Flury 1988; Mitteroecker and Bookstein 2009), we chose the RS and KP because they provide simple and intuitive measurements of overall similarity between two matrices and can be applied to covariance and correlation matrices.

The RS is based on the multivariate response to selection equation and consists of multiplying each matrix by 1000 normalized random-selection gradient vectors and correlating the resulting evolutionary-response vectors (normalized to a length of one) between each pair of matrices. The same vectors are used for each pair of matrices, and only the response to the same selection is correlated between each matrix pair. Therefore, a distribution of 1000 vector correlations of simulated evolutionary responses is obtained for each pair of matrices being compared. The mean vector correlation between the matrices' evolutionary-response vectors is a measure of similarity (Cheverud and Marroig 2007). The RS correlation, as any correlation, can vary from -1 to 1. Yet in real matrices usually it

TABLE 2. Sources of variation tested (x) and controlled (◆) for each xenarthran genus.

	Species		Sex	Intraspecific variation	Factors' interaction	Observations	Species studied (sample size)
	◆	x					
<i>Zaedyus</i>	◆	x				When both factors were considered together, only the univariate test was performed	<i>Z. pichiy</i> (40)
<i>Tolypeutes</i>		x		x		Only the univariate test was performed	<i>T. matacus</i> (28), <i>T. tricinctus</i> (14)
<i>Priodontes</i>		x					<i>P. maximus</i> (24)
<i>Euphractus</i>	◆	x	◆				<i>E. sexcinctus</i> (71)
<i>Dasypus</i>	◆	x	◆	x			<i>D. hybridus</i> (30), <i>D. kappleri</i> (25), <i>D. novemcinctus</i> (164)
<i>Chaetophractus</i>	◆	◆	◆	◆			<i>C. vellerosus</i> (47), <i>C. villosus</i> (45)
<i>Cabassous</i>	◆	x	◆	x			<i>C. centralis</i> (5), <i>C. tatouay</i> (15), <i>C. unicinctus</i> (27)
<i>Cyclopes</i>			◆				<i>C. didactylus</i> (91)
<i>Myrmecophaga</i>	◆	◆	◆			Tested on the set of distances 32D	<i>M. tridactyla</i> (67)
<i>Tamandua</i>	◆	x	◆	x		Tested on the set of distances 32D	<i>T. mexicana</i> (55), <i>T. tetradactyla</i> (83)
<i>Bradypus</i>	◆	x	◆	x		Tested on the set of distances 32D	<i>B. pygmaeus</i> (7), <i>B. tridactylus</i> (9), <i>B. variegatus</i> (120)
<i>Choloepus</i>	◆	◆	◆	◆		Tested on the set of distances 32D	<i>C. didactylus</i> (11), <i>C. hoffmanni</i> (34)
<i>Acratocnus</i>	◆	◆	◆	◆		Tested on the set of distances 25D; only the univariate test was performed	<i>A. odontotrigonus</i> (5), <i>A. yje</i> (3)
<i>Scelidotherium</i>							<i>S. leptocephalum</i> (16)
<i>Paramylodon</i>							<i>P. harlani</i> (37)
<i>Neocnus</i>							<i>N. comes</i> (6)

varies between 0 (the matrices have distinct structures) and 1 (the matrices share the same structure). Statistical significance was evaluated by comparing the observed RS value with an empirically derived distribution of the correlation of 1000 random vectors. If the observed value exceeded 95% of the empirically derived distribution, it was considered different from 0 (Cheverud 1996; Cheverud and Marroig 2007).

The KP consists in determining a projection **S** matrix that finds the best-matched pairs of orthogonal axes (PC) in the same k -dimensional subspace. The sum of the eigenvalues of the **S** matrix divided by the k dimensions represents the similarity of the two subspaces and is expressed between 0 and 1, where 0 means the two subspaces are dissimilar and 1 means they are strictly similar. The k dimensions were determined by the first k PCs of the integer part of half the original number of dimensions minus 1. For example, for the 35D data set, $k=16$ (integer part of $35/2 - 1$; Krzanowski 1979, 2000; Blows et al. 2004).

Due to the finite nature of the samples, covariance and correlation matrices are estimated with error and, consequently, the maximum structural similarity (r_{\max}) between two matrices is not 1. In fact, r_{\max} is the square root of the product of the repeatabilities (t) of the pair of matrices (see the t calculation below). Hence, it is possible to obtain an adjusted similarity (r_{adj}) as a proportion of r_{\max} and the observed similarity (r_{obs}) according to (Cheverud 1996):

$$r_{\text{adj}} = r_{\text{obs}}/r_{\max} \quad (1)$$

The t for both correlation and covariance matrices was determined using a bootstrap resampling procedure of self-correlation (Marroig and Cheverud 2001; Porto et al. 2009). For every evolutionary lineage, 1000 bootstrap resamplings of the original data were made, keeping sample size constant, after removing sources of variation (for details see Table 2). Correlation and covariance matrices were estimated for each of the resamples, and RS and KP were used to compare the original and the 1000 resample matrices. Mean RS and KP values represented t . It is important to keep in mind that the bootstrap resampling procedure should not be used when samples are

smaller than the number of traits (Chernick 2008; Chernick and LaBudde 2010; Puth et al. 2015). With small initial samples, use of the bootstrap resampling procedure on the matrix repeatability will produce overestimated values, as the bootstrap samples tend to be similar due to the small sample space, which will result in conservative corrections of the matrices' similarities. In any event, these high repeatabilities should not be interpreted as indicating a well-estimated matrix.

In addition, the covariance and correlation genus **P** matrices were compared with the covariance and correlation **G** and **P** matrices of the rodent *Calomys* (Garcia 2010) through the RS and KP methods. The rationale for comparing Xenarthran **P** matrices with **G** and **P** matrices of other taxa is that if the structural similarity between these matrices is high, the Cheverud conjecture is supported.

Recently, Aguirre et al. (2014) described methods for comparison of covariance matrices under the Bayesian paradigm, including the Bayesian RS and the Bayesian KP methods. The Cheverud conjecture between xenarthran genus **P** matrices was also evaluated using these methods.

Overall Magnitude of Integration.—The overall magnitude of integration was determined for the genera correlation **P** matrices (Pavlicev et al. 2009) by calculating the scale-independent coefficient of determination (r^2) that is the mean of squared correlation coefficients. The r^2 determines the overall level of correlation among all traits and potentially ranges from 0 (all traits are independent) to 1 (all traits are fully correlated; Cheverud et al. 1989; Porto et al. 2009).

It was recently reported the r^2 may be biased by the sampled population coefficient of variation (Young et al. 2010). To explore this issue, we calculated an adjusted coefficient of determination (r^2_{adj}). One thousand bootstrap resamplings for each genus (holding the sample size of the original data and taking into account the sources of undesired variation; Fig. 1; Table 2) were produced. For each resample, the r^2 was calculated. If the correlation between the r^2 and the mean coefficient of variation for all traits was significant ($p < 0.05$), the r^2 was regressed on the mean coefficient of

variation. Based on the regression equation for each evolutionary lineage, the r^2_{adj} was calculated at a mean coefficient of variation of 6% (Young et al. 2010; Porto et al. 2013). This procedure was done solely to evaluate the uncertainty of the r^2 estimates, since there is no biological explanation to support analyzing all evolutionary lineages at the same level of morphological variation (Porto et al. 2013).

Similarity and Distance Matrices.—The phylogenetic distance matrix was differently determined for the different data sets (35D, 32D, 28D, and 25D). For 35D and 32D, for which only extant taxa were considered, the phylogenetic distance between two taxa was the sum of the lengths of the corresponding branches of the adopted phylogeny (Möller-Krull et al. 2007). The branch length was the number of substitutions per site along the corresponding branch. For 28D and 25D the same procedure was adopted, but all branches had a length of 1, and the relationship within Folivora was considered a polytomy. For comparative reasons, the same procedure was also adopted for 35D and 32D. So, two phylogenetic distance matrices were estimated for 35D and 32D.

The morphological distance matrix was obtained for all data sets by calculating the squared Mahalanobis distance (D^2) between genera:

$$D_{ij}^2 = (\mu_i - \mu_j)' W^{-1} (\mu_i - \mu_j) \quad (2)$$

where μ_i and μ_j are the vector of trait means for the compared taxa, and W is the pooled within-group covariance matrix of all Xenarthra, controlling for the sources of undesired variation (for details see Table 2; Ackermann 2002). Morphological distances were moderate to weakly correlated with phylogenetic distances (Supplementary Table 2). Residual morphological distances were calculated by linearly regressing the morphological distances as a function of the phylogenetic distances. After a constant was added, the log-transformed residual morphological distance matrix was used in later analyses.

The overall magnitude-of-integration distance matrix was calculated for all data sets as the absolute differences in r^2 for each pair of matrices (Porto et al. 2009). The structural

similarity matrix for all data sets for the covariance and correlation matrices are the results of the RS and KP methods (Marroig and Cheverud 2001). Matrices were compared based on the cross-product of element entries followed by Mantel's test (15,000 permutations) for assessing statistical significance (Cheverud et al. 1989).

Statistical Analyses.—All analyses were done in the R Environment for Statistical Computing (R Development Core Team 2013) using the *EvolQG* package (Melo et al. 2015).

Results

The results presented here are mainly from the analyses of the 25D data set. However, whenever differences in the results between the data sets exist, they are also presented.

Testing the Cheverud Conjecture.—The observed similarity among the xenarthran genera covariance matrices were usually above 0.6 for both the RS and KP. The adjusted similarities were usually above 0.7 for the RS and frequently above 0.75 for the KP. In general, observed and adjusted similarities were higher for the correlation matrices than for the covariance matrices (Figs. 3 and 4; Supplementary Tables 3 and 4). The poorly sampled extinct sloths *Acratocnus* and *Neocnus* (Fig. 1) presented almost all original and adjusted RS similarities below 0.5. On the other hand, the KP similarities for these taxa showed higher similarities, especially the adjusted ones above 0.7. Besides the results for *Acratocnus* and *Neocnus*, adjusted similarities below 0.7 were found mainly for the RS of covariance matrices. These values were more frequent in cases in which one of the two taxa compared was the extinct sloths *Paramylodon* and *Scelidotherium* or the anteater *Myrmecophaga* or the armadillo *Prionodontes*. With the exception of the comparisons involving *Acratocnus* and *Neocnus*, all similarities were significant at $p < 0.01$. Most similarities (23 of 28) involving *Acratocnus* and 8 of 28 comparisons involving *Neocnus* were significant at $p < 0.05$. Lower RS and KP similarities for both covariance and correlation matrices were usually associated with lower

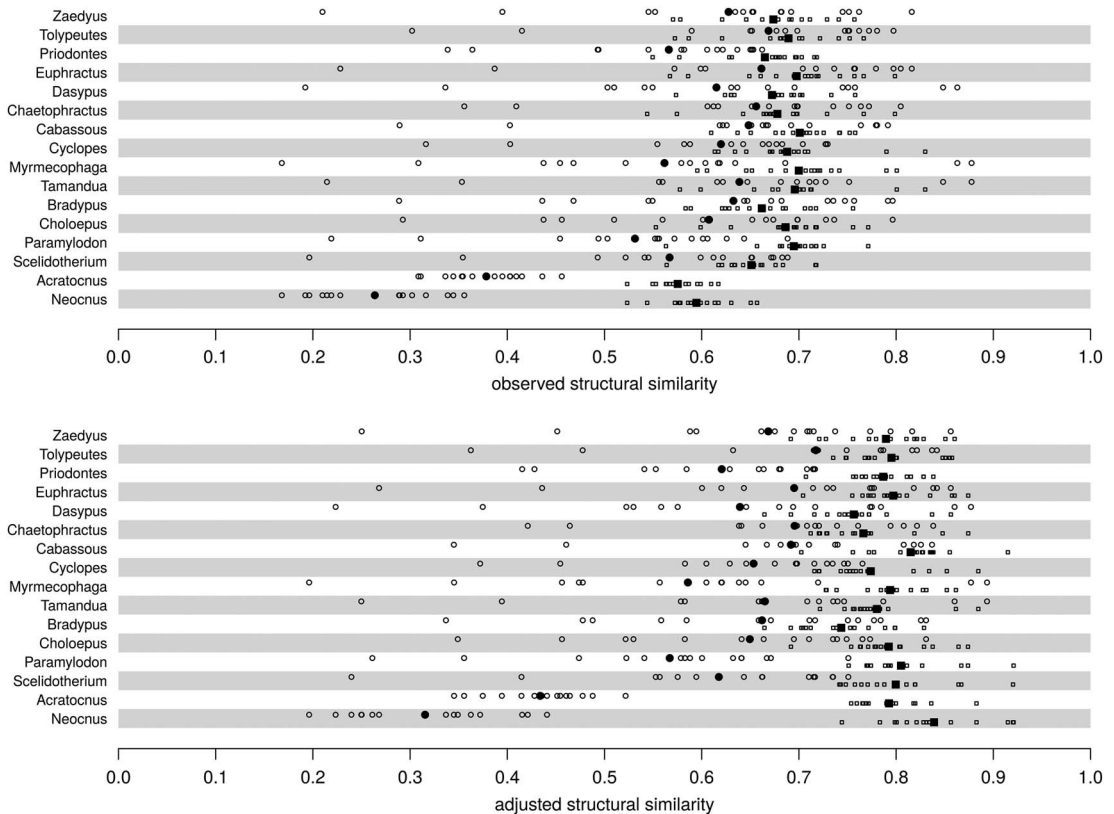


FIGURE 3. Structural similarity for the 25D data set between xenarthran genera covariance matrices. Open symbol, individual value; closed symbol, mean value; circle, RS; square, KP.

sample sizes from at least one of the matrices under comparison (Supplementary Figs. 1–4).

The adjusted structural similarities between xenarthran genus **P** matrices and the rodent **G** and **P** matrices were usually above 0.6, with several above 0.7 (Table 3). Similarities above 0.5 and below 0.65 were restricted to the RS and most often to the covariance matrices and were related to the armadillos *Priodontes* and *Dasypus*, the anteater *Myrmecophaga*, and the extinct sloth *Paramylodon*. Similarities below 0.4 were restricted to the RS and the poorly sampled extinct sloths *Acratocnus* and *Neocnus*. All similarities were significant at $p < 0.01$, except the ones involving *Acratocnus* and *Neocnus*.

At the subfamily, family, suborder, and order levels, similar results were found (Supplementary Tables 5 and 6). Observed and adjusted similarities were usually above 0.7 for both KP of covariance and correlation matrices and RS of correlation matrices. Several

adjusted similarities for the RS of covariance matrices were below 0.7 and most frequently involved Mylodontinae/dae (*Paramylodon*), Myrmecophaginae/dae (*Tamandua* and *Myrmecophaga*), Folivora (all sloths), and Vermilingua (all anteaters). Similarities based on correlation matrices were regularly higher than on the covariance matrices, especially for the RS. All similarities above the genus level were significant at $p < 0.01$.

Regarding the results obtained with the distinct sets of distances (35D, 32D, 28D, and 25D), conspicuous differences were observed between the 35D and 32D results, but only within taxa that had a markedly increased in sample size (Fig. 1; Supplementary Tables 7–18).

The results obtained using the Bayesian methods were congruent with the results presented above and generally show a pattern of high similarities between matrices (for details see the supplemental material).

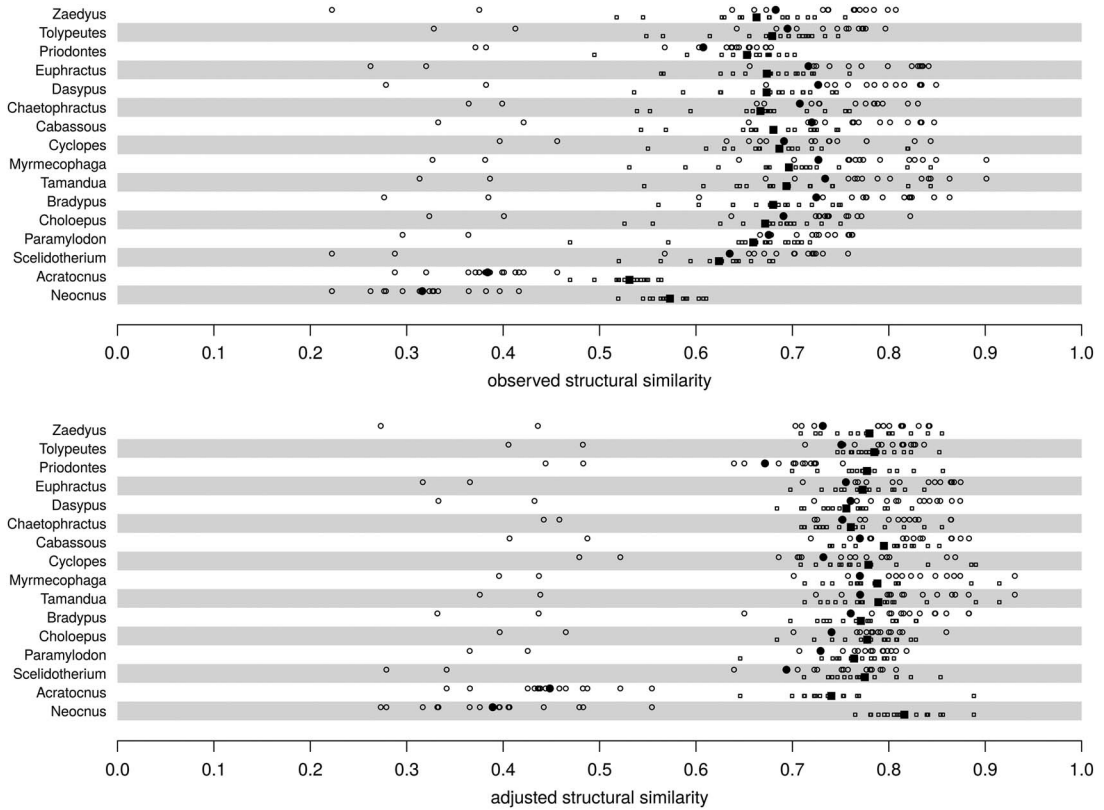


FIGURE 4. Structural similarity for the 25D data set between xenarthran genera correlation matrices. Open symbol, individual value; closed symbol, mean value; circle, RS; square, KP.

TABLE 3. Adjusted structural similarity for the covariance and correlation matrices for the 25D data set between the *Calomys* G and P matrix and the xenarthran P matrices. Bold values were significant at $p < 0.01$.

	Covariance matrix				Correlation matrix			
	RS		KP		RS		KP	
	Calomys (G)	Calomys (P)	Calomys (G)	Calomys (P)	Calomys (G)	Calomys (P)	Calomys (G)	Calomys (P)
<i>Zaedyus</i>	0.696	0.681	0.803	0.79	0.78	0.769	0.815	0.802
<i>Tolypeutes</i>	0.729	0.708	0.754	0.775	0.767	0.76	0.754	0.73
<i>Priodontes</i>	0.524	0.518	0.731	0.775	0.604	0.6	0.74	0.735
<i>Euphractus</i>	0.703	0.686	0.798	0.758	0.838	0.822	0.729	0.712
<i>Dasypus</i>	0.621	0.636	0.764	0.816	0.832	0.826	0.778	0.771
<i>Chaetophractus</i>	0.7	0.698	0.743	0.734	0.818	0.791	0.748	0.697
<i>Cabassous</i>	0.754	0.731	0.727	0.791	0.85	0.843	0.76	0.768
<i>Cyclopes</i>	0.714	0.693	0.704	0.689	0.769	0.75	0.704	0.683
<i>Myrmecophaga</i>	0.532	0.538	0.703	0.693	0.812	0.799	0.72	0.707
<i>Tamandua</i>	0.656	0.663	0.694	0.728	0.837	0.824	0.745	0.748
<i>Bradypus</i>	0.785	0.782	0.735	0.691	0.889	0.869	0.815	0.743
<i>Choloepus</i>	0.654	0.652	0.708	0.707	0.739	0.721	0.722	0.678
<i>Paramylodon</i>	0.608	0.603	0.733	0.713	0.759	0.738	0.755	0.734
<i>Scelidotherium</i>	0.669	0.655	0.72	0.717	0.709	0.709	0.808	0.77
<i>Acratocnus</i>	0.392	0.349	0.708	0.704	0.396	0.374	0.693	0.669
<i>Neocnus</i>	0.269	0.24	0.897	0.812	0.335	0.333	0.737	0.719

TABLE 4. Original (r^2) and adjusted (r^2_{adj}) overall magnitude of integration for the 25D data set of the Xenarthra.

	r^2	r^2_{adj}
<i>Zaedyus</i>	0.153	0.171
<i>Tolypeutes</i>	0.107	0.127
<i>Prionodontes</i>	0.108	0.144
<i>Euphractus</i>	0.208	0.217
<i>Dasyypus</i>	0.135	0.138
<i>Chaetophractus</i>	0.087	0.097
<i>Cabassous</i>	0.168	0.187
<i>Cyclopes</i>	0.102	0.111
<i>Myrmecophaga</i>	0.155	0.165
<i>Tamandua</i>	0.142	0.146
<i>Bradypus</i>	0.175	0.181
<i>Choloepus</i>	0.143	0.160
<i>Paramylodon</i>	0.124	0.147
<i>Scelidotherium</i>	0.239	–
<i>Acratocnus</i>	0.172	–
<i>Neocnus</i>	0.204	–

Overall Magnitude of Integration.—Before presenting the results, it is important to state that r^2 varies considerably with sample size used to estimate the correlation matrix. The r^2 estimated from less than 20–25 specimens is considerably overestimated (Supplementary Fig. 5). Consequently, in these circumstances the r^2 should be considered an upper limit for the population r^2 , instead of a good proxy for the real population r^2 . For this reason, the r^2 from the poorly sampled extinct sloths *Scelidotherium*, *Acratocnus*, and *Neocnus* were considered upper-limit r^2 estimates.

The r^2 values were all below 0.25 and, excluding the three extinct sloths mentioned above, ranged between 0.087 and 0.208 (Table 4). Pearson's product-moment correlation between r^2 and r^2_{adj} was high at 0.962 ($p < 0.001$). The r^2 across the different sets of distances varied markedly only between 35D and 32D for taxa with considerable variation in sample size (Fig. 1; Supplementary Table 19). One thousand bootstrap resamplings of the general linear model residuals of each genus showed that minor differences in the r^2 are due to sampling, but differences above ~ 0.1 cannot be explained solely by sampling (Supplementary Fig. 6).

Similarity and Distance Matrices.—The results for the phylogenetic distance matrix, determined by considering the number of substitutions per site as the branch length, and the phylogenetic distance matrix, determined by considering all branches as having a length of 1,

were virtually the same for the 35D and 32D data sets. For the sake of brevity, we chose to present only the results from the latter.

Structural similarity among xenarthran genus **P** matrices were correlated with phylogenetic history (Table 5). There was a weak correlation (~ 0.2) between the RS similarities and the phylogeny and a moderate correlation (~ 0.6) between KP similarities and phylogeny. The overall magnitude of the integration distance matrix was not correlated with the phylogenetic distance matrix. The log-transformed residual morphological distance matrix obtained by linearly regressing the morphological distances as a function of the phylogenetic distances presented a moderate correlation (~ 0.6) with the covariance matrix RS observed and adjusted similarities (Table 5). The overall magnitude of integration distance matrix was not correlated with the morphological distance matrix.

Regarding the results obtained with the distinct sets of distances (35D, 32D, 28D, and 25D), results were congruent between the data sets but with minor differences in the intensity of the correlations and number of significant correlations (Supplementary Table 20).

Discussion

The Cheverud Conjecture.—Xenarthran **P** matrices are not strictly identical or proportional. To satisfy these conditions, the product between one matrix and the inverse of the other should be a scalar multiple of the identity matrix (see Klingenberg and Monteiro 2005). However, two **P** matrices (or **G** matrices) will never satisfy this equation (Falconer and MacKay 1996; Cheverud and Marroig 2007; Porto et al. 2009). So, a more pertinent question is not whether the matrices are identical/proportional or not, but rather how much structural similarity is required between **G** or **P** matrices so they can be used in macroevolutionary studies (Roff 1996, 1997; Marroig and Cheverud 2001). There is no consensus among researchers about the minimum amount of structural similarity between **G** or **P** matrices needed to perform macroevolutionary studies (Cheverud 1988;

TABLE 5. Matrix correlation between original and adjusted structural similarity matrix and overall magnitude of integration distance matrix (r^2) and residual morphologic and phylogenetic distance matrices among xenarthran genera for the 25D data set. Bold values were significant at $p < 0.05$. The phylogenetic distance matrix was determined considering all branches having length of 1, and the residual morphological distance matrix was obtained based on this phylogenetic distance matrix.

	Covariance matrix				Correlation matrix				r^2
	RS		KP		RS		KP		
	Observed	Adjusted	Observed	Adjusted	Observed	Adjusted	Observed	Adjusted	
Phylogenetic	-0.194	-0.213	-0.614	-0.601	-0.148	-0.2	-0.587	-0.654	-0.109
Residual morphological	-0.563	-0.574	0.122	0.163	-0.120	-0.082	-0.021	-0.013	0.174

Arnold and Phillips 1999; Marroig and Cheverud 2001). However, at least for RS, it has been argued that if there is high (above 0.7) structural similarity between the matrices, their differences will have a small effect over the results of macroevolutionary quantitative genetics studies (Marroig and Cheverud 2001; Prôa et al. 2012). Furthermore, RS goes to the core of the evolutionary question by assessing whether two **G** or **P** matrices have similar evolutionary responses to the same selection gradients (Cheverud and Marroig 2007).

Overall, RS results indicate xenarthran **P** matrices tend to respond quite similarly to different selection gradients. In fact, xenarthran **P** matrices have in general high structural similarities (above 0.7), regardless of the taxonomic/phylogenetic level, the method (RS or KP) employed, and the type of matrix (covariance or correlation). Considering the deep evolutionary history of Xenarthra (Delsuc et al. 2012) and the broad phylogenetic context of this study, the high structural similarity among **P** matrices gives support to the idea that Xenarthra, in general, share a common genetic and developmental basis and, consequently, a common **G**-matrix structure. In other words, although there are differences in additive genetic variation between any evolutionary lineages (Falconer and MacKay 1996; Porto et al. 2009) and in skull development within Xenarthran evolutionary lineages (Hautier et al. 2011; Rager et al. 2014), those differences do not drastically alter the **G**-matrix structure. Accordingly, the **P** matrices are good proxies for their genetic counterparts, and the Cheverud conjecture holds true for this group. Further support to this interpretation is given by the generally high (above 0.7) adjusted similarities observed between xenarthran **P** matrices and rodent **G** and **P** matrices, irrespective of the method and the type of matrix analyzed. The discussion above fits perfectly well with previous findings that mammals in general share similar morphological cranial **G**- and **P**-matrices structures, regardless of the analyses being based on traditional morphometric data (Lofsvold 1986; Stepan 1997; Marroig and Cheverud 2001; Oliveira et al. 2009; Porto et al. 2009; Shirai and Marroig 2010; Haber 2015) or geometric

morphometric data (Goswami 2006, 2007; even though it is not yet well established if there is a clear and direct correspondence between results obtained from both types of data). This relative stability in mammalian **G** and **P** matrices is not fully comprehended, but in all likelihood some sort of shared stabilizing selection on the correlation pattern is involved in this large-scale stability (Jones et al. 2003; Melo and Marroig 2015). This shared stabilizing selection could be generated by similar functions being performed by the same homologous traits between lineages or by shared developmental paths between lineages resulting in similar internal stabilizing selection.

Nevertheless, there are some comparisons involving specific genera (the extinct sloths *Acratocnus*, *Neocnus*, *Scelidotherium*, and *Paramylodon*; the anteater *Myrmecophaga*; and the armadillo *Priodontes*) that did not show high similarities and should be discussed thoroughly. The most notable examples of lower similarities are related to the extinct sloths *Acratocnus* and *Neocnus*, which often presented adjusted RS similarities below 0.5. Considering both have poor sample sizes (eight and six specimens, respectively), it is likely these low similarities were to a large extent due to problems in the matrices' estimation related to sample size (Supplementary Figs. 1–4 and 7; Marroig et al. 2012), rather than differences inherent to their biological systems. The KP results reinforce this hypothesis, since the observed KP similarities involving these taxa were consistently smaller than the similarities involving other taxa (Figs. 3 and 4). Another point in favor of a sample size effect is that the RS similarities involving *Cyclopes* increased markedly between 35D and 32D, where sample size changed from 12 to 91 (Supplementary Tables 7–10). The armadillo *Priodontes* and the extinct sloth *Scelidotherium* also have small samples (24 and 16, respectively) and this probably contributed to some of the RS similarities involving them being below 0.7 (Supplementary Figs. 1–4 and 7; Marroig et al. 2012). In spite of this, it is remarkable that they still have several adjusted similarities above 0.7.

Sampling error cannot be the sole cause for the adjusted RS similarities below 0.7 found in

the comparisons involving the extinct sloth *Paramylodon* and the anteater *Myrmecophaga* for covariance matrices, since sample sizes for *Paramylodon*, *Myrmecophaga*, and all other Xenarthra but *Acratocnus*, *Neocnus*, *Priodontes*, and *Scelidotherium* ranged between reasonable and good (Fig. 1). Furthermore, sampling error should affect both correlation and covariance matrices in roughly similar ways, so the substantial increase in the RS similarities observed for the correlation matrices in these cases is not justified. The difference in RS similarities between the covariance matrix and the standardized correlation matrix suggests the lower RS similarities observed for the covariance matrices are related to a strong uneven distribution of trait variances (and associated covariances) across **P** matrices. Most commonly this is related to scale differences. Larger traits tend to present larger variances and smaller traits tend to show smaller variances. The same relationship can be extended to the associated covariances (Cheverud and Marroig 2007). This can be exemplified by the anteater *Myrmecophaga*, which has a very long rostrum and forehead in comparison with the posterior part of the braincase and therefore has relatively high (co)variances related to the distances that represent these regions (PMZS, PMZI, PMMT, NABR, and NAPNS; Supplementary Table 21). Another source of discrepancy in (co)variance occurs when the location of a landmark varies markedly across specimens within a population. This was only observed for the landmark PT for the extinct sloth *Paramylodon*. The suture between the frontal and parietal reaches the squamosal in different parts, more anteriorly or posteriorly depending on the specimen. As a consequence, the distances related to PT (BRPT, PTAPET, PTBA, PTEAM, PTTSP, and PTAS) have relatively large (co)variances, especially PTTSP (Supplementary Table 22).

In circumstances such as the ones described above, where some traits present considerably more (co)variation than others, the traits with more variation will correspond to axes of major variation in the multidimensional space and, consequently, will be highly correlated with the first PC (Krzanowski 2000). After the first PC, the other PCs can be interpreted as

successive orthogonal axes of major variation. Knowing the PCs is relevant, because their direction (i.e., eigenvectors) and variance (i.e., eigenvalues) may bias changes in mean phenotypes, especially the first PC, because it encompasses most of the available variation (Schluter 1996; Marroig and Cheverud 2005, 2010; Hansen and Voje 2011). Moreover, RS similarities are expected to be relatively low if the directions and/or the variances of the PCs are markedly different between the evolutionary lineages being compared. On the one hand, KP similarities show all Xenarthra share closely related subspaces. Additionally, the RS similarities for the correlation matrices—which have variance-standardized covariances—and for the reasonably/well-sampled taxa suggest Xenarthra have similar association patterns between traits. On the other hand, when we consider the covariance matrices for the reasonably/well-sampled taxa, the RS similarities indicate the PCs differ across some taxa in their direction/order of relevance or the amount of variance explained by them or both. For the extinct sloth *Paramylodon* and anteater *Myrmecophaga* covariance **P** matrices, few traits dominate the evolutionary response, because these traits present considerably higher variation. Evolutionary responses for these genera will always be biased by these few traits in relation to the others, exerting a strong attractor effect over the evolutionary-response vector. Consequently, they will diverge more often from other evolutionary lineages, and we now focus on the specific reasons for this divergence.

The discussion below refers to the results for covariance matrices, unless otherwise stated. Generally the first PC in skull data of mammals is size related both for covariance and correlation **P** matrices (Porto et al. 2013), and the only exception within Xenarthra is the extinct sloth *Paramylodon*, which has the second PC as size-related. The first PC for this genus is a contrast between the distance PTTSP and the distances BRPT, PTAPET, PTBA, PTEAM, and PTAS (Supplementary Table 23; notice this condition applies only for the covariance matrix, since the first PC for the correlation matrix is size related). The mean correlation of the *Paramylodon*'s first PC with the others is

0.47 (SD=0.1; minimum=0.28; maximum=0.65), while the mean correlation of the first PC of all Xenarthra but *Paramylodon* is 0.83 (SD=0.09; minimum=0.49; maximum=0.96; Supplementary Table 24). Thus, *Paramylodon*'s first PC in general has a different direction than the first PC from the other genera. This is a major reason why this genus has the largest number of adjusted RS similarities below 0.7 (not considering the results for *Acratocnus* and *Neocnus*). In fact, if the distance PTTSP is removed from the analysis, the first PC for *Paramylodon* also becomes size related, and the adjusted RS similarities involving this genus increase for several comparisons (Supplementary Table 25). This interpretation is reinforced when we compare the results involving this genus for the covariance and correlation matrices. Results for correlation matrices are considerably higher, since in this case *Paramylodon*'s first PC is, as for the other Xenarthra, size related. *Myrmecophaga* has a strong allometric first PC, which is mainly represented by the distances representing the length of the rostrum and forehead (PMZS, PMZI, PMMT, NABR, and NAPNS; Supplementary Table 23). The correlations of its first PC with those of the other taxa, excluding *Paramylodon*, are fairly high (mean=0.78; SD=0.13; minimum=0.49; maximum=0.96; Supplementary Table 24). Moreover, the first PC explains 66.7% of the total variance, the biggest apportioned variance explained by the first PC across Xenarthra (Supplementary Table 23). Thus, we hypothesized that although the direction of the first PC between *Myrmecophaga* and other Xenarthra is similar, the evolutionary responses will be more severely biased by the first PC in *Myrmecophaga*, and this bias is an important reason why this genus has several adjusted RS similarities below 0.7. To test our hypothesis, three covariance matrices (**C**₁, **C**₂, and **C**₃) were constructed using the following equation:

$$\mathbf{C} = \mathbf{V}^t \mathbf{\Lambda} \mathbf{V} \quad (3)$$

where **C** is a covariance matrix, **V** is a square matrix of normalized eigenvectors, **Λ** is a square diagonal matrix of eigenvalues, and the superscript **t** denotes matrix transpose. All **C**s have the same eigenvectors as *Myrmecophaga*

(i.e., \mathbf{V} is the normalized eigenvector matrix from *Myrmecophaga's* \mathbf{P} matrix). \mathbf{C}_1 has the eigenvalues from *Dasypus's* \mathbf{P} matrix, \mathbf{C}_2 has the eigenvalues from *Tamandua's* \mathbf{P} matrix, and \mathbf{C}_3 has the eigenvalues from *Bradypus's* \mathbf{P} matrix. Therefore, *Myrmecophaga's* \mathbf{P} matrix, \mathbf{C}_1 , \mathbf{C}_2 , and \mathbf{C}_3 have the same directions for the PCs but different total amounts of variation, and variances explained by each PC. Note, however, the RS is not influenced by differences in total amount of variation, since it compares selection responses for both matrices by their directions, which only depend on the relative variation between eigenvalues, and is not influenced by their norm, which depends on total variation. The adjusted RS similarities comparing the Cs and the Xenarthra matrices were determined and, regardless of the \mathbf{C} , were most often higher than the comparisons involving *Myrmecophaga* (Supplementary Table 26), suggesting the apportioned variance explained by the first PC contributes to the results observed in the RS similarities involving *Myrmecophaga*. However, the *Myrmecophaga* and Cs matrices show relatively low similarities with the sloths' matrices, suggesting in these cases not only the variance but also the direction of the first PC is a major contributor to the lower similarities.

This discussion is congruent with the results found at the subfamily, family, suborder, and order levels.

Overall Magnitude of Integration.—If compared to the theoretical maximum of 1, the overall magnitude of integration found across Xenarthra is low, since these values ranged between 0.09 and 0.24. However, the empirical distribution for a broad range of mammalian taxa suggests this statistic ranges between 0.04 and 0.6 (Oliveira et al. 2009; Porto et al. 2009; Shirai and Marroig 2010), at least when using linear distances. Although magnitude of integration statistics based on geometric morphometrics are available (e.g., Goswami et al. 2014), it is not yet clear whether these measurements of integration on linear distances and geometric morphometrics are directly comparable. Thus, xenarthran r^2 can be considered to be between low and moderate; the fossils analyzed are no exception, since *Paramylodon* had $r^2=0.124$, and the other

extinct taxa had r^2 below 0.239. More important than the values by themselves is the marked variation of the overall magnitude of integration across Xenarthra, with some genera having values around twice as big as others. Besides minor differences in the structure of the \mathbf{P} matrices, differences in the overall magnitude of integration help to illuminate the huge morphological diversity observed across extant and extinct Xenarthra. Even matrices that share exactly the same structure may respond in different directions to the same selection gradients if they have markedly distinct overall magnitudes of integration (Marroig et al. 2009; Porto et al. 2009). A high r^2 usually means traits are tightly correlated. As a result, the morphospace is restricted to one or few dimensions, and in turn, the responses to selection will be usually constrained by these few dimensions (i.e., only a few directions in the morphospace have variation for change). The opposite is true for a low r^2 , which will show weak intertrait associations and a morphospace represented by more dimensions (Lande 1979; Hansen and Houle 2008; Pavlicev et al. 2009; Shirai and Marroig 2010). Consequently, responses to selection will often be less constrained. Therefore, it is expected that genera such as the armadillo *Euphractus* and the sloth *Bradypus* will tend to be more constrained in their responses to selection than the armadillo *Chaetophractus* and the anteater *Cyclopes*, at least on a microevolutionary timescale. However, further analyses with appropriate statistics (Hansen and Houle 2008), which effectively measure how the \mathbf{G} or \mathbf{P} matrix influences evolution, are required to properly explore this issue.

Similarity and Distance Matrices.—There is a weak to moderate association between phylogenetic relationships and \mathbf{P} -matrix structure maintenance. KP similarity matrices were moderately correlated with the phylogenetic distance, suggesting morphological subspaces between closely related taxa are more aligned than between phylogenetically distant taxa. In contrast, RS similarity matrices were only weakly correlated with the phylogenetic distances, suggesting there is only a small tendency for closely related taxa to have more similar

evolutionary responses to the same selection gradients than do phylogenetically distant taxa. The log-transformed residual morphological distance matrix (obtained by the linear regression of the morphological distances as a function of the phylogenetic distances) was moderately correlated only with the RS similarities for the covariance matrices, suggesting differences in trait means have a limited influence over the similarity of the **P** matrices. The overall magnitude of integration distance matrix was not correlated with both the phylogenetic distance and the log-transformed residual morphological distance matrices, suggesting changes in overall magnitude of integration are to some extent not associated with phylogeny or morphological changes in character means.

Implications and Applications for Paleontology.—All the above discussions are important to guide evolutionary quantitative genetics analyses involving extinct lineages. We will first discuss biological and methodological reasons that may contribute to the observed structural differences in **G** or **P** matrices and then, later in this section, we will address when extant taxa **G** or **P** matrices may be used as surrogates for the extinct taxa **G** or **P** matrices.

The first concern is related to the ratio between sample size and number of traits (S/T). In order to estimate a covariance matrix, S/T should be above 1, otherwise the estimated matrix will not be a proper covariance matrix (i.e., it will not be a positive-definite matrix), and this may impose several limitations on further analyses (e.g., Lande and Arnold 1983; Marroig et al. 2012). Yet, when S/T is below 1 it is still possible to estimate a matrix, but as shown for the extinct sloths *Acratocnus* and *Neocnus*, matrices estimated from small samples may present unreliable results in some circumstances, since error will dominate the matrix estimation (Marroig et al. 2012). In these cases, results should be carefully scrutinized to determine whether they are biologically meaningful. This unreliability does not mean poorly estimated **P** matrices should not be determined, since they can be informative to some extent despite the estimate error. The adjusted KP similarities, for example, showed that those subsampled genera

share an aligned morphological subspace with other Xenarthra. It is also possible to determine a maximum overall magnitude of integration for these cases (Supplementary Fig. 5). The usual recommendation is to maximize S/T whenever possible, so the influence of error in matrix estimation is minimized (Sokal and Rohlf 1995), and the researcher should carefully consider the trade-off between sample size and number of traits. The relevance of error in matrix estimation is also conditioned by the eigenvalue distribution. Given a specified number of traits, when most of the total variation is explained by the first PCs, fewer specimens will be needed to reasonably estimate a matrix, while when the total variation is more evenly distributed across the eigenvalues, more specimens will be needed to reasonably estimate a matrix (Supplementary Fig. 7).

Other works dealing with mammalian skulls found that around 15 specimens should be enough to reasonably estimate a matrix, at least in the S/T below 1 situation (Goswami 2006; Goswami and Polly 2010). We agree with Goswami and Polly (2010) in that this value should not be generalized indiscriminately to other situations. On this topic, we would like to comment on two points. First, as shown above, matrix estimation accuracy depends on the number of traits and on the eigenvalue distribution of the matrix (Supplementary Figs. 5 and 7), thus different studies and populations may require more or fewer specimens to reasonably estimate a matrix. Second, assessment of the quality of the matrix estimation will be different depending on the type of matrix (covariance or correlation) and on the matrix comparison method (e.g., matrix correlation, RS, or KP). For instance, we based our discussion on the covariance matrix and on the RS and KP, while Goswami (2006) and Goswami and Polly (2010) based their discussions on matrix correlation of correlation matrices, and this difference in methodology leads to different similarity values that are not directly comparable.

Regarding the number of traits, the analyses with the distinct sets of distances (35D, 32D, 28D, and 25D) showed that changing the number of traits does not necessarily change

the results. Between 25D and 35D, 10 traits were added into the analyses, and no noticeable differences were observed in the results when sample sizes were constant across the distance sets. Our results are congruent with the findings of a recent study that used two distinct data sets, one with 107 interlandmark distances and another with 32 interlandmark distances (Haber 2015). Yet in some particular cases, removing only one trait (or a few) from the analyses may considerably change the results, as was the case when the PTTSP distance was removed from the covariance analyses and *Paramylodon's* covariance **P** matrix consequently became more similar to that of other Xenarthra.

There are at least three (co)variation and correlation particularities of traits that may more intensely affect the structural similarity of evolutionary lineages' matrices, and one should be aware of them. The first two cause strong, uneven distribution of trait variances and associated covariances: (1) the presence of some disproportionately large traits, as observed for the anteater *Myrmecophaga*, and (2) some traits varying markedly due to differences in landmark locations across specimens within a population, as is the case for the distances involving the landmark PT in the extinct sloth *Paramylodon*. As a result, in both cases relatively high variances and covariances will be associated with these traits, and these will exert an attractor effect over the evolutionary-response vector, diminishing the similarity of the evolutionary-response vector between these taxa and other evolutionary lineages. The same principle holds true for some disproportionately small traits, and this concern applies primarily to covariance matrices. The third element, which is relevant to both covariance and correlation matrices, is the pattern of association between traits. Theoretically, traits may have different patterns of association in different lineages (e.g., population A has a strong positive association between traits 1 and 2, while population B has a strong negative association), and this difference in patterns will have an impact in the similarity between matrices if several traits have different patterns of association between lineages. Thus, when considering fossils,

researchers should evaluate whether they might encounter one or more of these trait particularities in the analyses. Phylogenetic proximity and morphological similarity may also contribute to the resemblance between structure of **P** matrices and overall magnitude of integration (Marroig and Cheverud 2001; Goswami 2006; Oliveira et al. 2009; Porto et al. 2009; Haber 2015).

Another point of concern are the properties of the **G** or **P** matrices that are of interest for a particular study. The covariance matrix represents each trait's variation and the between-trait intensity and pattern of association, while the correlation matrix represents only the latter (Pavlicev et al. 2009). Most often, evolutionary studies rely on the information available in the covariance matrix (Hansen and Houle 2008), but in some circumstances analyzing the correlation matrix may be more relevant (Pavlicev et al. 2009). As an example of the impact of the two different matrices in evolutionary studies, take the RS similarities involving the extinct sloth *Paramylodon*. Considering only the variance-standardized covariances (i.e., correlation matrix) *Paramylodon* responds very similarly to other evolutionary lineages to random-selection vectors (mean adjusted RS similarities excluding *Acratocnus* and *Neocnus* is 0.78; SD=0.03), whereas if the variation for each trait is considered (i.e., covariance matrix), responses are far less aligned with other lineages (mean adjusted RS similarities, excluding *Acratocnus* and *Neocnus*, is 0.61; SD = 0.07). Also, depending on the particularities of the methods used to compare matrices, even the comparison between the same matrices may show different results. The contrast between the results produced by the RS and KP methods can be exemplified again by the results involving *Paramylodon's* covariance matrix. While the RS similarities involving this lineage indicate its responses to random-selection vectors are not strongly aligned with the responses from other lineages, the KP similarities show it shares a closely related morphological subspace with other lineages.

Now, the critical question to be answered is: When is it safe to use **G** or **P** matrices of extant lineages as surrogates for the **G** or **P** matrices of subsampled extinct lineages in evolutionary quantitative genetics analyses? There is good

overall evidence that therian mammals in general share a common genetic and phenotypic structure of variation and covariation (Lofsvold 1986; Stepan 1997; Marroig and Cheverud 2001; Goswami 2006, 2007; Oliveira et al. 2009; Porto et al. 2009; Shirai and Marroig 2010; Haber 2015), and the results for the Xenarthra are in line with this idea. This study and Haber (2015) also show this conclusion holds for different sets of distances. Therefore, for the therian mammals it is likely any given extant **G** or **P** matrix will suit well as a surrogate for extinct lineages. A conservative approach might be to select closely related phylogenetic or morphologically similar extant lineages as surrogates, since they may present more similar **G** or **P** matrices.

Nevertheless, special attention should be paid to extinct lineages showing either some disproportionately large (or small) traits or some traits varying markedly due to differences in the location of landmarks across specimens within a population. If the researcher suspects one of these two conditions applies to some traits of the lineage of interest, caution is required in the study of the covariance matrix and how it influences responses to selection. This concern is far less important if one is studying the properties expressed by the correlation matrix or if one is interested in the morphological subspace orientation. Also, attention should be paid to a trait's patterns of association in different lineages, since differences in these patterns may severely decrease similarities between covariance/correlation matrices. Unless an extant lineage with similar trait characteristics is known, there will not be a good extant surrogate. In these cases, the best option would be to run the analyses with some **G** or **P** matrices from extant lineages that have relatively different structures and evaluate whether these matrices bias the results or not. Indeed, using different extant matrices as surrogates may be an appropriate standard approach, since it takes into account the uncertainties related to the unknown **G** or **P** matrices of extinct lineages (e.g., Ackermann and Cheverud 2004). It is important to stress how a deep understanding of the methods used and the potential sources of variation is required to properly use extant **G** or **P** matrices as surrogates for the extinct ones.

Conclusions

The extant and extinct Xenarthra have similar **P**-matrix subspaces, as measured by the KP, and for taxa with at least adequate samples they tend to respond to selection gradients in similar ways. Two exceptions are the anteater *Myrmecophaga* and the extinct sloth *Paramylodon*, which have systematically lower adjusted RS similarities for the covariance matrices. In the case of *Myrmecophaga*, it is due to relatively minor differences in the PC directions and differences in the allocation of total variance in the first PC, which are consequences of the long rostrum and forehead in relation to the rest of the skull. In the case of *Paramylodon*, it is by virtue of a relatively intense departure of its first PC direction in relation to the other Xenarthra associated with differences in the position of the landmark PT among specimens. Xenarthran **P** matrices can be used as surrogates for their **G** matrices and, in general, extant taxa **P** matrices are good surrogates for extinct taxa **G** or **P** matrices. Nevertheless, the question of when it is reasonable to use extant **G** or **P** matrices as surrogates to extinct ones is not a trivial one with a straightforward answer. Guidelines were provided, and special attention is advised for traits that are either disproportionately large (or small) in relation to the other traits, or that are determined by at least one landmark that varies markedly within a population, or that have different patterns of association in different lineages.

Acknowledgments

The authors are thankful to people and institutions who provided generous help and/or access to collections: J. Galkin, R. Voss, and E. Westwig (American Museum of Natural History); M. Flannery and A. Goode (California Academy of Science); R. Fariña, D. Hernández, and A. Rojas (Facultad de Ciencias de la Universidad de la Republica); B. Patterson, W. Simpson, and S. Ware (Field Museum of Natural History); S. Aguiar and J. Ohana (Museu Paraense Emílio Goeldi); M. Thompson (Idaho Museum of Natural History); J. Dines and X. Wang (Los Angeles County Museum); A. Farrell and J. Harris

(La Brea Tar Pits and Museum); D. Flores and A. Kramarz (Museo Argentino de Ciencias Naturales); C. Cartelle, A. Vasconcelos, and L. Vilaboim (Museu de Ciências Naturais da Pontifícia Universidade Católica de Minas Gerais); T. Aguzzoli and D. Sanfelice (Museu de Ciências Naturais da Fundação Zoobotânica do Rio Grande do Sul); I. Olivares and M. Reguero (Museo de La Plata); S. Bermudez (Museo Paleontológico de Colonia del Sacramento); E. Gonzalez, A. Rinderknecht, and V. Scarabino (Museo Nacional de Historia Natural de Montevideo); S. Franco and J. A. de Oliveira (Museo Nacional); C. Conroy, E. Lacey, and J. Patton (University of California Museum of Vertebrate Zoology); A. Currant, P. Jenkins, and L. Tomsett (Natural History Museum); A. Barnosky and P. Holroyd (University of California Museum of Paleontology); K. Finn, R. Hulbert, B. MacFadden, C. McCaffery, and D. Reed (Florida Museum of Natural History); M. Brett-Surman, L. Gordon, and D. Lunde (Smithsonian National Museum of Natural History); J. Gualda, F. Nascimento, and M. de Vivo (Museu de Zoologia da Universidade de São Paulo); and P. Auricchio (Universidade Federal do Piauí). We are deeply indebted to Fábio Machado, Daniela Rossoni, Mark Hubbe, Lionel Hautier, P. David Polly, Anjali Goswami, and Jonathan Bloch for their insightful comments on a previous draft of this manuscript, to Nadia de Moraes-Barros for assisting with *Bradypus* species identification; and to Mary Safford Curioli for her editorial suggestions, which greatly improved the quality of our manuscript. This research was supported by grants and fellowships from Fundação de Amparo à Pesquisa do Estado de São Paulo (AH 2012/24937-9; DM 2014/26262-4; GM 2011/14295-7), Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior, and the Field Museum of Natural History.

Literature Cited

- Ackermann, R. R. 2002. Patterns of covariation in the hominoid craniofacial skeleton: implications for paleoanthropological models. *Journal of Human Evolution* 43:167–187.
- . 2003. Using extant morphological variation to understand fossil relationships: a cautionary tale. *South African Journal of Science* 99:255–258.
- . 2005. Variation in Neandertals: a response to Harvati (2003). *Journal of Human Evolution* 48:643–646.
- Ackermann, R. R., and J. M. Cheverud. 2004. Detecting genetic drift versus selection in human evolution. *Proceedings of the National Academy of Sciences USA* 101:17946–17951.
- Aguirre, J. D., E. Hine, K. McGuigan, and M. W. Blows. 2014. Comparing G: multivariate analysis of genetic variation in multiple populations. *Heredity* 112:21–29.
- Akesson, M., S. Bensch, and D. Hasselquist. 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.
- Anderson, R. P., and C. O. Handley. 2001. A new species of three-toed sloth (Mammalia: Xenarthra) from Panama, with a review of the genus *Bradypus*. *Proceedings of the Biological Society of Washington* 114:1–33.
- Armbruster, W. S., C. Pélabon, G. H. Bolstad, and T. F. Hansen. 2014. Integrated phenotypes: understanding trait covariation in plants and animals. *Philosophical Transactions of the Royal Society B* 369:20130245.
- Arnold, S. J. 1981. Behavioral variation in natural populations. I. Phenotypic, genetic and environmental correlations between chemoreceptive responses to prey in the garter snake. *Evolution* 35:489–509.
- . 1992. Constraints on phenotypic evolution. *American Naturalist* 140:S85–S107.
- Arnold, S., and P. Phillips. 1999. Hierarchical comparison of genetic variance-covariance matrices. II. Coastal-inland divergence in the garter snake, *Thamnophis elegans*. *Evolution* 53:1516–1527.
- Arnold, S. J., R. Buerger, P. A. Hohenlohe, B. C. Ajie, and A. G. Jones. 2008. Understanding the evolution and stability of the G-matrix. *Evolution* 62:2451–2461.
- Barton, N. H., and M. Turelli. 1989. Evolutionary quantitative genetics: how little do we know? *Annual Review of Genetics* 23:337–370.
- Bégin, M., and D. A. Roff. 2003. The constancy of the G matrix through species divergence and the effects of quantitative genetic constraints on phenotypic evolution: a case study in crickets. *Evolution* 57:1107–1120.
- Benton, M. J., and D. A. T. Harper. 2009. Introduction to paleobiology and the fossil record. Wiley-Blackwell, Chichester, U.K.
- Bergqvist, L. P., E. A. L. Abrantes, and L. S. Avilla. 2004. The Xenarthra (Mammalia) of São José de Itaboraí Basin (upper Paleocene, Itaboraian), Rio de Janeiro, Brazil. *Geodiversitas* 26:323–337.
- Berner, D., R. Kaeuffer, A. C. Grandchamp, J. A. M. Raeymaekers, K. Räsänen, and A. P. Hendry. 2011. Quantitative genetic inheritance of morphological divergence in a lake-stream stickleback ecotype pair: implications for reproductive isolation. *Journal of Evolutionary Biology* 24:1975–1983.
- Bininda-Emonds, O. R. P., M. Cardillo, K. E. Jones, R. D. E. MacPhee, R. M. D. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, and A. Purvis. 2007. The delayed rise of present-day mammals. *Nature* 446:507–512.
- Blows, M. W., S. F. Chenoweth, and E. Hine. 2004. Orientation of the genetic variance-covariance matrix and the fitness surface for multiple male sexually selected traits. *American Naturalist* 163:329–340.
- Cardini, A., and P. D. Polly. 2013. Larger mammals have longer faces because of size-related constraints on skull form. *Nature Communications* 4:2458.
- Cartelle, C., G. De Iuliis, and F. Pujos. 2008. A new species of Megalonychidae (Mammalia, Xenarthra) from the Quaternary of Poço Azul (Bahia, Brazil). *Comptes Rendus Palevol* 7: 335–346.
- Chernick, M.R. 2008. Bootstrap methods: a guide for practitioners and researchers. Wiley, Hoboken, N.J.

- Chernick, M. R., and L. A. LaBudde. 2010. Revisiting qualms about bootstrap confidence intervals. *American Journal of Mathematical and Management Sciences* 29:437–456.
- Cheverud, J. M. 1982. Phenotypic, genetic, and environmental morphological integration in the cranium. *Evolution* 36: 499–516.
- . 1984. Quantitative genetics and developmental constraints on evolution by selection. *Journal of Theoretical Biology* 110: 155–171.
- . 1988. A comparison of genetic and phenotypic correlations. *Evolution* 42:958–968.
- . 1995. Morphological integration in the saddle-back tamarin (*Saguinus fuscicollis*) cranium. *American Naturalist* 145:63–89.
- . 1996. Quantitative genetic analysis of cranial morphology in the cotton-top (*Saguinus oedipus*) and saddle-back (*S. fuscicollis*) tamarins. *Journal of Evolutionary Biology* 9:5–42.
- Cheverud, J. M., G. P. Wagner, and M. M. Dow. 1989. Methods for the comparative analysis of variation patterns. *Systematic Zoology* 38:201–213.
- Cheverud, J. M., and G. Marroig. 2007. Comparing covariance matrices: random skewers method compared to the common principal components model. *Genetics and Molecular Biology* 30:461–469.
- Clark, A. A. 2010. Ancient DNA from the extinct folivorous xenarthrans, or sloths, with specific attention toward the Greater Antillean Megalonychids and the Patagonian Mylodontid, *Mylodon darwini*. PhD Thesis. McMaster University, Ottawa.
- Delsuc, F., F. M. Catzeflis, M. J. Stanhope, and E. J. P. Douzery. 2001. The evolution of armadillos, anteaters and sloths depicted by nuclear and mitochondrial phylogenies: implications for the status of the enigmatic fossil Eurotamandua. *Proceedings of the Royal Society of London B* 268:1605–1615.
- Delsuc, F., M. Superina, M. Tilak, E. J. P. Douzery, and A. Hassanin. 2012. Molecular phylogenetics unveils the ancient evolutionary origins of the enigmatic fairy armadillos. *Molecular Phylogenetics and Evolution* 62:673–680.
- Eisenberg, J. F. 1989. *Mammals of the neotropics*, Vol. 1. The northern neotropics: Panama, Colombia, Venezuela, Guyana, Suriname, French Guiana. University of Chicago Press, Chicago.
- Eisenberg, J. F., and K. H. Redford. 1999. *Mammals of the neotropics*, Vol. 3. The central neotropics: Ecuador, Peru, Bolivia, Brazil. University of Chicago Press, Chicago.
- Elbroch, M. 2006. *Animal skulls: a guide to North American species*. Stackpole, Mechanicsburg, PA.
- Engelmann, G. F. 1985. The phylogeny of the Xenarthra. Pp. 51–64 in G. G. Montgomery, ed. *The evolution and ecology of armadillos, sloths and vermilinguas*. Smithsonian Institution Press, Washington, D.C.
- Falconer, D. S., and T. F. C. MacKay. 1996. *Introduction to quantitative genetics*. Longman, New York.
- Fisher, R. A. 1930. *The genetic theory of natural selection*. Clarendon Press, Oxford.
- Flury, B. 1988. *Common principal components and related multivariate models*. Wiley, New York.
- Garcia, G. 2010. Análise comparativa dos padrões de covariação genética e fenotípica no crânio e mandíbula de *Calomys expulsius* (Rodentia: Muridae). Master's dissertation. Universidade de São Paulo, São Paulo.
- Gardner, A. L. 2007. *Mammals of South America*, Vol. 1. Marsupials, xenarthrans, shrews, and bats. University of Chicago Press, Chicago.
- Gaudin, T. J. 2004. Phylogenetic relationships among sloths (Mammalia, Xenarthra, Tardigrada): the craniodental evidence. *Zoological Journal of the Linnean Society* 140:255–305.
- Goloboff, P. A., S. A. Catalano, J. M. Mirande, C. A. Szumik, J. S. Arias, M. Kallersjo, and J. S. Farris. 2009. Phylogenetic analysis of 73060 taxa corroborates major eukaryotic groups. *Cladistics* 25:211–230.
- Goswami, A. 2006. Morphological integration in the carnivoran skull. *Evolution* 60:169–183.
- . 2007. Phylogeny, diet, and cranial integration in Australodelphian marsupials. *PLoS ONE* 2:e995.
- Goswami, A., and P. D. Polly. 2010. Methods for studying morphological integration, modularity and covariance evolution. In J. Alroy, and G. Hunt, eds. *Quantitative Methods in Paleobiology*. Paleontological Society Short Course, October 30th, 2010. Paleontological Society Papers 16: 213–243.
- Goswami, A., J. B. Smaers, C. Soligo, and P. D. Polly. 2014. The macroevolutionary consequences of phenotypic integration: from development to deep time. *Philosophical Transactions of the Royal Society of London B* 369:20130254.
- Haber, A. 2015. The evolution of morphological integration in the ruminant skull. *Evolutionary Biology* 42:99–114.
- Hallgrímsson, B., H. Jamniczky, N. M. Young, C. Rolian, T. E. Parsons, J. C. Boughner, and R. S. Marcucio. 2009. Deciphering the palimpsest: studying the relationship between morphological integration and phenotypic covariation. *Evolutionary Biology* 36:355–376.
- Hansen, T. F., and D. Houle. 2008. Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology* 21:1201–1219.
- Hansen, T. F., and K. L. Voje. 2011. Deviation from the line of least resistance does not exclude genetic constraints: a comment on Berner et al. (2010). *Evolution* 65:1821–1822.
- Hautier, L., V. Weisbecker, A. Goswami, F. Knight, N. Kardjilov, and R. J. Asher. 2011. Skeletal ossification and sequence heterochrony in xenarthran evolution. *Evolution and Development* 13:460–476.
- Hoganson, J. W., and H. G. McDonald. 2007. First report of Jefferson's ground sloth (*Megalonyx jeffersonii*) in North Dakota: paleobiogeographical and paleoecological significance. *Journal of Mammalogy* 88:73–80.
- House, C. M., and L. W. Simmons. 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.
- Jones, A. G., S. J. Arnold, and R. Bürger. 2003. Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution* 57:1747–1760.
- . 2004. Evolution and stability of the G-matrix on a landscape with a moving optimum. *Evolution* 58:1639–1654.
- Klingenberg, C. P., and L. R. Monteiro. 2005. Distances and directions in multidimensional shape spaces: implications for morphometric applications. *Systematic Biology* 54:678–688.
- Kohn, L. A., and W. R. Atchley. 1988. How similar are genetic correlation structures? Data from mice and rats. *Evolution* 42:467–481.
- Krzanowski, W. J. 1979. Between-groups comparison of principal components. *Journal of the American Statistical Association* 74:703–707.
- . 2000. *Principles of multivariate analysis: a user's perspective*, rev. ed. Clarendon, Oxford.
- Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* 33:402–416.
- Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–1226.
- Lessells, C. M., and P. T. Boag. 1987. Unrepeatable repeatabilities: a common mistake. *Auk* 104:116–121.
- Lofsvold, D. 1986. Quantitative genetics of morphological differentiation in *Peromyscus*. I. Tests of the homogeneity of genetic covariance structure among species and subspecies. *Evolution* 40:559–573.

- MacPhee, R. D. E., J. L. White, and C. A. Woods. 2000. New megalonychid sloths (Phyllophaga, Xenarthra) from the Quaternary of Hispaniola. *American Museum Novitates* 3303:1–32.
- Marqu ez, E. J., R. Cabeen, R. P. Woods, and D. Houle. 2012. The measurement of local variation in shape. *Evolutionary Biology* 39:419–439.
- Marroig, G., and J. M. Cheverud. 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* 55:2576–2600.
- . 2005. Size as a line of least evolutionary resistance: diet and adaptive morphological radiation in New World monkeys. *Evolution* 59:1128–1142.
- . 2010. Size as a line of least resistance II: direct selection on size or correlated response due to constraints? *Evolution* 64:1470–1488.
- Marroig, G., M. de Vivo, and J. M. Cheverud. 2004. Cranial evolution in sakis (*Pithecia*, Platyrrhini) II: evolutionary processes and morphological integration. *Journal of Evolutionary Biology* 17:144–155.
- Marroig, G., L. T. Shirai, A. Porto, F. B. de Oliveira, and V. De Conto. 2009. The evolution of modularity in the mammalian skull II: evolutionary consequences. *Evolutionary Biology* 36:136–148.
- Marroig, G., D. A. R. Melo, and G. Garcia. 2012. Modularity, noise, and natural selection. *Evolution* 66:1506–1524.
- McAfee, R. K. 2009. Reassessment of the cranial characters of *Glossotherium* and *Paramylodon* (Mammalia: Xenarthra: Mylodontidae). *Zoological Journal of the Linnean Society* 155: 885–903.
- McDonald, H. G. 2005. Paleoecology of extinct xenarthrans and the Great American Biotic Interchange. *Bulletin of the Florida Museum of Natural History* 45:313–333.
- McDonald, H. G., C. R. Harrington, and G. De Iuliis. 2000. The ground sloth, *Megalonyx*, from Pleistocene deposits of the Old Crow Basin, Yukon, Canada. *Artic* 53:213–220.
- McDonald, H. G., and S. Pelikan. 2006. Mammoths and mylodonts: exotic species from two different continents in North American Pleistocene faunas. *Quaternary International* 142–143:229–241.
- McGuigan, M. C. 2006. Studying phenotypic evolution using multivariate quantitative genetics. *Molecular Ecology* 15:883–896.
- Melo, D., G. Garcia, A. Hubbe, A. P. Assis, and G. Marroig. 2015. EvolQC—an R package for evolutionary quantitative genetics, Version 1 [referees: 1 approved, 1 approved with reservations]. *F1000Research* 4:925.
- Melo, D., and G. Marroig. 2015. Directional selection can drive the evolution of modularity in complex traits. *Proceedings of the National Academy of Sciences USA* 112:470–475.
- Mi o-Boilini, A. R., and A. A. Carlini. 2009. The Scelidotheriinae Ameghino, 1904 (Phyllophaga, Xenarthra) from the Ensenadan-Lujanian Stage/Ages (Early Pleistocene to Early-Middle Pleistocene–Early Holocene) of Argentina. *Quaternary International* 210:93–101.
- Mitteroecker, P., and F. L. Bookstein. 2009. The ontogenetic trajectory of the phenotypic covariance matrix, with examples from craniofacial shape in rats and humans. *Evolution* 63: 727–737.
- M ller-Krull, M., F. Delsuc, G. Churakov, C. Marker, M. Superina, J. Brosius, E. J. P. Douzery, and J. Schmitz. 2007. Retroposed elements and their flanking regions resolve the evolutionary history of xenarthran mammals (armadillos, anteaters, and sloths). *Molecular Biology and Evolution* 24:2573–2582.
- Murphy, W. J., E. Eizirik, S. J. O’Brien, O. Madsen, M. Scally, C. J. Douady, E. Teeling, O. A. Ryder, M. J. Stanhope, W. W. de Jong, and M. S. Springer. 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294:2348–2351.
- O’Leary, M. A., J. I. Bloch, J. J. Flynn, T. J. Gaudin, A. Giallombardo, N. P. Giannini, S. L. Goldberg, B. P. Kraatz, Z.-X. Luo, J. Meng, X. Ni, M. J. Novacek, F. A. Perini, Z. S. Randall, G. W. Rougier, E. J. Sargis, M. T. Silcox, N. B. Simmons, M. Spaulding, P. M. Velazco, M. Weksler, J. R. Wible, and A. L. Cirranello. 2013. The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science* 339:662–667.
- Oliveira, F. B., A. Porto, and G. Marroig. 2009. Covariance structure in the skull of Catarrhini: a case of pattern stasis and magnitude evolution. *Journal of Human Evolution* 56: 417–430.
- Pavlicev, M., J. M. Cheverud, and G. P. Wagner. 2009. Measuring morphological integration using eigenvalue variance. *Evolutionary Biology* 36:157–170.
- Porto, A., F. B. D. Oliveira, L. T. Shirai, V. D. Conto, and G. Marroig. 2009. The evolution of modularity in the mammalian skull I: morphological integration patterns and magnitudes. *Evolutionary Biology* 36:118–135.
- Porto, A., L. T. Shirai, F. B. de Oliveira, and G. Marroig. 2013. Size variation, growth strategies, and the evolution of modularity in the mammalian skull. *Evolution* 67:3305–3322.
- Pr a, M., P. O’Higgins, and L. R. Monteiro. 2012. Type I error rates for testing genetic drift with phenotypic covariance matrices: a simulation study. *Evolution* 67:185–195.
- Puth, M., M. Neuh user, and G. D. Ruxton. 2015. On the variety of methods for calculating confidence intervals by bootstrapping. *Journal of Animal Ecology* 84:892–897.
- R Development Core Team. 2013. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rager, L., L. Hautier, A. Forasiepi, A. Goswami, and M. S nchez-Villagra. 2014. Timing of cranial suture closure in placental mammals: phylogenetic patterns, intraspecific variation, and comparison with marsupials. *Journal of Morphology* 275: 125–140.
- Redford, K. H., and J. F. Eisenberg. 1992. Mammals of the neotropics, Vol. 2. The Southern Cone: Chile, Argentina, Uruguay, Paraguay. University of Chicago Press, Chicago.
- Reusch, T., and W. U. Blanckenhorn. 1998. Quantitative genetics of the dung fly *Sepsis cynipsea*: Cheverud’s conjecture revisited. *Heredity* 81:111–119.
- Roff, D. A. 1995. The estimation of genetic correlations from phenotypic correlations: a test of Cheverud’s conjecture. *Heredity* 74:481–490.
- . 1996. The evolution of genetic correlations: an analysis of pattern. *Evolution* 50:1392–1403.
- . 1997. *Evolutionary quantitative genetics*. Chapman and Hall, London.
- Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. *Evolution* 50:1766–1774.
- Shaw, F. H., R. G. Shaw, G. S. Wilkinson, and M. Turelli. 1995. Changes in genetic variances and covariances: G whiz!. *Evolution* 49:1260–1267.
- Shirai, L. T., and G. Marroig. 2010. Skull modularity in neotropical marsupials and monkeys: size variation and evolutionary constraint and flexibility. *Journal of Experimental Zoology B* 314B:663–683.
- Simpson, G. G. 1980. *Splendid isolation: the curious history of South American mammals*. Yale University Press, New Haven, Conn.
- Sokal, R. R., and F. J. Rohlf. 1995. *Biometry*. Freeman, New York.
- Steppan, S. J. 1997. Phylogenetic analysis of phenotypic covariance structure. I. Contrasting results from matrix correlation and common principal component analyses. *Evolution* 51:571–586.
- Steppan, S. J., P. C. Phillips, and D. Houle. 2002. Comparative quantitative genetics: evolution of the G matrix. *Trends in Ecology and Evolution* 17:320–327.

- Stock, C. 1942. A ground sloth in Alaska. *Science* 95:552–553.
- Turelli, M. 1988. Phenotypic evolution, constant covariances and the maintenance of additive variance. *Evolution* 43: 1342–1347.
- Van der Linde, K., and D. Houle 2009. Inferring the nature of allometry from geometric data. *Evolutionary Biology* 36: 311–322.
- Vizcaino, S. F., M. S. Bargo, and R. A. Fariña. 2008. Form, function, and paleobiology in xenarthrans. Pp. 86–99 in S. F. Vizcaino, and W. J. Loughry, eds. *The biology of the Xenarthra*. University Press of Florida, Gainesville.
- Wagner, G. P., and L. Altenberg. 1996. Complex adaptations and the evolution of evolvability. *Evolution* 50:967–976.
- Walker, J.A. 2000. Ability of geometric morphometric methods to estimate a known covariance matrix. *Systematic Biology* 49:686–696.
- Webster, M., and M. L. Zelditch. 2011. Evolutionary lability of integration in Cambrian ptychoparioid trilobites. *Evolutionary Biology* 38:144–162.
- Wilkinson, G. S., K. Fowler, and L. Partridge. 1990. Resistance of genetic correlation structure to directional selection in *Drosophila melanogaster*. *Evolution* 44:1990–2003.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.
- Young, N. M., G. P. Wagner, and B. Hallgrímsson. 2010. Development and the evolvability of human limbs. *Proceedings of the National Academy of Sciences USA* 107:3400–3405.